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 75 No. 11 November 2023

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 75 • NOVEMBER 2023 • NO. 11

Special Articles

Rheumatology Education and Practice

Pediatrics

Cover image: The image on the cover (from Collado et al, page 2277) shows tenosynovitis on Doppler ultrasound imaging of the finger of a 6-year-old child. The presence of tendon sheath effusion and peritendinous Doppler signal within the synovial sheath in the flexor tendon is shown.

Arthritis Care & Research
Vol. 75, No. 11, November 2023, pp 2227–2238
DOI 10.1002/acr.25175

American College of Rheumatology
Empowering Rheumatology Professionals

2023 American College of Rheumatology and American Association of Hip and Knee Surgeons Clinical Practice Guideline for the Optimal Timing of Elective Hip or Knee Arthroplasty for Patients With Symptomatic Moderate-to-Severe Osteoarthritis or Advanced Symptomatic Osteonecrosis With Secondary Arthritis for Whom Nonoperative Therapy Is Ineffective

Charles P. Hannon,[1*] Susan M. Goodman,[2*] Matthew S. Austin,[3*] Adolph Yates Jr. [4*] Gordon Guyatt,[5] Vinay K. Aggarwal,[6] Joshua F. Baker,[7] Phyllis Bass,[8] Delamo Isaac Bekele,[9] Danielle Dass,[10] Hassan M. K. Ghomrawi,[11] David S. Jevsevar,[12] C. Kent Kwoh,[13] Claudette M. Lajam,[6] Charis F. Meng,[2] Larry W. Moreland,[14] Linda I. Suleiman,[11] Jesse Wolfstadt,[15] Kimberly Bartosiak,[1] Nicholas A. Bedard,[9] Jason L. Blevins,[2] Anna Cohen-Rosenblum,[16] P. Maxwell Courtney,[3] Ruth Fernandez-Ruiz,[2] Elizabeth B. Gausden,[2] Nilasha Ghosh,[2] Lauren K. King,[17] Alexa Simon Meara,[18] Bella Mehta,[2] Reza Mirza,[5] Adam J. Rana,[19] Nancy Sullivan,[20] Marat Turgunbaev,[21] Katherine D. Wysham,[22] Kevin Yip,[2] Linda Yue,[2] Michael G. Zywiel,[23] Linda Russell,[2] Amy S. Turner,[21] and Jasvinder A. Singh[24*]

Guidelines and recommendations developed and/or endorsed by the American College of Rheumatology (ACR) are intended to provide guidance for patterns of practice and not to dictate the care of a particular patient. The ACR considers adherence to the recommendations within this guideline to be voluntary, with the ultimate determination regarding their application to be made by the clinician in light of each patient's individual circumstances. Guidelines and recommendations are intended to promote beneficial or desirable outcomes but cannot guarantee any specific outcome. Guidelines and recommendations developed and endorsed by the ACR are subject to periodic revision as warranted by the evolution of medical knowledge, technology, and practice. ACR recommendations are not intended to dictate payment or insurance decisions, and drug formularies or other third-party analyses that cite ACR guidelines should state this. These recommendations cannot adequately convey all uncertainties and nuances of patient care.

The American College of Rheumatology is an independent, professional, medical, and scientific society that does not guarantee, warrant, or endorse any commercial product or service.

Objective. To develop evidence-based consensus recommendations for the optimal timing of hip and knee arthroplasty to improve patient-important outcomes including, but not limited to, pain, function, infection, hospitalization, and death at 1 year for patients with symptomatic and radiographic moderate-to-severe osteoarthritis or advanced symptomatic osteonecrosis with secondary arthritis of the hip or knee who have previously attempted nonoperative therapy, and for whom nonoperative therapy was ineffective, and who have chosen to undergo elective hip or knee arthroplasty (collectively referred to as TJA).

Methods. We developed 13 clinically relevant population, intervention, comparator, outcomes (PICO) questions. After a systematic literature review, the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach was used to rate the quality of evidence (high, moderate, low, or very low), and evidence tables were created. A Voting Panel, including 13 physicians and patients, discussed the PICO

questions until consensus was achieved on the direction (for/against) and strength (strong/conditional) of the recommendations.

Results. The panel conditionally recommended against delaying TJA to pursue additional nonoperative treatment including physical therapy, nonsteroidal antiinflammatory drugs, ambulatory aids, and intraarticular injections. It conditionally recommended delaying TJA for nicotine reduction or cessation. The panel conditionally recommended delay for better glycemic control for patients who have diabetes mellitus, although no specific measure or level was identified. There was consensus that obesity by itself was not a reason for delay, but that weight loss should be strongly encouraged, and the increase in operative risk should be discussed. The panel conditionally recommended against delay in patients who have severe deformity or bone loss, or in patients who have a neuropathic joint. Evidence for all recommendations was graded as low or very low quality.

Conclusion. This guideline provides evidence-based recommendations regarding the optimal timing of TJA in patients who have symptomatic and radiographic moderate-to-severe osteoarthritis or advanced symptomatic osteonecrosis with secondary arthritis for whom nonoperative therapy was ineffective to improve patient-important outcomes, including pain, function, infection, hospitalization, and death at 1 year. We acknowledge that the evidence is of low quality primarily due to indirectness and hope future research will allow for further refinement of the recommendations.

INTRODUCTION

Patients who have osteoarthritis (OA) and advanced symptomatic osteonecrosis (ON) with secondary arthritis can benefit from nonoperative treatment, e.g., physical therapy, nonsteroidal antiinflammatory drugs (NSAIDs), braces, intraarticular injections, and weight reduction (1–4). However, none of these treatments are disease modifying, and progressive pain and loss of function lead many patients to choose arthroplasty when nonoperative therapy has lost efficacy. While projected increases in the utilization of total joint arthroplasty (TJA), including total hip (THA) or knee arthroplasty (TKA), vary widely (from estimates of >4 million people in the US by 2030, to models projecting a slower rise, with a plateau in 2009 [5]), there is consensus that utilization will increase (6–11). Both procedures have demonstrated success in reducing pain, restoring function, and improving quality of life for patients who have radiographic moderate-to-severe OA or advanced symptomatic ON with secondary arthritis after insufficient relief from nonoperative treatments (12,13). As the volume of these procedures continues to rise, the comparative value of these surgeries versus nonoperative treatment has been questioned (10). Nonoperative treatments include, but are not limited to, activity modification, analgesic medications such as NSAIDs or acetaminophen, physical therapy, intraarticular injections, bracing, weight loss, and gait aids (2).

For this guideline, our population consists of patients who have moderate-to-severe pain and loss of function and moderate-to-severe radiographic OA or ON with secondary arthritis, using standard radiographic measures such as Kellgren/Lawrence (K/L) grade (14), and who have also completed ≥1 trials of appropriate nonoperative therapy and elected to undergo TJA after a shared decision-making process with their physician. This does not include patients who have mild radiographic OA or ON with secondary arthritis, patients who

This article is simultaneously published in Arthritis & Rheumatology and Journal of Arthroplasty.

Supported by the American College of Rheumatology and the American Association of Hip and Knee Surgeons.

[1]Charles P. Hannon, MD, MBA, Kimberly Bartosiak, MD: Washington University School of Medicine, St. Louis, Missouri; [2]Susan M. Goodman, MD, Charis F. Meng, MD, Jason L. Blevins, MD, Ruth Fernandez-Ruiz, MD, MS, Elizabeth B. Gausden, MD, MPH, Nilasha Ghosh, MD, MS, Bella Mehta, MBBS, MS, MD, Kevin Yip, MD, Linda Yue, MD, Linda Russell, MD: Hospital for Special Surgery and Weill Cornell Medicine, New York, New York; [3]Matthew S. Austin, MD, P. Maxwell Courtney, MD: Rothman Orthopaedic Institute, Philadelphia, Pennsylvania; [4]Adolph Yates Jr., MD: University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania; [5]Gordon Guyatt, MD, Reza Mirza, MD: McMaster University, Hamilton, Ontario, Canada; [6]Vinay K. Aggarwal, MD, Claudette M. Lajam, MD: NYU Langone Health, New York, New York; [7]Joshua F. Baker, MD, MSCE: University of Pennsylvania, Philadelphia; [8]Phyllis Bass, MSML, MBA: Cypress, Texas; [9]Delamo Isaac Bekele, MBBS, Nicholas A. Bedard, MD: Mayo Clinic, Rochester, Minnesota; [10]Danielle Dass: Sterling, Virginia; [11]Hassan M. K. Ghomrawi, PhD, MPH, Linda I. Suleiman, MD: Feinberg School of Medicine of Northwestern University, Chicago, Illinois; [12]David S. Jevsevar, MD, MBA: OrthoVirginia, Richmond, Virginia; [13]C. Kent Kwoh, MD: University of Arizona College of Medicine, Tucson; [14]Larry W. Moreland, MD: University of Colorado Anschutz Medical Campus, Aurora; [15]Jesse Wolfstadt, MD, MSc: Sinai Health, University of Toronto, Toronto, Ontario, Canada; [16]Anna Cohen-Rosenblum, MD, MSC: LSU Health, New Orleans, Louisiana; [17]Lauren K. King, MD, MSc: University of Toronto, Toronto, Ontario, Canada; [18]Alexa Simon Meara, MD: The Ohio State Wexner Medical Center, Columbus, Ohio; [19]Adam J. Rana, MD: Maine Medical Partners, Portland; [20]Nancy Sullivan, BA: ECRI Institute, Plymouth Meeting, Pennsylvania; [21]Marat Turgunbaev, MD, MPH, Amy S. Turner: American College of Rheumatology, Atlanta, Georgia; [22]Katherine D. Wysham, MD: VA Puget Sound Health Care System and University of Washington School of Medicine, Seattle; [23]Michael G. Zywiel, MD, MSc, FRCSC: Schroeder Arthritis Institute, Toronto Western Hospital, University of Toronto, Toronto, Ontario, Canada; [24]Jasvinder A. Singh, MD, MPH: University of Alabama at Birmingham and Veterans Affairs Medical Center, Birmingham, Alabama.

*Drs. Hannon, Goodman, Austin, Yates, and Singh contributed equally to this work.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25175.

Address correspondence via email to Charles P. Hannon, MD, MBA, at charles.p.hannon@gmail.com.

Submitted for publication March 10, 2023; accepted in revised form June 14, 2023.

SUMMARY

- The decision of when to proceed with total joint arthroplasty (TJA) in patients who have symptomatic and moderate-to-severe radiographic osteoarthritis (OA) or advanced symptomatic osteonecrosis (ON) with secondary arthritis for whom nonoperative therapies were ineffective should be made by the physician and patient through a shared decision-making process during which the unique risks and benefits for the individual patient are considered.
- In patients who have moderate-to-severe symptomatic OA or advanced symptomatic ON with secondary arthritis who are indicated for TJA and for whom nonoperative therapy has been ineffective, we conditionally recommend proceeding directly to surgery without delay for additional nonoperative treatment of the joint problem.
- For patients who have obesity and moderate-to-severe symptomatic OA or advanced symptomatic ON with secondary arthritis who are indicated for TJA, we conditionally recommend against delaying surgery to meet a rigid weight or body mass index threshold. Patients should be educated on the increased risk of medical and surgical complications due to their obesity as well as counseled on how to lose weight.
- In patients with diabetes mellitus and moderate-to-severe symptomatic OA or advanced symptomatic ON with secondary arthritis who are indicated for TJA, we conditionally recommend delaying surgery to allow for improved glycemic control.
- In patients with nicotine dependence and moderate-to-severe symptomatic OA or advanced symptomatic ON with secondary arthritis who are indicated for TJA, we conditionally recommend delaying TJA to achieve nicotine cessation or decreased use of nicotine products.
- While all recommendations in this guideline are conditional based at least in part on the quality of evidence, we have systematically reviewed all the evidence available to date, which can be used to make treatment decisions, and the consensus was high among the expert panel.

have minimal pain and/or disability, or patients who have not tried some form of nonoperative therapy. Prior to being indicated for TJA, patients' medical comorbidities and prior nonoperative treatments are evaluated. However, patients may have their procedure postponed if they did not try specific treatments for nonoperative arthritis or to pursue medical optimization (15,16). While nonoperative treatment has benefits for most patients who have OA or advanced symptomatic ON with secondary arthritis (1–4), there is no consensus on the effectiveness of specific additional nonoperative treatments after nonoperative therapy has been ineffective in patients in the defined population for this guideline, those who have radiographically moderate-to-severe OA or ON with secondary arthritis of the hip or knee and moderate-to-severe pain or loss of function who have completed ≥1 trials of appropriate nonoperative therapy (17).

Patients who have certain risk factors, such as obesity, diabetes mellitus, and nicotine use, may also have surgical treatment delayed by hospital policy or third-party payers in order to meet specific criteria to mitigate their surgical risk. However, while these factors are clearly associated with increased risk for adverse events, it is unknown whether delaying surgery in order to achieve a specific glycemic end point, weight or body mass index (BMI) target, or absolute nicotine cessation leads to improved outcomes after TJA (18–22).

The purpose of this clinical practice guideline was to develop consensus on evidence-based recommendations for the optimal timing of TJA in patients with symptomatic moderate-to-severe OA or advanced symptomatic ON with secondary arthritis for whom nonoperative therapy has been ineffective and who elected to undergo TJA, and to evaluate benefits of delays of surgery for additional nonoperative arthritis treatments or to achieve specific targets for medical optimization. This guideline is intended for use during a shared decision-making process with this defined group of patients and their physicians after nonoperative therapies have ceased to be effective; this is not a guideline on the efficacy of nonoperative therapies in patients who have OA or ON with secondary arthritis who are not candidates for THA or TKA. Although patients who have inflammatory arthritis may also have OA, either primary or secondary, they also have moderate systemic inflammatory disease activity and are likely to be taking immunosuppressant medications at the time of surgery, the management of which was felt to be beyond the scope of this guideline. This was the focus of the 2022 American College of Rheumatology (ACR)/American Association of Hip and Knee Surgeons (AAHKS) Guideline for the Perioperative Management of Antirheumatic Medication in Patients With Rheumatic Diseases Undergoing Elective Total Hip or Total Knee Arthroplasty (23).

METHODS

This guideline follows the ACR guideline development process and ACR policy guiding management of conflicts of interest and disclosures https://rheumatology.org/clinical-practice-guidelines, which includes Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology (24,25). Supplementary Appendix 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25175, includes a detailed description of the methods. Briefly, the core leadership team drafted 13 clinical population, intervention, comparator, outcomes (PICO) questions (see Supplementary Appendix 2, available on the *Arthritis Care & Research* website at

http://onlinelibrary.wiley.com/doi/10.1002/acr.25175). The literature review team performed systematic searches with the guidance and oversight of a medical research librarian, based on the PICO questions, on September 27, 2021, and later updated on June 19, 2022; in total, 8,283 abstracts were identified. For the purpose of this guideline, our defined population is patients who have radiographically moderate-to-severe OA or advanced symptomatic ON with secondary arthritis of the hip or knee and moderate-to-severe pain or loss of function for whom nonoperative therapy was ineffective. Radiographic severity may be measured by validated grading systems such as K/L or Tonnis (14,26).

After abstract and full-text review, 176 papers were included to serve as the evidence base for the development of recommendations. The literature review team then graded the quality of evidence (high, moderate, low, very low) and produced the evidence report (see Supplementary Appendix 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25175). A Patient Panel of 8 patients, who were either candidates for or had prior TJA, was convened and moderated by a rheumatologist and ACR staff (LR, AT, RP). Patients reviewed the evidence report and provided their perspectives and preferences for consideration by the Voting Panel. The evidence was reviewed, and recommendations were formulated and voted on by an expert Voting Panel consisting of rheumatologists, orthopedic surgeons, and patients.

Consensus required ≥70% agreement on both direction (for or against) and strength (strong or conditional) of each recommendation, as per ACR practice. A recommendation could be either in favor of or against the proposed intervention and either strong or conditional. According to GRADE, a recommendation was categorized as strong if the panel was very confident that the benefits of an intervention clearly outweigh the harms or burdens (or vice versa); a conditional recommendation denoted uncertainty regarding the balance of benefits and harms, such as when the evidence quality is low or very low, when the decision was sensitive to individual patient preferences, or when costs were expected to impact the decision. Thus, conditional recommendations referred to decisions in which incorporation of patient preferences was a particularly essential element of decision-making. Rosters of the Core Leadership Team, Literature Review Team, and Voting Panel are included in Supplementary Appendix 4, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25175.

Target population and guiding principles

These recommendations are for patients who have radiographically moderate-to-severe OA or advanced symptomatic ON with secondary arthritis of the hip or knee, using standard radiographic grading such as K/L or Tonnis, and moderate-to-severe pain or loss of function who have been indicated for elective TJA through a shared decision-making process with their physician and have completed and did not improve with ≥1 trials of appropriate nonoperative therapy such as physical therapy, NSAIDs, and/or intraarticular injections (e.g., glucocorticoids or viscosupplementation). This does not include patients who have mild radiographic OA, patients who have minimal pain and/or disability, or patients who have not tried nonoperative therapy. This guideline does not address arthroplasty for patients who have rheumatic diseases, as they were the focus of the recent ACR/AAHKS guideline for the perioperative management of antirheumatic medications in patients undergoing total hip and total knee arthroplasty (23).

A conditional recommendation means that the panel has inferred that the majority of informed patients would choose the recommended course of action, but that an appreciable minority would not. A shared decision-making process with full consideration of patient preferences and individualized risk estimates should determine the appropriate course of action.

For recommendations regarding modifying risk factors prior to surgery, including BMI, glycemic control, and nicotine dependence, patients should be educated on the increased risk of medical and surgical complications associated with their specific condition. Patients should be counseled on effective methods to modify the risk factors (e.g., weight loss, improved glycemic control, nicotine cessation) and be provided resources to assist them through that process. However, it is recognized that not all patients have the medical, financial, or social resources or support available to them to modify some or all these risk factors.

RECOMMENDATIONS

All recommendations in this guideline are conditional due to the low or very low quality of evidence (Table 1). There are no strong recommendations in this guideline, although there was high or unanimous consensus for all recommendations.

In our defined population, we conditionally recommend proceeding to TJA without delay over delaying arthroplasty 3 months.

There should be no mandate that patients wait 3 months prior to TJA as an arbitrary cool-down period. The recommendation is conditional because there may be exceptions, and the evidence supporting the recommendation is indirect and very low quality. Prior to presenting to an orthopedic surgeon and being indicated for TJA, patients in the defined population have already attempted nonoperative treatment for an extended period. Further delay to TJA may lead to

Table 1. Recommendations for defined population*

Recommendation	Certainty of evidence	Based on the evidence report of the following PICOs	Page numbers of evidence tables in the Supplementary Appendix†
In our defined population, we **conditionally** recommend proceeding to TJA without delay over delaying arthroplasty 3 months.	Very low	1	1–7
In our defined population, we **conditionally** recommend proceeding to TJA without delay over delaying arthroplasty for a trial of physical therapy.	Low	2	8–37
In our defined population, we **conditionally** recommend proceeding to TJA without delay over delaying surgical treatment for a trial of NSAIDs.	Very low	3	38–46
In our defined population, we **conditionally** recommend proceeding to TJA without delay over delaying surgical treatment for a trial of braces and/or ambulatory aids.	Very low	4	47–53
In our defined population, we **conditionally** recommend proceeding to TJA without delay over delaying surgical treatment for a trial of intraarticular glucocorticoid injections.	Very low	5	54–63
In our defined population, we **conditionally** recommend proceeding to TJA without delay over delaying surgical treatment for a trial of viscosupplementation injections.	Very low	6	64–76
In our defined population with a BMI of ≥50, we **conditionally** recommend proceeding to TJA without delaying to achieve weight reduction to a BMI of <50.	Very low	7	77–130
In our defined population with a BMI of 40–49, we **conditionally** recommend proceeding to TJA without delaying to achieve weight reduction to a BMI of <40.	Very low	8	77–130
In our defined population with a BMI of 35–39, we **conditionally** recommend proceeding to TJA without delaying to achieve weight reduction to a BMI of <35.	Very low	9	77–130
In our defined population with poorly controlled diabetes mellitus, we **conditionally** recommend delaying TJA to improve glycemic control.	Very low	10	131–156
In our defined population with nicotine dependence, we **conditionally** recommend delaying TJA for nicotine use reduction/cessation.	Low	11	157–180
In our defined population with bone loss with deformity or severe ligamentous instability, we **conditionally** recommend proceeding to TJA without delay over delaying TJA for optimization of non–life-threatening conditions.	There were no studies that either directly or indirectly answered our PICO question.	12	181
In our defined population with a neuropathic joint, we **conditionally** recommend proceeding to TJA without delay over delaying for optimization of non–life-threatening conditions.	There were no studies that either directly or indirectly answered our PICO question.	13	181

* The defined population is patients with radiographically moderate-to-severe osteoarthritis or osteonecrosis of the hip or knee using standard radiographic grading such as Kellgren/Lawrence or Tonnis, and for patients with moderate-to-severe pain or loss of function who have been indicated for elective total joint arthroplasty (TJA) through a shared decision-making process with their physician and have completed trials of ≥1 appropriate nonoperative therapy. BMI = body mass index; NSAIDs = nonsteroidal antiinflammatory drugs; PICO = population, intervention, comparator, outcomes.
† In Supplementary Appendix 3, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25175.

increased pain, loss of function, and worsening of medical comorbidities due to limited mobility. Patients may elect to delay surgery due to personal reasons (e.g., work or family obligations) or ongoing medical comorbidities that can be optimized prior to surgery. In these cases, patients may consider proceeding with nonoperative treatment (excluding intraarticular injections in some cases; see below) to provide pain relief while awaiting surgery.

In our defined population, we conditionally recommend proceeding to TJA without delay over delaying arthroplasty for a trial of physical therapy.

In patients who are indicated for TJA, mandated physical therapy is not recommended to delay or avoid surgery. While physical therapy may provide benefit in knee and hip OA (2), and physical therapy may be of benefit in anticipation of

arthroplasty as a form of prerehabilitation to improve the outcome of surgery (27), delaying surgery for physical therapy may cause increased pain due to the severity of an individual's disease (28). However, nonambulatory patients, patients recovering from medical comorbidities (e.g., stroke) that may limit their rehabilitation postoperatively, or patients who have major lower extremity muscular weakness may benefit from delaying TJA for physical therapy to help improve postoperative outcomes. This recommendation does not apply directly to prerehabilitation, such as a preoperative individualized exercise and lifestyle modification program. Observational studies provided additional evidence for prehabilitation alone and prehabilitation versus usual care for patients with knee or hip OA or knee or hip ON with secondary arthritis awaiting TKA/THA. These studies had small sample sizes and provided indirect comparisons, sometimes with lack of precision in effect estimates, so the evidence supporting the recommendation is indirect or of low quality. Moreover, the included randomized controlled trials either did not have a surgical arm or randomized patients on surgical waiting lists. The effect of physical therapy ranged from insignificant to borderline significant with small effect sizes. This recommendation is conditional because there may be exceptions to this recommendation, and the evidence supporting the recommendation is indirect and low quality. The exceptions listed for the first recommendation above, including delay for personal reasons or other ongoing medical comorbidities, apply to this recommendation as well.

In our defined population, we conditionally recommend proceeding to TJA without delay over delaying surgical treatment for a trial of NSAIDs.

The NSAIDs are one of the mainstays of nonoperative treatment for OA and can provide pain relief for patients with mild disease. Oral NSAIDs are, however, associated with adverse events (e.g., peptic ulcer disease, acute kidney injury, increased cardiovascular risk, and bleeding) (29). Delaying TJA for treatment with oral NSAIDs may cause increased harm to the patient with limited clinical benefit. This recommendation is conditional because there may be exceptions to this, and the evidence supporting it is indirect and very low quality. The exceptions listed for the first recommendation above, including delay for personal reasons or other ongoing medical comorbidities, apply to this recommendation as well.

In our defined population, we conditionally recommend proceeding to TJA without delay over delaying surgical treatment for a trial of braces and/or ambulatory aids.

This recommendation is conditional because there may be exceptions to this recommendation, and the evidence supporting the recommendation is indirect and very low quality. The exceptions listed for the first recommendation above, including delay for personal reasons or other ongoing medical comorbidities, apply to this recommendation as well. Patients who are recovering from another lower limb surgery (e.g., contralateral THA or TKA) may benefit from delaying TJA and using an ambulatory aid during the recovery period. However, delaying TJA for treatment with a brace or ambulatory aid can place a burden on the patient given the need for education on the proper use of ambulatory aids such as canes, as improper use may lead to altered gait mechanics, increased pain, and worsened function (30–32).

In our defined population, we conditionally recommend proceeding to TJA without delay over delaying surgical treatment for a trial of intraarticular glucocorticoid injections.

This recommendation is conditional because there may be exceptions to this recommendation, and the evidence supporting the recommendation is indirect and very low quality. The exceptions listed for the first recommendation above, including delay for personal reasons or other ongoing medical comorbidities, apply to this recommendation as well. Patients who have an acute flare of their OA or other inflammatory arthropathy (e.g., gout, calcium pyrophosphate deposition disease) may be interested in delaying TJA for treatment with a glucocorticoid injection to provide immediate pain relief. There are, however, potential harms associated with delaying surgery for glucocorticoid injection treatment, particularly in patients with diabetes mellitus who have an increased risk of hyperglycemia with intraarticular glucocorticoids or the increased risk of joint infection if the surgery is performed within 3 months of the intraarticular injection (33,34).

In our defined population, we conditionally recommend proceeding to TJA without delay over delaying surgical treatment for a trial of viscosupplementation injections.

This recommendation is conditional because there may be exceptions to this recommendation, and the evidence supporting the recommendation is indirect and very low quality. The data on viscosupplementation for patients who are otherwise candidates for TJA were very limited. Viscosupplementation may place an unnecessary burden on the patient, with limited benefit on pain and function (30,35,36). The exceptions listed for the first recommendation above, including delay for personal reasons or other

ongoing medical comorbidities, apply to this recommendation as well.

In our defined population with a BMI of ≥50, we conditionally recommend proceeding to TJA without delaying to achieve weight reduction to a BMI of <50.

In our defined population with a BMI of 40–49, we conditionally recommend proceeding to TJA without delaying to achieve weight reduction to a BMI of <40.

In our defined population with a BMI of 35–39, we conditionally recommend proceeding to TJA without delaying to achieve weight reduction to a BMI of <35.

Recommendations 7–9 are conditional because there may be exceptions to these recommendations, and the evidence supporting a preoperative weight reduction and a rigid BMI or weight threshold is indirect and very low quality. A majority of the studies supporting all 3 recommendations were based on comparing TJA outcomes in patients who underwent bariatric surgical procedures to the outcomes of those who did not, which are confounded by effects of bariatric surgery, including malnutrition and metabolic syndrome, or comparing outcomes in patients with obesity to outcomes in patients who had a lower BMI. It is well-established that a greater BMI in TJA patients is associated with greater medical and surgical risks, particularly periprosthetic joint infection (21,22). Patients who had an elevated BMI should be informed of these risks when undergoing surgery at their current weight and should be strongly encouraged to reduce weight prior to TJA, if possible, to mitigate such risk; it is not clear, however, that postponing TJA for weight reduction improves outcomes. Additionally, pain and function improvements are similar for those who have a BMI of ≥35 compared to patients without obesity (37). Although weight reduction may be used as a criterion for TJA, the use of absolute BMI or rigid thresholds is discouraged. Not all patients have the necessary medical, financial, or social support and resources to effectively lose weight at all or within a suitable timeframe. In addition, patients in whom weight loss is unlikely and who would benefit markedly from the increased mobility afforded by TJA in improving their quality of life should have the information needed to engage in a shared decision-making process with their surgeon. The shared decision-making process educates the patient about their role in deciding among treatment options and helps them understand the expected outcomes and risks, including the increase in technical challenges for the surgeon, associated with TJA in patients with obesity. This process helps patients understand the pros and cons and make the decision that is right for them.

In our defined population with poorly controlled diabetes mellitus, we conditionally recommend delaying TJA to improve glycemic control.

It is well-established that patients who have poor glycemic control have an increased risk of poor outcomes after TJA (19). There is likely a benefit to delaying TJA to improve glycemic control; however, the optimal measure and optimal threshold of glycemic control to predict surgical outcomes is not known. Measures of glycemic control include, but are not limited to, glycosylated hemoglobin (HbA_{1c}), fructosamine, and fasting glucose. Thus, we do not recommend a specific measure or threshold, but recommend improved glycemic control overall. This recommendation is conditional because the evidence supporting the recommendation is indirect and very low quality.

In our defined population with nicotine dependence, we conditionally recommend delaying TJA for nicotine use reduction/cessation.

Nicotine use is associated with both increased medical and surgical risks in TJA (18,20). Similar to BMI and poor glycemic control, patients should be educated on these risks and counseled to modify the risk prior to TJA through nicotine use reduction or cessation. In addition, patients should be provided resources to assist with their nicotine use reduction or cessation. For these patients presenting with nicotine dependence, there is a potential benefit of delaying TJA for nicotine use reduction or cessation. This recommendation is conditional because there are exceptions, and the quality of evidence supporting the recommendation is low. The decision to proceed with TJA should not be contingent on complete nicotine cessation. Instead, the patient should be educated about the increased surgical risks associated with nicotine use and ideally engage in nicotine-reduction strategies.

In our defined population with bone loss with deformity or severe ligamentous instability, we conditionally recommend proceeding to TJA without delay over delaying TJA for optimization of non–life-threatening conditions.

There was no evidence for this recommendation; thus, the recommendation is based on clinician and patient opinion and experiences. In these patients, delaying TJA may lead to increased instability and increased juxtaarticular bone loss or deformity, which may increase the technical difficulty of the procedure as well as increase the risk of failure and need for revision. Although patients who have severe bone loss, deformity, or

instability have an increased risk of revision or reoperation, this risk will likely only increase over time, with further delay in surgery. Thus, timely TJA should be performed in these cases when medically appropriate. This recommendation is conditional because of the very low quality of evidence.

In our defined population with a neuropathic joint, we conditionally recommend proceeding to TJA without delay over delaying for optimization of non–life-threatening conditions.

There was no evidence for this recommendation; thus the recommendation is based on clinician and patient opinion and experiences. Patients who have neuropathic joints in the early stages of their disease may not have major pain or loss of function but may have severe joint destruction. As the disease progresses, patients develop pain, and the extent of bone loss and joint destruction worsens. These procedures are more technically challenging and often necessitate the use of more constrained implants typically reserved for revision arthroplasty. Proceeding with operative treatment in these cases is recommended because delaying surgery increases the technical difficulty of the procedure and does not improve outcomes after the procedure (38). The recommendation is conditional because of the very low quality of the evidence in addition to the rare exceptions that may apply. The exceptions listed for the first recommendation above, including delay for personal reasons or other ongoing medical comorbidities, apply to this recommendation as well. In addition, there may be a benefit to delaying TJA in patients whose underlying condition associated with the neuropathic joint is not known to allow for further diagnostic workup.

DISCUSSION

This guideline provides evidence-based recommendations regarding the optimal timing of elective TJA in patients who have symptomatic moderate-to-severe OA or advanced symptomatic ON with secondary arthritis who have chosen to undergo surgical treatment after a shared-decision making process with their physician after nonoperative therapy has lost efficacy. Further recommendations regarding the timing of TJA in patients with specific medical comorbidities and risk factors are also provided. The evidence for each PICO question was very low quality except for physical therapy and nicotine cessation, which had low quality of evidence, primarily due to indirectness, as the studies that would address our questions directly would compare results in patients randomized to immediate arthroplasty versus those delayed for the proposed intervention. We included observational studies but acknowledge that they describe associations of outcomes with the conditions of interest and were rated down for risk of bias, imprecision, as well as indirectness. No recommendations were supported by high or moderate quality evidence.

There are many existing appropriateness criteria, insurance coverage determination policies, and other guidelines that comment on the indications for elective TJA (17,39–47). After the patient elects to proceed with TJA, third parties evaluate the medical necessity of the procedure using these criteria (15–22). These guidelines and policies focus on the general diagnosis of OA or ON with secondary arthritis and prompt a dichotomous choice of nonoperative versus operative treatment. Coverage determination policies are utilized by insurance companies to determine if patients have met the policy indications for TJA and are cited to delay surgical treatment in favor of continued nonoperative management or medical risk-factor modification prior to surgery. However, coverage determination policies are often not based on evidence studying patients in our defined population with symptomatic moderate-to-severe radiographic OA or advanced symptomatic ON with secondary arthritis who have passed the threshold for TJA indication after a shared decision-making process with their physician. A prior review of the literature cited by 4 of the major commercial payers' coverage determination policies found that <10% of the literature cited in these policies discussed the effectiveness of nonoperative treatments specifically in our defined population of patients who have moderate-to-severe OA or advanced symptomatic ON with secondary arthritis who were indicated for TJA (15). Coverage determination policies are further limited because they are rarely created from a formal systematic review process (15,16). In contrast, clinical practice guidelines are based on a formal systematic review of the current state of the scientific literature and provide evidence-supported, consensus-driven best practices for operative and nonoperative treatment of OA or ON with secondary arthritis of the hip and knee that may predict optimal outcomes (48). This guideline is the first to provide evidence-based recommendations developed from a systematic review on the efficacy of these nonoperative treatments in our defined population of patients indicated for elective TJA.

The Voting Panel recommended against delaying TJA in our defined population for additional nonoperative treatment including physical therapy, NSAIDs, braces or ambulatory aids, as well as intraarticular injections. Importantly, our defined population consists of patients who have moderate-to-severe symptomatic and radiographic OA or advanced symptomatic ON with secondary arthritis who already unsuccessfully tried a course of nonoperative treatment prior to indication for TJA. The results from this systematic review found that the efficacy of additional nonoperative treatments in these patients indicated for TJA is limited. However, it is not uncommon for patients to have their surgical procedure delayed by a third party for additional nonoperative treatment, creating a major barrier to care. In an 8-year follow-up study of 3,417 knees deemed appropriate for TKA, Ghomrawi et al found that only 9% underwent a timely TKA (defined as within 2 years of meeting appropriateness criteria) (6). In this cohort, 91% were considered potentially appropriate for TKA but delayed their

surgery. This delay in elective arthroplasty may lead to further pain and limitations in physical function and subsequently increased risk of disability and chronic disease (6,49). Patients who are indicated for TJA also prefer to proceed directly with surgical treatment. In a survey of 200 patients scheduled for TJA in a 3-month period, 93% stated they would not want to delay TJA for mandatory physical therapy (50). Our Patient Panel agreed. Both the Patient Panel and Voting Panel highlighted the clinical and economic value of timely TJA, which leads to improved pain, function, quality of life, and satisfaction for patients.

As noted, TJA is the only approved definitive therapy for moderate-to-severe symptomatic OA of the hip or knee, yet racial disparities in arthroplasty utilization have persisted for decades (51). Rigid cutoffs for BMI, HbA_{1c}, or smoking status could increase disparities in arthroplasty utilization by decreasing eligibility among vulnerable populations and those with lower household income or social status (52). Pooled data from 21,294 adults who were ≥50 years of age from the 1999–2014 National Health and Nutrition Examination Survey demonstrated that fewer non-Hispanic Black patients and those with lower household incomes would be eligible for TJA if the criteria were a BMI of <40, an HbA_{1c} level of <8%, or complete nicotine cessation (52).

The Patient Panel was instrumental in the development of this guideline and provided valuable insight into how best to apply these recommendations in the clinical setting. In particular, the Patient Panel stressed the importance of the shared decision-making process when indicating a patient for TJA. Each patient is unique in terms of their goals, preferences, risk tolerance, social support, socioeconomic status, medical and psychiatric comorbidities, and disease severity. It should be left to the shared decision-making process for the patient and their physician to determine whether and when to proceed with TJA. This shared decision-making process should comprehensively include a discussion of the unique risks and benefits of the procedure for the individual patient. Patients who have medical or surgical risk factors as described in this guideline should be counseled as to their increased risks, and preoperative attempts to modify these risk factors through efforts such as weight loss, glycemic control, or smoking cessation should be encouraged. However, both the Voting and Patient Panels did not support universal thresholds or inflexible cutoffs for these modifications (e.g., BMI or HbA_{1c}) because they limit access to care, particularly for racial and ethnic minority populations, and do not consider the unique medical, surgical, and social situation of each patient (53). Although lower BMI cutoffs and HbA_{1c} cutoffs may result in fewer complications in a small number of patients, the larger impact is increasingly limited access to complication-free THA and TKA for many more patients (52,54–58). This practice could result in increased health care disparities. In addition, the Patient Panel stressed the importance of providing patients with ample resources to assist with modifying their risk factors, recognizing that some patients have less access to resources than others to meet preoperative goals. The Voting Panel made all recommendations conditional because they also recognized that there are exceptions to these recommendations, such as a delay for personal reasons due to family or work obligations. It is important that these unique circumstances be considered during the shared decision-making process.

The major potential limitation to this guideline is the indirectness and low quality of the available evidence. Moderate- and high-quality studies addressing these PICO questions will be challenging to perform. Direct evidence for our questions would entail randomizing patients indicated for TJA to receive surgical treatment or delay for a trial of additional nonsurgical treatment such as physical therapy, intraarticular injections, or bracing, and then assessing long-term patient-important outcomes. Patients and surgeons may have concerns about participating in studies in which patients could be randomized to delayed surgery or not being offered resources for risk factor modification such as poor glycemic control. It may be difficult to complete studies with enough power to demonstrate the effectiveness of risk factor modification in part due to the relatively low rate of specific complications associated with TJA, even in high-risk patients. Nevertheless, future research is clearly needed in this distinct patient population. Quasi-experimental study designs may be more fit to answer some of these questions. Another limitation of the guideline is that we grouped several separate populations for our PICO questions (e.g., knee OA, hip OA, hip ON with secondary arthritis, and knee ON with secondary arthritis) based on a clinical consensus from the orthopedic surgeons and the rheumatologists on the Core Team and lack of knowledge of the proportion of cases of ON with secondary arthritis that would be included in the literature review, with an understanding that subgroups might be created if evidence of differences in clinical outcomes was found by joint type (hip versus knee) and pathology (OA versus ON with secondary arthritis). No such evidence was found, and therefore, these were treated as a group despite the clear heterogeneity of the populations. Additional cost to the patient and cost effectiveness of nonoperative treatments were considered when the recommendations were made, but these were made based on a priori assumptions because there was a lack of cost effectiveness data on these treatments in our defined population. We did not include specialists in nonoperative therapy, such as physical therapists, as this guideline is intended for those patients who had more advanced and symptomatic disease for whom nonoperative therapy is no longer helpful; however, absence of their perspective is recognized as a limitation. Also, our population is those patients who have attempted nonoperative treatment and for whom that treatment is no longer effective. The determination that a treatment is no longer effective is individualized as determined by a shared decision-making process between a patient and the physician.

We did not include patients who have rheumatoid arthritis (RA) in this guideline, which is a limitation but was beyond our

scope, as questions regarding the timing of surgery in patients who have RA prioritize medication management to decrease infection risk, which was the focus of the updated 2022 ACR/AAHKS Guideline for the Perioperative Management of Antirheumatic Medication in Patients With Rheumatic Diseases Undergoing Elective Total Hip or Total Knee Arthroplasty (23). Although patients who have RA may receive TJA for primary or secondary OA, their perioperative management is driven by their systemic inflammatory disease.

This guideline has several strengths. The recommendations were made and voted on by a multidisciplinary collaboration group of orthopedic surgeons, rheumatologists, and patients who have undergone or are scheduled to undergo elective TJA who provided their expertise and insights. In addition, the GRADE methodology is well validated and was utilized to make these consensus-based recommendations (24,25). In addition, there was high consensus for most of the recommendations, with over one-half of the recommendations unanimously agreed upon.

In conclusion, this guideline provides evidence-based recommendations regarding the optimal timing of elective TJA in patients who have symptomatic moderate-to-severe OA or advanced symptomatic ON with secondary arthritis for whom nonoperative treatment has been ineffective and who have chosen to undergo surgical treatment after a shared decision-making process with their physician. Further recommendations regarding the timing of TJA in patients who have specific medical comorbidities and risk factors are also provided. Through a systematic review process incorporating the insight, expertise, and experience of expert clinicians and patients, consensus recommendations were made based on the best available evidence for this specific cohort of patients. We acknowledge that the data supporting these recommendations are of low quality and hope that future research will allow for further refinement and strengthening of the recommendations for the benefit of patients who suffer from moderate-to-severe OA or ON with secondary arthritis.

ACKNOWLEDGMENTS

We thank Dr. Cara A. Cipriano for her review of the evidence. We thank the patients who (along with authors Phyllis Bass and Danielle Dass) participated in the Patient Panel meeting: Stephanie Dodunski, Lawrence "Rick" Phillips, Richard F. Seiden, Catherine Simons, Joni Sprung, and Madonna J. Weiss. We thank the ACR staff, including Regina Parker for assistance in coordinating the administrative aspects of the project, and Cindy Force for assistance with manuscript preparation. We thank Janet Waters for her assistance in developing the literature search strategy, as well as for performing the initial literature search and update searches.

AUTHOR CONTRIBUTIONS

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

Study conception and design. Hannon, Goodman, Austin, Yates, Aggarwal, Bass, Jevsevar, Lajam, Blevins, Courtney, Gausden, Russell, Turner, Singh.

Acquisition of data. Hannon, Goodman, Austin, Yates, Aggarwal, Bass, Dass, Ghomrawi, Jevsevar, Lajam, Meng, Bedard, Blevins, Cohen-Rosenblum, Courtney, Fernandez-Ruiz, Gausden, Ghosh, King, Meara, Mehta, Mirza, Sullivan, Turgunbaev, Wysham, Yip, Yue, Zywiel, Russell, Turner, Singh.

Analysis and interpretation of data. Hannon, Goodman, Austin, Yates, Guyatt, Baker, Bass, Bekele, Dass, Jevsevar, Kwoh, Lajam, Meng, Moreland, Suleiman, Wolfstadt, Bartosiak, Blevins, Cohen-Rosenblum, Fernandez-Ruiz, Gausden, Mehta, Rana, Sullivan, Turgunbaev, Yip, Yue, Zywiel, Russell, Turner, Singh.

REFERENCES

1. Abbott JH, Robertson MC, Chapple C, et al. Manual therapy, exercise therapy, or both, in addition to usual care, for osteoarthritis of the hip or knee: a randomized controlled trial. 1: clinical effectiveness. Osteoarthritis Cartilage 2013;21:525–34.
2. Kolasinski SL, Neogi T, Hochberg MC, et al. 2019 American College of Rheumatology/Arthritis Foundation guideline for the management of osteoarthritis of the hand, hip, and knee [review]. Arthritis Care Res (Hoboken) 2020;72:149–62.
3. Van Baar ME, Assendelft WJ, Dekker J, et al. Effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a systematic review of randomized clinical trials [review]. Arthritis Rheum 1999;42:1361–9.
4. Zacharias A, Green RA, Semciw AI, et al. Efficacy of rehabilitation programs for improving muscle strength in people with hip or knee osteoarthritis: a systematic review with meta-analysis [review]. Osteoarthritis Cartilage 2014;22:1752–73.
5. Daugberg L, Jakobsen T, Nielsen PT, et al. A projection of primary knee replacement in Denmark from 2020 to 2050. Acta Orthop 2021;92:448–51.
6. Ghomrawi HM, Mushlin AI, Kang R, et al. Examining timeliness of total knee replacement among patients with knee osteoarthritis in the U.S.: results from the OAI and MOST longitudinal cohorts. J Bone Joint Surg Am 2020;102:468–76.
7. Inacio MC, Paxton EW, Graves SE, et al. Projected increase in total knee arthroplasty in the United States: an alternative projection model. Osteoarthritis Cartilage 2017;25:1797–803.
8. Kurtz S, Ong K, Lau E, et al. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. J Bone Joint Surg Am 2007;89:780–5.
9. Kurtz SM, Lau E, Ong K, et al. Future young patient demand for primary and revision joint replacement: national projections from 2010 to 2030. Clin Orthop Relat Res 2009;467:2606–12.
10. Kurtz SM, Ong KL, Lau E, et al. Impact of the economic downturn on total joint replacement demand in the United States: updated projections to 2021. J Bone Joint Surg Am 2014;96:624–30.
11. Maradit Kremers H, Larson DR, Crowson CS, et al. Prevalence of total hip and knee replacement in the United States. J Bone Joint Surg Am 2015;97:1386–97.
12. Koenig L, Zhang Q, Austin MS, et al. Estimating the societal benefits of THA after accounting for work status and productivity: a Markov model approach. Clin Orthop Relat Res 2016;474:2645–54.
13. Pivec R, Johnson A, Mears S, et al. Hip Arthroplasty [review]. Lancet 2012;380:1768–77.
14. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957;16:494–502.

15. Austin MS, Ashley BS, Bedard NA, et al. What is the level of evidence substantiating commercial payers' coverage policies for total joint arthroplasty? J Arthroplasty 2021;36:2665–2673.e8.
16. Chen AF, Bloomfield MR, Lichstein PM, et al. What is the level of evidence substantiating the Medicare local coverage determinations? J Arthroplasty 2015;30:356–60.
17. American Academy of Orthopaedic Surgeons. Osteoarthritis of the knee: clinical practice guideline on management of osteoarthritis of the knee (3rd edition). URL: https://www.aaos.org/quality/quality-programs/lower-extremity-programs/osteoarthritis-of-the-knee/.
18. Duchman KR, Gao Y, Pugely AJ, et al. The effect of smoking on short-term complications following total hip and knee arthroplasty. J Bone Joint Surg Am 2015;97:1049–58.
19. Na A, Middleton A, Haas A, et al. Impact of diabetes on 90-day episodes of care after elective total joint arthroplasty among Medicare beneficiaries. J Bone Joint Surg Am 2020;102:2157–65.
20. Tischler EH, Matsen Ko L, Chen AF, et al. Smoking increases the rate of reoperation for infection within 90 days after primary total joint arthroplasty. J Bone Joint Surg Am 2017;99:295–304.
21. Wagner ER, Kamath AF, Fruth K, et al. Effect of body mass index on reoperation and complications after total knee arthroplasty. J Bone Joint Surg Am 2016;98:2052–60.
22. Wagner ER, Kamath AF, Fruth KM, et al. Effect of body mass index on complications and reoperations after total hip arthroplasty. J Bone Joint Surg Am 2016;98:169–79.
23. Goodman SM, Springer BD, Chen AF, et al. 2022 American College of Rheumatology/American Association of Hip and Knee Surgeons guideline for the perioperative management of antirheumatic medication in patients with rheumatic diseases undergoing elective total hip or total knee arthroplasty. Arthritis Rheumatol 2022;74:1464–73.
24. Andrews JC, Schünemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. J Clin Epidemiol 2013;66:726–35.
25. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924–6.
26. Busse J, Gasteiger W, Tönnis D. A new method for roentgenologic evaluation of the hip joint: the hip factor. Arch Orthop Unfallchir 1972;72:1–9. In German.
27. Gränicher P, Mulder L, Lenssen T, et al. Prehabilitation improves knee functioning before and within the first year after total knee arthroplasty: a systematic review with meta-analysis [review]. J Orthop Sports Phys Ther 2022;52:709–25.
28. Skou ST, Roos EM, Laursen MB, et al. Total knee replacement and non-surgical treatment of knee osteoarthritis: 2-year outcome from two parallel randomized controlled trials. Osteoarthritis Cartilage 2018;26:1170–80.
29. Simon LS, Hochberg MC. Non-steroidal anti-inflammatory drugs. In: Doherty M, Hunter DJ, Bijlsma H, et al, eds. Oxford textbook of osteoarthritis and crystal arthropathy. 3rd ed. Oxford (UK): Oxford University Press; 2016.
30. Nin DZ, Chen YW, Talmo CT, et al. Drivers of unequal healthcare costs in the nonoperative treatment of late-stage knee osteoarthritis prior to primary total knee arthroplasty. J Arthroplasty 2022;37:1967–72.e1.
31. Alkjaer T, Larsen PK, Pedersen G, et al. Biomechanical analysis of rollator walking. Biomed Eng Online 2006;5:2.
32. Mahoney J, Euhardy R, Carnes M. A comparison of a two-wheeled walker and a three-wheeled walker in a geriatric population. J Am Geriatr Soc 1992;40:208–12.
33. Habib G, Chernin M, Sakas F, et al. The Impact of intra-articular depot betamethasone injection on insulin resistance among diabetic patients with osteoarthritis of the knee: a case-control study. J Clin Rheumatol 2018;24:193–6.
34. Lai Q, Cai K, Lin T, et al. Prior intra-articular corticosteroid injection within 3 months may increase the risk of deep infection in subsequent joint arthroplasty: a meta-analysis. Clin Orthop Relat Res 2022;480:971–9.
35. Pereira TV, Jüni P, Saadat P, et al. Viscosupplementation for knee osteoarthritis: systematic review and meta-analysis. BMJ 2022;378:e069722.
36. Rutjes AW, Jüni P, da Costa BR, et al. Viscosupplementation for osteoarthritis of the knee: a systematic review and meta-analysis [review]. Ann Intern Med 2012;157:180–91.
37. Collins JE, Donnell-Fink LA, Yang HY, et al. Effect of obesity on pain and functional recovery following total knee arthroplasty. J Bone Joint Surg Am 2017;99:1812–8.
38. Tibbo ME, Chalmers BP, Berry DJ, et al. Primary total knee arthroplasty in patients with neuropathic (Charcot) arthropathy: contemporary results. J Arthroplasty 2018;33:2815–20.
39. American Academy of Orthopaedic Surgeons. Appropriate use criteria. URL: https://www.orthoguidelines.org/go/auc/.
40. Centers for Medicare & Medicaid Services. Local Coverage Determination (LCD): major joint replacement (hip and knee) (L33618). URL: https://www.cms.gov/medicare-coverage-database/view/lcd.aspx?lcdId=33618&ver=23.
41. Clearinghouse NG. Osteoarthritis. Care and management in adults. URL: http://www.guideline.gov/content.aspx?id=47862.
42. American Academy of Orthopaedic Surgeons. Surgical management of osteoarthritis of the knee. URL: https://www.aaos.org/quality/quality-programs/lower-extremity-programs/surgical-management-of-osteoarthritis-of-the-knee/.
43. Knee arthroplasty and hip arthroplasty. In: Inpatient and surgical care. Milliman Care Guidelines LLC; 2011.
44. Total joint replacement, knee and hip & removal and replacement, total joint replacement knee and hip. In: InterQual® Procedures Adult Criteria. McKesson Corporation; 2011.
45. American Academy of Orthopaedic Surgeons. Management of osteoarthritis of the hip: evidence-based clinical practice guideline. 2017. URL: https://www.aaos.org/globalassets/quality-and-practice-resources/osteoarthritis-of-the-hip/oa-hip-cpg_6-11-19.pdf.
46. Escobar A, Quintana JM, Aróstegui I, et al. Development of explicit criteria for total knee replacement. Int J Technol Assess Health Care 2003;19:57–70.
47. Riddle DL, Jiranek WA, Hayes CW. Use of a validated algorithm to judge the appropriateness of total knee arthroplasty in the United States: a multicenter longitudinal cohort study. Arthritis Rheumatol 2014;66:2134–43.
48. Riddle DL, Perera RA, Jiranek WA, et al. Using surgical appropriateness criteria to examine outcomes of total knee arthroplasty in a United States sample. Arthritis Care Res (Hoboken) 2015;67:349–57.
49. Turkiewicz A, Kiadaliri AA, Englund M. Cause-specific mortality in osteoarthritis of peripheral joints. Osteoarthritis Cartilage 2019;27:848–54.
50. Grace TR, Eralp I, Khan IA, et al. Are patients with end-stage arthritis willing to delay arthroplasty for payer-mandated physical therapy? J Arthroplasty 2022;37:S27–31.
51. Singh JA, Lu X, Rosenthal GE, et al. Racial disparities in knee and hip total joint arthroplasty: an 18-year analysis of national Medicare data. Ann Rheum Dis 2014;73:2107–15.
52. Wang AY, Wong MS, Humbyrd CJ. Eligibility criteria for lower extremity joint replacement may worsen racial and socioeconomic disparities. Clin Orthop Relat Res 2018;476:2301–8.
53. Dlott CC, Metcalfe T, Jain S, et al. Preoperative risk management programs at the top 50 orthopaedic institutions frequently enforce strict

cutoffs for BMI and hemoglobin A1c which may limit access to total joint arthroplasty and provide limited resources for smoking cessation and dental care. Clin Orthop Relat Res 2023;481:39–47.

54. Arnold N, Anis H, Barsoum WK, et al. Preoperative cut-off values for body mass index deny patients clinically significant improvements in patient-reported outcomes after total hip arthroplasty. Bone Joint J 2020;102-B:683–92.
55. Carender CN, DeMik DE, Elkins JM, et al. Are body mass index cutoffs creating racial, ethnic, and gender disparities in eligibility for primary total hip and knee arthroplasty? J Arthroplasty 2022;37:1009–16.
56. DeMik DE, Kohler JG, Carender CN, et al. What is the impact of body mass index cutoffs on total hip arthroplasty complications? J Arthroplasty 2022;37:1320–25.e1.
57. DeMik DE, Muffly SA, Carender CN, et al. What is the impact of body mass index cutoffs on total knee arthroplasty complications? J Arthroplasty 2022;37:683–7.e1.
58. Roth A, Anis HK, Emara AK, et al. The potential effects of imposing a body mass index threshold on patient-reported outcomes after total knee arthroplasty. J Arthroplasty 2021;36:S198–208.

Arthritis Care & Research
Vol. 75, No. 11, November 2023, pp 2239–2241
DOI 10.1002/acr.25202

American College of Rheumatology
Empowering Rheumatology Professionals

EDITORIAL

The Devil Is in the Detail: Clinical Practice Guideline for the Optimal Timing of Elective Hip or Knee Arthroplasty

Gillian A. Hawker

What is the optimal timing for hip or knee total joint arthroplasty (TJA) in people with symptomatic and moderate-to-severe radiographic osteoarthritis (OA) or osteonecrosis (ON) who have not responded to nonoperative treatments? The new clinical practice guidelines by the American College of Rheumatology (ACR) and American Academy of Hip and Knee Surgeons (AAHKS) (1), published in this issue of *Arthritis Care & Research*, sought to provide evidence-based recommendations to address this question.

This is an important topic. With rising prevalence of symptomatic knee OA, including at younger ages, the rates and associated costs of TJA are skyrocketing (2). Policies that restrict access to TJA have emerged to combat rising TJA costs, ideally while also ensuring access and quality of care. The new recommendations aim to inform shared decision-making (SDM) between patients and clinicians regarding TJA while also challenging health insurance coverage policies that restrict timely access to TJA. The conditional recommendations are based on low or very low-quality evidence. Thus, they provide little substance to guide TJA shared decision-making, other than underscoring the importance of SDM in the context of TJA. However, they do serve to highlight the tremendous burden on patients imposed by advanced, symptomatic hip and knee arthritis and the injustice of imposing undue delays to receipt of TJA to candidates deemed "appropriate" for surgery.

In essence, the guideline recommends that people who are deemed "appropriate" for TJA at surgical consultation should not have their surgery delayed for additional trials of nonoperative treatments, e.g., physical therapy, nonsteroidal antiinflammatory drugs, ambulatory aids, and intraarticular injections. At face value, this seems reasonable. But, as is often the case, the devil is in the details. Whether or not these recommendations will shift coverage policy makers' stance is unclear for the following reasons:

First, the authors are careful to state, and restate, that their recommendations do not apply to patients who are not appropriate for TJA (e.g., those with mild radiographic OA or ON, with minimal pain or disability, or who have not tried some form of nonoperative therapy). However, a widely accepted and standardized approach to determining patient appropriateness for TJA does not currently exist (3).

Appropriate care is broadly defined as that which provides net benefit to the patient (4,5). There is general consensus that TJA is appropriate in people who have a demonstrable need for surgery (moderate-to-severe OA based on symptoms and imaging who have tried and had inadequate response to nonsurgical therapies), are fit for surgery (potential surgical benefits outweigh risks), are ready and willing to undergo TJA, and who have realistic surgical expectations (3,5–8). However, research suggests that these criteria are not consistently considered or assessed (6). In our recently published cohort study of 1,273 people with knee OA undergoing primary, elective TKA, only 62% had received formal or informal exercise/physical activity interventions, WOMAC OA Index and Knee Injury and Osteoarthritis Outcome Score Physical Function Shortform scores covered the full range, from very low to very high, one-fifth reported their knee symptoms as "acceptable," and one-third had Kellgren-Lawrence grades 1 or 2 on radiographs (5). At one year post-surgery, only 78% had experienced meaningful improvement in their knee symptoms and were satisfied with their surgical results. These findings are consistent with those of other studies (9–11) and indicate that there may be individuals undergoing TJA who do not need it, want it, and/or have a low probability of benefit from it. The guidelines on optimal timing of TJA do not address this issue and should not be conflated with recommendations for patient appropriateness for TJA.

Second, as noted above, the recommendations are conditional based on low or very low-quality evidence. Further, some recommendations, for example, improved glycemic control before operating in patients with diabetes mellitus, lack specific guidance regarding the target presurgical status that should be

Gillian A. Hawker, MD, MSc: University of Toronto, Toronto, Ontario, Canada.

Author disclosures are available online at https://onlinelibrary.wiley.com/doi/10.1002/acr.25202.

Address correspondence via email to Gillian A. Hawker, MD, MSc, at g.hawker@utoronto.ca.

Submitted for publication May 3, 2023; accepted in revised form May 18, 2023.

achieved. This makes operationalizing the recommendations challenging in clinical practice. The poor-quality evidence reflects a paucity of clinical trials that have compared TJA to other treatments or delayed surgery. Only one randomized controlled trial has been performed comparing TJA to a program comprised of recommended nonsurgical therapies (12). Both improvement in pain and function were greater at 1 year follow-up in the surgical group compared with the nonsurgical group, but the surgical group was also at higher risk for adverse events, including deep venous thrombosis and return to the operating room for manipulation under anesthesia. These findings are consistent with outcomes from many longitudinal observational cohort studies, which show that most TJA recipients experience substantial, sustained improvement in joint pain, mobility, and emotional status, but that about one in five TJA recipients will not, resulting in dissatisfaction with their results (13,14). Long wait times exist for TJA in many countries, but the data on the impact of waiting is mixed; some have shown progression of symptom severity while others have not (15,16). Thus, the impact of delayed surgery is unclear. That said, among people who are truly appropriate for surgery, it seems reasonable to ensure they receive it in a timely fashion.

How might these recommendations help patients?

Importantly, they place TJA decision-making firmly in the realm of the patient and their health care team. Given that "appropriate" care is that which provides the best possible outcome to the patient, our job as clinicians is to provide our patients the best available evidence to inform their treatment decisions, respecting their personal preferences, values, and circumstances, e.g., age, comorbidities, social support, and employment status (3). That is, through a process of SDM. SDM has been shown to improve patient satisfaction and adherence to therapy and may also reduce undesired care (17).

When nonsurgical therapies no longer adequately control patients' symptoms, impacting their quality of life, guidelines recommend that TJA be considered (18–20). But whether or not to consider TJA is one of the trickiest treatment decisions people with symptomatic knee OA must make as TJA is an elective, preference-sensitive procedure, performed to improve quality of life. In prior work, we have shown that the patient's willingness to undergo TJA is the strongest predictor of their subsequent receipt of surgery (21). TJA willingness reflected patients' perceptions of their candidacy for surgery, including perceived OA severity, coping efficacy, and perceived risks and benefits of surgery (21–23). These beliefs and perceptions are influenced by gender, race/ethnicity/religion, health literacy, social network factors, and socioeconomic status (21). In a recently published study, we showed that preoperative measures of TKA readiness, willingness, and TJA expectations significantly enhanced ability to discriminate those who did versus did not go on to experience a good TKA outcome compared with measures of knee pain and disability alone (5). Thus, consistent with current TJA clinical practice guidelines (18,19,24), decisions regarding TJA, including appropriateness and optimal timing, should be determined through SDM. Decision aids have been developed to support patient decision-making regarding TJA (25), but none currently exist to support patient-surgeon SDM regarding TJA.

It's time to bridge this gap

In prior qualitative research, we showed that hospital administrators want greater transparency regarding patient selection for TJA (26), but agreed that clinical decisions should be left to the patient and health care professionals. Arthroplasty surgeons agreed that a decision support tool would be useful to inform referral for TJA, identification of appropriate candidates for TJA, and to enhance their accountability to the system (27). It's time to develop, implement, and evaluate such a tool in clinical practice. An implementation science approach, whereby all stakeholders come together to understand barriers and facilitators to SDM in the context of TJA decision-making and to develop and test the feasibility and effectiveness of strategies to overcome identified barriers, is needed.

The guideline presented in this issue on optimal timing of TJA applies to the care of patients who are "appropriate" for TJA. Until we can demonstrate to payors and policy makers that those being recommended for surgery are appropriate, policies that restrict access to TJA are likely to persist. To ensure timely access to this highly effective procedure for those who need it, want it, and have high likelihood of benefiting from it, we need to work together to put in place a standardized approach to assessment of patient appropriateness for surgery and ensure access to all effective nonsurgical therapies to all people living with symptomatic knee OA or ON.

Summary of main points:

- The ACR recommendations apply to people who are "appropriate for surgery," yet no standardized approach exists to establish patient appropriateness.
- The recommendations are based on low-quality evidence, limiting their usefulness in clinical practice, and the evidence against surgical delay is mixed.
- To advance timely access to TJA, strategies are required to improve the use of SDM for TJA in surgical practice. Enhanced SDM has potential to reduce undesired care, improve patient satisfaction, and address policy makers' concerns regarding TJA overuse.

AUTHOR CONTRIBUTIONS

Dr. Hawker drafted the article, revised it critically for important intellectual content, and approved the final version to be published.

REFERENCES

1. Hannon CP, Goodman SM, Austin MS, et al. 2023 American College of Rheumatology (ACR) and American Association of Hip and Knee Surgeons (AAHKS) clinical practice guideline for the optimal timing of elective hip or knee arthroplasty for patients with symptomatic moderate to severe osteoarthritis or osteonecrosis who have failed nonoperative therapy. Arthritis Care Res (Hoboken) 2023;75:2227–38.
2. Sloan M, Premkumar A, Sheth NP. Projected volume of primary total joint arthroplasty in the U.S., 2014 to 2030. J Bone Joint Surg Am 2018;100:1455–60.
3. Pacheco-Brousseau L, Stacey D, Desmeules F, et al. Instruments to assess appropriateness of hip and knee arthroplasty: a systematic review [review]. Osteoarthritis Cartilage 2023;31:874–64.
4. Sanmartin C, Murphy K, Choptain N, et al. Appropriateness of healthcare interventions: concepts and scoping of the published literature [review]. Int J Technol Assess Health Care 2008;24:342–9.
5. Hawker GA, Bohm E, Dunbar MJ, et al. Patient appropriateness for total knee arthroplasty and predicted probability of a good outcome. RMD Open 2023;9:e002808.
6. Frankel L, Sanmartin C, Conner-Spady B, et al. Osteoarthritis patients' perceptions of "appropriateness" for total joint replacement surgery. Osteoarthritis Cartilage 2012;20:967–73.
7. Escobar Martinez A, Perera RA, Riddle DL. Development and underlying structure of a second-generation appropriateness classification system for total knee arthroplasty. Arthritis Care Res (Hoboken) 2021;73:801–9.
8. American Academy of Orthopaedic Surgeons. Management of osteoarthritis of the knee appropriate use criteria. 2022. URL: https://www.aaos.org/oakauc.
9. Ghomrawi HM, Alexiades M, Pavlov H, et al. Evaluation of two appropriateness criteria for total knee replacement. Arthritis Care Res (Hoboken) 2014;66:1749–53.
10. Escobar A, Bilbao A, Bertrand ML, et al. Validation of a second-generation appropriateness classification system for total knee arthroplasty: a prospective cohort study. J Orthop Surg Res 2021;16:227.
11. Karunaratne S, Harris IA, Trevena L, et al. Observing the use of knee arthroplasty appropriateness tools in clinical practice: do appropriateness criteria tools predict surgeon decision-making? Osteoarthritis Cartilage 2021;29:1275–81.
12. Skou ST, Roos EM, Laursen MB, et al. A randomized, controlled trial of total knee replacement. N Engl J Med 2015;373:1597–606.
13. Beswick AD, Wylde V, Gooberman-Hill R, et al. What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients. BMJ Open 2012;2:e000435.
14. Most J, Hoelen TA, Spekenbrink-Spooren A, et al. Defining clinically meaningful thresholds for patient-reported outcomes in knee arthroplasty. J Arthroplasty 2022;37:837–44.e3.
15. Ho KW, Pong G, Poon WC, et al. Progression of health-related quality of life of patients waiting for total knee arthroplasty. J Eval Clin Pract 2021;27:69–74.
16. Hoogeboom TJ, van den Ende CH, van der Sluis G, et al. The impact of waiting for total joint replacement on pain and functional status: a systematic review [review]. Osteoarthritis Cartilage 2009;17:1420–7.
17. Bomhof-Roordink H, Gärtner FR, Stiggelbout AM, et al. Key components of shared decision making models: a systematic review. BMJ Open 2019;9:e031763.
18. 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 [review]. Arthritis Care Res (Hoboken) 2020;72:149–62.
19. National Institute for Health and Care Excellence. Osteoarthritis in over 16s: diagnosis and management. NICE guideline [NG226]. URL: www.nice.org.uk/guidance/ng226.
20. Bannuru RR, Osani MC, Vaysbrot EE, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis [review]. Osteoarthritis Cartilage 2019;27:1578–89.
21. Hawker GA, Guan J, Croxford R, et al. A prospective population-based study of the predictors of undergoing total joint arthroplasty. Arthritis Rheum 2006;54:3212–20.
22. Hawker GA, Wright JG, Badley EM, et al. Perceptions of, and willingness to consider, total joint arthroplasty in a population-based cohort of individuals with disabling hip and knee arthritis. Arthritis Rheum 2004;51:635–41.
23. Hawker GA. Who, when, and why total joint replacement surgery? The patient's perspective [review]. Curr Opin Rheumatol 2006;18:526–30.
24. Singh JA, Dowsey MM, Dohm M, et al. Achieving consensus on total joint replacement trial outcome reporting using the OMERACT filter: endorsement of the final core domain set for total hip and total knee replacement trials for endstage arthritis. J Rheumatol 2017;44:1723–6.
25. Bansback N, Trenaman L, MacDonald KV, et al. An online individualised patient decision aid improves the quality of decisions in patients considering total knee arthroplasty in routine care: a randomized controlled trial. Osteoarthr Cartil Open 2022;4:100286.
26. Clavel N, De Coster C, Pomey MP, et al. Appropriateness for total joint replacement: perspectives of decision-makers. Healthc Policy 2016;11:80–92.
27. Frankel L, Sanmartin C, Hawker G, et al. Perspectives of orthopaedic surgeons on patients' appropriateness for total joint arthroplasty: a qualitative study. J Eval Clin Pract 2016;22:164–70.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 75, No. 11, November 2023, pp 2242–2244
DOI 10.1002/acr.25161

EDITORIAL

Advancing Rheumatology Through Medical Education Research

Eli M. Miloslavsky[1] and Beth L. Jonas[2]

Medical education research has undergone tremendous advances over the past half century. This field of inquiry has evolved, from evaluating curricular innovations at a single center with outcomes centering on learner perception to using both quantitative and qualitative methods to study topics ranging from the learning environment to patient outcomes resulting from educational interventions (1). In this issue of *Arthritis Care & Research*, Leverenz and colleagues (2) describe the development and evaluation of RheumMadness, an online educational intervention designed to educate and engage several levels of learners across multiple institutions. Modeled on the successful NephMadness concept developed in nephrology, Leverenz et al created a tournament of rheumatology concepts where recent advances in rheumatology (e.g., avacopan in antineutrophil cytoplasmic antibody–associated vasculitis) compete among each other via a voting system. Their work utilizes multiple advances in medical education, including collaborative learning and constructivist activities (rheumatology fellows worked in teams to create "scouting reports" which educated participants about each concept), gamification (using a tournament with winners to engage learners), and the use of social media (to enhance discussion about the concepts being taught and connect learners across institutions). Theory-based evaluation of RheumMadness using the Community of Inquiry framework further enhances their contribution by providing a nuanced view of its impact and future steps for improvement. This thoughtfully designed, theory-based project overcomes the barrier of testing curricula in small training programs and serves as one example of educational scholarship progress in our field.

With progress also comes the promise of medical education research in helping to overcome some of the biggest challenges facing rheumatology today. This potential has long been recognized within rheumatology, as it was one of the first medicine subspecialties to support education scholars through the Rheumatology Research Foundation's Clinician Scholar Educator Award, which began in 1999 (3). Herein we highlight the critical role we anticipate medical education will have in impacting areas important to rheumatology such as workforce training and retention, transitions across career phases, interprofessional patient care, and diversity, equity, and inclusion.

Advances in the science of rheumatology over the last decades have led to the discovery of innovative diagnostic tools and exceptionally effective therapies for many of our complex diseases. The American College of Rheumatology (ACR) 2015 workforce study revealed a significant projected deficit of both adult and pediatric rheumatology providers in the US (4); one that is likely to grow due to increasing demands for rheumatology expertise, the changing demographics of the workforce, geographic maldistribution of professionals, and accelerated attrition due to the modern-day challenges of medical practice. Developing a robust and competent workforce skilled in the rapidly evolving science and practice of rheumatology is critical for the care of a growing population of patients with rheumatic and musculoskeletal diseases. Our learners must be knowledgeable, collaborative, adaptive, and comfortable with the uncertainty that surrounds many of our diseases. Developing the next generation of rheumatology providers falls squarely with our academic clinician educators.

Work is ongoing to increase the number and geographic distribution of training slots, but we cannot underestimate the impact and value of the quality of rheumatology instruction and assessment in workforce development. Medical education scholarship in rheumatology is a critical tool for the development of a well-trained rheumatology workforce, enhancing both the skills of the faculty and the trainees. Advances in the scholarship of teaching and assessment have led to new and innovative evidence-based strategies including Competency Based Medical Education and the Milestones Competency Assessment tool, which have been adapted for rheumatology training by rheumatology clinical scholars (5,6). These tools are critical in assessing trainee growth,

[1]Eli M. Miloslavsky, MD: Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts; [2]Beth L. Jonas, MD: University of North Carolina School of Medicine, Chapel Hill

Author disclosures are available online at https://onlinelibrary.wiley.com/doi/10.1002/acr.25161.

Address correspondence via email to Eli M. Miloslavsky, MD, at Emiloslavsky@mgh.harvard.edu.

Submitted for publication April 25, 2023; accepted in revised form May 18, 2023.

providing feedback, and ensuring that fellowship programs graduate competent and effective physicians. Active learning strategies (problem-based learning, team-based learning, flipped classroom, and other modalities) have been incorporated into undergraduate medical education (UME), but less so in graduate medical education (GME). In a meta-analysis of 225 studies comparing student performance in undergraduate science, technology, engineering, and mathematics (STEM) courses under traditional lecturing versus active learning, active learning was associated with increased examination performance by about a half of a standard deviation and lecturing was associated with an increased failure rate by 55% (7). Yet, adoption of active learning has been very slow even among faculty who are knowledgeable about the benefits, citing lack of time and adequate training as primary barriers to change (8). The challenge for rheumatology educational scholars is to develop the professional skills necessary to support active learning, effective feedback, and deliberate practice and to disseminate this knowledge to clinical teaching faculty.

Transitions between career phases (from student to resident to fellow to attending) have been a major challenge for most physicians, likely impacting critical outcomes such as knowledge and skill acquisition, professional identity formation, growth mindset and burnout. Notably, these attributes are also important for retaining practitioners in our field. Recent focus on such transitions has identified many barriers in our education system including use of different evaluation instruments in UME and GME, bias and variable reliability and accuracy of faculty evaluations, and lack of transparency in communicating trainee progress between training phases, among others (9). Recent recommendations from the Coalition for Physician Accountability serve as an important first step to address the transitions challenge (10). The recommendations include developing a common framework of competencies across UME and GME, reducing bias in assessment, and promoting coaching across the medical education continuum that supports life-long learning and professional identity formation. Medical education research will play a critical role in supporting this process. Developing robust assessment methods requires an understanding of education theory, assessment frameworks, and psychometrics. A nuanced understanding of how professional identity is formed and how the learning environment, mentoring, coaching, curricula, and assessment influence its development will help determine the best methods to support the development of physicians who can best serve our society (11). Finally, elucidating the complex interplay between professional identity formation, training structure, resilience, and burnout will be critical to retaining and sustaining a robust and productive workforce.

While GME programs are focused on training rheumatology physicians, there are shared challenges across the spectrum of rheumatic disease professionals in recruiting, training, and supporting health professionals in the rheumatology workforce. The practice of rheumatology relies on interdisciplinary teams for the most effective care with contributions of physicians, advanced practice providers, nurses, occupational and physical therapists, and other professionals. National (Institute of Medicine) and international (World Health Organization) bodies have called for the use of interprofessional education to promote collaborative learning in preparation for interdisciplinary care in practice (12,13). There is growing evidence that effective interprofessional care improves the quality of care and patient outcomes by improving learners' attitudes, knowledge, and skills in the health care environment (14). Much like challenges to implementing active learning strategies, while there is empirical evidence that interprofessional education improves outcomes, there is much work to do preparing our educational workforce to embrace and implement these strategies. Medical education research on effective interdisciplinary musculoskeletal education is one example from the field of rheumatology education that can lead to more effective interprofessional care (15).

Promotion of diversity, equity, and inclusion is critical to realizing better care for our patients. For example, race concordant care has been associated with improved health care outcomes (16); however, in 2019, only 5.2% of US physicians identified as Black and 6.9% as Hispanic (17). Within adult rheumatology, physicians underrepresented in medicine may be even fewer, with only 0.8% of respondents to the 2015 ACR Workforce Survey identifying as Black and 8.5% as Hispanic (3). Notably, diversity among rheumatology fellows has also lagged behind other medicine subspecialties (18). Addressing the lack of workforce diversity is one way to reduce disparities in care access and quality.

Medical education research is a critical tool in recognizing the barriers to advancing diversity in our workforce and understanding the impact of interventions. While training and retaining a more diverse workforce has long been recognized as critical, addressing racial bias during training is also important to enhancing the diversity of our workforce. For example, race and sex differences have been described in medical school clerkship evaluations and grading (19). Racial bias likely extends to United States Medical Licensing Examination testing and Alpha Omega Alpha membership, two factors critical in the residency match process (20). Moreover, mistreatment has been demonstrated to be common during medical school and has been associated with lower empathy, higher burnout, and career regret (21). Interventions targeting teaching methods, assessment, and the learning environment are likely to play a key role in overcoming these barriers. Studying the efficacy of promising interventions, such as implicit bias training, inclusion of structural competency in training curricula, understanding effective teaming, and developing interventions to reduce compassion fatigue, hold the key to achieving diversity goals in our field.

Recognizing and supporting medical education scholarship holds promise in addressing some of the most pressing challenges facing rheumatology today. The work by Leverenz et al, as well as other education researchers in our field, represent the

progress and the potential of this discipline. However, just like all other science research, effective medical education research requires the acquisition of research skills, dedicated time, and effective mentorship. Investment in these areas, as the Rheumatology Research Foundation did with the development of the Clinician Scholar Educator Award, will pay dividends for providers, patients, and society.

AUTHOR CONTRIBUTIONS

Both authors drafted the article, revised it critically for important intellectual content, approved the final version to be published, and take responsibility for the integrity of the data and the accuracy of the data analysis.

REFERENCES

1. Norman G. Fifty years of medical education research: waves of migration [review]. Med Educ 2011;45:785–91.
2. Leverenz DL, Garren AU, Katz G, et al. RheumMadness: Creating an online community of inquiry in rheumatology. Arthritis Care Res (Hoboken) 2023;75:2248–58.
3. Berman JR, O'Rourke KS, Kolasinski SL, et al. Rheumatology Research Foundation clinician scholar educator award: fifteen years promoting rheumatology educators and education. Arthritis Care Res (Hoboken) 2016;68:1591–7.
4. Battafarano DF, Ditmyer M, Bolster MB, et al. 2015 American College of Rheumatology workforce study: supply and demand projections of adult rheumatology workforce, 2015–2030. Arthritis Care Res (Hoboken) 2018;70:617–26.
5. Carraccio C, Englander R, Van Melle E, et al. Advancing competency-based medical education: a charter for clinician-educators. Acad Med 2016;91:645–9.
6. Liebowitz JE, Torralba KD, Kolfenbach J, et al. Rheumatology milestones 2.0: a roadmap for competency-based medical training of rheumatology fellows in the 21st century. Arthritis Care Res (Hoboken) 2022;74:675–85.
7. Freeman S, Eddy SL, McDonough M, et al. Active learning increases student performance in science, engineering, and mathematics. Proc Natl Acad Sci U S A 2014;111:8410–5.
8. Bucklin BA, Asdigian NL, Hawkins JL, et al. Making it stick: use of active learning strategies in continuing medical education. BMC Med Educ 2021;21:44.
9. Miloslavsky EM, Dua AB. The transition from residency to fellowship: enhancing training by increasing transparency. Arthritis Rheumatol 2022;74:1625–7.
10. Coalition for Physician Accountability. Initial summary report and preliminary recommendations of the Undergraduate Medical Education to Graduate Medical Education Review Committee (UGRC). 2021. URL: https://physicianaccountability.org/wp-content/uploads/2021/04/UGRC-Initial-Summary-Report-and-Preliminary-Recommendations-1.pdf.
11. Toh RQ, Koh KK, Lua JK, et al. The role of mentoring, supervision, coaching, teaching and instruction on professional identity formation: a systematic scoping review. BMC Med Educ 2022;22:531.
12. Institute of Medicine, Board on Global Health, Committee on Measuring the Impact of Interprofessional Education on Collaborative Practice and Patient Outcomes. Measuring the impact of interprofessional education on collaborative practice and patient outcomes. Washington (DC): National Academies Press; 2015.
13. World Health Organization. Framework for action on interprofessional education and collaborative practice. 2010. URL: https://www.who.int/publications/i/item/framework-for-action-on-interprofessional-education-collaborative-practice.
14. Reeves S, Fletcher S, Barr H, et al. A BEME systematic review of the effects of interprofessional education: BEME Guide No. 39 [review]. Med Teach 2016;38:656–68.
15. Battistone MJ, Barker AM, Durning SJ. Interprofessional musculoskeletal education: a review of national initiatives from the Department of Veterans Affairs [review]. Rheum Dis Clin North Am 2020;46:135–53.
16. Jetty A, Jabbarpour Y, Pollack J, et al. Patient-physician racial concordance associated with improved healthcare use and lower healthcare expenditures in minority populations. J Racial Ethn Health Disparities 2022;9:68–81.
17. Salsberg E, Richwine C, Westergaard S, et al. Estimation and comparison of current and future racial/ethnic representation in the US health care workforce. JAMA Netw Open 2021;4:e213789.
18. Huppert LA, Santhosh L, Babik JM. Trends in US internal medicine residency and fellowship applications during the COVID-19 pandemic vs previous years. JAMA Netw Open 2021;4:e218199.
19. Low D, Pollack SW, Liao ZC, et al. Racial/ethnic disparities in clinical grading in medical school. Teach Learn Med 2019;31:487–96.
20. Boatright D, Ross D, O'Connor P, et al. Racial disparities in medical student membership in the Alpha Omega Alpha honor society. JAMA Intern Med 2017;177:659–65.
21. Dyrbye LN, Satele D, West CP. Association of characteristics of the learning environment and US medical student burnout, empathy, and career regret. JAMA Netw Open 2021;4:e2119110.

Arthritis Care & Research
Vol. 75, No. 11, November 2023, pp 2245–2247
DOI 10.1002/acr.25120

American College of Rheumatology
Empowering Rheumatology Professionals

EDITORIAL

Current State of Ultrasound Training in US Rheumatology Fellowships

Ana Valle[1] and Shereen N. Mahmood[2]

Numerous articles have advocated for the routine use of rheumatology musculoskeletal ultrasound (MSUS) based on its ability to improve patient outcomes with earlier diagnoses, utility in treatment monitoring, and targeted procedures. Unlike other imaging modalities, MSUS allows for expedient bilateral, dynamic testing while being noninvasive, affordable, and limiting exposure to undue radiation. While it is a powerful teaching tool for patients to understand their disease and its progression, it is purported to be invaluable in fellows' education so that they can visualize disease processes and correlate them to physical examination and pathology findings. After multiple years of ultrasound advocacy in the US, it is time to reevaluate how ultrasound training offered by rheumatology fellowships has evolved so we can consider its future trajectory within fellowship education and address barriers that may shape its prevalence.

In the US, the use of MSUS within rheumatology has tremendously expanded over the past two decades. In the early 2000s, MSUS training in the US relied on invited international experts teaching courses or conferences (1). By 2008, the Ultrasound School of North American Rheumatologists (USSONAR) was established as a US-based working group to train other rheumatologists, establish competencies, and carry out ultrasound-related research (2). Shortly after, the American College of Rheumatology (ACR) partnered with USSONAR in 2011 to provide ultrasound education as faculty development for interested fellowship programs and, one year later, the rheumatology musculoskeletal ultrasound (RhMSUS) board certification came to fruition (3,4). The increasing momentum behind ultrasound use was palpable, and a 2013 ACR survey of fellowship program directors found that 60% of respondents included MSUS teaching at some level (if not in a formalized curriculum), and a quarter of those who had not were planning to implement ultrasound into their curriculum in the near future (1). In 2014, the ACR held its first ultrasound symposium, which highlighted fellowship programs with MSUS curricula, such as University of Southern California and Loma Linda University, who considered MSUS an essential component of clinical anatomy (5). By 2016, Torralba et al reported 108 of 113 (94%) rheumatology fellowship program directors reported teaching MSUS and 41% had a dedicated, implemented MSUS curriculum (6). While the implementation of MSUS was moving quickly, it likely stalled during the COVID-19 pandemic as fellowships relied heavily on remote learning. Our own review of the rheumatology fellowship websites in the summer of 2022 revealed that most programs, 113 of 141 or approximately 80%, advertised a curricular ultrasound component, yet only 14% (20 of 141) of programs offered a pathway to ultrasound certification through USSONAR.

As one can see, given the value in education and patient care, many strides to incorporate MSUS in training have occurred over a relatively short time. Yet, as a specialty, we have not come to a consensus on the goal of MSUS training during fellowship. Torralba et al discuss the vital role of MSUS as a tool for fellows to learn clinical anatomy (5). Other studies imply MSUS implementation has a unique role in patient care which suggests the goal is MSUS competency to improve patient outcomes (3,7). Most program directors also believe the ACR should champion MSUS certification for fellows, suggesting that while certification may not be a necessary component of MSUS fellowship training, all trainees should equally have access to pursue certification even if it may be outside of their fellowship program (6).

While many fellowship programs have developed their own ultrasound criteria, the ACR has published and updated several scanning protocols, including one specifically for trainees, in an effort for standardization (8,9). We, however, have minimal learner feedback regarding the utility of these protocols. There is a stark need for curriculum evaluation so that there is evidence supporting the methods we are using to educate fellows on the topic of MSUS. Reasons for the lack of evaluation may include disparate resources which limit MSUS implementation and practice.

[1]Ana Valle, MD, MHS: Montefiore Medical Center, Bronx, New York; [2]Shereen N. Mahmood, MD, RhMSUS: Albert Einstein College of Medicine, Bronx, New York.

Author disclosures are available online at https://onlinelibrary.wiley.com/doi/10.1002/acr.25120.

Address correspondence via email to Ana Valle, MD, MHS, at avalle1@montefiore.org.

Submitted for publication February 28, 2023; accepted in revised form March 23, 2023.

Faculty with MSUS experience and/or accreditation as well as clinics with ultrasound machine access vary across states. Additionally, trainees have limited time, which is often focused on other board-tested material. Furthermore, as MSUS appears more frequently in pathology definitions, diagnostic criteria, and taking into account its utility in treatment monitoring and outcomes, we must delineate when and to what degree ultrasound competency should be a part of rheumatology board certification. As a specialty, we must have an explicit conversation regarding the objectives and goals of MSUS training during fellowship, including pathways to MSUS certification and board certification knowledge requirements, and only afterward can we discuss standardization.

When we look toward our European colleagues, we see similar patterns of ultrasound interest and education incorporation without standardized curriculum and learner evaluation. Taggart et al, Gutiérrez et al, and others have reported their own MSUS curriculum (10–12). Yet evaluation of the EULAR courses and the rheumatology competency assessment (COMPASS) by trainees are difficult to find. Grigoriou et al revealed that while a large quantity of rheumatology trainees (42%) had formal ultrasound training, the frequency of financial support and certification is much lower, with 12% and 10%, respectively (13).

In the US, we must also consider the medical organizations that influence our MSUS incorporation. Currently, the Accreditation Council of Graduate Medical Education does not regard exposure to or proficiency in ultrasound as a core competency despite its universal explosive interest over the last decade. Approximately a decade ago, sports medicine fellowships were in a similar predicament in which MSUS was viewed by many educators and providers as crucial, yet was not included in any national requirements. However, the American Medical Society for Sports Medicine advocated for ultrasound as a vital component of sports medicine education that led to a specialty specific MSUS coined "SPORTS US," which is now a requirement for all sports medicine fellowship programs (14). Also, rheumatologists currently receive MSUS reimbursements by the Centers for Medicare and Medicaid Services (CMS) and in recent years there has been growing concerns due to diminishing MSUS reimbursements that may ultimately cease (15). This downward reimbursement trend comes at a time when we continue to see growth in the value of MSUS to patient care, given its ability for early diagnosis and treatment monitoring that translates into earlier treatment interventions—ultimately minimizing long-term disability and health care costs. This decrease in reimbursements has led to advocacy from rheumatologists and the ACR to CMS, which kept payment cuts at bay, yet only for this year (15).

We believe it is imperative to evaluate the current state of ultrasound training within fellowship so we can have a proactive and decisive voice in the future role of ultrasound within the field of rheumatology. The priority we place on the mission, quality, and standardization of ultrasound education now will determine the integration of MSUS within routine clinical care in the next generation of rheumatologists. Discussions of national MSUS requirements for rheumatology fellowship programs are crucial so that ultrasound grows from an interest to integral part of rheumatologic care. As more rheumatologists integrate MSUS into their patient encounters, we hope it will increase the numbers of those who advocate for CMS reimbursements for MSUS and perhaps even solidify the importance of this reimbursement in the future. These steps are necessary to ensure MSUS is not treated as a passing fad or niche learning tool but rather an instrument that improves patient outcomes.

AUTHOR CONTRIBUTIONS

Both authors drafted the article, revised it critically for important intellectual content, approved the final version to be published, and take responsibility for the integrity of the data and the accuracy of the data analysis.

REFERENCES

1. Cannella AC, Kissin EY, Torralba KD, et al. Evolution of musculoskeletal ultrasound in the United States: implementation and practice in rheumatology. Arthritis Care Res (Hoboken) 2014;66:7–13.
2. Ultrasound School of North American Rheumatologists. URL: https://www.ussonar.org/about.
3. Samuels J, Abramson SB, Kaeley GS. The use of musculoskeletal ultrasound by rheumatologists in the United States. Bull NYU Hosp Jt Dis 2010;68:292–8.
4. The Rheumatologist. Introducing RhMSUS–Musculoskeletal Ultrasound Certification in Rheumatology. 2012. URL: https://www.the-rheumatologist.org/article/introducingrhmsusmusculoskeletal-ultrasound-certification-in-rheumatology/.
5. Torralba KD, Villaseñor-Ovies P, Evelyn CM, et al. Teaching of clinical anatomy in rheumatology: a review of methodologies. Clin Rheumatol 2015;34:1157–63.
6. Torralba KD, Cannella AC, Kissin EY, et al. Musculoskeletal ultrasound instruction in adult rheumatology fellowship programs. Arthritis Care Res (Hoboken) 2020;72:859–70.
7. Widener BB, Cannella A, Martirossian L, et al. Modern landscapes and strategies for learning ultrasound in rheumatology. Rheum Dis Clin North Am 2020;46:61–71.
8. Cannella AC, Kissin EY, Torralba KD, et al. 2019 Rheumatologic ultrasound (RhUS) curriculum supplement to the American College of Rheumatology 2015 core curriculum outline. URL: https://www.rheumatology.org/Portals/0/Files/Rhumatologic-Ultrasound-Curriculum-Supplement.pdf.
9. Yinh J, Torralba KD, Choi KS, et al. North American musculoskeletal ultrasound scanning protocol of the hip, knee, ankle, and foot: update of a Delphi consensus study. Clin Rheumatol 2021;40:4233–42.
10. Taggart A, Filippucci E, Wright G, et al. Musculoskeletal ultrasound training in rheumatology: the Belfast experience. Rheumatology (Oxford) 2006;45:102–5.
11. Gutiérrez M, Di Geso L, Rovisco J, et al. Ultrasound learning curve in gout: a disease-oriented training program. Arthritis Care Res (Hoboken) 2013;65:1265–74.

12. Miguel C, De Miguel E, Batlle-Gualda E, et al. Teaching enthesis ultrasound: experience of an ultrasound training workshop. Rheumatol Int 2012;32:4047–52.
13. Grigoriou A, Kiely P. AB0935 rheumatologists and musculoskeletal ultrasound: discrepancy between attitudes, desire and reality in routine practice in South East England. Ann Rheum Dis 2016;75: 1221.
14. American Medical Society for Sports Medicine. F.A.Q. Sports ultrasound: accreditation and certification. URL: https://www.amssm.org/MemberFiles/FAQ_Sports_US.pdf.
15. Saag KG, American College of Rheumatology. Final ACR comments CY23 MPFS and QPP 09.06.22. 2022. URL: https://www.rheumatology.org/Portals/0/Files/ACR-Comments-CY23-Medicare-Physician-Fee-Schedule-Quality-Payment-Program.pdf.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 75, No. 11, November 2023, pp 2248–2258
DOI 10.1002/acr.25108

RheumMadness: Creating an Online Community of Inquiry in Rheumatology

David L. Leverenz,[1] Akrithi U. Garren,[2] Guy Katz,[3] Didem Saygin,[4] Allen Witt,[5] Robert Harper,[6] Matthew A. Sparks,[1] and Lisa Criscione-Schreiber[1]

Objective. To evaluate the educational impact of RheumMadness, an online tournament of rheumatology concepts grounded in social constructivist theory, as viewed through the community of inquiry (CoI) framework.

Methods. The curricular scaffold of RheumMadness was a bracket of 16 rheumatology concepts competing as "teams" in a tournament. Participants could create and review "scouting reports" about each team, listen to a RheumMadness podcast, discuss on social media, and submit a bracket predicting tournament outcomes according to the perceived importance of each team. Engagement was measured with direct analytics and through self-report on a survey. The survey also assessed participants' educational experience using an adapted 34-item CoI survey, which describes the cognitive, social, and teaching presences in a learning activity.

Results. One hundred brackets were submitted. On average, each scouting report was viewed 92 times, each podcast episode was downloaded 163 times, and 486 tweets were sent about #RheumMadness from 105 users. The survey received 58 of 107 responses (54%). Respondent agreement with prompts related to each CoI presence was: 70.3% cognitive, 61.7% social, 84.9% teaching. Reported engagement in RheumMadness correlated strongly with overall CoI survey scores ($r = 0.72$, $P < 0.001$).

Conclusion. RheumMadness created an online CoI that fostered social constructivist learning about rheumatology.

INTRODUCTION

According to the theory of social constructivism, the development of meaning through social interaction is fundamental to human learning (1). Collaborative learning activities rooted in social constructivism emphasize active knowledge co-creation through discourse and reflection (2). Social media has expanded the reach of social constructivist initiatives, enabling learners from diverse educational settings to share knowledge, challenge prior assumptions, and make new cognitive and social connections (3).

We were inspired to explore collaborative learning in rheumatology after seeing the impact of NephMadness, an online tournament for the global nephrology community (4–6). In NephMadness, a bracket of "teams" representing key concepts in nephrology compete against each other through multiple head-to-head matchups, similar to a March Madness college basketball tournament. Each year, approximately 1,000 participants submit a NephMadness bracket, attempting to predict the winners of each matchup against the choices of a Blue Ribbon Panel of experts to determine the most important, impactful, and/or exciting concept in the tournament (7). The tournament generates passionate discussion on social media, and nephrology communities around the world host NephMadness parties to celebrate the tournament. NephMadness has inspired the creation of numerous social media–based educational initiatives for nephrology, including a Twitter journal club, the Nephrology Social Media Collective, and the Nephrology Simulator (8–10). Clearly, NephMadness is not just an educational game, it is the birthplace of an online learning community in nephrology. We were inspired to create a similar community in rheumatology, so we partnered with NephMadness leadership to create RheumMadness.

Dr. Leverenz's work was supported by the Rheumatology Research Foundation Clinician Scholar Educator award to create RheumMadness (grant 2833413).

[1]David L. Leverenz, MD, Med, Matthew A. Sparks, MD, Lisa Criscione-Schreiber, MD, Med: Duke University School of Medicine, Durham, North Carolina; [2]Akrithi U. Garren, MD: MedStar Health and Georgetown Washington Hospital Center, Washington, DC; [3]Guy Katz, MD: Massachusetts General Hospital, Boston; [4]Didem Saygin, MD: University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania; [5]Allen Witt, MD: Durham Veterans Affairs Medical Center, Durham, North Carolina; [6]Robert Harper, EdD: University of Cincinnati College of Medicine, Cincinnati, Ohio.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.25108&file=acr25108-sup-0001-Disclosureform.pdf.

Address correspondence via email to David L. Leverenz, MD, Med, at David.Leverenz@Duke.edu.

Submitted for publication September 21, 2022; accepted in revised form February 23, 2023.

SIGNIFICANCE & INNOVATIONS

- An online tournament of rheumatology concepts successfully fostered collaborative learning by incorporating curricular elements based in social constructivist learning.
- The community of inquiry (CoI) framework offered a unique glimpse into the cognitive, social, and teaching presences experienced by RheumMadness participants.
- Higher engagement with RheumMadness was associated with a more meaningful educational experience, as described by the CoI.
- Engagement with didactic curricular elements (the scouting reports and podcast series) correlated most with the cognitive presence, whereas engagement on social media correlated most with the social presence.

While NephMadness provided the curricular model for RheumMadness, we needed a sound theoretical framework to support our efforts. Learning theory is crucial in social media–based education for illuminating the educational perspectives of course leaders, justifying the use of technology, and choosing an assessment strategy that aligns with intended outcomes (11–14). The structure of RheumMadness draws upon the concept of gamification, in which game elements are applied to traditionally nongame contexts to increase engagement, motivation, and emotional connection to the educational activity (15). Our primary goal was to use this game to inspire collaborative learning. Therefore, we based our assessment strategy on the community of inquiry (CoI) framework, which offers a useful foundation for evaluating online social constructivist activities (16).

In the CoI, a learner's educational experience is described as the function of 3 domains: the cognitive, social, and teaching presences (17). The philosophical basis of the CoI derives from the work of John Dewey, who pioneered the concept of inquiry-based learning as a way to foster deep and meaningful educational experiences (16). Dewey succinctly summarized this perspective when he wrote, "the quality of mental process, not the production of correct answers, is the measure of educative growth" (18). Therefore, the chief purpose of the CoI framework is to describe the quality of a learner's mental process as they experience an educational activity, thus providing insight into how meaningful the educational experience is to the learner (16). Validation studies show that the 3 CoI presences are distinct yet interconnected, acting as key mediators of one another (see Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25108, which shows a Venn diagram conceptualizing the 3 interconnected CoI presences) (17,19–22). The CoI presences also correlate positively with course satisfaction, perceived learning, and grades (17,23,24). Despite this evidence, few medical education initiatives have used this tool (25). We chose the CoI framework because it aligned with our social constructivist perspective and offered a clear lens for viewing the educational experience of RheumMadness participants.

In this project, we created and assessed the educational impact of RheumMadness, an online collaborative learning initiative in rheumatology modeled after NephMadness and grounded in social constructivism and the CoI framework. Our aim was to engage participants in collaborative learning and inspire knowledge co-creation in rheumatology. In this article, we explore the interaction between participants' engagement in RheumMadness and their educational experience as described by the 3 presences within the CoI framework.

MATERIALS AND METHODS

Participants and setting. RheumMadness is intended for all learners interested in rheumatology, including practicing rheumatologists, advanced practice providers, fellows, internal medicine residents, medical students, and patients. Social constructivist activities are centered in learners' zone of proximal development, where peers and instructors must work together to grow in knowledge (26). Therefore, we designed RheumMadness so that participants create much of the learning themselves through collaboration and discussion.

We recruited participants through social media (99 promotional tweets from RheumMadness leadership from July 1, 2020 through the end of bracket submissions March 26, 2021), direct emails to colleagues throughout the US (345 emails about RheumMadness sent by DLL from July 1, 2020 through March 26, 2021), and inclusion in the American College of Rheumatology Fellow-in-Training and Faculty newsletters in January 2021. A timeline of recruitment efforts and key dates for each curricular element in RheumMadness is shown in Figure 1. The Duke University Institutional Review Board exempted RheumMadness as educational research.

Curriculum structure. The RheumMadness tournament took place in March 2021, consisting of a bracket containing 16 rheumatology concepts competing as individual teams (see Supplementary Figure 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25108, which shows the full bracket). We based each team on 1 or 2 articles published within the preceding 2 years. For example, the systemic lupus erythematosus region contained 2 teams: "Belimumab for Lupus Nephritis," based on the BLISS-LN trial (27), and "Anifrolumab," based on the TULIP-1 and TULIP-2 trials (28,29). Participants were invited to complete their own brackets, attempting to predict which teams would progress through the tournament based on each topic's current and/or future implications for patients, providers, and researchers. These criteria were intentionally vague to draw out different perspectives

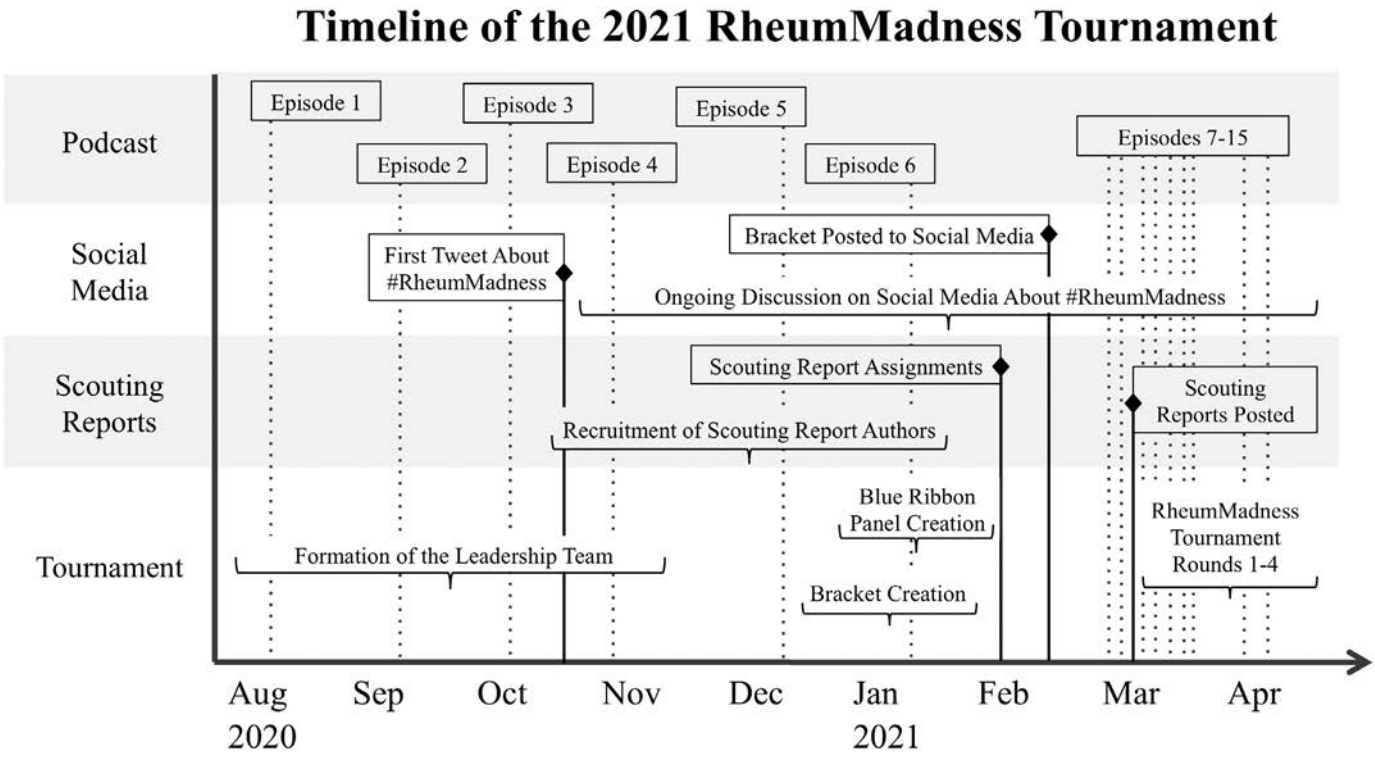

Figure 1. Timeline of the development and implementation of the 2021 RheumMadness tournament.

and inspire debate. The outcome of each matchup was decided by a Blue Ribbon Panel of rheumatologists, chosen by RheumMadness leadership with the intention of including members with diverse areas of expertise and practice settings. The final panel included 7 members: 1 rheumatology fellow, 1 private practice rheumatologist, and 5 academic rheumatologists (2 Assistant Professors, 2 Associate Professors, and 1 Professor) (30). Panel members reviewed the scouting reports and original articles on which the teams were based, after which they voted to determine which team won the matchup according to the same criteria provided to the participants. Participants received points for each correct prediction that matched the choices of the panel.

In keeping with social constructivism, the bracket of teams served as the scaffold upon which participants built knowledge as they engaged with other curricular elements in RheumMadness (1,11). These elements included scouting reports reviewing each team in the tournament, a RheumMadness podcast series, and organic social media–based discussions. The scouting reports were collaboratively written by approximately 40 adult rheumatology fellows from 14 training programs in the US and peer reviewed by the RheumMadness leadership team. In NephMadness, scouting reports are written primarily by members of the NephMadness leadership team, whereas in RheumMadness we asked learners to create the scouting reports to emphasize the social constructivist principle of knowledge co-creation by the learners, while also helping to increase engagement in the tournament. We recruited scouting report authors through direct emails to fellowship program directors, who then coordinated with their fellows to accept or decline the invitation. Scouting report topics were assigned to each fellowship program by the RheumMadness leadership team. In addition, the leadership team provided extensive instructions to the authorship teams, including an explainer video, an example scouting report, and a blank scouting report template. Each scouting report followed the same structure: 1) topic overview; 2) current and future implications for patients, providers, and researchers; 3) chances the topic would win in the first round and the tournament as a whole; and 4) hyperlinks to primary literature. Within these specifications, each program was free to determine how fellows collaborated together to write the reports. Once completed, the reports were peer reviewed by the RheumMadness leadership team and then posted to the RheumMadness website for participants to use as a learning resource to inform their bracket winner choices (31).

Podcasts can foster a sense of connection in professional communities (32). Therefore, we created a RheumMadness podcast series with 15 episodes from August 2020 through April 2021. The podcast was a unique addition to the NephMadness curricular model, as NephMadness does not have a podcast series. In the RheumMadness podcast series, our leadership team discussed concepts in the tournament to provide direct teaching, familiarize listeners with the idea of rheumatology concepts competing as teams, and contextualize the type of conversation we hoped RheumMadness might foster. We also posted

audio versions of the scouting reports to appeal to different learning preferences.

Regarding social media–based discussion, participants were encouraged to read and discuss posts about RheumMadness on Twitter using the hashtag #RheumMadness. This practice is similar to NephMadness, which uses the hashtag #NephMadness. The purpose of social media–based discussion from a social constructivist perspective is to encourage collaborative learning, discussion, and knowledge co-creation among participants (1). We continually monitored and fostered this discussion through "Tweetorials," polls, and other posts. In addition, we created a private RheumMadness Facebook group only for rheumatology trainees.

Assessment instruments. Our research aims were to describe participant engagement with RheumMadness, to understand the perceived educational experience of RheumMadness participants as examined through the CoI framework, and to analyze the relationship between engagement with RheumMadness and perceived educational experience, hypothesizing that participants with higher reported engagement would report a more meaningful experience. Data pertaining to these aims were collected through web-based engagement analytics and a survey.

To measure engagement with RheumMadness, we used Google Analytics to monitor scouting report utilization from the date the scouting reports were posted (March 15, 2021) through tournament completion (April 5, 2021). We used Symplur to monitor the Twitter hashtag #RheumMadness when the majority of social media activity occurred (February 8, 2021 to April 8, 2021). Facebook analytics tracked engagement on the RheumMadness Facebook Group. Finally, podcast downloads were measured from the month the podcast began (August 2020) through the last month of the tournament (April 2021).

The survey assessed participant demographic information, self-reported engagement with RheumMadness, perceived experience of the CoI presences, and satisfaction (see Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25108, which contains the full survey). Medical students, residents, and advanced practice provider trainees were asked to indicate how RheumMadness impacted their interest in rheumatology. All participants were asked to estimate engagement with the podcast episodes and scouting reports using the following 5-point scale: 1 = none, 2 = some, 3 = approximately half, 4 = many, 5 = all. In addition, participants were asked to estimate their frequency of reading and writing social media posts about RheumMadness using the following 5-point scale: 1 = never, 2 = rarely (just once or twice), 3 = occasionally (approximately once per week), 4 = a moderate amount (several times per week), 5 = a great deal (approximately every day).

The survey also asked participants to respond to 34 prompts assessing the CoI presences. We adapted these prompts from the 34-item CoI survey, a validated tool for describing the 3 CoI presences and their subdomains in online education courses (20,22). According to the CoI framework, the cognitive presence is composed of 4 subdomains, representing 4 stages of cognitive processing: 1) knowledge triggering (identification of an issue or problem), 2) exploration (searching for relevant information), 3) integration (constructing meaning and synthesizing ideas), and 4) resolution (application of new knowledge). The social presence is comprised of 3 equally important subdomains: affective expression (the display of personality and emotions), open communication (a sense of trust), and group cohesion (a perception of effective collaboration). The teaching presence brings the cognitive and social presences together within the educational context; its 3 subsubdomains are course design (planning course structure), facilitation (engaging participants in social and cognitive components of the course), and direct instruction (teaching led by course leadership) (17). Of the 34 total prompts in the CoI portion of the survey, 12 address the cognitive presence, 9 address the social presence, and 13 address the teaching presence (see Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25108, for survey prompts related to each subdomain). Each CoI prompt uses the following 5-point scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree.

We distributed the survey by email to all RheumMadness participants who submitted a bracket and to members of the Blue Ribbon Panel, and additionally we shared the survey on social media. Responses from social media were excluded if they did not report submitting a bracket. We also excluded incomplete survey responses, defined as surveys with no responses to questions on engagement, the modified CoI questions, or satisfaction. All surveys were anonymous. Participants consented to the anonymous use of their responses in data analysis and reporting.

Data analysis. We used descriptive statistics to analyze direct and self-reported engagement with RheumMadness (research aim 1) and participants' experience of the CoI presences on the survey (aim 2). We also analyzed the content of tweets about #RheumMadness and free-text comments on the survey to identify statements that illustrated participants' experience with each CoI presence according to an established coding template for CoI-based content analysis of course transcripts (16).

Next, we analyzed the interaction between participant engagement with RheumMadness and perceived educational experience on the survey (aim 3). We defined 2 groups based on survey results: full participants were those who listened to at least half the podcast episodes, read/listened to at least half the scouting reports, and read social media posts about #RheumMadness at least once per week. The remaining respondents were defined as partial participants. The cutoffs for defining full versus partial participants were based on our estimation as a leadership team

of what constituted substantial engagement in each of the major curricular elements in RheumMadness. These cutoffs were defined prior to data analysis. We used Fisher's exact test to compare the proportion of survey responses from full versus partial participants indicating disagreement (score 1–2), neutrality (score 3), or agreement (score 4–5) with questions related to each CoI presence. Missing responses were considered neutral (score 3).

To further explore this relationship, we calculated Pearson's correlation coefficients between weighted averages of survey responses about engagement with each curricular element and the CoI presences (scale of 1–5). We analyzed 4 engagement variables: podcast, scouting reports, social media, defined as the average of the 2 questions about reading and writing social media posts, and overall engagement, defined as the average of the preceding 3 variables. The 4 CoI variables were derived from the average responses to prompts relating to the cognitive, social, and teaching presences, and overall CoI, defined as the average response to all 34 CoI prompts and weighted so that each of the 3 presences contributed equally to the overall CoI score. This analysis required 16 correlations, and thus statistical significance was set at a *P* value of 0.003 (Bonferroni correction).

RESULTS

We received 100 bracket submissions. The 16 scouting reports received 1,472 total page views (average of 92 views per report, range 58–170). On Twitter, the hashtag #RheumMadness was used in 486 tweets from 105 users. The top terms from Twitter participants were "scouting reports," "bracket," "round," "can't wait," "gout guidelines," "VEXAS," "avacopan" (the winning team), and "RheumBoss" (the winning participant). The RheumMadness Facebook group garnered 81 members; all posts were from RheumMadness leaders. The podcast received 2,449 downloads (163 per episode, range 119–241). The outcome of each tournament matchup according to the Blue Ribbon

Table 1. Characteristics and reported engagement in RheumMadness curricular elements of 58 postsurvey respondents, after exclusions*

	All respondents (n = 58)	Full participants (n = 19)	Partial participants (n = 39)
Position/training level			
Practicing physician	31 (53.4)	11 (57.9)	20 (51.3)
Fellow	19 (32.8)	6 (31.6)	13 (33.3)
Resident	4 (6.9)	2 (10.5)	2 (5.1)
Other	4 (6.9)	0 (0.0)	4 (10.2)
Country of residence			
US	50 (86.2)	16 (84.2)	34 (87.2)
Non-US	8 (13.8)	3 (15.8)	5 (12.8)
Unique roles in RheumMadness			
Participated in scouting report creation	14 (24.1)	4 (21.1)	10 (25.6)
Blue Ribbon Panel member	7 (12.1)	4 (21.1)	3 (7.7)
Podcast episodes listened to			
None	22 (37.9)	0 (0.0)	22 (56.4)
Some	14 (24.1)	0 (0.0)	14 (35.9)
Approximately half	9 (15.5)	6 (31.6)	3 (7.7)
Many	9 (15.5)	9 (47.4)	0 (0.0)
All	4 (6.9)	4 (21.1)	0 (0.0)
Scouting reports reviewed			
None	7 (12.1)	0 (0.0)	7 (17.9)
Some	15 (25.9)	0 (0.0)	15 (38.5)
Approximately half	5 (8.6)	3 (15.8)	2 (5.1)
Many	13 (22.4)	6 (31.6)	7 (17.9)
All	18 (31.0)	10 (52.6)	8 (20.5)
Social media: frequency of reading posts			
Never	9 (15.5)	0 (0.0)	9 (23.1)
Rarely (just once or twice)	7 (12.1)	0 (0.0)	7 (17.9)
Occasionally (approximately once/week)	11 (19.0)	4 (21.1)	7 (17.9)
A moderate amount (several times/week)	19 (32.8)	10 (52.6)	9 (23.1)
A great deal (approximately every day)	12 (20.7)	5 (26.3)	7 (17.9)
Social media: frequency of writing posts			
Never	29 (50.0)	3 (15.8)	26 (66.7)
Rarely (just once or twice)	6 (10.3)	2 (10.5)	4 (10.3)
Occasionally (approximately once/week)	10 (17.2)	5 (26.3)	5 (12.8)
A moderate amount (several times/week)	12 (20.7)	9 (47.4)	3 (7.7)
A great deal (approximately every day)	1 (1.7)	0 (0.0)	1 (2.6)

* Values are the number (%).

Panel and participant selections is shown in Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25108.

We directed the survey to 107 recipients (100 participants and 7 Blue Ribbon Panel members). There were 64 responses, 61 from email and 3 from social media. We excluded 5 incomplete responses and 1 response from social media that did not submit a bracket. After exclusions, our response rate was 58 of 107 (54%). Respondent characteristics are shown in Table 1.

Regarding self-reported engagement, 22 survey respondents (37.9%) listened to approximately half or more of the podcasts, 36 (62.1%) reviewed approximately half or more of the scouting reports, 42 (72.4%) read social media posts approximately once per week or more, and 23 (39.7%) wrote a social media post approximately once per week or more. There were 19 respondents (32.8%) who met our definition of full participants, and the remaining 39 (67.2%) were partial participants (Table 1). Engagement was similar between practicing rheumatologists, fellows, and other participants (see Supplementary Table 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25108).

Among all participants, 23 (39.7%) reported they were satisfied, and 35 (60.3%) were very satisfied with the RheumMadness tournament; none were neutral or dissatisfied. Respondent satisfaction with each individual curricular element is shown in Supplementary Table 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25108. Of the 51 respondents who reported a preferred method for reviewing the scouting reports, 39 (76.5%) preferred the website, 6 (11.8%) preferred the podcast, and 13 (25.5%) preferred both.

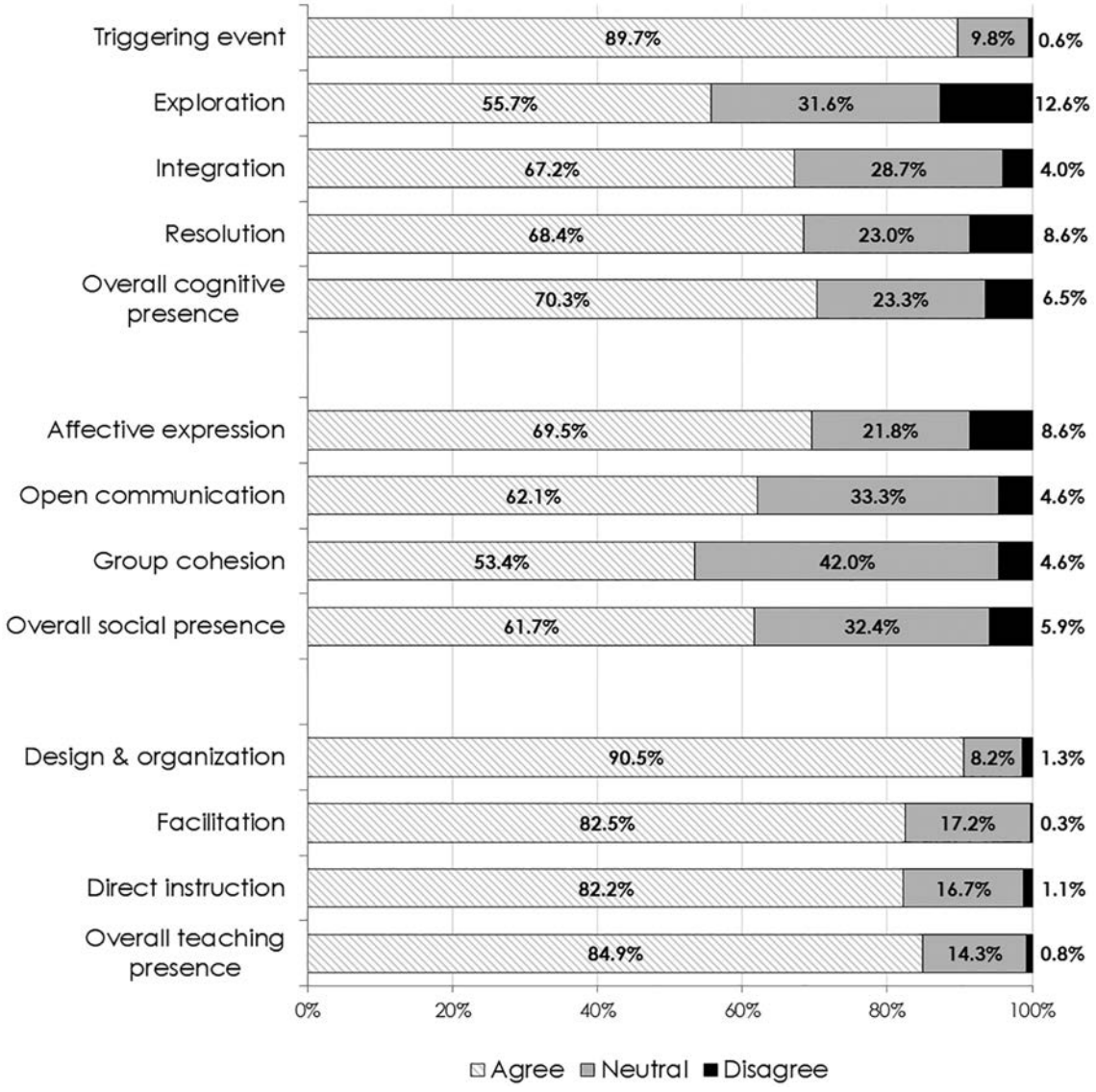

Figure 2. Community of inquiry (CoI) survey results among 58 respondents. The bars represent the proportion of responses indicating agreement, neutrality, or disagreement with prompts relating to each CoI subdomain and overall presence.

Among the 6 early trainee respondents, 5 (83%) indicated that RheumMadness increased their interest in rheumatology and 1 (17%) did not respond to this question.

In the CoI portion of the survey, 8 of 1,972 responses were blank and considered neutral. The majority of responses indicated agreement with prompts for the cognitive presence (70.3%), social presence (61.7%), and teaching presence (84.9%) (Figure 2). Within each presence, the subdomains with the highest agreement were triggering for cognitive (89.7%), affective expression for social (69.5%), and design and organization for teaching (90.5%). The only subdomain with >10% disagreement was cognitive exploration (12.6%). Representative quotes from Twitter and the survey for each subdomain are shown in Table 2.

Table 2. Representative quotes from Twitter and the survey for each community of inquiry subdomain in RheumMadness*

	Indicators	Quotes
Cognitive presence		
Triggering	Interest and excitement, sense of puzzlement, cognitive dissonance	T: "Well I'm ridiculously excited for this fun. Brackets look a little like Duke v some division 3 teams…"
		S: "In my mind, the key value of the program was to highlight new literature articles. As a trainee, much of the knowledge and practice patterns begin with some degree of dogmatic approach, and having a greater sense of the current literature adds value and excitement."
Exploration	Search for and exchange information	S: "I felt compelled to read more into the different topics prior to submitting my bracket to make sure I put together the best prediction possible but ultimately I learned a lot more than expected."
Integration	Connect ideas together, gain new understanding	T: "A trial that can reverse a black box warning? Think FAST. Trials that tell us to do less, not more. Hello SEMIRA! (okay, you too Avacopan). Defense wins games, just sayin' #RheumMadness."
		S: "Seeing the themes of research was helpful for understanding the field, not only by category (e.g., PEXIVAS and ADVOCATE for ANCA, gout articles, etc.) but also by broader theme, like the anti-steroid sentiments more broadly."
Resolution	Apply new information	T: "If you or a family/friend have gout, as you watch the hoops finals tonight, rest easy that in the rheumatology brackets, the study proving febuxostat does NOT increase risk of heart attack made it to the finals. FDA should remove or change the boxed warning about this."
Social presence		
Affective expression	Use of emoticons, humor, self-disclosure	T: "Bracket busted! The round of #Entheseal Eight" shocked me (dropping to 27th place) & 250 million humans who suffer knee arthritis & whose physical therapy trial lost in the 2nd round. Still, SO educational."
Open communication	Risk-free communication, recognize others	T: "I can only #TrashTalk because you are all so sharp. Those scouting reports are [fire emoji]! And you should always expect the kidney docs to bring a little salt(iness) to the game…come over and reciprocate the rivalry at #NephMadness! #RheumMadness."†
Group cohesion	Collaboration, sense of belonging	S: "This was such an innovative and exciting way to engage with other rheumatologists and explore papers from the past year. As a former academic turn[ed] private practitioner, this is such an engaging and imaginative way to bring fellows together and have them shine with the outside rheumatology community."
Teaching presence		
Design and organization	Establish timelines, define discussion topics	T: "Only a few hours left to submit #RheumMadness brackets! When will results be revealed? Round 1: Sat, 3/27 at 2pm ET, Round 2: Mon, 3/29 at 8pm ET, Round 3: Sat, 4/3 at 2pm ET, Round 4: Mon, 4/5 at 8pm ET. Results by email & social media."†
Facilitation	Focus the discussion, encourage participation	T: "Still can't get over how good the #RheumMadness scouting reports are. Common themes: most teams are convinced they will win their first round match-up [thinking emoji]… Are we being blinded by the bright and shiny AAV/SLE trials? I say yes."†
Direct instruction	Present new content, summarize discussions, provide feedback	T: (responding to a participant comment on IgG4-RD criteria). "Love this! These criteria were a huge undertaking and an amazing step for our field! Yes, these were validated for research. But I find them really useful clinically! Esp. exclusion criteria… so helpful to know what the disease isn't! Who knew fever is rare??"†

* We also show the indicators used to identify these quotes, based on methodology for community of inquiry–based content analysis (ref. 16).
AAV = antineutrophil cytoplasmic antibody (ANCA)–associated vasculitis; ET = eastern time; FDA = US Food and Drug Administration; PEXIVAS = Plasma EXchange and glucocorticoids for the treatment of anti-neutrophil cytoplasmic antibody-associated VAScultis trial; S = survey comment; SEMIRA = Steroid EliMination in Rheumatoid Arthritis trial; SLE = systemic lupus erythematosus; T = Twitter comment.
† Indicates comment from RheumMadness leadership.

Table 3. Comparison of community of inquiry survey results between full versus partial participants*

	Full participants			Partial participants		
	Disagree	Neutral	Agree	Disagree	Neutral	Agree
Cognitive presence						
Triggering	0 (0)	0 (0)	57 (100)	1 (0.9)	17 (14.5)	99 (84.6)
Exploration	6 (10.5)	11 (19.3)	40 (70.2)	16 (13.7)	44 (37.6)	57 (48.7)
Integration	3 (5.3)	9 (15.8)	45 (78.8)	4 (3.4)	41 (35.0)	72 (61.5)
Resolution	0 (0)	9 (15.8)	48 (84.2)	15 (12.9)	31 (26.5)	71 (60.7)
Overall cognitive†	9 (3.9)	29 (12.7)	190 (83.3)	36 (7.7)	133 (28.4)	299 (63.9)
Social presence						
Affective expression	0 (0)	8 (14.0)	49 (86.0)	15 (12.8)	30 (25.6)	72 (61.5)
Open communication	0 (0)	10 (17.5)	47 (82.5)	8 (6.8)	48 (41.0)	61 (52.1)
Group cohesion	3 (5.3)	13 (22.8)	41 (71.9)	5 (4.3)	60 (51.3)	52 (44.4)
Overall social†	3 (1.8)	31 (18.1)	137 (80.1)	28 (8.0)	138 (39.3)	185 (52.7)
Teaching presence						
Design/organization	0 (0)	2 (2.6)	74 (97.4)	3 (1.9)	17 (10.9)	136 (87.2)
Facilitation	0 (0)	4 (3.5)	110 (96.5)	1 (0.4)	56 (23.9)	177 (75.6)
Direct instruction	0 (0)	4 (7.0)	53 (93.0)	2 (1.7)	25 (21.4)	90 (76.9)
Overall teaching†	0 (0)	10 (4.0)	237 (96.0)	6 (1.2)	98 (19.3)	403 (79.5)

* Values are the number (%). The table shows the proportion of responses relating to each community of inquiry presence and subdomain. Statistical comparison was performed only for the overall presences, not the subdomains. Of 1,972 potential responses, 8 were left blank. These all occurred in the partial participant group. Blank responses were considered neutral.
† Tested for difference in response frequency between full and partial participants; all showed $P < 0.001$.

Comparing the CoI experience of full versus partial participants, a significant difference was found in the proportion of responses indicating agreement, neutrality, and disagreement for all 3 CoI presences ($P < 0.001$), with full participants indicating more agreement and less neutrality/disagreement within each presence (Table 3). There was also a strong positive correlation between overall CoI scores and overall engagement with RheumMadness (r = 0.72, $P < 0.001$) (Figure 3). The cognitive presence correlated moderately with scouting report (r = 0.49) and podcast (r = 0.48) engagement but not with social media. The social presence correlated strongly with social media (r = 0.78) and moderately with podcast engagement (r = 0.45) but not with scouting reports. The teaching presence correlated moderately with all curricular elements.

	Correlation (r)	p value
Overall Community of Inquiry (CoI) Score		
Overall Engagement	0.72	<0.001*
Podcast Engagement	0.54	<0.001*
Scouting Reports Engagement	0.49	<0.001*
Social Media Engagement	0.65	<0.001*
Cognitive Presence		
Overall Engagement	0.57	<0.001*
Podcast Engagement	0.48	<0.001*
Scouting Reports Engagement	0.49	<0.001*
Social Media Engagement	0.33	0.0106
Social Presence		
Overall Engagement	0.64	<0.001*
Podcast Engagement	0.45	<0.001*
Scouting Reports Engagement	0.31	0.0165
Social Media Engagement	0.78	<0.001*
Teaching Presence		
Overall Engagement	0.60	<0.001*
Podcast Engagement	0.43	<0.001*
Scouting Reports Engagement	0.46	<0.001*
Social Media Engagement	0.50	<0.001*

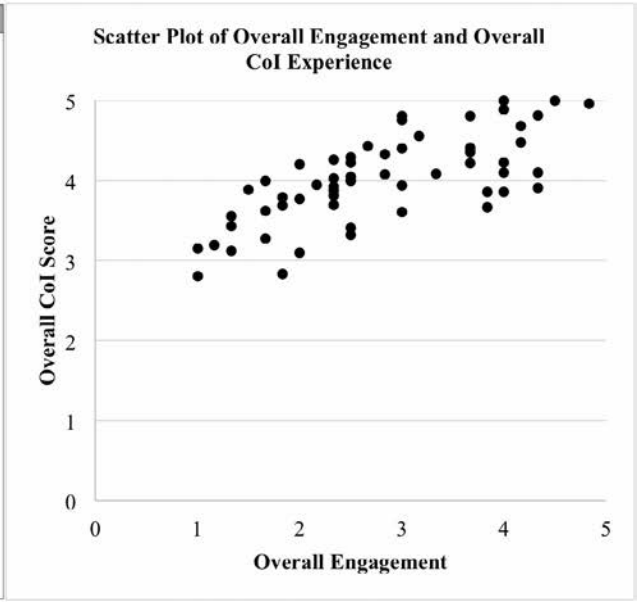

Figure 3. Correlation between reported engagement in RheumMadness and the Community of inquiry presences. Correlation strength key: strong (≥0.6), moderate (0.40–0.59), weak (0.20–0.39). * = statistically significant after Bonferroni correction for 16 tests, setting P value for significance at 0.003125.

DISCUSSION

In the first year of RheumMadness, we fostered collaborative learning among rheumatology fellows, practicing rheumatologists, and other interested learners through a host of social constructivist curricular elements. Our results describe the engagement and learning experience of RheumMadness participants according to the CoI framework.

In keeping with social constructivism, fellow participants in the RheumMadness learning community created the scouting reports reviewing each team in the tournament. These reports were a central learning tool for participants, as each report was viewed an average of 92 times on the website, the majority of survey respondents reported reviewing at least half of the reports, and the term "scouting reports" was a top term in tweets about #RheumMadness. Learners at all levels engaged with these reports, including practicing rheumatologists, who comprised approximately half of tournament participants.

The CoI survey results offer a unique perspective into the educational experience of RheumMadness participants. As with other online social constructivist learning activities, RheumMadness was best at stimulating knowledge triggering within the cognitive presence. However, we were surprised to find that knowledge exploration was experienced less than the higher-order cognitive subdomains of integration and resolution. We suspect this occurred because each team was based on 1 or 2 predefined articles. In future years, we will center teams on more general concepts to determine whether this strategy encourages participants to engage in more knowledge exploration.

Although the majority of respondents agreed with prompts relating to the social presence, it was the least developed of the 3 CoI presences in RheumMadness. Based on our survey results, fostering more open communication and group cohesion is critical for further growth of the social presence within our learning community. The primary avenue for participants to communicate with each other in RheumMadness was on social media (specifically Twitter). Although 72.4% of survey respondents reported reading social media posts at least once per week, only 39.7% actually wrote a social media post at least once per week, and 50.0% never posted on social media. This finding suggests that some participants felt hesitant to engage with each other on this particular platform. Additional analyses are needed to further explore these findings and encourage more open communication and group cohesion in future tournaments.

Participants who engaged more fully in RheumMadness reported higher levels of each CoI presence. Furthermore, engagement with curricular elements containing primarily didactic content, the scouting reports and podcast series, correlated most with the cognitive presence. In contrast, social media engagement correlated strongly with the social presence but not with the cognitive presence. This finding suggests that social media discussions helped participants connect as a community, but that these discussions were not the primary avenue for learning about the concepts in the tournament. In particular, only 4 of the 7 Blue Ribbon Panel members were active on social media during the tournament, and we did not require panel members to explain their determinations. The lack of communication from the panel could have stunted the opportunity to develop socially constructed knowledge as the tournament progressed. Overall, these results inform future efforts to expand and optimize each curricular element in RheumMadness to further stimulate all CoI presences, leading to an even more meaningful learning experience for participants.

Our study has several limitations. The survey response rate of 54% raises the possibility of nonresponse bias. In particular, our study relies heavily on self-reported engagement, and those who responded to the survey may by nature be more likely to engage in a learning activity. We did not collect separate demographic or engagement information from participants during the bracket submission process; therefore, we cannot directly compare the characteristics of participants who responded versus those who did not respond to the survey. However, the substantial engagement seen on direct measures of each curricular element at least implies that survey respondents did not overestimate their engagement. For instance, the 16 scouting reports received 1,472 total page views on the website. Assuming the 58 survey respondents are representative of the full group of 107 participants, their self-reported engagement with the scouting reports would correspond with approximately 1,003 page views (see Supplementary Table 4, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25108). A similar analysis of self-reported podcast engagement on the survey corresponds to only 518 podcast downloads, far fewer than the 2,449 total downloads recorded on direct metrics (see Supplementary Table 5, at http://onlinelibrary.wiley.com/doi/10.1002/acr.25108). We hypothesize that this discordance suggests that many learners engaged with RheumMadness content on the podcast but did not ultimately submit a bracket and were not included in the survey population. Thus, our data may provide an incomplete picture of the full reach of RheumMadness, and future assessments will attempt to capture the educational experience of these learners. Finally, we did not ask participants to report how they heard about RheumMadness, and thus whether direct emails, social media, or other communication strategies were most effective in recruiting participants is unknown.

Despite these limitations, our findings open the door to numerous future directions. First, continued growth is necessary to ensure that RheumMadness becomes a sustainable curriculum in rheumatology education. Participant engagement in RheumMadness roughly matched that of the first year of NephMadness in 2013 which resulted in "a few dozen" bracket entries and 484 tweets about NephMadness from 77 users (5). NephMadness has now grown to receive approximately 1,000 bracket submissions each year, demonstrating the potential of this curricular model to stimulate dramatic growth. As the RheumMadness

tournament grows, performing additional analyses to further explore its educational impact will be possible. We plan to study the educational experience of fellows involved in creating the scouting reports to determine the value of integrating this activity into the curriculum of rheumatology training programs. Qualitative analyses will provide additional insight into the factors that stimulate or prohibit knowledge exploration, open communication, and group cohesion. In addition, future analyses will compare the educational experience of RheumMadness participants at different training levels, and explore the impact of RheumMadness on interest in rheumatology, provider behavior, and patient outcomes. Ultimately, our goal is to optimize RheumMadness such that all participants, from practicing rheumatologists to medical students, are inspired to build knowledge and deepen social connections through this learning community.

In conclusion, the first year of RheumMadness fostered collaborative learning and social connection among an international group of practicing rheumatologists, fellows, and other interested learners. We achieved this outcome by following the successful NephMadness model and grounding the multimodal curriculum of RheumMadness in social constructivist learning theory. In addition, the CoI framework demonstrated that RheumMadness did more than simply trigger interest in rheumatology concepts or help colleagues meet over the internet. Rather, our participants worked together to integrate and apply knowledge, express their personalities, and cohere as a group throughout the world. We believe the social constructivist principles of RheumMadness and the CoI framework can be applied in other disciplines to foster knowledge co-construction and deep social connection across the spectrum of learners in medical education.

ACKNOWLEDGMENTS

This work represents a Master's project by David Leverenz as part of the University of Cincinnati and Cincinnati Children's Hospital Medical Center Master's of Education in medical education program. We wish to thank the NephMadness leadership team for allowing us to use their model, providing guidance, and promoting the tournament on social media. Finally, we would like to thank the numerous volunteer participants who made this possible, including the 7-member Blue Ribbon Panel and the over 40 rheumatology fellows from 14 different training programs who created scouting reports for the tournament.

AUTHOR CONTRIBUTIONS

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

Study conception and design. Leverenz, Garren, Katz, Saygin, Harper, Sparks, Criscione-Schreiber.

Acquisition of data. Leverenz, Garren, Katz, Saygin, Harper, Sparks, Criscione-Schreiber.

Analysis and interpretation of data. Leverenz, Garren, Katz, Saygin, Witt, Harper, Sparks, Criscione-Schreiber.

REFERENCES

1. Powell KC, Kalina CJ. Cognitive and social constructivism: developing tools for an effective classroom. Education 2009;130:241–50.
2. Palincsar AS. Social constructivist perspectives on teaching and learning. Annu Rev Psychol 1998;49:345–75.
3. Schrader DE. Constructivism and learning in the age of social media: changing minds and learning communities. New Dir Teach Learn 2015;2015:23–35.
4. Sparks MA, Lerma EV, Kupin W, et al. NephMadness 2015: nephrology as a cornerstone of medicine. Am J Kidney Dis 2015;65:375–7.
5. Sparks MA, Topf JM. NephMadness after 5 years: a recap and game plan for the future. Am J Kidney Dis 2018;71:299–301.
6. Topf JM, Burgner A, Farouk S, et al. NephMadness: lessons from seven years on the leading edge of social media medical education. Eur Med J 2019;7:48–53.
7. AJKD Blog. #NephMadness 2021: thank you. 2021. URL: https://ajkdblog.org/2021/04/16/nephmadness-2021-thank-you/.
8. Colbert GB, Topf J, Jhaveri KD, et al. The social media revolution in nephrology education. Kidney Int Reports 2018;3:519–29.
9. Farouk S, Hilburg R, Sparks M. Design, dissemination, and assessment of NephSIM: a mobile-optimized nephrology teaching tool. J Grad Med Educ 2019;11:708–12.
10. Ramakrishnan M, Sparks MA, Farouk SS. Training the public physician: the Nephrology Social Media Collective internship. Semin Nephrol 2020;40:320–7.
11. Flynn L, Jalali A, Moreau KA. Learning theory and its application to the use of social media in medical education. Postgrad Med J 2015;91:556–60.
12. Sterling M, Leung P, Wright D, et al. The use of social media in graduate medical education: a systematic review. Acad Med 2017;92:1043–56.
13. Chan TM, Dzara K, Dimeo SP, et al. Social media in knowledge translation and education for physicians and trainees: a scoping review. Perspect Med Educ 2020;9:20–30.
14. Sandars J, Patel R, Goh P, et al. The importance of educational theories for facilitating learning when using technology in medical education. Med Teach 2015;37:1039–42.
15. Rutledge C, Walsh CM, Swinger N, et al. Gamification in action: theoretical and practical considerations for medical educators. Acad Med 2018;93:1014–20.
16. Garrison DR, Anderson T, Archer W. Critical inquiry in a text-based environment: computer conferencing in higher education. Internet High Educ 1999;2:87–105.
17. Richardson JC, Arbaugh JB, Cleveland-Innes M, et al. Using the community of inquiry framework to inform effective instructional design. In: Moller L, Huett J, editors. The next generation of distance education. Boston (MA): Springer; 2012. p. 97–125.
18. Dewey J. Democracy and education: an introduction to the philosophy of education. New York: Macmillan; 1916.
19. Stenbom S. A systematic review of the Community of Inquiry survey. Internet High Educ 2018;39:22–32.
20. Arbaugh JB, Cleveland-Innes M, Diaz SR, et al. Developing a community of inquiry instrument: testing a measure of the Community of Inquiry framework using a multi-institutional sample. Internet High Educ 2008;11:133–6.
21. Garrison DR, Cleveland-Innes M, Fung TS. Exploring causal relationships among teaching, cognitive and social presence: student perceptions of the community of inquiry framework. Internet High Educ 2010;13:31–6.
22. Bangert AW. Building a validity argument for the community of inquiry survey instrument. Internet High Educ 2009;12:104–11.

23. Picciano AG. Beyond student perceptions: issues of interaction, presence, and performance in an online course. J Asynchronous Learn Networks 2002;6:21–40.
24. Richardson JC. Examining social presence in online courses in relation to students' perceived learning and satisfaction. State University of New York at Albany; 2001. URL: file:///C:/Users/ca051923/Downloads/RichardsonSwan%20JALN7(1)%20(1).pdf.
25. Smadi O, Chamberlain D, Shifaza F, et al. A Community of Inquiry lens into nursing education: the educators' experiences and perspectives from three Australian universities. Nurse Educ Pract 2021;54:103114.
26. Doolittle PE. Vygotsky's Zone of Proximal Development as a theoretical foundation for cooperative learning. J Excell Coll Teach 1997;8:83–103.
27. Furie R, Rovin BH, Houssiau F, et al. Two-year, randomized, controlled trial of belimumab in lupus nephritis. N Engl J Med 2020;383:1117–28.
28. Furie RA, Morand EF, Bruce IN, et al. Type I interferon inhibitor anifrolumab in active systemic lupus erythematosus (TULIP-1): a randomised, controlled, phase 3 trial. Lancet Rheumatol 2019;1:e208–19.
29. Morand EF, Furie R, Tanaka Y, et al. Trial of anifrolumab in active systemic lupus erythematosus. N Engl J Med 2020;382:211–21.
30. Leverenz D. RheumMadness. 2021 Blue Ribbon Panel. 2021. URL: https://sites.duke.edu/rheummadness/2021/02/10/2021-blue-ribbon-panel/.
31. Leverenz D. RheumMadness. The scouting reports are here! URL: https://sites.duke.edu/rheummadness/2021/03/15/the-scouting-reports-are-here/.
32. Riddell J, Robins L, Brown A, et al. Independent and interwoven: a qualitative exploration of residents' experiences with educational podcasts. Acad Med 2020;95:89–96.

Arthritis Care & Research
Vol. 75, No. 11, November 2023, pp 2259–2266
DOI 10.1002/acr.25156

AMERICAN COLLEGE of RHEUMATOLOGY
Empowering Rheumatology Professionals

BRIEF REPORT

Long-Term Maintenance of Clinical Responses by Individual Patients With Polyarticular-Course Juvenile Idiopathic Arthritis Treated With Abatacept

Hermine I. Brunner,[1] Nikolay Tzaribachev,[2] Ingrid Louw,[3] Inmaculada Calvo Penades,[4] Francisco Avila-Zapata,[5] Gerd Horneff,[6] Ivan Foeldvari,[7] Daniel J. Kingsbury,[8] Maria Eliana Paz Gastanaga,[9] Carine Wouters,[10] Johannes Breedt,[11] Robert Wong,[12] Margarita Askelson,[13] Joe Zhuo,[12] Alberto Martini,[14] Daniel J. Lovell,[1] and Nicolino Ruperto[15] for the Paediatric Rheumatology International Trials Organisation (PRINTO) and the Pediatric Rheumatology Collaborative Study Group (PRCSG) investigators

Objective. To investigate the frequency and trajectories of individual patients with polyarticular-course juvenile idiopathic arthritis (JIA) achieving novel composite end points on abatacept.

Methods. Data from a clinical trial of subcutaneous abatacept (NCT01844518) and a post hoc analysis of intravenous abatacept (NCT00095173) in patients with polyarticular-course JIA were included. Three end points were defined and evaluated: combined occurrence of low disease activity (LDA) measured by the Juvenile Arthritis Disease Activity Score; 50% improvement in American College of Rheumatology criteria for JIA (ACR50); and patient-reported outcomes. Patient-reported outcomes included visual analog scale score of minimal pain (pain-min) and Childhood Health Assessment Questionnaire disability index score of 0 (C-HAQ DI0). In this post hoc analysis, maintenance of month 13 and 21 end points (LDA+pain-min, LDA+C-HAQ DI0, and ACR50+pain-min) in those who achieved them at month 4 was determined.

Results. Composite end points (LDA+pain-min, LDA+C-HAQ DI0, and ACR50+pain-min) were achieved at month 4 (44.7%, 19.6%, and 58.9% of the 219 patients treated with subcutaneous abatacept, respectively). Of those who achieved LDA+pain-min at month 4, 84.7% (83 of 98) and 65.3% (64 of 98) maintained LDA+pain-min at months 13 and 21, respectively. The proportions of patients meeting LDA+pain-min outcomes increased from 44.7% (98 of 219) at month 4 to 54.8% (120 of 219) at month 21. The frequency of patients who met an LDA+C-HAQ DI score of 0 increased from 19.6% (43 of 219) at month 4 to 28.8% (63 of 219) at month 21.

Conclusion. Among individual patients with polyarticular-course JIA treated with abatacept who achieved 1 of the combined clinical and patient-reported outcomes composite end points, many maintained them over 21 months of abatacept treatment.

INTRODUCTION

Juvenile idiopathic arthritis (JIA) is the term used to describe a group of noninfectious inflammatory conditions of unknown etiology with onset prior to age 16 years resulting in chronic arthritis for a minimum duration of 6 weeks (1,2). JIA may be associated with extraarticular features such as uveitis, fever, and rashes (1,2). Children and adolescents with JIA often experience

Supported by Bristol Myers Squibb.

[1]Hermine I. Brunner, MD, MSc, MBA, Daniel J. Lovell, MD, MPH: Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; [2]Nikolay Tzaribachev, MD, PhD: PRI Research, Bad Bramstedt, Germany; [3]Ingrid Louw, MBChB, MMED: Panorama Medical Centre, Cape Town, South Africa; [4]Inmaculada Calvo Penades, MD: Hospital Universitario y Politécnico de La Fe, Valencia, Spain; [5]Francisco Avila-Zapata, MD: Hospital Star Medica Merida, Yucatan, Mexico; [6]Gerd Horneff, MD: Asklepios Klinik, Sankt Augustin, Germany, and University of Cologne, Cologne, Germany; [7]Ivan Foeldvari, MD: Hamburger Zentrum für Kinder- und Jugendrheumatologie, Hamburg, Germany; [8]Daniel J. Kingsbury, MD: Randall Children's Hospital at Legacy Emanuel, Portland, Oregon; [9]Maria Eliana Paz Gastanaga, MD: Clínica Anglo Americana, Lima, Peru; [10]Carine Wouters, MD, PhD: Leuven University Hospital, Leuven, Belgium; [11]Johannes Breedt, MBcHB DOH: University of Pretoria, Pretoria, South Africa; [12]Robert Wong, MD, Joe Zhuo, PhD: Bristol Myers Squibb, Princeton, New Jersey; [13]Margarita Askelson, MS: Bristol Myers Squibb, Braine-L'Alleud, Belgium; [14]Alberto Martini, MD: Università degli Studi di Genova, Genoa, Italy; [15]Nicolino Ruperto, MD, MPH: IRCCS Istituto Giannina Gaslini, UOSID Centro Trial, Genoa, Italy.

Author disclosures are available online at https://onlinelibrary.wiley.com/doi/10.1002/acr.25156.

Address correspondence via email to Hermine I. Brunner, MD, MSc, MBA, at Hermine.Brunner@cchmc.org.

Submitted for publication July 25, 2022; accepted in revised form May 11, 2023.

SIGNIFICANCE & INNOVATIONS

- The analysis of data from a phase 3 multicenter study and a post hoc analysis confirms that individual children age 2–17 years with polyarticular-course juvenile idiopathic arthritis (JIA) treated with subcutaneous (SC) or intravenous (IV) abatacept achieved composite end points comprised of both a clinically meaningful end point and a meaningful patient-reported outcome end point.
- Moreover, individual children with polyarticular-course JIA treated with SC or IV abatacept who achieved novel composite end points can maintain or further improve these responses/end points over 21 months.

poor health-related quality of life and carry the risk of permanent joint damage, especially if joint inflammation remains poorly treated (1,2). Abatacept selectively modulates T-cell costimulation and has been found to be effective and well tolerated in patients with polyarticular-course JIA when administered by the intravenous (IV) or subcutaneous (SC) route (3,4). We have previously shown that the clinical benefits in patients with polyarticular-course JIA can be maintained for 7 years with IV abatacept treatment (5) and for over 24 months with SC abatacept (3). A preliminary assessment of patients with polyarticular-course JIA treated with SC abatacept examined the maintenance of clinical response over 2 years and treatment response by individual patients and noted that the majority achieved and maintained efficacy end points over time (6). Treatment with IV or SC abatacept has also led to substantial improvements in patient-reported outcomes, such as chronic pain and functional ability (3,5).

In recent years, treat-to-target strategies have been recommended for the treatment of polyarticular-course JIA (7). In support of implementing treat-to-target therapeutic strategies, clinicians could benefit from information pertaining to the persistence of treatment responses in individual patients. The results of a recent study that evaluated disease activity and patient-reported outcomes in the same patients with polyarticular-course JIA using machine learning suggested that both clinical and patient-reported outcomes show similar trajectories over time.

The main goals of this post hoc analysis were to investigate the frequency and trajectories of achieving treatment goals in individual patients with polyarticular-course JIA, as well as the simultaneous achievement of low disease activity (LDA) in combination with highly favorable patient-reported outcomes in response to SC abatacept treatment and subsequent maintenance for up to 21 months.

PATIENTS AND METHODS

Compliance with research ethics standards. Studies included in this post hoc analysis were conducted in accordance with the Declaration of Helsinki, the International Conference on Harmonization Guidelines for Good Clinical Practice, and local regulations. At each site, an individual institutional review board or independent ethics committee approved the protocol, consent forms, and any other written information provided to patients or their legal representatives. Written consent was obtained from all participants.

Data sets and study details. Data presented are from analyses of 2 abatacept studies (3,4). First, data from a post hoc analysis of a 24-month, single-arm, open-label, multicenter phase 3 trial of weekly weight-tiered SC abatacept in patients with polyarticular-course JIA who had an inadequate response/intolerance to ≥1 disease-modifying antirheumatic drug (NCT01844518) (3). Second, additional data were included from a previous post hoc analysis of a double-blind, randomized, placebo-controlled withdrawal trial of IV abatacept in patients with JIA age 6–17 years (NCT00095173) (4). Patients who failed to achieve an improvement of 30% in American College of Rheumatology criteria for JIA (ACR30) were discontinued from the study. All patients remaining after month 4 continued abatacept treatment.

In both abatacept studies, 6 ACR JIA criteria core set variables were measured: number of active joints; number of joints with limitation of motion; physician's global assessment of disease activity measured using a visual analog scale (VAS); parent's global assessment of patient overall well-being measured using a VAS; cross-culturally adapted and validated versions of the Childhood Health Assessment Questionnaire disability index (C-HAQ DI) (8); and a laboratory marker of inflammation (either C-reactive protein [CRP] or erythrocyte sedimentation rate). The C-HAQ DI measures physical function limitations on a scale of 0–3 across 8 domains of disability components, with higher values indicating greater disability.

Composite end points. In this analysis, we aimed to assess the ability of individual patients to simultaneously achieve both a clinical efficacy end point and a patient-reported outcome end point over time. While clinical end points such as the Juvenile Arthritis Disease Activity Score in 27 joints (JADAS-27) are valuable, it is also important to assess meaningful improvements in patient-reported outcomes for each child. However, the evaluated values of pain (measured on a 0–100 mm VAS [pain-VAS], with higher values indicating greater pain) and C-HAQ DI scores are not included in the JADAS-27 score, and although the ACR JIA criteria response measures include the C-HAQ DI score, they do not include a pain-VAS. To assess a patient-reported outcome variable independent of the efficacy variable, the patient-reported outcomes evaluated here included pain, as reduction in pain is a priority for patients with polyarticular-course JIA (9), along with components of the ACR JIA criteria core set variables (3). Therefore, combined clinical and patient-reported outcome composite end points were devised for this study, and the following

3 composite end points were then evaluated in individual patients: LDA (defined as a JADAS-27 score using a CRP level of ≤3.8) (10–12) plus minimal pain (LDA+pain-min); LDA plus absence of disability (LDA + a C-HAQ DI score of 0 [C-HAQ DI0]); and a 50% improvement from baseline to month 4 in ACR JIA criteria (ACR50) plus minimal pain (ACR50+pain-min).

Definitions of favorable clinical and patient-reported outcomes considered in composite end points. Favorable patient-reported outcomes were defined as the absence of disability measured by the C-HAQ DI0 and no more than minimal chronic pain (a pain-VAS score of <35 mm) (13). Favorable clinical outcomes considered were LDA and ACR50 (8).

Statistical analyses. Data were analyzed using SAS, version 9.4. Descriptive statistics and Kaplan-Meier analyses were performed to determine the proportion of patients achieving composite end points (LDA+pain-min, LDA+C-HAQ DI0, and ACR50+pain-min) at month 4 (selected as a time point to match the follow-up time for the primary end point of the SC abatacept study [NCT01844518]) and the maintenance of these responses at months 13 and 21 (7 and 26 for IV). Months 13 and 21 were the closest time points to year 1 and year 2 milestones where data were collected, respectively (some month 24 efficacy data were inadvertently not collected by investigators at some sites).

The proportions of patients achieving responses were assessed in the intent-to-treat (ITT) population, defined as all treated patients (patients with missing data were imputed as nonresponders). For the continuous patient-reported outcome variables, an "as observed" (missing values were not imputed) analysis was conducted.

Heat maps and Sankey diagrams were used to evaluate individual patients as either composite end point responders or nonresponders over the course of study. Patients with missing values (including patients who discontinued due to lack of efficacy) were considered as nonresponders for the ACR50+pain-min end point. Bar graphs were used to summarize proportions of patients meeting composite end points at month 4 and continuing to meet these end points at months 13 and 21. Time-to-achieve composite end points are shown using Kaplan-Meier plots. We also evaluated the proportion of patients who achieved ACR50+pain-min with LDA+pain-min and LDA+C-HAQ DI0 end points. The results presented in this study are for the overall population of the SC abatacept study (combining the 2 age cohorts). Results for the individual cohorts from the SC abatacept study (cohort 1, patients age 6–17 years and cohort 2, patients age 2–5 years) and IV study are reported in Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25156.

RESULTS

Patients and clinical response. Baseline characteristics of patients in the SC and IV abatacept trials included in this analysis are shown in Table 1 and Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25156. Before or by month 4, 7 of the 219 patients (3.2%) discontinued from the study due to lack of efficacy with open-label SC abatacept treatment. By month 4, LDA was achieved by 46.1% (101 of 219) of patients, and an ACR50 response was achieved by 57.1% (165 of 219).

Composite end points in overall study population. Figure 1A shows the proportion of patients treated with SC abatacept who achieved composite end points at month 4 and continued to meet these same end points at months 13 and 21. Of the 44.7% (98 of 219) who achieved LDA+pain-min at month 4, 84.7% (83 of 98) maintained this status at month 13, and 65.3% (64 of 98) maintained this at month 21. Of the 58.9% (129 of 219) who achieved ACR50+pain-min at month 4, 84.5% (109 of 129) maintained this at month 13, and 73.6% (95 of 129) maintained this at month 21. Supplementary Figure 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25156, shows comparable analyses of data from the phase 3 trial of IV abatacept. In data from this trial, 24.7% (47 of 190) of patients achieved LDA+pain-min at month 4; 66.0% (31 of 47) maintained this status at month 7, and 48.9% (23 of 47) maintained it at month 26. Similar to the SC trial, lower proportions of patients achieved LDA+C-HAQ DI0 at month 4 (data not shown).

Figure 1B shows time to achievement of all 3 composite end points in patients treated with SC abatacept. There are marked differences in the median time to achieving composite end points ranging from 1.9 (ACR50+pain-min) to 21.5 months (LDA+C-HAQ DI0).

Composite end points in individual patients. Figure 2 shows 3 heat maps displaying the individual responder status over time for all patients treated with SC abatacept who met composite end points at month 4. Overall, the majority of patients who achieved LDA+pain-min (Figure 2A), LDA+C-HAQ DI0 (Figure 2B), and ACR50+pain-min (Figure 2C) at month 4 maintained this status at month 13 (87.8%, 84.0%, and 81.4%, respectively) and month 21 (72.4%, 72.0%, and 60.5%, respectively).

The Sankey diagrams shown in Supplementary Figure 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25156, provide a summation of the course of individual patients treated with SC abatacept meeting composite end points. The proportion of patients achieving LDA+pain-min increased from 44.7% (98 of 219) at month 4 to 54.8% (120 of 219) at month 21. Patients who were

Table 1. Baseline demographic and disease characteristics in the subcutaneous abatacept trial*

Characteristic	Cohort 1 (6–17 years) (n = 173)	Cohort 2 (2–5 years) (n = 46)	Overall population (n = 219)
Age, median (IQR) years	13.0 (10.0–15.0)	4.0 (3.0–5.0)	11.0 (2.0–17.0)†
Female	136 (78.6)	28 (60.9)	164 (74.9)
Weight, median (IQR) kg	45.0 (31.5–57.0)	18.0 (15.0–21.1)	37.4 (12.0–146.3)†
Weight categories, kg			
<25	18 (10.4)	43 (93.5)	61 (27.9)
25 to <50	74 (42.8)	3 (6.5)	77 (35.2)
≥50	81 (46.8)	0	81 (37.0)
Race‡			
White	144 (83.2)	44 (95.7)	188 (85.8)
Black/African American	14 (8.1)	1 (2.2)	15 (6.8)
Other	15 (8.7)	1 (2.2)	16 (7.3)
Disease duration, median (IQR) years	2.0 (0.0–4.0)	0.5 (0.0–1.0)	1.0 (0–15)†
<2	102 (59.0)	42 (91.3)	144 (65.8)
2 to <5	37 (21.4)	4 (8.7)	41 (18.7)
5 to ≤10	30 (17.3)	0	30 (13.7)
>10	4 (2.3)	0	4 (1.8)
JIA categories			
Polyarthritis RF negative	94 (54.3)	29 (63.0)	123 (56.2)
Polyarthritis RF positive	46 (26.6)	3 (6.5)	49 (22.4)
Extended oligoarthritis	19 (11.0)	10 (21.7)	29 (13.2)
Systemic arthritis	5 (2.9)	0	5 (2.3)
Psoriatic arthritis	0	4 (8.7)	4 (1.8)
Enthesitis-related arthritis	4 (2.3)	0	4 (1.8)
Undifferentiated or persistent oligoarthritis§	5 (2.9)	0	5 (2.3)
JIA-ACR core set variables			
No. of active joints, median (IQR)	10.0 (6.0–19.0)	7.0 (6.0–12.0)	9.0 (6–17)
No. of joints with LOM, median (IQR)	8.0 (4.0–15.0)	8.0 (4.0–11.0)	8 (4–14)
PhGA median (IQR) mm	48.0 (31.0–67.0)	50.0 (3.50–6.00)	48 (32.0–65.0)
P-well VAS score, median (IQR) mm	47.8 (24.1–68.0)¶	42.1 (17.9–54.7)	47.2 (21.8–65.6)
C-HAQ DI, median (IQR)	0.9 (0.4–1.5)¶	1.2 (0.8–1.6)	1.0 (0.5–1.6)
CRP, median (IQR) mg/dl#	0.2 (0.1–0.9)	0.1 (0.1–1.4)	0.2 (0.1–1.0)
JADAS-27 CRP, median	19.1¶	16.1	18.1
JADAS-71 CRP, median (IQR)	21.0 (13.5–30.3)**	18.1 (14.0–23.1)	19.9 (13.8–28.1)
Pain VAS score, median, mm	49	39.5	–
Methotrexate use at baseline	136 (78.6)	37 (80.4)	173 (79.0)
Methotrexate dose at baseline, median (IQR) mg/m^2/week	11.6 (9.7–14.4)	13.3 (10.9–15.3)	–
Route of methotrexate administration			
Oral	76 (55.9)	18 (48.6)	–
Parenteral††	60 (44.1)	19 (51.4)	–
Oral corticosteroid use at baseline‡‡	56 (32.4)	9 (19.6)	66 (30.1)
Oral prednisone (or equivalent) dose at baseline, median (IQR) mg/kg/day	0.1 (0.1–0.2)§§	0.2 (0.2–0.4)¶¶	–
Prior biologic use##	46 (26.6)	10 (21.7)	56 (25.6)

* Values are the number (%) unless indicated otherwise. ACR = American College of Rheumatology; C-HAQ DI = Childhood Health Assessment Questionnaire disability index; CRP = C-reactive protein; JADAS-27 = Juvenile Arthritis Disease Activity Score in 27 joints; JADAS-71 = Juvenile Arthritis Disease Activity Score in 71 joints; JIA = juvenile idiopathic arthritis; LOM = limitation of motion; PhGA = physician global assessment of disease activity; P-well = parent's global assessment of well-being; RF = rheumatoid factor; VAS = visual analog scale.
† Values are the median (minimum, maximum).
‡ Race and ethnicity were self-reported from a fixed set of categories.
§ Protocol violation.
¶ N = 172.
# Normal range for CRP: ≤0.6 mg/dl.
** N = 171.
†† Includes subcutaneous and intramuscular.
‡‡ Prednisone or prednisolone.
§§ N = 52.
¶¶ N = 8.
Adalimumab, etanercept, and tocilizumab.

LDA+pain-min responders maintained their response over time (see Supplementary Figure 3A). Likewise, the proportion meeting the LDA+C-HAQ DI0 end point increased from 19.6% (43 of 219) at month 4 to 28.8% (63 of 219) at month 21, while only a few patients reaching this composite end point at month 4 lost it later (see Supplementary Figure 3B). Responders for the ACR +pain-min end point increased from 58.9% (129 of 219) at month 4 to 63.5% (139 of 219) at month 21 (see Supplementary

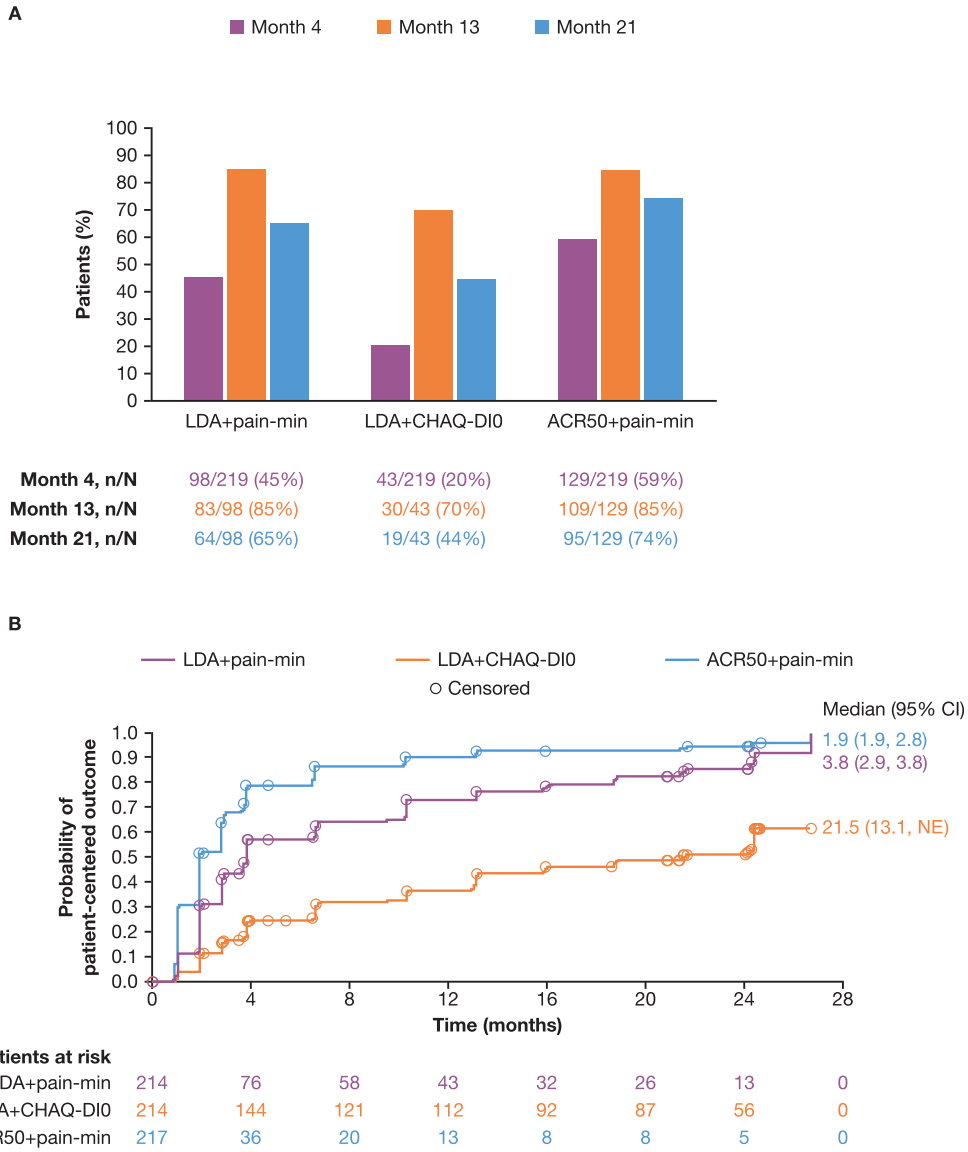

Figure 1. Proportions of patients meeting composite end points at month 4 and maintaining response at months 13 and 21 (**A**) and Kaplan-Meier plots for the time to achievement of composite end points in patients treated with subcutaneous abatacept (**B**). For panel A, the percentage at months 13 and 21 is based on the number of patients who achieved response at month 4 (denominator). For panel B, the month was calculated using the actual days since abatacept treatment/30 and rounded to 1 decimal. Patients without the combined event are censored at the last assessment for the combined event. The number at month 0 is the number of treated patients with the combined event at day 1. Patients who have the event at baseline are excluded from the analysis. 95% CI = 95% confidence interval; ACR50+pain-min = 50% improvement in Juvenile Idiopathic Arthritis–American College of Rheumatology criteria plus minimal pain; LDA = low disease activity; LDA+pain-min = LDA plus minimal pain; LDA+CHAQ-DI0 = LDA plus Childhood Health Assessment Questionnaire disability index score of 0; NE = not evaluable.

Figure 3C). However, achievement of this end point was less well maintained compared with the other composite end points (LDA+pain-min and LDA+C-HAQ DI0).

DISCUSSION

In this post hoc analysis of data from 2 phase 3 multicenter studies of abatacept in patients with polyarticular-course JIA, we explored the achievement of select combined clinical and patient-reported outcome composite end points on treatment initiation. The majority of individual patients who achieved the composite end points at month 4 maintained these responses through month 21. These findings attest to the efficacy of abatacept in patients with polyarticular-course JIA with benefits on several aspects of health-related quality of life, namely patient well-being, pain, and functional ability. Notably, disease flares are a major source of patient concern. Additionally, disease worsening may adversely impact a patient's family (14). The findings from

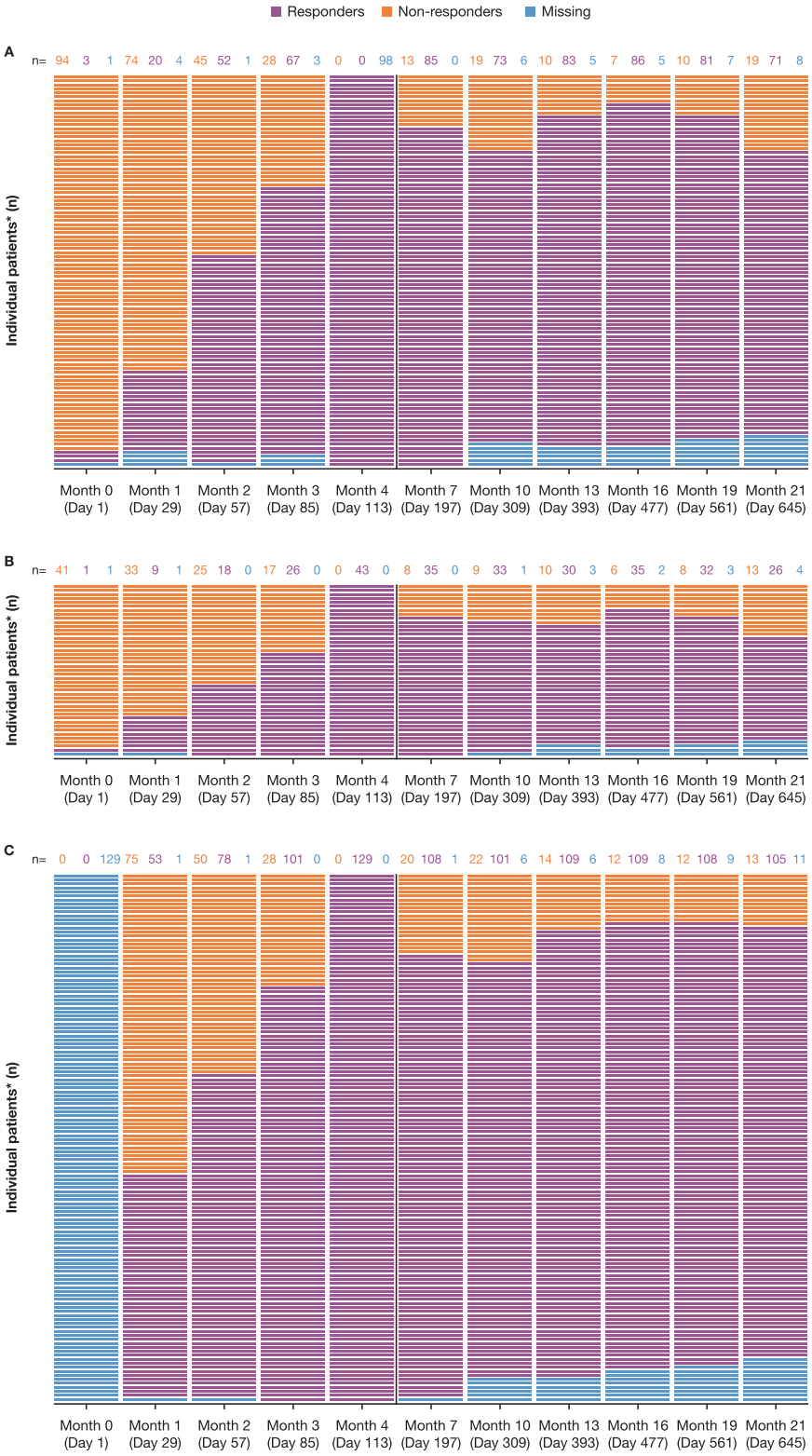

Figure 2. Heat maps of individual patients treated with subcutaneous abatacept who met composite end points at month 4 and their responder status over time: LDA+pain-min (**A**), LDA+CHAQ-DI0 (**B**), and ACR50+pain-min (**C**). Responders are patients who met composite end points. Patients with missing data are imputed as nonresponders. Each bar represents the outcomes achieved over time by a single individual patient. * = each horizontal row represents an individual patient. ACR50+pain-min = 50% improvement in Juvenile Idiopathic Arthritis–American College of Rheumatology criteria plus minimal pain; LDA = low disease activity; LDA+pain-min = LDA plus minimal pain; LDA+CHAQ-DI0 = LDA plus Childhood Health Assessment Questionnaire-disability index score of 0.

the study of SC abatacept are supported by data from the IV abatacept trial, which also showed that the stringent composite end points were achieved by individual patients by month 4, most notably for ACR50+pain-min. Once achieved, composite end points were generally maintained through month 21. In patients receiving SC abatacept, changes over time showed that individual patients who achieved composite end points early maintained them through month 21.

SC abatacept is known to be beneficial in treating children with polyarticular-course JIA with respect to clinical and patient-reported outcomes. Abatacept administered intravenously has been shown to maintain clinical efficacy (ACR30) and patient-reported outcome (mean C-HAQ DI) responses over a 5-year follow-up period (5). However, individual patients can achieve and lose response during a clinical trial, which may not be reflected in group-level data. Therefore, it is important to ascertain if individual children can not only achieve optimal traditional clinical outcomes and patient-reported outcome end points but also sustain them over time. The present research builds on previous population/aggregate analyses in which children with polyarticular-course JIA were successfully treated with SC abatacept (3,5) to show that individual children can achieve and maintain rigorous efficacy end points over time. Similarly, the results from individual patients treated with IV abatacept support the sustainability of composite end points (4).

While the present study reports the possible trajectory of an individual patient who achieves early composite end points, efforts to identify patients who are most likely to achieve an initial treatment response are ongoing. The identification of distinct patient groups as defined by disease manifestation or trajectories of progression, and of prognostic factors for response to abatacept, may help treatment plans for individuals with JIA.

One of the potential limitations of this study may be that we newly defined composite end points. However, the stringent end points we chose are well founded based on current knowledge (3,9–13,15). Additionally, we avoided any thresholds of combined clinical and patient-reported outcome assessments that would be unlikely or impossible to be shared by the same individual (e.g., ACR30 and pain-VAS of 0 mm). The pairing of other clinical and patient-reported outcome end points may either show similar or different results. Furthermore, although this study does not use the latest proposed JADAS-27 cutoffs, the use of previously well-established cutoffs, which were endorsed by professional organizations and used during the interim period of the medical community's transition to the more recent cutoffs, is scientifically valid.

These novel composite end points may be used in future treat-to-target studies, with appropriate input from clinicians and additional validation within a more generalizable JIA population. The data from this post hoc analysis must be interpreted in the context of the initial study populations being a single-arm, open-label SC abatacept trial and a withdrawal trial of patients who achieved an initial response to IV abatacept.

This study demonstrated that individual children with polyarticular-course JIA treated with SC or IV abatacept who achieved stringent composite end points maintain these end points over 21 months. This information may support the development of further treat-to-target strategies and aid discussions among families and care providers for children and adolescents with polyarticular-course JIA.

ACKNOWLEDGMENTS

The authors thank the patients and all the investigators who participated in the study. The authors acknowledge the support of Mary Swingle as the protocol manager. Professional medical writing and editorial assistance was provided by Robert Coover, PhD, and Fiona Boswell, PhD, at Caudex and was funded by Bristol Myers Squibb.

AUTHOR CONTRIBUTIONS

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

Study conception and design. Brunner, Avila-Zapata, Horneff, Wong, Zhuo, Martini, Lovell, Ruperto.

Acquisition of data. Brunner, Tzaribachev, Louw, Penades, Avila-Zapata, Horneff, Foeldvari, Kingsbury, Gastanaga, Wouters, Breedt, Askelson, Martini, Lovell, Ruperto.

Analysis and interpretation of data. Brunner, Tzaribachev, Louw, Avila-Zapata, Horneff, Foeldvari, Wouters, Wong, Askelson, Zhuo, Martini, Lovell, Ruperto.

ROLE OF THE STUDY SPONSOR

Bristol Myers Squibb 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 Bristol Myers Squibb.

ADDITIONAL DISCLOSURES

Authors Zhuo and Askelson are employees of Bristol Myers Squibb.

REFERENCES

1. Ringold S, Angeles-Han ST, Beukelman T, et al. 2019 American College of Rheumatology/Arthritis Foundation guideline for the treatment of juvenile idiopathic arthritis: therapeutic approaches for non-systemic polyarthritis, sacroiliitis, and enthesitis. Arthritis Care Res (Hoboken) 2019;71:717–34.
2. Ravelli A, Martini A. Juvenile idiopathic arthritis [review]. Lancet 2007; 369:767–78.
3. Brunner HI, Tzaribachev N, Vega-Cornejo G, et al. Subcutaneous abatacept in patients with polyarticular-course juvenile idiopathic arthritis: results from a phase III open-label study. Arthritis Rheumatol 2018;70:1144–54.
4. Ruperto N, Lovell DJ, Quartier P, et al. Abatacept in children with juvenile idiopathic arthritis: a randomised, double-blind, placebo-controlled withdrawal trial. Lancet 2008;372:383–91.

5. Lovell D, Ruperto N, Mouy R, et al. Long-term safety, efficacy, and quality of life in patients with juvenile idiopathic arthritis treated with intravenous abatacept for up to seven years. Arthritis Rheumatol 2015;67:2759–70.
6. Ruperto N, Brunner H, Tzaribachev N, et al. OP0056 maintenance of clinical response in individual children with juvenile idiopathic arthritis treated with subcutaneous abatacept. Ann Rheum Dis 2019;78 Suppl 2:99–100.
7. Klein A, Minden K, Hospach A, et al. Treat-to-target study for improved outcome in polyarticular juvenile idiopathic arthritis. Ann Rheum Dis 2020;79:969–74.
8. Giannini EH, Ruperto N, Ravelli A, et al. Preliminary definition of improvement in juvenile arthritis. Arthritis Rheum 1997;40:1202–9.
9. Stinson JN, Luca NJ, Jibb LA. Assessment and management of pain in juvenile idiopathic arthritis [review]. Pain Res Manag 2012;17: 391–6.
10. Nordal EB, Zak M, Aalto K, et al. Validity and predictive ability of the juvenile arthritis disease activity score based on CRP versus ESR in a Nordic population-based setting. Ann Rheum Dis 2012;71:1122–7.
11. Consolaro A, Ruperto N, Bazso A, et al. Development and validation of a composite disease activity score for juvenile idiopathic arthritis. Arthritis Rheum 2009;61:658–66.
12. Bazso A, Consolaro A, Ruperto N, et al. Development and testing of reduced joint counts in juvenile idiopathic arthritis. J Rheumatol 2009;36:183–90.
13. Hirschfeld G, Zernikow B. Cut points for mild, moderate, and severe pain on the VAS for children and adolescents: what can be learned from 10 million ANOVAs? Pain 2013;154:2626–32.
14. Guzman J, Gómez-Ramírez O, Jurencak R, et al. What matters most for patients, parents, and clinicians in the course of juvenile idiopathic arthritis? A qualitative study. J Rheumatol 2014;41: 2260–9.
15. Ruperto N, Ravelli A, Pistorio A, et al. Cross-cultural adaptation and psychometric evaluation of the Childhood Health Assessment Questionnaire (CHAQ) and the Child Health Questionnaire (CHQ) in 32 countries. Review of the general methodology. Clin Exp Rheumatol 2001;4 Suppl 23:S1–9.

Arthritis Care & Research
Vol. 75, No. 11, November 2023, pp 2267–2276
DOI 10.1002/acr.25134

American College of Rheumatology
Empowering Rheumatology Professionals

Improving Outcomes of Pediatric Lupus Care Delivery With Provider Goal-Setting Activities and Multidisciplinary Care Models

Joyce C. Chang,[1] Shreya A. Varghese,[2] Edward M. Behrens,[3] Sabrina Gmuca,[3] Jane S. Kennedy,[3] Emily J. Liebling,[3] Melissa A. Lerman,[3] Jay J. Mehta,[3] Beth H. Rutstein,[3] David D. Sherry,[3] Cory J. Stingl,[3] Lehn K. Weaver,[3] Pamela F. Weiss,[3] and Jon M. Burnham[3]

Objective. The present study was undertaken to evaluate high-quality care delivery in the context of provider goal-setting activities and a multidisciplinary care model using an electronic health record (EHR)–enabled pediatric lupus registry. We then determined associations between care quality and prednisone use among youth with systemic lupus erythematosus (SLE).

Methods. We implemented standardized EHR documentation tools to autopopulate a SLE registry. We compared pediatric Lupus Care Index (pLCI) performance (range 0.0–1.0; 1.0 representing perfect metric adherence) and timely follow-up 1) before versus during provider goal-setting activities and population management, and 2) in a multidisciplinary lupus nephritis versus rheumatology clinic. We estimated associations between pLCI and subsequent prednisone use adjusted for time, current medication, disease activity, clinical features, and social determinants of health.

Results. We analyzed 830 visits by 110 patients (median 7 visits per patient [interquartile range 4–10]) over 3.5 years. The provider-directed activity was associated with improved pLCI performance (adjusted β 0.05 [95% confidence interval (95% CI) 0.01, 0.09]; mean 0.74 versus 0.69). Patients with nephritis in multidisciplinary clinic had higher pLCI scores (adjusted β 0.06 [95% CI 0.02, 0.10]) and likelihood of timely follow-up than those in rheumatology (adjusted relative risk [RR] 1.27 [95% CI 1.02, 1.57]). A pLCI score of ≥0.50 was associated with 0.72-fold lower adjusted risk of subsequent prednisone use (95% CI 0.53, 0.93). Minoritized race, public insurance, and living in areas with greater social vulnerability were not associated with reduced care quality or follow-up, but public insurance was associated with higher risk of prednisone use.

Conclusion. Greater attention to quality metrics is associated with better outcomes in childhood SLE. Multidisciplinary care models with population management may additionally facilitate equitable care delivery.

INTRODUCTION

There is a need to identify strategies to improve outcomes of children with pediatric-onset systemic lupus erythematosus (pSLE) and related conditions and to ensure equitable care delivery. In adults with SLE, delivery of recommended care processes has been associated with better outcomes, including lower damage accrual (1). However, for youth with pSLE, considerable variation in care process completion exists (2,3), and literature on methods to standardize and evaluate adherence to care processes in the pediatric setting remains sparse (4). Our center has previously developed a composite index of 13 recommended pSLE care metrics, the pediatric Lupus Care Index (pLCI), to assess care quality in pSLE across 3 domains: clinical assessment, comorbidity management, and population management (5). Using the pLCI, we identified provider-level variation in performance as well as areas in need of practice-level improvement, which informed the design of a Maintenance of Certification

Dr. Chang's work was supported by the NIH (grant K23-HL-148539). Drs. Chang and Burnham's work was supported by GSK (investigator-sponsored research grant).

[1]Joyce C. Chang, MD, MSCE: Children's Hospital of Philadelphia and the University of Pennsylvania, Philadelphia, and Boston Children's Hospital and Harvard Medical School, Boston, Massachusetts; [2]Shreya A. Varghese, MS: Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; [3]Edward M. Behrens, MD, Sabrina Gmuca, MD, MSCE, Jane S. Kennedy, MSW, Emily J. Liebling, MD, Melissa A. Lerman, MD, PhD, MSCE, Jay J. Mehta, MD, MS, Beth H. Rutstein, MD, MSCE, David D. Sherry, MD, Cory J. Stingl, MD, Lehn K. Weaver, MD, PhD, Pamela F. Weiss, MD, MSCE, Jon M. Burnham, MD, MSCE: Children's Hospital of Philadelphia and the University of Pennsylvania, Philadelphia.

Author disclosures are available online at https://onlinelibrary.wiley.com/doi/10.1002/acr.25134.

Address correspondence via email to Jon M. Burnham, MD, MSCE, at burnhams@chop.edu.

Submitted for publication December 16, 2022; accepted in revised form April 13, 2023.

SIGNIFICANCE & INNOVATIONS

- Provider-level, self-directed goal setting activities can improve performance on care quality metrics for children with lupus.
- Multidisciplinary care models with social work support and population management strategies are associated with better care quality and timely follow-up care.
- Greater adherence to a composite index of care quality metrics for pediatric lupus is associated with reduced likelihood of any prednisone use at each subsequent visit.
- Standardized documentation of lupus characteristics and clinical assessments can facilitate both measurement of high-quality care delivery as well as research.

(MOC) activity to improve pLCI performance using goal-setting activities and self-evaluations. At the same time, we established a multidisciplinary care model for youth with lupus nephritis with a separate population management strategy.

One of the major challenges of evaluating programmatic interventions is that data collection methods commonly employed for research and quality improvement efforts are labor intensive and often unsustainable (6). In an effort to address this challenge, we developed lupus-specific, electronic health record (EHR)–enabled tools to standardize clinical documentation with embedded discrete data that could be used to autopopulate an observational pSLE research registry. The tools were designed to overcome limitations of billing code and prescription databases by capturing clinical disease manifestations, disease activity measures, and provider-documented medication instructions within the context of routine clinical workflow.

The objectives of this study were to leverage an EHR-enabled pSLE research registry 1) to evaluate the effectiveness of 2 different programmatic changes (an MOC activity and a multidisciplinary lupus nephritis care model) to improve high-quality care delivery; and 2) to determine whether pLCI performance is associated with relevant clinical outcomes. We hypothesized that provider self-evaluation and goal setting can improve pLCI performance while population management strategies help ensure timely follow-up. We also hypothesized that higher pLCI scores are associated with reduced likelihood of any prednisone use, a frequent cause of treatment-related morbidity, among children with pSLE.

MATERIALS AND METHODS

Study design. This was a retrospective analysis of a prospective observational database of youth with SLE and mixed connective tissue disease (MCTD) followed at our tertiary care pediatric center. An exemption for secondary use of clinical data and waiver of informed consent was granted by the institutional review board (IRB 19-016207).

Data source. We extracted data from our EHR-based pSLE research registry from December 2018 to July 2022. In December 2018, we implemented EHR tools at our center to standardize documentation of patient-level pSLE manifestations and treatment history (lupus history form), as well as visit-level data for each clinical encounter (lupus visit form), including medication instructions, disease activity, disease damage, and target assessments. Discrete data elements were embedded into the standardized documentation tools, which autopopulated a quality improvement dashboard and the pSLE research registry. The research registry additionally interfaced with an EHR-based steroid registry and billing data on hospital and rheumatology visit encounters in real-time (see Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25134). The EHR-based steroid registry uses an internally validated algorithm to determine when patients first met criteria for chronic glucocorticoid prescriptions (≥15 days) and when at least 18 months had elapsed since the last active prescription (7).

Setting. Coinciding with the development of EHR-based tools for pSLE, our center established a Lupus Program in 2018 with input from providers, patients and families, and representatives from patient advocacy groups. As of September 2018, programmatic components included a multidisciplinary lupus nephritis clinic with 2 rheumatologists, 2 nephrologists, a dedicated social worker, and a psychologist. Each month, the multidisciplinary team met to discuss population management, including outreach to patients in need of follow-up visits. To ensure delivery of high-quality care for all patients with pSLE and MCTD seen in the rheumatology clinic, we began an MOC activity in July 2020 to improve performance on metrics in the previously published pLCI (Figure 1), which is a composite measure of 13 quality indicators across 3 domains (standardized clinical assessment, comorbidity assessment and prevention, and population management) (5). Components of the MOC activity included a population management strategy, self-directed evaluation for individual providers, and goal-setting activities. Clinicians reviewed their pLCI performance, identified opportunities for improvement, and set performance goals for the next 3 months (see Supplementary Figure 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25134). To obtain MOC credit, 3 goal-setting activities were required. Four of 11 providers completed 2 activities, 4 of 11 completed 3 activities, and 3 of 11 completed 4 activities over 12 months.

The baseline visit for each patient was defined as the first visit occurring any time after EHR tool implementation in December 2018. Index visits for both pre-MOC (December 2018 to July 2020) and MOC activity periods (July 2020 to July 2022) were

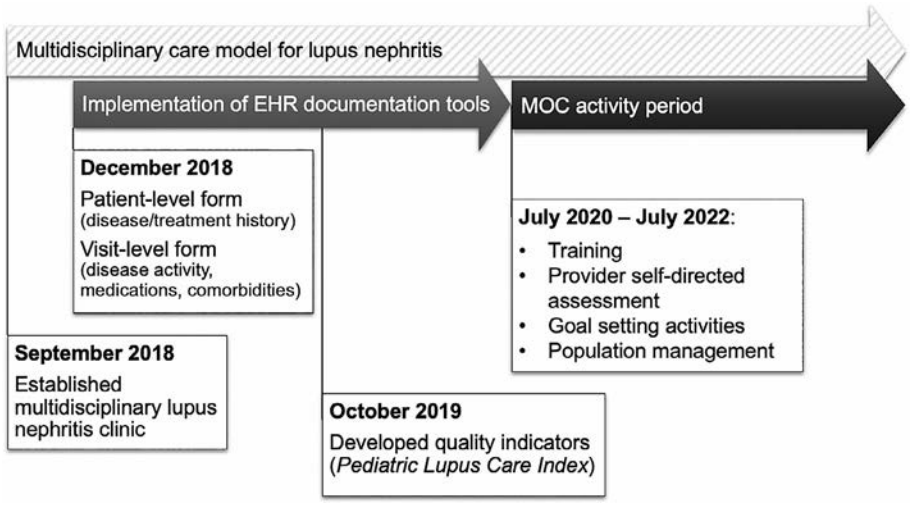

Figure 1. Timeline of interventions, beginning from the initial establishment of a multidisciplinary lupus nephritis clinic and implementation of electronic health record (EHR) documentation templates to standardize lupus clinical assessments at each rheumatology visit and patient-level summaries of disease and treatment histories. A set of quality indicators for pediatric lupus care delivery was developed in 2019 focusing on 3 domains: clinical assessment; comorbidity prevention; and population management. A Maintenance of Certification (MOC) activity was subsequently initiated at our center in July 2020 to improve performance on these quality indicators.

defined as the first visit for each patient occurring in the corresponding activity period. Patients with <2 outpatient rheumatology visits during the observation period were excluded from analysis (see Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25134).

Measures. The primary outcome for high-quality care delivery was total pLCI score (5), modified to exclude influenza vaccination (due to incomplete source documentation of influenza vaccinations outside the health system), and documentation of adrenal insufficiency in the problem list from the numerator and denominator (due to non-rheumatology providers managing the problem list and challenges resolving timestamps). Thus, the denominator for the modified pLCI was 11 for patients with a diagnosis of SLE and 9 for those with MCTD, where completion of 11 of 11 or 9 of 9 metrics (100% adherence) constituted a pLCI of 1.0, respectively. As a secondary outcome, we defined timely follow-up care as <120 days between clinic visits to evaluate the population management components of the interventions. The primary exposure was the MOC activity (before versus during the MOC activity period). For the subgroup of patients with lupus nephritis, we also evaluated associations between exposure to the multidisciplinary clinic model and care delivery outcomes. To evaluate longitudinal relationships between pLCI and improved clinical outcomes, namely a lower likelihood of any glucocorticoid requirement, we assessed prednisone use at each subsequent visit as a binary outcome (started/continued versus not prescribed/discontinued). Prednisone use was determined by discrete, provider-entered prednisone instructions embedded in the lupus visit form when available or otherwise determined using the prescription-based EHR steroid registry (see Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25134).

Covariates. Patient-level factors included age at the baseline visit, sex, race and ethnicity (as reported in the medical record), Social Vulnerability Index (SVI) derived from census tract codes, and insurance type. Due to the known spatial polarization of Black neighborhoods as well as Hispanic neighborhoods (irrespective of race) in Philadelphia where our center was located (8), we analyzed mutually exclusive race and ethnicity categories as follows: Asian alone or in combination; Black or African American alone or in combination; Hispanic ethnicity with any other race; Non-Hispanic other/unknown race; and Non-Hispanic White race. We also considered major organ manifestations, including history of nephritis, central nervous system involvement, and serositis. Visit-level time-varying factors considered included recent disease diagnosis (duration <6 months), prednisone use, use of disease-modifying antirheumatic drugs (DMARDs), and Systemic Lupus Erythematosus Disease Activity Index 2000 (SLEDAI-2K) scores categorized into active disease (SLEDAI-2K score >4), low/inactive disease (SLEDAI-2K score ≤4), or not assessed (9). Rules applied for calculating SLEDAI scores in the setting of missing SLEDAI components are described in Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25134.

Statistical analysis. Characteristics of patients at their index visit before versus during the MOC activity were compared using standard descriptive statistics, including chi-square or Fisher's exact tests for categorical variables and Student's *t*-tests

or Wilcoxon's rank sum tests for continuous variables, as appropriate.

We used linear mixed-effects models to estimate differences in continuous outcomes and modified (robust) Poisson models to estimate relative risks for binary outcomes, adjusted for time since the baseline visit, patient-level factors (age, sex, race and ethnicity, SVI, insurance, major organ manifestations), visit-level factors (disease duration <6 months from diagnosis, current prednisone use), as well as within-subject random effects. We also ran separate models restricted to SLE patients, additionally adjusted for disease activity and/or DMARD use. Separate subgroup analyses were conducted in patients with lupus nephritis to estimate differences in outcomes by clinic setting (multidisciplinary versus rheumatology only).

We considered pLCI quartiles as potential cut points for predicting subsequent prednisone use, and model performance was compared using Akaike and Bayesian information criteria to select a minimum threshold at which care quality may associate with differential outcomes. We conducted a sensitivity analysis limited to visits for which provider-entered medication instructions were available compared with results including prescription-based registry data. To address potential nonrandom missingness, we simulated the potential range of point estimates if prednisone use or nonuse was assumed for all visits missing prednisone use data. All analyses were conducted using Stata, version 16.0.

RESULTS

Of 133 cases of pSLE or MCTD in the EHR-based pSLE registry, 128 had both patient-level and visit-level form data available. A total of 110 patients had at least 2 outpatient rheumatology visits during the observation period, 74 of which had follow-up extending through both pre-MOC and MOC periods. There was a median of 7 visits per subject (interquartile range [IQR] 4–10), comprising a total of 830 outpatient rheumatology visits and 720 follow-up intervals. The standardized EHR documentation tool was used in 79% of visits before the MOC activity and 87% of visits during the MOC activity. Data contributing to both SLEDAI scores and medication usage was captured for 76% of pre-MOC visits and 80% of visits during MOC.

Evaluating performance in high-quality care delivery during an MOC activity. Demographic and clinical characteristics at the index visits in both pre-MOC and MOC activity periods were stable over both periods as shown in Table 1. Approximately one-half of patients were publicly insured, over one-fourth lived in neighborhoods with the highest social vulnerability, 39% were Black, 20% were Asian, and 11–13% reported Hispanic ethnicity. One-third had been diagnosed with lupus nephritis. Hydroxychloroquine use was nearly universal, and a majority (72–73%) of patients had been treated with mycophenolate.

pLCI performance over time. Median unadjusted pLCI scores were 0.7 (IQR 0.5–0.8) during the pre-MOC period when EHR documentation tools were available for use and 0.8 (IQR 0.6–0.9) during the MOC activity period (see Supplementary Figure 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25134). On average, pLCI scores increased over time by 0.04 per year of follow-up (95% confidence interval [95% CI] 0.01, 0.07), adjusted for sociodemographic characteristics, nephritis, neurologic involvement, disease duration, and prednisone use. The MOC activity period was additionally associated with a modest but statistically significant 0.05 unit average increase in pLCI scores (95% CI 0.01, 0.09) (marginal mean 0.74 versus 0.69). Insurance status, race and ethnicity, and SVI were not significantly associated with pLCI scores on either unadjusted or adjusted analyses (Table 2). A history of nephritis and current prednisone use were associated with higher average pLCI scores, while visits occurring within 6 months of initial diagnosis were associated with lower scores. During the MOC activity period, improvements in pLCI were driven by completion of clinical assessments (Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index, physician global, target attestation, lupus characteristics review, as well as comorbidity assessment [blood pressure, lipid, and vitamin D]) (Figure 2).

Timely outpatient rheumatology follow-up. There was no significant increase in timely follow-up during the MOC activity versus pre-MOC period in adjusted models (67% versus 61%; adjusted relative risk [RR] 1.10 [95% CI 0.94, 1.29]) (Table 3). Upon restricting the analysis to patients with SLE, additional adjustment for SLEDAI scores and DMARD use yielded similar results (65% versus 60%; RR 1.07 [95% CI 0.90, 1.28]) (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25134). Hispanic ethnicity, younger age, disease duration <6 months, and current prednisone use were associated with a higher likelihood of timely follow-up (Table 3). Of note, higher SVI, Black race, and insurance status were not associated with a lower likelihood of timely follow-up during the observation period.

Lupus nephritis care quality in the context of a multidisciplinary care model. In the subgroup of 35 patients with lupus nephritis (comprising a total of 252 visits), we similarly observed a 0.07-unit adjusted increase in pLCI associated with the MOC activity period (95% CI 0.02, 0.12; $P = 0.01$). A total of 19 patients with lupus nephritis (of which 47% had proliferative disease, 37% pure membranous) were evaluated in the multidisciplinary clinic at least once during the observation period; the remaining 16 patients (69% proliferative, 19% membranous) were followed exclusively in general rheumatology clinic. Those evaluated in multidisciplinary clinic were on average younger

Table 1. Patient characteristics at index visits before and during goal-setting activity*

	Pre-MOC (n = 88)	MOC (n = 96)†	*P*
Disease type			
SLE	76 (86)	84 (88)	0.82
MCTD/overlap syndrome	12 (14)	12 (12)	
Age at SLE diagnosis, mean ± SD years	13.4 ± 3.5	13.4 ± 3.4	0.94
Age at index visit, mean ± SD years	16.2 ± 3.1	16.6 ± 3.4	0.37
Female sex	70 (80)	79 (82)	0.64
Disease duration, median (IQR) years	1.8 (0.3–4.7)	2.6 (0.3–4.8)	0.34
Race			
Asian alone or in combination‡	17 (20)	19 (20)	0.84
Black alone or in combination‡	34 (39)	37 (39)	
Other	10 (11)	15 (16)	
White	26 (30)	24 (25)	
Unknown	0 (0)	1 (1)	
Hispanic ethnicity	10 (11)	12 (13)	0.83
Insurance			
Public (Medicaid)	45 (52)	50 (52)	0.86
Private	40 (46)	45 (47)	
Self-pay/other/uninsured	2 (2)	1 (1)	
Social Vulnerability Index			
Lowest	22 (25)	29 (30)	0.87
Medium low	19 (22)	19 (20)	
Medium high	23 (26)	23 (24)	
Highest	23 (26)	25 (26)	
Unknown	1 (1)	0 (0)	
Multidisciplinary clinic setting at index visit	6 (7)	12 (13)	0.19
Historical lupus manifestations			
Arthritis	43 (49)	48 (51)	0.77
Serositis	7 (8)	8 (9)	0.93
Nephritis	29 (33)	30 (32)	0.84
Neuropsychiatric	5 (6)	5 (5)	0.93
dsDNA antibody positive	65 (74)	70 (76)	0.73
Hypocomplementemia	59 (68)	67 (73)	0.46
Lupus treatments (ever use)			
Hydroxychloroquine	86 (98)	95 (99)	0.51
Mycophenolate	63 (72)	70 (73)	0.84
Methotrexate	31 (35)	32 (33)	0.79
Rituximab	30 (34)	29 (30)	0.57
Cyclophosphamide	14 (16)	13 (14)	0.65
Azathioprine	8 (9)	9 (9)	0.95
Belimumab	5 (6)	6 (6)	0.87
Calcineurin inhibitor	4 (5)	5 (5)	1.00
Disease status and treatment at index visit (current use)			
SLEDAI-2K score, median (IQR)	0 (0–4)	0 (0–4)	0.82
Any DMARD use§	55 (93)	51 (88)	0.33
Prednisone use	26 (31)	25 (29)	0.75
Prednisone dose in users, median (IQR) mg/day	10 (10–30)	10 (5–30)	0.59

* Values are the number (%) unless indicated otherwise. Comparison of individual patient characteristics at each index visit, defined as the first visit occurring in each period before (December 2018 to June 2020) or during the MOC activity (July 2020 to July 2022), using Student's *t*-test or rank sum test for continuous variables and chi-square or Fisher's exact tests (n < 5) for categorical variables. DMARD = disease-modifying antirheumatic drug; dsDNA = double-stranded DNA; IQR = interquartile range; MCTD = mixed connective tissue disease; MOC = Maintenance of Certification; SLE = systemic lupus erythematosus; SLEDAI-2K = Systemic Lupus Erythematosus Disease Activity Index 2000.
† Samples not mutually exclusive; n = 74 individuals had follow-up extending through both periods.
‡ Includes 3 individuals reporting multiple races (Asian category: Asian and White; Black category: Black and American Indian, Black and Asian). Asian category is inclusive of Asian (n = 20) and Indian (n = 2) as recorded in the medical record.
§ DMARDs include mycophenolate, azathioprine, methotrexate, calcineurin inhibitors, sirolimus, and belimumab.

(mean ± SD age 14.7 ± 3.5 years versus 17.1 ± 2.7 years; $P = 0.03$), with higher disease activity at the baseline visit (median SLEDAI-2K score 6 [IQR 2–11] versus 0 [IQR 0–2], $P < 0.01$), and the majority were publicly insured (63% versus 25%; $P = 0.06$). The median pLCI at each multidisciplinary clinic visit was 0.91 (IQR 0.82–1.00) (n = 118) compared to 0.81 (IQR 0.60–0.90)

Table 2. Factors associated with the pediatric Lupus Care Index over time*

	Unadjusted†		Adjusted	
	β (95% CI)	*P*	β (95% CI)	*P*
Month of follow-up	0.007 (0.006, 0.008)	<0.01	0.003 (0.001, 0.006)	<0.01
MOC activity period	0.04 (–0.01, 0.08)	0.12	0.05 (0.01, 0.09)	0.04
Age at baseline visit	0.002 (–0.01, 0.01)	0.72	–0.01 (–0.02, 0.01)	0.24
Male sex	0.06 (–0.03, 0.15)	0.19	0.02 (–0.06, 0.10)	0.60
Race and ethnicity				
Asian alone or in combination	0.03 (–0.07, 0.14)	0.56	0.04 (–0.05, 0.13)	0.36
Black alone or in combination	0.06 (–0.03, 0.15)	0.18	0.04 (–0.05, 0.12)	0.38
Hispanic, other/White race	0.07 (–0.05, 0.19)	0.23	0.07 (–0.04, 0.17)	0.20
Other/unknown race, Non-Hispanic	0.02 (–0.12, 0.16)	0.77	0.02 (–0.10, 0.14)	0.80
White, Non-Hispanic	Ref.			
Social Vulnerability Index				
Lowest	Ref.			
Medium low	0.01 (–0.09, 0.11)	0.88	0.02 (–0.06, 0.11)	0.56
Medium high	0.02 (–0.08, 0.11)	0.73	0.02 (–0.07, 0.11)	0.68
Highest	0.04 (–0.05, 0.13)	0.36	0.06 (–0.04, 0.15)	0.23
Insurance				
Private	Ref.			
Medicaid	0.03 (–0.04, 0.1)	0.36	–0.02 (–0.09, 0.04)	0.48
Self-pay/uninsured	0.10 (–0.12, 0.32)	0.38	0.02 (–0.17, 0.20)	0.87
History of nephritis	0.16 (0.09, 0.22)	<0.01	0.08 (0.02, 0.14)	0.01
History of neurologic involvement	0.12 (–0.03, 0.27)	0.11	0.07 (–0.05, 0.19)	0.24
Recent diagnosis (<6 months)‡	–0.09 (–0.14, –0.05)	<0.01	–0.12 (–0.17, –0.08)	<0.01
Current prednisone use‡	0.14 (0.1, 0.18)	<0.01	0.13 (0.09, 0.18)	<0.01

* Mixed effects linear regression with patient-level random intercept (n = 683 visits for 107 unique patients). Three patients and 37 visits were dropped due to missing data for ≥1 demographic characteristics (1 patient) or prednisone use (34 visits). 95% CI = 95% confidence interval; MOC = Maintenance of Certification; Ref. = reference.
† All univariable analyses are additionally adjusted for time (month of follow-up).
‡ Time-varying covariates.

(n = 134) for rheumatology clinic visits ($P < 0.01$, unadjusted). On average, multidisciplinary visits were associated with a 0.06 higher pLCI compared to rheumatology clinic visits (95% CI 0.02, 0.10; $P = 0.01$; marginal mean 0.87 versus 0.82), adjusted for time, sociodemographic factors, disease duration, disease activity, and prednisone use. Furthermore, patients who received care in the multidisciplinary clinic at any time during follow-up had a 0.10 higher average pLCI compared to those receiving care exclusively in the rheumatology clinic (95% CI 0.03, 0.16; $P < 0.01$; marginal mean 0.88 versus 0.78). This was driven by greater completion of disease characteristics review, pneumococcal vaccination, assessment of disease activity and disease damage, as well as follow-up within 180 days (data not shown).

As with the full cohort, there was no statistically significant increase in timely outpatient follow-up within 120 days for the lupus nephritis subgroup during the MOC activity versus pre-MOC periods (77% versus 65%; adjusted RR 1.17 [95% CI 0.86, 1.61]). However, being seen in the multidisciplinary clinic was associated with significantly higher likelihood of timely follow-up compared to being seen in rheumatology clinic (adjusted RR 1.27 [95% CI 1.02, 1.57], $P = 0.03$). Black race and Hispanic ethnicity were also independently associated with greater likelihood of timely follow-up compared to non-Hispanic White race (adjusted RR 1.50 [95% CI 1.11, 2.02], $P = 0.01$, and 1.37 [95% CI 1.02, 1.84], $P = 0.03$, respectively). Moreover, patients with lupus nephritis living in census tracts with medium-high social vulnerability were also 1.45 times more likely to have timely follow-up compared to those from census tracts with the lowest SVI (95% CI 1.01, 2.08; $P = 0.05$) (see Supplementary

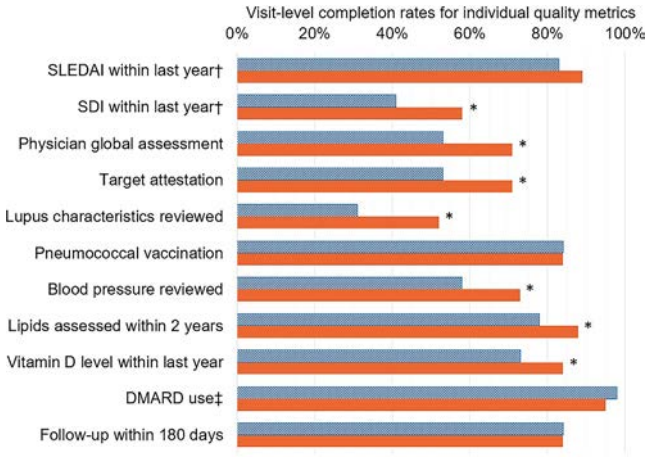

Figure 2. Bar graph representing the total proportion of visits at which each quality metric was met, both before (hatched blue) and during (solid orange) the Maintenance of Certification (MOC) activity. * = $P < 0.05$, unadjusted; † = Denominator excludes visits for patients with mixed connective tissue disease; ‡ = Denominator limited to visits meeting criteria for chronic steroid use. DMARD = disease-modifying antirheumatic drug; SDI = Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index; SLEDAI = Systemic Lupus Erythematosus Disease Activity Index.

Table 3. Factors associated with timely outpatient rheumatology follow-up*

	Adjusted RR (95% CI)	*P*
Month of follow-up	0.99 (0.98, 1.00)	0.02
MOC activity period	1.10 (0.94, 1.29)	0.24
Age at baseline visit, years	0.96 (0.93, 0.99)	0.01
Male sex	0.93 (0.74, 1.17)	0.55
Race and ethnicity		
Asian alone or in combination	1.07 (0.81, 1.41)	0.64
Black alone or in combination	1.21 (0.91, 1.61)	0.20
Hispanic White/other	1.34 (1.04, 1.72)	0.02
Other/unknown race, Non-Hispanic	0.63 (0.36, 1.12)	0.12
White, Non-Hispanic (ref.)	–	
Social Vulnerability Index		
Lowest (ref.)	–	
Medium low	1.07 (0.81, 1.42)	0.64
Medium high	0.95 (0.69, 1.29)	0.73
Highest	1.06 (0.81, 1.38)	0.67
Insurance		
Medicaid	1.06 (0.87, 1.28)	0.58
Self-pay/uninsured	1.10 (0.88, 1.38)	0.41
History of nephritis	1.14 (0.97, 1.33)	0.11
History of neurologic manifestations	1.03 (0.76, 1.38)	0.87
Recent diagnosis (<6 months)†	1.29 (1.09, 1.54)	0.00
Current prednisone use†	1.18 (1.01, 1.39)	0.04

* Estimates from modified robust Poisson models with subject-level random effects (n = 107 patients with systemic lupus erythematosus/mixed connective tissue disease; 683 of 730 visits with complete data). 95% CI = 95% confidence interval; MOC = Maintenance of Certification; Ref. = reference; RR = relative risk.
† Time-varying covariate.

Table 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25134).

Association between high-quality care delivery and clinical outcomes. Medication instructions were entered by providers in the lupus visit form for 106 of 110 patients for a median of 90% of visits per patient during the observation period (IQR 73–100%), corresponding to a visit-level completion rate of 669 of 830 (81%). Current use of prednisone was documented in 279 of 669 (41.7%) visits with any medication template entry, and a discrete prednisone dose was also entered for 254 of 279 (91%) instances of documented prednisone use. For the remaining 161 visits without provider-entered medication instructions, active (n = 23) versus inactive (n = 99) prednisone prescriptions were obtained from the EHR-based steroid registry for 122 of 161 (76%) visits, yielding 39 of 830 (5%) visits for which prednisone use data were missing. No patients with nephritis were missing prednisone use data. Shorter follow-up time and lower pLCI performance were associated with missing prednisone data (see Supplementary Table 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25134).

In the total SLE/MCTD cohort, a pLCI score of ≥0.5 versus <0.5 (lower quartile) was associated with an adjusted RR of 0.62 (95% CI 0.48, 0.82; $P < 0.01$) for subsequent prednisone use at the next visit, adjusted for current prednisone use, disease duration, baseline major organ involvement, and sociodemographic factors. Upon restricting the analysis to SLE patients only and further adjusting for current disease activity, a pLCI score of ≥0.5 was associated with a 0.72-fold lower risk of subsequent prednisone use (95% CI 0.53, 0.93; $P = 0.01$). Compared to privately insured patients, those with public insurance were 1.4 times more likely to require prednisone at the subsequent visit (Table 4). In contrast, medium-high SVI was significantly associated with a 0.66-fold reduction in risk of prednisone use. In addition, the lack of SLEDAI score assessment at each visit was also independently associated with a 1.25-fold (95% CI 1.04, 1.49) and 1.42-fold (95% CI 1.16, 1.73) higher likelihood of subsequent prednisone use compared to an assessment of active disease (SLEDAI score >4) or low disease activity (SLEDAI score ≤4), respectively. Results were similar in a sensitivity analysis limited to visits in which provider-entered prednisone use instructions were available (adjusted RR 0.70 for a pLCI score of ≥0.5 versus <0.5 [95% CI 0.50, 0.98], $P = 0.04$). Assuming prednisone use at all visits missing prednisone data versus nonuse yielded a range of

Table 4. Association between pediatric Lupus Care Index (pLCI) performance and subsequent prednisone use*

	RR (95% CI)	*P*
Month of follow-up	1.00 (1.00, 1.01)	0.12
pLCI score ≥0.5†	0.72 (0.54, 0.96)	0.03
Age at baseline visit	1.01 (0.98, 1.04)	0.49
Male sex	1.25 (0.97, 1.60)	0.08
Social Vulnerability Index		
Lowest	Ref.	
Medium low	1.04 (0.81, 1.33)	0.75
Medium high	0.66 (0.48, 0.90)	0.01
Highest	0.85 (0.69, 1.05)	0.14
Race and ethnicity		
Asian alone or in combination	0.98 (0.73, 1.33)	0.92
Black alone or in combination	1.20 (0.91, 1.60)	0.20
Hispanic, other/White race	1.23 (0.93, 1.63)	0.15
Other/unknown race, Non-Hispanic	1.38 (0.97, 1.98)	0.08
White, Non-Hispanic	Ref.	
Insurance		
Medicaid	1.42 (1.17, 1.72)	<0.01
Self-pay/uninsured	0.94 (0.61, 1.44)	0.77
History of nephritis	1.13 (0.91, 1.40)	0.26
History of neurologic manifestations	1.45 (1.02, 2.06)	0.04
History of serositis	1.21 (1.00, 1.47)	0.05
Recent diagnosis (<6 months)†	1.40 (1.19, 1.64)	0.00
Current prednisone use†	9.36 (6.12, 14.31)	0.00
Current disease activity†		
SLEDAI-2K score >4	Ref.	
SLEDAI-2K score ≤4	0.88 (0.75, 1.04)	0.14
SLEDAI-2K score not assessed	1.25 (1.04, 1.49)	0.02

* Factors associated with any prednisone use at each subsequent visit in 92 patients with systemic lupus erythematosus only (n = 591 visits). The range for the pLCI is 0.0–1.0. 95% CI = 95% confidence interval; Ref. = reference; RR = relative risk; SLEDAI-2K = Systemic Lupus Erythematosus Disease Activity Index 2000.
† Time-varying covariate.

possible RRs from 0.82 (95% CI 0.63, 1.07) to 0.68 (95% CI 0.51, 0.92), respectively.

DISCUSSION

At our center, uptake of pSLE-specific EHR documentation tools was high in the context of interventions to improve high-quality care. Without any manual chart abstraction, we were able to evaluate changes over time in care quality performance and clinical outcomes in the setting of 2 programmatic changes. We observed small but statistically significant improvements in care quality metric performance over time with a relatively low-intensity intervention involving provider self-directed goal-setting activities. At the same time, a complex intervention involving a multidisciplinary care model for patients with lupus nephritis was associated with higher care quality performance and more timely follow-up care. Importantly, in the context of these interventions, there were no significant disparities in care quality metrics by race and ethnicity, insurance status, or neighborhood-level social vulnerability indices. Finally, we also demonstrated a longitudinal relationship between high-quality care delivery and reduced likelihood of any prednisone use, suggesting that improving adherence to care quality metrics may also improve clinical outcomes.

While there is a significant body of literature describing care process measures for SLE, very few studies have been able to address how improving recommended care processes translates to better clinical outcomes (1). A previous study in children with SLE did not find that adherence to individual quality metrics is associated with decreased damage (3). However, by using a composite index of care quality, we demonstrate that adherence to at least 50% of care quality indicators is associated with significant reductions in the risk of subsequent prednisone use. Interestingly, the association between lack of SLEDAI assessment at a given visit and a higher likelihood of subsequent prednisone use raises the question of whether even more frequent disease activity assessment, such as in a treat-to-target framework (10), could further reduce prednisone exposure and related toxicity. As a composite measure, pLCI performance represents overall attention to best practices; therefore, the relationship between pLCI scores and prednisone use may be driven by factors other than the individual care processes. Of note, public insurance remained associated with a higher risk of subsequent prednisone use, independent of care quality. This suggests that either the pLCI does not fully capture processes driving prednisone use, or that systems and individual-level factors mediate the relationship between insurance status and prednisone use. Studies have demonstrated that Black, Asian, and Hispanic adults with SLE receive higher maximum prednisone doses and over longer periods of time compared to their White counterparts (11,12). Frequent outpatient care was protective, with Black and Hispanic individuals having fewer visits (12). In our cohort, there were no significant disparities in timely rheumatology follow-up, which may have mitigated differences in prednisone use by race and ethnicity but does not explain differences by insurance status. It is possible that delays in initial access to subspecialty care or use of steroid-sparing agents drive this difference. Qualitative research may elucidate reasons for higher prednisone exposure among publicly insured children with SLE (13).

Measurement of health care quality has been proposed as an important mechanism through which disparities in care can be identified and ameliorated (14). We did not observe racial or socioeconomic disparities in quality metric performance during the study period, which spanned introduction of a dedicated lupus social worker, the multidisciplinary lupus nephritis clinic, the MOC activities, as well as the COVID-19 pandemic. This is notable, particularly as the pandemic both exposed and exacerbated existing disparities in many other contexts (15,16) and was ongoing throughout the MOC activity, potentially even attenuating improvements in care delivery outcomes. Furthermore, living in areas with medium-high social vulnerability was unexpectedly associated with a lower risk of subsequent prednisone use compared to living in areas with the lowest social vulnerability, which may reflect efforts to ensure timely follow-up care, particularly among patients with lupus nephritis with greater prednisone exposure. While we cannot determine which of the complex interventions contributed to maintaining equitable care delivery, our data suggest that interventions combining social work support with standardized clinical assessment and/or multidisciplinary care models to eliminate health disparities should be systematically studied. Hybrid effectiveness–implementation designs may be particularly well suited to the evaluation of these types of complex interventions (17).

While multidisciplinary care models are considered standard of care in oncology and have attained recognition at policy levels (18,19), published literature on the implementation of multidisciplinary care models for SLE remains sparse. One study demonstrated that a multidisciplinary care model for adults with lupus nephritis was associated with decreased time to kidney biopsy and improved performance on select quality metrics (20). In our study, we similarly observed better pLCI performance associated with a multidisciplinary care model for pediatric lupus nephritis, particularly with respect to routine assessment of disease activity and damage, which is a key component of treat-to-target approaches (10). As being seen in dedicated lupus clinics or by providers with higher volumes of lupus patients is also associated with higher care quality performance (21), it is not possible to directly attribute higher pLCI performance to the multidisciplinary components, such as collaboration, coordination, colocalization, interdisciplinary integration of knowledge, or patient-centered care (22). Nevertheless, our findings support the growing body of literature that suggests that patients with lupus would benefit from multidisciplinary chronic care models that are routinely implemented for other complex chronic conditions.

In addition to higher care quality, we also observed a greater likelihood of timely follow-up for patients with lupus nephritis associated with care in the multidisciplinary clinic compared to the general rheumatology clinic. Although population management was also a component of the MOC activity for all rheumatology providers, each provider could choose whether or not to set population management as a goal. In contrast, the multidisciplinary care team had a distinct population management strategy, in which the team confirmed the intended follow-up timeframe after each visit and also tracked and reviewed visit intervals on a monthly basis for all patients ever evaluated in the multidisciplinary clinic. The social worker played a key role in recontacting families prior to each visit and evaluating transportation needs or other barriers. Albeit challenging to quantify, the psychologist likely also facilitated a bidirectional relationship between better care delivery and patient engagement by addressing mental health needs in real-time. In this context, patients belonging to historically marginalized racial groups, living in areas with greater social vulnerability, or with public insurance were equally, if not more likely, to receive timely follow-up care. This has important implications for health equity, as access to appropriate outpatient care has been associated with receipt of recommended care (23) and has also been hypothesized to mediate disparities in renal outcomes by insurance status (24,25) and by race (11). As such, population management strategies that integrate dedicated social work support may help mitigate differential access to timely rheumatology follow-up and reduce health disparities.

Strengths of our study include availability of comprehensive, longitudinal data in a pediatric lupus cohort with racial, ethnic, and socioeconomic diversity. Use of lupus-specific EHR documentation tools enabled collection of discrete data at the point of care, eliminating manual review of unstructured data. These tools can potentially be implemented across institutions using a common data model and improve upon current examples of EHR-based learning health networks by providing access to disease activity scores and clinical phenotypes (26–28). There are also disadvantages of EHR-enabled registries that present limitations to the current study. Data completeness is dependent on uptake of EHR forms, which may not be easily replicated in other settings. In addition, data programming requirements limit spread to practices with limited information technology resources. However, as these efforts align with policies regarding "meaningful use" of EHRs, investment by institutions is warranted (29). Additional limitations of our study include incomplete generalizability, as academic centers tend to have higher performance on recommended SLE care (30), also evidenced by nearly 100% adherence to hydroxychloroquine prescribing in our study. Missing data were present and could have biased our point estimates to a certain degree, albeit unlikely to change the direction of the effect. Analysis of the impact of interventions on other relevant glucocorticoid outcomes, such as cumulative prednisone dose or tapering, would have required additional chart review; albeit this data infrastructure would still substantially reduce manual review and be more accurate than electronic prescriptions (31). Due to low damage scores in our cohort and inclusion of damage assessment in the pLCI, we could not evaluate relationships between pLCI and damage accrual. However, future studies combining EHR-enabled registries with prospective data could evaluate relationships between care quality and longer term outcomes. Last, changes over time cannot be directly attributed to any particular intervention in observational studies. We could not account for mitigation efforts and shielding behaviors during the COVID-19 pandemic, which could have decreased rates of follow-up and metric completion both in the months prior to and during the MOC activity. Reassuringly, the trends in rates we observed remained relatively stable over the study period.

In conclusion, an EHR-enabled pediatric lupus registry with standardized clinical documentation enables evaluation of longitudinal disease and care delivery outcomes without manual review of medical records. In a real-world population of youth with SLE, we demonstrated improvements in delivery of high quality care associated with provider-directed goal setting activities as well as a multidisciplinary care model for patients with lupus nephritis and implementation of population management strategies. In the context of these interventions, we did not observe any racial or socioeconomic disparities in timely outpatient rheumatology care or receipt of recommended care processes. Routine automated assessment of care processes and disease status can serve as an important means to ensure equitable care delivery and evaluate interventions designed to deliver comprehensive, patient-centered care.

AUTHOR CONTRIBUTIONS

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

Study conception and design. Chang, Burnham.

Acquisition of data. Chang, Varghese, Behrens, Gmuca, Kennedy, Liebling, Lerman, Mehta, Rutstein, Sherry, Stingl, Weaver, Weiss, Burnham.

Analysis and interpretation of data. Chang, Burnham.

ROLE OF THE STUDY SPONSOR

GSK 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 GSK.

REFERENCES

1. Yazdany J, Trupin L, Schmajuk G, et al. Quality of care in systemic lupus erythematosus: the association between process and outcome measures in the Lupus Outcomes Study. BMJ Qual Saf 2014;23:659–66.
2. Mina R, Harris JG, Klein-Gitelman MS, et al. Initial benchmarking of the quality of medical care in childhood-onset systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2016;68:179–86.

3. Harris JG, Maletta KI, Kuhn EM, et al. Evaluation of quality indicators and disease damage in childhood-onset systemic lupus erythematosus patients. Clin Rheumatol 2017;36:351–9.
4. Smitherman EA, Huang B, Furnier A, et al. Quality of care in childhood-onset systemic lupus erythematosus: report of an intervention to improve cardiovascular and bone health screening. J Rheumatol 2020;47:1506–13.
5. Burnham JM, Cecere L, Ukaigwe J, et al. Factors associated with variation in pediatric systemic lupus erythematosus care delivery. ACR Open Rheumatol 2021;3:708–14.
6. Austin JM, Kachalia A. The state of health care quality measurement in the era of COVID-19: the importance of doing better. JAMA 2020; 324:333–4.
7. Basiaga ML, Burrows EK, Denburg MR, et al. Variation in preventive care in children receiving chronic glucocorticoid therapy. J Pediatrics 2016;179:226–32.
8. The Pew Charitable Trusts, Philadelphia Research Initiative. A city transformed: the racial and ethnic changes in Philadelphia over the last 20 years. 2011. URL: https://www.pewtrusts.org/~/media/legacy/uploadedfiles/wwwpewtrustsorg/reports/philadelphia_research_initiative/PhiladelphiaPopulationEthnicChangespdf.pdf.
9. Gladman DD, Ibañez D, Urowitz MB. Systemic Lupus Erythematosus Disease Activity Index 2000. J Rheumatol 2002;29:288–91.
10. Smolen JS, Breedveld FC, Burmester GR, et al. Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. Ann Rheum Dis 2016;75:3–15.
11. Alarcón GS, McGwin G Jr, Bartolucci AA, et al. Systemic lupus erythematosus in three ethnic groups: IX. Differences in damage accrual. Arthritis Rheum 2001;44:2797–806.
12. Chandler MT, Santacroce LM, Costenbader KH, et al. Racial differences in persistent glucocorticoid use patterns among Medicaid beneficiaries with incident systemic lupus erythematosus. Semin Arthritis Rheum 2023;58:152122.
13. Lewandowski LB, Watt MH, Schanberg LE, et al. Missed opportunities for timely diagnosis of pediatric lupus in South Africa: a qualitative study. Pediatr Rheumatol Online J 2017;15:14.
14. Arora S, Yazdany J. Use of quality measures to identify disparities in health care for systemic lupus erythematosus [review]. Rheum Dis Clin North Am 2020;46:623–38.
15. Lopez L III, Hart LH III, Katz MH. Racial and ethnic health disparities related to COVID-19. JAMA 2021;325:719–20.
16. Perry BL, Aronson B, Pescosolido BA. Pandemic precarity: COVID-19 is exposing and exacerbating inequalities in the American heartland. Proc Natl Acad Sci 2021;118:e2020685118.
17. Curran GM, Bauer M, Mittman B, et al. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012;50:217–26.
18. Selby P, Popescu R, Lawler M, et al. The value and future developments of multidisciplinary team cancer care [review]. Am Soc Clin Oncol Educ Book 2019;39:332–40.
19. Borras JM, Albreht T, Audisio R, et al. Policy statement on multidisciplinary cancer care. Eur J Cancer 2014;50:475–80.
20. Garg S, Singh T, Panzer SE, et al. Multidisciplinary lupus nephritis clinic reduces time to renal biopsy and improves care quality. ACR Open Rheumatol 2022;4:581–86.
21. Arora S, Nika A, Trupin L, et al. Does systemic lupus erythematosus care provided in a lupus clinic result in higher quality of care than that provided in a general rheumatology clinic? Arthritis Care Res (Hoboken) 2018;70:1771–7.
22. Struckmann V, Leijten FR, van Ginneken E, et al. Relevant models and elements of integrated care for multi-morbidity: results of a scoping review [review]. Health Policy 2018;122:23–35.
23. Yazdany J, Feldman CH, Liu J, et al. Quality of care for incident lupus nephritis among Medicaid beneficiaries in the United States. Arthritis Care Res (Hoboken) 2014;66:617–24.
24. Ward MM. Medical insurance, socioeconomic status, and age of onset of endstage renal disease in patients with lupus nephritis. J Rheumatol 2007;34:2024–7.
25. Ward MM. Access to care and the incidence of endstage renal disease due to systemic lupus erythematosus. J Rheumatol 2010;37: 1158–63.
26. Marsolo K, Margolis PA, Forrest CB, et al. A digital architecture for a network-based learning health system: integrating chronic care management, quality improvement, and research. EGEMS (Wash DC) 2015;3:1168.
27. Wenderfer S, Chang J, Goodwin Davies A, et al. Using a multi-institutional pediatric learning health system to identify systemic lupus erythematosus and lupus nephritis: development and validation of computable phenotypes. Clin J Am Soc Nephrol 2022:17;65–74.
28. Hammam N, Evans M, Bell CF, et al. Evaluating the use of glucocorticoids among belimumab-treated patients with systemic lupus erythematosus in real-world settings using the Rheumatology Informatics System for Effectiveness registry. ACR Open Rheumatol 2022;4: 883–9.
29. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010;363:501–4.
30. Aggarwal I, Li J, Trupin L, et al. Quality of care for the screening, diagnosis, and management of lupus nephritis across multiple healthcare settings. Arthritis Care Res (Hoboken) 2020;72:888–96.
31. Joseph RM, van Staa TP, Lunt M, et al. Exposure measurement error when assessing current glucocorticoid use using UK primary care electronic prescription data. Pharmacoepidemiol Drug Saf 2019;28: 179–86.

Arthritis Care & Research
Vol. 75, No. 11, November 2023, pp 2277–2284
DOI 10.1002/acr.25159

American College of Rheumatology
Empowering Rheumatology Professionals

OMERACT International Consensus for Ultrasound Definitions of Tenosynovitis in Juvenile Idiopathic Arthritis: Systematic Literature Review and Delphi Process

Paz Collado,[1] María Victoria Martire,[2] Stefano Lanni,[3] Orazio De Lucia,[4] Peter Balint,[5] Severine Guillaume-Czitrom,[6] Cristina Hernandez-Diaz,[7] Nina Krafft Sande,[8] Silvia Magni-Manzoni,[9] Clara Malattia,[10] Linda Rossi-Semerano,[11] Johannes Roth,[12] Tracy Ting,[13] Patricia Vega-Fernandez,[13] Daniel Windschall,[14] Maria Antonietta D'Agostino,[15] and Esperanza Naredo[16] for the OMERACT Ultrasound Group

Objective. Synovitis and tenosynovitis are present in juvenile idiopathic arthritis (JIA), both as joint pain and/or inflammation, making them difficult to detect on physical examination. Although ultrasonography (US) allows for discrimination of the 2 entities, only definitions and scoring of synovitis in children have been established. This study was undertaken to produce consensus-based US definitions of tenosynovitis in JIA.

Methods. A systematic literature search was performed. Selection criteria included studies focused on US definition and scoring systems for tenosynovitis in children, as well as US metric properties. Through a 2-step Delphi process, a panel of international US experts developed definitions for tenosynovitis components (step 1) and validated them by testing their applicability on US images of tenosynovitis in several age groups (step 2). A 5-point Likert scale was used to rate the level of agreement.

Results. A total of 14 studies were identified. Most used the US definitions developed for adults to define tenosynovitis in children. Construct validity was reported in 86% of articles using physical examination as a comparator. Few studies reported US reliability and responsiveness in JIA. In step 1, experts reached a strong group agreement (>86%) by applying adult definitions in children after one round. After 4 rounds of step 2, the final definitions were validated on all tendons and at all locations, except for biceps tenosynovitis in children <4 years old.

Conclusion. The study shows that the definition of tenosynovitis used in adults is applicable to children with minimal modifications agreed upon through a Delphi process. Further studies are required to confirm our results.

INTRODUCTION

Juvenile idiopathic arthritis (JIA) is characterized by chronic inflammatory processes primarily targeting the synovial tissue of joints and tendon sheaths. The gold standard for synovitis detection has traditionally been the clinical assessment of swollen joints by physicians. However, distinguishing tenosynovitis from underlying synovitis, an abnormality which can also be present, may be challenging based only on clinical examination, especially for small joints (1,2). Understanding the exact location of inflammation is crucial to optimize therapeutic decision-making, particularly

[1]Paz Collado, MD, PhD: Hospital Universitario Severo Ochoa and Universidad Internacional Alfonso X El Sabio, Madrid, Spain; [2]María Victoria Martire, MD: Hospital San Roque de Gonnet, La Plata, Buenos Aires, Argentina; [3]Stefano Lanni, MD: Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy; [4]Orazio De Lucia, MD: ASST Centro Traumatologico Ortopedico G. Pini-CTO, Milan, Italy; [5]Peter Balint, MD, PhD: National Institute of Musculoskeletal Diseases, Budapest, Hungary; [6]Severine Guillaume-Czitrom, MD: University Hospital of Bicetre, APHP LeKremlin-Bicêtre, France; [7]Cristina Hernandez-Diaz, MD: Instituto Nacional de Rehabilitación Luis Guillermo Ibarra and Hospital Juárez de México, Ciudad de México, Mexico; [8]Nina Krafft Sande, MD: Oslo University Hospital, Oslo, Norway; [9]Silvia Magni-Manzoni, MD, PhD: IRCCS Bambino Gesù Children's Hospital, Rome, Italy; [10]Clara Malattia, MD, PhD: UOC Clinica Pediatrica e Reumatologia, IRCCS Istituto Giannina Gaslini, and University of Genoa, Genoa, Italy; [11]Linda Rossi-Semerano, MD: Centre Hospitalier Universitaire de Bicêtre, APHP, National Reference Centre for Auto-Inflammatory Diseases and Amyloidosis of Inflammatory origin, Le Kremlin-Bicêtre, France; [12]Johannes Roth, MD, PhD: Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; [13]Tracy Ting, MD, Patricia Vega-Fernandez, MD: Cincinnati Children's Hospital Medical Center and University of Cincinnati, Cincinnati, Ohio; [14]Daniel Windschall, MD: Clinic of Paediatric and Adolescent Rheumatology, St.-Josef-Stift Sendenhorst, Northwest German Center for Rheumatology, Sendenhorst, Germany, and Martin-Luther-University Halle-Wittenberg, Halle, Germany; [15]Maria Antonietta D'Agostino, MD, PhD: Fondazione Policlinico Universitario Agostino Gemelli IRCCS and Università Cattolica del Sacro Cuore, Rome, Italy; [16]Esperanza Naredo, MD, PhD: Hospital General Universitario Gregorio Marañón and Universidad Autónoma de Madrid, Madrid, Spain.

Author disclosures are available online at https://onlinelibrary.wiley.com/doi/10.1002/acr.25159.

Address correspondence via email to Paz Collado, MD, PhD, at paxko10@gmail.com.

Submitted for publication February 28, 2023; accepted in revised form May 16, 2023.

SIGNIFICANCE & INNOVATIONS

- Musculoskeletal ultrasonography (US) is an important tool in the assessment of disease activity in childhood arthritis.
- To assess disease activity, precise definitions for different pathologic findings are an essential prerequisite for the reliable use of this technology in the pediatric ages.
- To date, the pediatric subgroup of the Outcome Measures in Rheumatology US Working Group has completed validation processes for US definitions of normal joint components and synovitis. Now, US definitions for tenosynovitis in children have just been developed and validated through an international consensus process.

for local injections. Musculoskeletal ultrasonography (US) provides an objective assessment of inflammation in peripheral joints (3). US is a versatile, multiplanar, and inexpensive bedside imaging modality with high patient acceptability, and provides direct visualization for local steroid injections (4,5).

With the advent of the treat-to-target concept and the availability of novel therapies, objective and sensitive monitoring of treatment efficacy is of upmost importance. In this perspective validated study, US definitions of elementary lesion components of synovitis and tenosynovitis represent a useful adjunct to clinical practice.

The Outcome Measures in Rheumatology (OMERACT) US Pediatric subgroup was formed to standardize the use of US in JIA. To date, US definitions for normal joint components and for synovitis have been developed (6,7,8). The JIA subgroup of the OMERACT US Working Group has recently completed a validation process for tenosynovitis according to the OMERACT Filter 2.1 Instrument Selection Algorithm (OFISA) (9). The purpose of this study was to define and validate the definition of the elementary lesions of US tenosynovitis in children with JIA.

MATERIALS AND METHODS

Study design. The study consisted of 2 phases: 1) a systematic literature review focused on the identification of studies on US as an outcome measure for the diagnosis and monitoring of tenosynovitis in children with JIA, and 2) a consensus process to develop and validate definitions for the elementary lesions of tenosynovitis in children.

Systematic literature review: search strategy, study selection, and data extraction. We searched in 3 databases (PubMed, Embase, and Cochrane) from their inceptions to September 1, 2022, for studies that assessed tenosynovitis in children <18 years old with JIA. We followed the Patient/Population, Intervention, Comparison, and Outcomes methodology (musculoskeletal ultrasound [US], comparison: other imaging techniques; and outcomes: diagnosis of tenosynovitis) to define the setting.

The search strategy included the following combination of subject headings and search terms: "ultrasonography," "echography," "Doppler," "juvenile idiopathic arthritis," "juvenile arthritis," "tenosynovitis," "tendinopathy," "diagnosis," "follow-up," "therapy management." To ensure completeness of the search, a manual review of the references of included studies was performed. However, no studies were found through this additional review. The limits used were original articles, English language, humans only, and subjects ≤18 years old. Exclusion criteria consisted of 1) reviews, editorials, letters, case reports, and abstracts of scientific congresses; and 2) studies with a mixed patient population (i.e., adult and children). Review of studies for inclusion was conducted by 2 separate authors (PC and VM) and a third author (EN) was appointed to resolve any discrepancies.

Data were extracted from articles fulfilling the selection criteria, with particular focus on the definition used, scoring system applied, and the metric properties of US evaluated. Data were then recorded using a predetermined form that was previously designed (10) for this purpose. Extracted data included author, publication year, study design, JIA subtype, number of patients and controls, tendons examined, definitions of US tenosynovitis in children with JIA (if present), the global description of the US technique, the US mode used (i.e., B-mode/grayscale alone, Doppler mode alone, or a combination of both), and the scoring systems used for the scanned tendons: 1) binary (yes/no) for the presence of synovial hypertrophy or effusion or power Doppler alone in the tendon sheath, or 2) semiquantitative. Information about the construct and criterion validity, reliability, and discriminant validity of US was also recorded. Each included study was analyzed to determine whether the measurement properties of US fulfilled the criteria according to the OFISA (9).

Systematic literature review: quality assessment of included studies. To analyze the study quality, the same methodology as in a previous systematic literature review of synovitis in JIA was used (11). It included a set of 6 predefined criteria: 1) was the recruitment of patients well-defined in the methods section; 2) was there a description of normal US anatomy of pediatric tendons; 3) was there a description of the US scanning technique (settings used, type of machine and protocol of scanning); 4) was there a description of the blinding attempted for observers; 5) was there a description of US tenosynovitis and the scoring system mentioned; and 6) was the comparator adequately explained (baseline and/or follow-up) and the results presented in their entirety? Quality was reported on a scale of 0–6, with higher results indicating higher quality.

Consensus process. A 2-step consensus process on US-defined tenosynovitis in children was carried out to develop the definitions (step 1) as well as to validate them by testing their applicability on US images representing various degrees of tenosynovitis in various age groups (step 2).

In the first step, based on the information obtained from the systematic literature review, the preliminary proposal was to evaluate whether the consensus definitions developed and used in adults with rheumatoid arthritis (RA) would be suitable for children with JIA (12). A questionnaire based on the statements developed for RA (12) was sent to a panel of 37 international experts in pediatric musculoskeletal US from 14 countries that are part of the OMERACT US group. The participants were asked to rate their level of agreement with each statement on a 5-point Likert scale (1 = strongly disagree; 5 = strongly agree). Agreement was achieved if ≥80% of responders scored a statement as either 4 or 5. Additionally, comments were allowed at the end of each statement.

The aim of the second step was to validate each statement (i.e., normal tendon structure, normal tendon synovial sheath, normal retinaculum and finger flexor pulley, tendon sheath effusion, tenosynovial hypertrophy, and US definitions of tenosynovitis on both B-mode and power Doppler) by testing their applicability on US images showing tenosynovitis in children from JIA patients at various ages. A total of 18 task force members, who participated actively in the previous task, were instructed to acquire standardized US images according to the literature (13). Tendons often involved in JIA were preselected (biceps, hand, wrist, and ankle tendons). The participants collected at least 1 B-mode and power Doppler image per tendon, representative of tenosynovitis, in both transverse and longitudinal planes. Additionally, they indicated the age group (toddler and preschool ages 2–4 years, young children ages 5–8 years, preadolescent ages 9–12 years, and teenager ages 13–16 years) in the saved image as agreed upon in a previous exercise (13). Only anonymized images were used. Participating centers did not require ethics approval for this web-based exercise. The US equipment used for image collection differed from one participating center to another.

Table 1. Description of the studies reporting ultrasound validity in the assessment of tenosynovitis in juvenile idiopathic arthritis*

Author, year (ref.)	Study type	No. of patients	No. of patients in control group	Joint region	Tendon assessed	Reliability	Quality score
Karmazyn et al, 2007 (15)	CS	20	12	MCP	FFT	NA	6
Magni-Manzoni et al, 2009 (16)	CS	32	0	Elbow,wrist, IP, MCP, knee, ankle, toes	NA	NA	2
Rooney et al, 2009 (1)	CS	34	0	Ankle	TPT, PT	NA	2
Pascoli et al, 2010 (2)	CS	42	0	Ankle	TPT, PT	Interobserver	2
Laurell et al, 2011 (4)†	PL	30	0	Ankle	TPT, PT, FDL, FHL, TA, EHL, EDL	NA	3
Laurell et al, 2012 (5)†	PL	11	0	Wrist	APL, EPB, ECR, EPL, EDC, EDM, ECU, FCR, FDD	NA	5
Hendry et al, 2012 (17)	CS‡	30	0	Ankle, IP, MTP	TPT, PT, FDL, FHL	NA	3
Magni-Manzoni et al, 2013 (18)	PL	39	39	Elbow, wrist, IP, MCP, knee, ankle, toes	Not specified	NA	5
Collado et al, 2014 (19)	CS	34	0	Shoulder, elbow, wrist, MCP, hand, IP, knee, ankle, MTP	EDC, FFT, TFT	Interobserver	5
Peters et al, 2017 (20)	CS	244	0	Ankle, wrist	TPT, PT, FDL TA, FHL, APL, EPB, ECR, EPL, EDC, TFT, FFT, BT	NA	2
Ventura-Ríos et al, 2018 (3)	CS	30	0	Wrist, 2MCP, 3MCP	EDC, FFT	Interobserver/ intraobserver	5
Lanni et al, 2021 (21)	CS	78	0	Ankle	TA, EDL, FDL, PT, EHL, TPT, FHL	NA	4
Collado et al, 2022 (22)	CS	28	54	Ankle	TA, EHL, EDL, TPT, FDL, FHL, PT	NA	5
Della Paolera et al, 2022 (23)†	CS	48	0	Ankle	TPT, PT, FDL, FHL, EDL, EHL	NA	2

* The comparator used for construct validity was clinical examination in all studies except Peters et al (ref. 20), in which a comparator was not available. APL = abductor pollicis longus; BT = biceps tendons; CS = cross-sectional; ECU = extensor carpi ulnaris; ECR = extensor carpi radialis; EDC = extensor digitorum communis; EDL = extensor digitorum longus; EDM = extensor digit minimi; EHL = extensor hallucis longus; EPB = extensor pollicis brevis; EPL = extensor pollicis longus; FCR = Flexor carpi radialis; FDD = flexor digitorum superficialis and profundus; FDL = flexor digitorum longus; FFT = finger flexor tendon; FHL = flexor hallucis longus; FPL = flexor pollicis longus; IP = interphalangeal; MCP = metacarpophalangeal; MTP = metatarsophalangeal joint; NA = not available; PL = prospective longitudinal; PT = peroneal tendons; TA = tibialis anterior; TFT = toe flexor tendon; TPT = tibialis posterior tendon.
† Sensitivity to change was measured in the indicated study.
‡ Cross-sectional (CS) here indicates a section of a phase II randomized controlled trial, the Foot Arthritis trial.

The convenor (PC) collected images from the participants and sent them back a representative selection, asking to assess each image and rate applicability of statements using a 5-point Likert scale. An agreement of ≥70% was considered mandatory for consensus. The answers from each round of the questionnaire that did not reach the agreement threshold were revised and modified according to the experts' comments, then resent to the panel for the next round until agreement was reached for all statements. Similarly, images not reaching the agreement threshold were replaced with new images provided by the participants.

Statistical analysis. Descriptive statistics were calculated from the responses to the questionnaires. The results from the Delphi process were presented as the percentage of responders who scored a statement as either 4 or 5.

RESULTS

Systematic literature review process. The study selection process is shown in a Preferred Reporting Items for Systematic reviews and Meta-Analysis flow diagram (14) (Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25159/abstract).

Summary of included studies. The features and metric properties of the 14 studies included in the systematic literature review are shown in Tables 1 and 2 (1–5,15–23). The main objective in most studies was the assessment of synovitis, whereas tenosynovitis was the second objective. The designs of the studies are shown in Table 1; the majority of which were cross-sectional (15–16,19–23). The number of patients included in each study was relatively small (range 11–42), except for 2 that included 78 and 244 patients, respectively (20,21). There were only 3 studies that included a control group (15,18,22). The ankle tendons were most studied (11 of 14 [78%]), particularly the posterior tibialis (1,2,4,18–23), while the biceps tendon was rarely investigated (20).

US definitions and scoring systems of tenosynovitis in children reported are shown in Table 2. Most articles (1–5,17,19,21) included the US definitions for adult RA to define tenosynovitis in children (24), and a tenosynovitis scoring system was included in 9 articles (64%). Different machine brands and frequencies of the transducers were used (from 5–20 MHz) (Table 2). Except for

Table 2. US definition and description of tenosynovitis or its elementary components evaluated and correlating tenosynovitis scoring systems*

Author, year (ref.)	Definition of tenosynovitis included in the study	Scoring system	Equipment (Doppler setting)
Karmazyn et al, 2007 (15)	"Fluid and vascularity within TS"	Semiquantitative (CD)	Philips 5000 (CD: 1.7–2.5 cm/second; maximized gain levels until color noise was outside the vessel wall)
Magni-Manzoni et al, 2009 (16)	OMERACT†	Binary	GE Logiq 9
Rooney et al, 2009 (1)	OMERACT†	NA	SonoSite 180 Plus or Esaote MyLab 25
Pascoli et al, 2010 (2)	OMERACT†	NA	Esaote MyLab 25 (PRF 700 Hz, LF, highest gain not displaying background artefact)
Laurell et al, 2011 (4)	OMERACT†	NA	GE Logiq 9 (PRF 600 Hz, LF, color gain just below the level at which noise appeared)
Laurell et al, 2012 (5)	OMERACT†	NA	GE Logiq 9 (PRF 600 Hz, LF, color gain just below the level at which noise appeared)
Hendry et al, 2012 (17)	OMERACT†	Binary	Siemens Acuson Antares
Magni-Manzoni et al, 2013 (18)	OMERACT†	Binary	GE Logiq 9
Collado et al, 2014 (19)	OMERACT†	Binary	GE Logiq E (PRF 600 Hz, LF, highest PD gain not showing signal under the bony cortex)
Peters et al, 2017 (20)	"TS thickening with echogenic fluid, increased CD signal, or a combination"	Binary	NA
Ventura-Ríos et al, 2018 (3)	OMERACT†	NA	GE Logiq E R6 (PRF 500 Hz, LF, highest gain not showing signal under the bony cortex)
Lanni et al, 2021 (21)	"The presence of swelling in the related tendon area"	Binary	Esaote MyLab alpha (PRF 480–700 Hz, LF, highest gain just below the level not displaying color noise in the underlying bone)
Collado et al, 2022 (22)	"Abnormal TS thickening with or without PD"	Binary	GE Logiq E (PRF 600 Hz, LF, gain was adjusted until the background signal was removed)
Della Paolera et al, 2022 (23)	"Thickened TS with fluid, Doppler signal may be present"	Binary	NA

* Doppler imaging was used in all studies except in Peters et al, 2017 (ref. 20). CD = color Doppler; LF = lowest filter; NA = not applicable; PD = power Doppler; PRF = pulse repetition frequency; TS = tendon sheath.
† The Outcome Measures in Rheumatology Clinical Trials (OMERACT) definitions for ultrasound (US) tenosynovitis in adults (ref. 24).

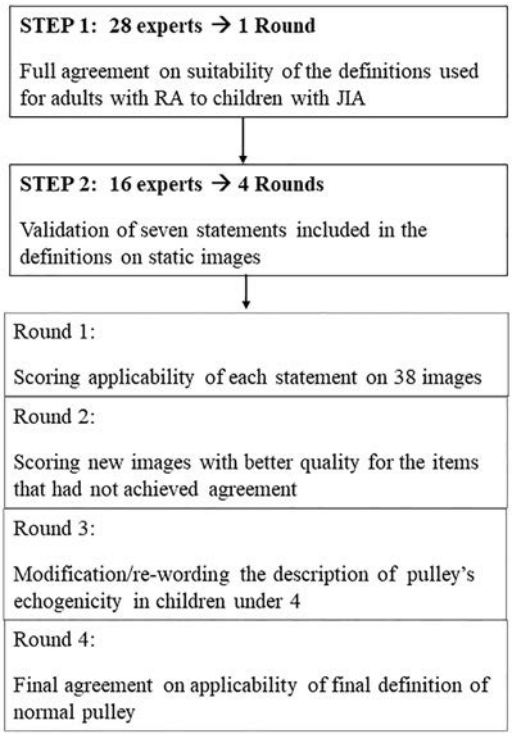

Figure 1. Workflow showing the consensus process to develop and validate the definitions of tenosynovitis in children (step 1 and step 2). JIA = juvenile idiopathic arthritis; RA = rheumatoid arthritis.

2 articles, all studies included power Doppler mode, but only 8 detailed the settings (2–5,15,19,21,22).

The metric properties of the studies are shown in Tables 1 and 2. Construct validity was the most common aspect of US validity reported, which included clinical examination and patient-reported outcome measures as comparators in 12 and 6 studies, respectively. Overall, the results of those studies show that US detected tenosynovitis more often than clinical examination (Supplementary Table 1, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25159/abstract). Criterion validity of US in relation to histology was not studied. The study by Karmazyn et al was the only one that described a scoring system for Doppler mode (15). The reliability and ability of US to detect changes over time in JIA were reported in 3 studies (2,3,19) and in 3 other studies, respectively (4,5,23).

Quality assessment of included studies. Quality scores are shown in Table 1 (Detailed scores are shown in Supplementary Table 1, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25159/abstract). Of the 14 studies selected, only 6 achieved a minimum quality score ≥4 (3,5,15,18,19,22).

Consensus process findings. The workflow of the consensus process to develop and validate definitions on static images is shown in Figure 1.

First step. A total of 28 of 37 participants responded to the first questionnaire (75% response rate). For all statements group agreement was ≥86% in the first round.

Second step. Once group agreement was reached in step 1, 7 statements related to the consensus definitions in adults with RA were validated in a 4-round web-based exercise testing its applicability in tenosynovitis in children. A total of 18 experts were invited, of whom 16 (89%) participated in all rounds.

For the first and the second rounds, a set of 38 US images were assessed by the participants who rated their agreement on applicability of each statement. The 7 statements reached a strong group agreement for the ankle tendons for all age groups (range 80–100%) in the first round, but there was no agreement for the remaining tendons. The second round provided new images for the items that had not achieved agreement. Except for the normal pulley in the youngest children, group agreement was reached for the tendons of the wrist (73–100%) and the finger (73–100%). Despite the ability of US to display the normal finger pulleys at the level of metacarpophalangeal joints in children, some comments were raised about the pulley's echogenicity in young children after evaluating the images in the second round. The percentage of group agreement per statement after the second round is shown in Table 3.

For the third round, the description of a normal pulley was reworded as follows: "a linear hyperechoic structure, although it could artifactually appear hypoechoic." This modification achieved a strong group agreement (99.8%) in the fourth and final round.

Table 3. Group agreement for tenosynovitis definition at different ages in various joints*

	Definition per each statement						
Age, years	1	2	3	4	5	6	7
Shoulder							
2–4 years	100	86	NA	56	60	62	56
5–8 years	86	86	NA	77	100	100	93
9–12 years	86	80	NA	71	71	80	80
13–15 years	94	88	NA	93	93	100	93
Wrist							
2–4 years	93	100	75	86	93	100	100
5–8 years	100	71	100	100	100	100	100
9–12 years	86	75	73	80	93	100	93
13–15 years	86	93	86	73	86	100	100
Finger							
2–4 years	87	73	68	100	93	73	75
5–8 years	80	80	67	93	100	100	93
9–12 years	73	73	87	87	100	100	100
13–15 years	93	93	87	93	73	93	73
Ankle							
2–4 years	87	73	78	100	93	100	100
5–8 years	87	80	80	86	86	93	93
9–12 years	87	80	80	86	100	100	100
13–15 years	87	80	80	100	87	100	93

* Values are the percentage of answers that scored grade 4 or 5 for each of the images on a 5-point Likert scale in the second round.

Table 4. Description of the final validated US definitions*

Statement	Definition
Normal structure; tendon (definition 1)	Hyperechoic (relative to subdermal fat) fibrillar pattern (i.e., hyperechoic parallel lines in long plane and hyperechoic dots in transverse plane)
Normal structure; tendon synovial sheath (definition 2)	A thin regular hypoechoic (relative to tendon fibers) halo surrounding (in transverse plane), thin regular hypoechoic lines above and below the tendon structure (in long plane) at anatomical sites where synovial sheaths are known to exist and which can be distinguished from pulleys and retinaculae
Normal structure; retinaculum (wrist and ankle level) and pulleys (finger flexor level) (definition 3)	Annular pulley appeared as a focal hyperechoic (or hypoechoic depending on the US insonation angle) fibrillar structure relative to the adjacent flexor tendon that can be detected overlying the parietal synovial sheath of the digital flexor tendon at its expected anatomical location
Elementary lesion; tendon sheath effusion (definition 4)	Presence of abnormal anechoic or hypoechoic (relative to tendon fibers) material within the synovial sheath, either localized (e.g., in the synovial sheath cul-de-sacs) or surrounding the tendon that is displaceable and seen in 2 perpendicular planes
Elementary lesion; tenosynovial hypertrophy (definition 5)	Presence of abnormal hypoechoic (relative to tendon fibers) tissue within the synovial sheath that is not displaceable and poorly compressible and seen in 2 perpendicular planes
Tenosynovitis on B-mode (definition 6)	Abnormal anechoic or hypoechoic (relative to tendon fibers) tendon sheath widening, which can be related to both the presence of tenosynovial abnormal effusion or hypertrophy
Tenosynovitis on Doppler (definition 7)	The presence of peritendinous Doppler signal within the synovial sheath, seen in 2 perpendicular planes, excluding normal feeding vessels (i.e., vessels at the mesotenon or vincula) only if the tendon shows peritendinous synovial sheath widening on B-mode

* US = ultrasound.

After 4 rounds, the final definitions were validated on all tendons and at all locations, except for biceps tenosynovitis in children <4 years old, which showed no agreement regarding its applicability, due to a lack of images available for this location and age group. The final version of the validated US definitions for elementary lesions in tenosynovitis in children is shown in Table 4. An US image illustrating some of these lesions that correspond to the definition is shown in Figure 2.

DISCUSSION

Definitions for the US appearance of tenosynovitis in children were developed through a consensus process and validated in web-based exercises. This study is an important step toward a more reliable use of musculoskeletal US in children as an outcome measure of disease activity (25).

Consistent with the previous systematic literature review on synovitis (11), US examination of the foot and hand remains a priority in JIA, given the frequency of involvement and the challenges of the clinical assessment (1,2,18,22). In contrast, biceps tendon involvement is less well studied, since inflammation affecting the shoulder joint or tendons is less reported in children compared to adults. Like the systematic literature review of synovitis, the current systematic literature review showed a moderate quality of existing studies illustrating a need for ongoing research on the validity of US in tenosynovitis in children.

As reflected in the systematic literature review, the absence of definitions of tenosynovitis in children for JIA over the years has led several authors to use existing definitions that were developed for adults (24). However, their applicability to children might require some considerations. To address this issue, the OMERACT US Working Group has conducted the current Delphi process. The group unanimously agreed that the definitions used in adults can be applied in the pediatric population in the first round of step 1. In step 2, instead of a de novo development of definitions, the adult definitions were assessed for their suitability in children on static images. The final definitions for tenosynovitis in children were validated for all tendons and ages, except biceps

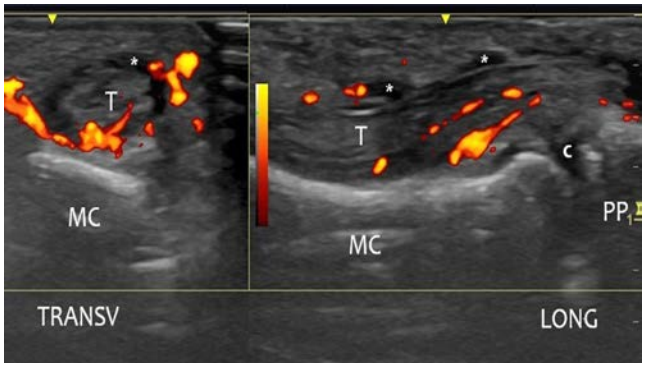

Figure 2. Tenosynovitis on Doppler ultrasound imaging of the finger of a 6-year-old child. The image shows the presence of tendon sheath effusion (*) and peritendinous Doppler signal within the synovial sheath in the flexor tendon of the second finger, seen in 2 perpendicular planes. c = unossified hyaline cartilage of the secondary ossification nucleus of the phalanx; MC = metacarpal bone; PP = proximal phalanx; T = tendon.

tenosynovitis in children ages 2–4 years, since no images of tenosynovitis were available for this location and age group. This may reflect limited involvement of the shoulder in JIA but may also indicate that this is an understudied area.

A group agreement for the US-defined normal finger pulley in children <8 years old was reached after 4 rounds. It highlights the difficulty to clearly distinguish this structure in healthy children, even on high-quality images, or the fact that pulleys became more detectable on US in children who are engaged in physical activities like rock climbing, which often increases visibility of pulleys due to changes in the supporting structures of the hands and fingers (26–29). Some limitations should be noted; we validated the US definition of tenosynovitis in children only on static images, and most participants were experts in US. However, performing an international validation study in children in real time is a challenge because it is not feasible to gather children grouped by age to be scanned several times by different sonographers in a day.

Although children differ significantly from adults in their bone anatomy (mostly related to maturation of bones), the present study demonstrated that the definitions of tenosynovitis used in adults are applicable to children with minimal modifications determined through a Delphi process. Further studies are required to confirm our results as well as to evaluate metric properties of US in the assessment of tenosynovitis in children with JIA.

AUTHOR CONTRIBUTIONS

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

Study conception and design. Collado, D'Agostino, Naredo.

Acquisition of data. Collado, Martire, Lanni, De Lucia, Balint, Guillaume-Czitrom, Hernandez-Diaz, Sande, Magni-Manzoni, Malattia, Rossi-Semerano, Roth, Ting, Vega-Fernandez, Windschall, D'Agostino, Naredo.

Analysis and interpretation of data. Collado, Martire, Lanni, De Lucia, Balint, Guillaume-Czitrom, Hernandez-Diaz, Sande, Magni-Manzoni, Malattia, Rossi-Semerano, Roth, Ting, Vega-Fernandez, Windschall, D'Agostino, Naredo.

REFERENCES

1. Rooney ME, McAllister C, Burns JF. Ankle disease in juvenile idiopathic arthritis: ultrasound findings in clinically swollen ankles. J Rheumatol 2009;36:1725–9.
2. Pascoli L, Wright S, McAllister C, et al. Prospective evaluation of clinical and ultrasound findings in ankle disease in juvenile idiopathic arthritis: importance of ankle ultrasound. J Rheumatol 2010;37:2409–14.
3. Ventura-Ríos L, Faugier E, Barzola L, et al. Reliability of ultrasonography to detect inflammatory lesions and structural damage in juvenile idiopathic arthritis. Pediatr Rheumatol Online J 2018;16:58.
4. Laurell L, Court-Payen M, Nielsen S, et al. Ultrasonography and color Doppler in juvenile idiopathic arthritis: diagnosis and follow-up of ultrasound-guided steroid injection in the ankle region. A descriptive interventional study. Pediatr Rheumatol Online J 2011;9:4.
5. Laurell L, Court-Payen M, Nielsen S, et al. Ultrasonography and color Doppler in juvenile idiopathic arthritis: diagnosis and follow-up of ultrasound guided steroid injection in the wrist region. A descriptive interventional study. Pediatr Rheumatol J 2012;10:11.
6. Roth J, Jousse-Joulin S, Magni-Manzoni S, et al. Definitions for the sonographic features of joints in healthy children. Arthritis Care Res (Hoboken) 2015;67:136–42.
7. Collado P, Windschall D, Vojinovic J, et al. Amendment of the OMERACT ultrasound definitions of joints' features in healthy children when using the DOPPLER technique. Pediatr Rheumatol Online J 2018;16:23.
8. Roth J, Ravagnani V, Backhaus M, et al. Preliminary definitions for the sonographic features of Synovitis in children. Arthritis Care Res (Hoboken) 2017;69:1217–23.
9. Outcome Measures in Rheumatology. The OMERACT handbook. URL: https://omeracthandbook.org/handbook.
10. Joshua F, Lassere M, Bruyn GA, et al. Summary findings of a systematic review of the ultrasound assessment of synovitis [review]. J Rheumatol 2007;34:839–47.
11. Collado P, Jousse-Joulin S, Alcalde M, et al. Is ultrasound a validated imaging tool for the diagnosis and management of synovitis in juvenile idiopathic arthritis? A systematic literature review [review]. Arthritis Care Res (Hoboken) 2012;64:1011–9.
12. Naredo E, D'Agostino MA, Wakefield R, et al. Reliability of a consensus-based ultrasound score for tenosynovitis in rheumatoid arthritis. Ann Rheum Dis 2013;72:1328–34.
13. Collado P, Vojinovic J, Nieto JC, et al. Toward standardized musculoskeletal ultrasound in pediatric rheumatology: normal age-related ultrasound findings. Arthritis Care Res (Hoboken) 2016;68:348–56.
14. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.
15. Karmazyn B, Bowyer SL, Schmidt KM, et al. US findings of metacarpophalangeal joints in children with idiopathic juvenile arthritis. Pediatr Radiol 2007;37:475–82.
16. Magni-Manzoni S, Epis O, Ravelli A, et al. Comparison of clinical versus ultrasound-determined synovitis in juvenile idiopathic arthritis. Arthritis Rheum 2009;61:1497–504.
17. Hendry GJ, Gardner-Medwin J, Steultjens MP, et al. Frequent discordance between clinical and musculoskeletal ultrasound examinations of foot disease in juvenile idiopathic arthritis. Arthritis Care Res (Hoboken) 2012;64:441–7.
18. Magni-Manzoni S, Scirè CA, Ravelli A, et al. Ultrasound-detected synovial abnormalities are frequent in clinically inactive juvenile idiopathic arthritis, but do not predict a flare of synovitis. Ann Rheum Dis 2013;72:223–8.
19. Collado P, Gamir ML, Lopez-Robledillo JC, et al. Detection of synovitis by ultrasonography in clinically inactive juvenile idiopathic arthritis on and off medication. Clin Exp Rheumatol 2014;32:597–603.
20. Peters SE, Laxer RM, Connolly BL, et al. Ultrasound-guided steroid tendon sheath injections in juvenile idiopathic arthritis: a 10-year single-center retrospective study. Pediatr Rheumatol Online J 2017;15:22.
21. Lanni S, Marafon DP, Civino A, et al. Comparison between clinical and ultrasound assessment of the ankle region in children with juvenile idiopathic arthritis. Arthritis Care Res (Hoboken) 2021;73:1180–6.
22. Collado P, González-Fernández ML. The pediatric foot: prevalence and differentiation of sonographic and podiatric findings in juvenile arthritis and healthy children. Clin Exp Rheumatol 2022;40:655–61.
23. Della Paolera S, Pastore S, Zabotti A, et al. Ultrasonographic assessment for tenosynovitis in juvenile idiopathic arthritis with ankle involvement: diagnostic and therapeutic significance. Children (Basel) 2022;9:509.

24. Wakefield RJ, Balint PV, Szkudlarek M, et al. Musculoskeletal ultrasound including definitions for ultrasonographic pathology. J Rheumatol 2005;32:2485–7.
25. Colebatch-Bourn AN, Edwards CJ, Collado P, et al. EULAR-PReS points to consider for the use of imaging in the diagnosis and management of juvenile idiopathic arthritis in clinical practice [review]. Ann Rheum Dis 2015;74:1946–57.
26. Cohen MJ, Kaplan L. Histology and ultrastructure of the human flexor tendon sheath J Hand Surg Am 1987;12A:25–9.
27. Doyle JR. Palmar and digital flexor tendon pulleys [review]. Clin Orthop Relat Res 2001;383:84–96.
28. Hauger O, Chung CB, Lektrakul N, et al. Pulley system in the fingers: normal anatomy and simulated lesions in cadavers at MR imaging, CT, and US with and without contrast material distention of the tendon sheath. Radiology 2000;217:201–12.
29. Boutry N, Titéca M, Demondion X, et al. High-frequency ultrasonographic examination of the finger pulley system. J Ultrasound Med 2005;24:1333–9.

Arthritis Care & Research
Vol. 75, No. 11, November 2023, pp 2285–2294
DOI 10.1002/acr.25136

American College of Rheumatology
Empowering Rheumatology Professionals

Health Equity Implications of Missing Data Among Youths With Childhood-Onset Systemic Lupus Erythematosus: A Proof-of-Concept Study in the Childhood Arthritis and Rheumatology Research Alliance Registry

Jennifer M. P. Woo,[1] Faith Simmonds,[2] Anne Dennos,[3] Mary Beth F. Son,[4] Laura B. Lewandowski,[2] and Tamar B. Rubinstein,[5] for the CARRA Registry investigators

Objective. Health disparities in childhood-onset systemic lupus erythematosus (SLE) disproportionately impact marginalized populations. Socioeconomically patterned missing data can magnify existing health inequities by supporting inferences that may misrepresent populations of interest. Our objective was to assess missing data and subsequent health equity implications among participants with childhood-onset SLE enrolled in a large pediatric rheumatology registry.

Methods. We evaluated co-missingness of 12 variables representing demographics, socioeconomic position, and clinical factors (e.g., disease-related indices) using Childhood Arthritis and Rheumatology Research Alliance Registry childhood-onset SLE enrollment data (2015–2022; n = 766). We performed logistic regression to calculate odds ratios (ORs) and 95% confidence intervals (95% CIs) for missing disease-related indices at enrollment (Systemic Lupus Erythematosus Disease Activity Index 2000 [SLEDAI-2K] and/or Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index [SDI]) associated with data missingness. We used linear regression to assess the association between socioeconomic factors and SLEDAI-2K at enrollment using 3 analytic methods for missing data: complete case analysis, multiple imputation, and nonprobabilistic bias analyses, with missing values imputed to represent extreme low or high disadvantage.

Results. On average, participants were missing 6.2% of data, with over 50% of participants missing at least 1 variable. Missing data correlated most closely with variables within data categories (i.e., demographic). Government-assisted health insurance was associated with missing SLEDAI-2K and/or SDI scores compared to private health insurance (OR 2.04 [95% CI 1.22, 3.41]). The different analytic approaches resulted in varying analytic sample sizes and fundamentally conflicting estimated associations.

Conclusion. Our results support intentional evaluation of missing data to inform effect estimate interpretation and critical assessment of causal statements that might otherwise misrepresent health inequities.

INTRODUCTION

Pervasive health disparities in childhood-onset systemic lupus erythematosus (SLE) affect the quality of life and disease outcomes across marginalized populations. Studies have found that Black and Hispanic/Latino people with childhood-onset SLE disproportionately develop more severe disease manifestations (1). In addition, low socioeconomic status is associated with high disease activity and organ damage among people with SLE compared to those with high socioeconomic status (2). Despite the known disparities in disease severity and outcomes in pediatric rheumatology, the most affected populations continue to be underrepresented in research studies. In a review of randomized clinical trials for SLE in adults, individuals who identified as

Supported by the Intramural Research Programs of the NIH, National Institute of Environmental Health Sciences and the NIH, National Institute of Arthritis and Musculoskeletal and Skin Diseases.

[1]Jennifer M. P. Woo, PhD, MPH: National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina; [2]Faith Simmonds, BS, Laura B. Lewandowski, MD, MS: National Institute of Arthritis and Musculoskeletal and Skin Diseases, Bethesda, Maryland; [3]Anne Dennos, MSPH: Duke University, Durham, North Carolina; [4]Mary Beth F. Son, MD: Boston Children's Hospital, Boston, Massachusetts; [5]Tamar B. Rubinstein, MD, MS: Children's Hospital at Montefiore and Albert Einstein College of Medicine, Bronx, New York.

Drs. Lewandowski and Rubinstein contributed equally to this work.

Author disclosures are available online at https://onlinelibrary.wiley.com/doi/10.1002/acr.25136.

Address correspondence via email to Jennifer M. P. Woo, PhD, MPH, at jennifer.woo@nih.gov.

Submitted for publication September 1, 2022; accepted in revised form April 18, 2023.

SIGNIFICANCE & INNOVATIONS

- Health disparities in systemic lupus erythematosus (SLE) contribute to poor outcomes in historically marginalized populations. Patient registries may have data that are missing in patterns that can unintentionally obscure important findings related to health disparities in childhood-onset SLE.
- Our study is the first to use multiple analytic methods to assess the effects of missing data on disease measures in epidemiologic research in childhood-onset SLE.
- Using data from the Childhood Arthritis and Rheumatology Research Alliance Registry, we demonstrate how socioeconomically patterned missing data and the use of varying analytic methods to handle missing data can impact sample size and effect estimates; researchers should integrate these practices in the interpretation of study results to ensure that findings across demographics are not misrepresented.

members of minoritized racial and ethnic groups comprised 73% of prevalent SLE cases and only 45% of clinical trial participants (3).

Because childhood-onset SLE is a relatively rare disease, multisite patient registries are essential for observational research to gain knowledge on the disease and its outcomes. The Childhood Arthritis and Rheumatology Research Alliance (CARRA) Registry is the largest childhood-onset SLE registry to date and represents over 70 sites across North America and Israel (see Appendix A for members of the CARRA Registry). The CARRA Registry has supported multiple published studies on childhood-onset SLE in the past decade (4–14). While the CARRA Registry represents a diverse cohort of patients, a better understanding of the patterns of missing data in the registry is required to ensure that knowledge gained from the registry is generalizable and representative of the diverse patient population affected by childhood-onset SLE.

Missing data that are socioeconomically, racially, or ethnically patterned can obscure existing health inequities by underestimating underlying associations or by contributing to unsupported inferences that fail to represent the target population. Because of the potential impact of misinterpreting biased estimates, there is a growing emphasis in epidemiologic research on the use of quantitative bias analysis methods to evaluate the impact of missing data, measurement error, selection bias, and other mechanisms that can contribute to biased estimates (15,16). A recent systematic review by Lauper et al (2021) reported that 83% of longitudinal observational studies reported in key rheumatology journals between 2008 and 2019 failed to report missing data on covariates and almost half relied on complete case analysis (17). When looking at prior CARRA research studies, we found that while many reported missing data, only a few studies used quantitative bias analysis methods, such as sensitivity analyses and multiple imputation (11,12,18–24).

No current studies have comprehensively assessed patterns of missing data in the CARRA Registry or have evaluated how missing data may impact the interpretation of study results. By assessing the socioeconomic patterning of missing data related to disease outcomes in the childhood-onset SLE population, researchers can better determine whether they should consider additional methods to address missingness when using CARRA Registry data to minimize the potential for perpetuating existing health inequities and to improve generalizability of results. In this study, we aim to identify patterns of missing data in the CARRA Registry, assess whether missing data are more prevalent among minoritized racial and ethnic and/or lower socioeconomic groups, and to examine the health equity implications that might arise when missing data are not addressed in analyses.

PATIENTS AND METHODS

CARRA Registry. The CARRA Registry is a prospective observational registry of persons with childhood-onset rheumatic disease designed to evaluate therapeutic safety among the study population (25). Enrollment began in July 2015 and is ongoing among 74 pediatric rheumatology clinical sites across the US, Canada, and Israel and included over 12,000 participants as of February 2022.

A total of 925 CARRA Registry participants were diagnosed with childhood-onset SLE using the 2012 Systemic Lupus International Collaborating Clinics (SLICC) diagnostic criteria. We include in this proof-of-concept study of missing data 766 participants who also fulfilled at least 4 of 11 American College of Rheumatology (ACR) criteria for SLE (1997) or had biopsy-proven lupus nephritis with at least 2 additional ACR criteria prior to age 19 years (26). We use these generally stricter guidelines to standardize the potential impact of missing diagnosis criteria, which is more difficult to address when using the SLICC diagnostic criteria, as they allow for a varying number of criteria to confirm a childhood-onset SLE diagnosis (i.e., 4 of 17 criteria without lupus nephritis or lupus nephritis and antinuclear antibody or double-stranded DNA positivity). All participants were diagnosed with childhood-onset SLE within 24 months of enrollment or had a new diagnosis of lupus nephritis if their original childhood-onset SLE diagnosis was >24 months prior. At enrollment, participants and/or parents/guardians completed self- or medical staff–administered questionnaires regarding sociodemographics, symptom onset, and patient reported outcomes. Pediatric rheumatology staff completed corresponding questionnaires regarding clinical manifestations, including disease symptoms, laboratory values, and physician global assessment of disease activity.

CARRA Registry participants (and/or a parent or guardian when required) provided written informed consent at registry

enrollment. Our current analyses were conducted using de-identified data and were considered exempt by the Institutional Review Board at the National Institutes of Health.

Socioeconomic factors. Data on socioeconomic factors were collected at enrollment and included information on self-identified race (White, Black, Asian, Middle Eastern or North African, Native American, American Indian, or Alaska Native, Native Hawaiian or other Pacific Islander, multiple races, or races not otherwise specified) and ethnicity (Hispanic, not Hispanic), social constructs that are rooted in practices of structural racism rather than biologic differences between racialized groups (27–30); annual household income (in US or Canadian dollars, <$25,000, $25,000–49,999, $50,000–74,999, $75,000–99,999, $100,000–150,000, >$150,000, prefer not to answer, unknown); highest level of parental/guardian education completed (elementary/middle school, some high school, graduated high school or General Educational Development Test, college, including junior college or technical school, graduate school, prefer not to answer); and insurance status (private, government-assisted [e.g., Medicare, Medicaid, military health care, state-specific plan (non-Medicaid), Indian Health Services], other [e.g., non-US insurance, other], none).

Clinical factors and covariates. Measures of SLE disease activity and severity were collected at enrollment. Measures included the Systemic Lupus Erythematosus Disease Activity Index 2000 (SLEDAI-2K) for SLE disease activity over the past 30 days (possible score: 0 to 105), and the SLICC/ACR Damage Index (SDI) for SLE-related damage (possible score: 0 to 43). Additional clinical variables assessed were age at symptom onset and age at first visit with a pediatric rheumatologist. Disease duration was calculated as the time from symptom onset to the time of enrollment (years). Covariates were collected at enrollment, including participant sex assigned at birth (male, female, other) and self-identified gender (male, female, other). Models were also adjusted for participant age (continuous) and enrollment site.

Statistical analysis. Descriptive statistics were used to describe patterns of missingness among CARRA Registry participants. Tetrachoric correlation coefficients, which are measures of the strength of correlation between 2 binary variables (range –1.0 [high negative correlation] to 1.0 [high positive correlation]), were estimated using PROC FREQ and the PLCORR option to evaluate patterns of co-missingness among variables (missing versus not missing). We used logistic regression models to estimate odds ratios (ORs) and 95% confidence intervals (95% CIs) to evaluate the association between socioeconomic factors and the odds of missing clinical data. For this analysis, missing data for all variables were coded as a discrete category so as to not be dropped from the models.

In our comparative study, we used linear regression models to test the association between socioeconomic factors and disease activity (SLEDAI-2K) and disease damage (SDI) for youths with childhood-onset SLE at the time of enrollment, adjusting for age, sex assigned at birth, disease duration at enrollment, time since the first pediatric rheumatology visit, and study site. First, we used complete case analysis, excluding all participants missing any data.

Next, we used multiple imputation by chained equations (PROC MI with FCS command) to obtain 20 imputed data sets, imputing values for missing socioeconomic factors, covariates, and clinical assessments. The variables included in the imputation models were: SLEDAI-2K, SDI, race and ethnicity, sex, annual household income, parental/guardian educational attainment, insurance status, disease duration, and time between symptom onset and first visit with a pediatric rheumatologist, as well as SLICC diagnostic criteria as an auxiliary variable. The multiple imputation data sets were then used to obtain pooled estimates of socioeconomic impact on disease activity and severity using PROC MIANALYZE. A detailed description of the multiple imputation process and example code are included in Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25136.

Finally, we conducted a nonprobabilistic bias analysis to evaluate the impact of potential differential missingness related to socioeconomic factors (16). Using fixed bias parameter analysis methods, we ran separate models in which we assigned missing socioeconomic factors to represent extreme high disadvantage or extreme low disadvantage and examined how estimates of disease activity and severity changed at these extremes. In extreme high disadvantage models, missing values were imputed to non-Hispanic Black/African American, selected based on the persistent impact of racial residential segregation and marginalization in the US (31), household income less than $25,000 annually, parental educational attainment less than high school, and no health insurance. In extreme low disadvantage models, missing values were imputed to non-Hispanic White, annual household income $150,000 and above, parental educational attainment of graduate school, and private health insurance. All statistical analyses were completed using SAS software, version 9.4.

RESULTS

Overall, we included 766 participants with childhood-onset SLE in the analyses. Characteristics of the study cohort are described in Table 1. The mean age at enrollment into the CARRA Registry was 14.2 ± 3.0 years and approximately 86% were assigned female sex at birth. Furthermore, nearly 74% (n = 555) did not identify as non-Hispanic White, 35% of participants reported an annual household income of <$50,000, and 56% of participants with known insurance status had nonprivate insurance or no insurance. There were 37 unique patterns of missing data, with >50% of participants missing data on at least 1 variable

Table 1. Characteristics of CARRA Registry participants diagnosed with childhood-onset systemic lupus erythematosus by level of individual data missingness (n = 766)*

Characteristic	Overall	≤10% missing (n = 688)	>10% missing (n = 78)
Demographic factors			
Sex assigned at birth, no. (%)			
Male	106 (14)	97 (14)	9 (11)
Female	660 (86)	591 (86)	69 (89)
Gender, no. (%)			
Male	107 (14)	98 (14)	9 (12)
Female	650 (86)	586 (86)	64 (88)
Other	1 (0)	1 (0)	0 (0)
Missing	8	3	5
Age, years	14.2 ± 3.0	14.2 ± 3.0	14.2 ± 3.0
Socioeconomic factors			
Race and ethnicity, no. (%)			
Asian, Native Hawaiian, or Pacific Islander	87 (12)	80 (12)	7 (11)
Hispanic	196 (26)	178 (26)	18 (28)
Middle Eastern or North African	9 (1)	7 (1)	2 (3)
Native American, American Indian, or Alaska Native	7 (1)	6 (1)	1 (2)
Non-Hispanic Black, African American, African, or Afro-Caribbean	210 (28)	196 (29)	14 (22)
Non-Hispanic White	195 (26)	180 (26)	15 (24)
Race not previously mentioned	13 (2)	13 (2)	0 (0)
Multiracial	33 (4)	27 (4)	6 (10)
Missing	16	1	15
Household income in $, no. (%)			
<25,000	104 (16)	98 (16)	6 (11)
25,000–49,999	122 (19)	119 (20)	3 (5)
50,000–74,999	78 (12)	78 (13)	0 (0)
75,000–99,999	59 (9)	59 (10)	0 (0)
100,000–150,000	68 (10)	67 (11)	1 (2)
>150,000	66 (10)	61 (10)	5 (9)
Prefer not to answer	154 (24)	114 (19)	40 (73)
Missing	115	92	23
Parent/guardian educational attainment, no. (%)			
Less than high school	68 (11)	63 (11)	5 (10)
High school or GED	172 (27)	159 (28)	13 (26)
College	279 (44)	259 (45)	20 (40)
Graduate School	108 (17)	96 (17)	12 (24)
Missing	139	111	28
Insurance status, no. (%)			
Private	316 (44)	293 (44)	23 (41)
Government-assisted†	318 (44)	294 (44)	24 (44)
Other, including non-US insurance	72 (10)	67 (10)	5 (9)
No insurance	20 (3)	18 (3)	2 (4)
Missing	40	16	24
Clinical factors			
SLEDAI-2K	7.1 ± 7.3	7.0 ± 7.3	8.1 ± 7.7
SDI	0.3 ± 0.8	0.3 ± 0.8	0.4 ± 1.0
Lupus nephritis, no. (%)	390 (51)	353 (51)	37 (48)
Age at symptom onset	12.7 ± 3.1	12.7 ± 3.1	12.8 ± 3.0
Age at first pediatric rheumatology visit	13.0 ± 3.0	13.0 ± 3.0	13.0 ± 3.0
Disease duration from symptom onset‡	1.5 ± 1.8	1.5 ± 1.9	1.3 ± 1.5

* Values are the mean ± SD unless indicated otherwise. Missing not otherwise noted: age (n = 8), SLEDAI-2K (n = 90), SDI (n = 12), age of symptom onset (n = 8), age of first pediatric rheumatology visit (n = 8). CARRA = Childhood Arthritis and Rheumatology Research Alliance; GED = General Educational Development Test; SDI = Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index; SLEDAI-2K = Systemic Lupus Erythematosus Disease Activity Index 2000.
† Medicare, Medicaid, military health care, state-specific plan (non-Medicaid), Indian Health Services.
‡ Disease duration defined as time from symptom onset to enrollment.

(see Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25136). On average, participants were missing approximately 6.2% (range 0–70%) of the variables we investigated, and 10% of participants (n = 78) were missing >10% of included variables (Table 1). Missingness was observed across demographic, socioeconomic, and clinical variables, with socioeconomic variables missing at the highest frequency (Figure 1). The variable with the

greatest level of missing data was household income, with 35% of participants (n = 269) either selecting "Prefer not to answer" or not providing a response. The distributions of participants with known data across variables of interest were similar for most variables, except for household income, when comparing those who were missing ≤10% of variables and those missing >10% of variables (Table 1).

Missing data were most highly correlated with variables of the same type (i.e., demographic, socioeconomic, or clinical). Figure 1 represents the correlation of missing data between variables. Missingness of SLEDAI-2K score was perfectly correlated with missingness of age, age at symptom onset, and age at first visit with a pediatric rheumatologist (tetrachoric correlation coefficients = 1.0). Missingness of gender was also highly correlated with missingness of parental highest educational attainment (tetrachoric correlation coefficient = 0.82).

We next evaluated the characteristics associated with missing either a SLEDAI-2K and/or SDI score. Reporting government-assisted health insurance was associated with 2.0 times greater odds of missing a clinical score compared to reporting private health insurance (OR = 2.04 [95% CI 1.22, 3.41]) (Table 2). Furthermore, each year increase in age was associated with an 8% decrease in the odds of missing a clinical score (OR 0.92 [95% CI 0.85, 0.99]). Race and ethnicity, household income, and parental educational attainment were not statistically significantly associated with missing SLEDAI-2K and/or SDI score.

The sample distributions of characteristics for each method used to address missing data are described in Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25136. As expected, the greatest variation in study samples was observed between the 2 samples used in the nonprobabilistic bias analysis, in which missing values were imputed to the most and least disadvantaged groups. Results from the 3 statistical methods used to address missing data and the association between socioeconomic factors and the SLEDAI-2K score at enrollment are reported in Table 3. In complete case analysis, 385 participants were dropped from analysis due to missing data, leaving an analytic sample of 381. Due to the nature of multiple imputation, each imputed data set reflected data from all 766 participants. Finally, in nonprobabilistic bias analyses, 90 participants were dropped due to missing data on model covariates (age, disease duration, and time since first seen by a pediatric rheumatologist) and SLEDAI-2K score, since only the socioeconomic factors were imputed for the purpose of this study.

The variation in SLEDAI-2K scores associated with socioeconomic factors differed across methods of addressing missing data, suggesting inverse, positive, or no association. For example, when compared to private health insurance, government-assisted health insurance was suggestive of an inverse association with SLEDAI-2K scores in complete case analysis (β = −1.60 [95% CI −3.76, 0.56]) after controlling for potential confounding due to age, sex assigned at birth, disease duration, time since first seen by a pediatric rheumatologist, and study site. Further, in similarly adjusted models using multiple imputation, there was no apparent association between government-assisted health insurance compared to

	Demographics			Socioeconomic Factors				Clinical Factors				
	Sex assigned at birth	Gender	Age	Race and ethnicity	Household income	Parental educational attainment	Insurance	SLEDAI-2K	SDI	Lupus nephritis	Age symptom onset	Age first PR visit
Missing, N(%)	0 (0.0)	8 (1.0)	8 (1.0)	16 (2.1)	269 (35.1)	12 (1.6)	40 (5.2)	90 (11.8)	13 (1.7)	0 (0.0)	8 (1.0)	8 (1.0)
Sex assigned at birth	1.00											
Gender	0	1.00										
Age	0	-0.79	1.00									
Race and ethnicity	0	0	0	1.00								
Household income	0	-0.11	0.14	0.49	1.00							
Parental educational attainment	0	0.82	0	0	-0.11	1.00						
Insurance	0	0.37	0.37	0.32	0.09	0.27	1.00					
SLEDAI-2K	0	0	1.00	0.02	-0.01	0	0.08	1.00				
SDI	0	0	0.73	0	0.07	0	0.58	0.47	1.00			
Lupus nephritis	0	0	0	0	0	0	0	0	0	1.00		
Age symptom onset	0	-0.79	1.00	0	0.14	0	0.37	1.00	0.73	0	1.00	
Age first PR visit	0	-0.79	1.00	0	0.14	0	0.37	1.00	0.73	0	1.00	1.00

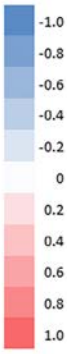

Figure 1. Patterns of missing data in the Childhood Arthritis and Rheumatology Research Alliance Registry, tetrachoric correlations (missing versus not missing), ranging from perfect negative correlation (−1.0, blue) to perfect positive correlation (1.0, red). PR = pediatric rheumatologist; SDI = Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index; SLEDAI-2K = Systemic Lupus Erythematosus Disease Activity Index 2000.

Table 2. Association between demographic and socioeconomic factors and the odds of missing SLEDAI-2K and/or SDI at enrollment among participants with childhood-onset SLE enrolled in the CARRA Registry (n = 758)*

Factor	OR (95% CI)†
Sex assigned at birth	
Male	Ref.
Female	1.98 (0.89, 4.42)
Gender	
Male	Ref.
Female	2.03 (0.91, 4.53)
Other	–
Missing	–
Age	0.92 (0.85, 0.99)
Race and ethnicity	
Asian/NHPI	0.52 (0.22, 1.26)
Hispanic	0.86 (0.47, 1.55)
Non-Hispanic Black/African American	0.63 (0.34, 1.18)
Non-Hispanic other‡	1.01 (0.44, 2.30)
Non-Hispanic White	Ref.
Missing	0.91 (0.19, 4.22)
Household income, $	
<25,000	2.57 (0.91, 7.31)
25,000–49,999	1.77 (0.61, 5.11)
50,000–74,999	0.76 (0.21, 2.78)
75,000–99,999	0.41 (0.08, 2.22)
100,000–150,000	1.98 (0.63, 6.21)
>150,000	Ref.
Prefer not to answer	1.92 (0.69, 5.36)
Missing	1.24 (0.40, 3.83)
Parental educational attainment	
<High school	0.93 (0.41, 2.11)
High school graduate or GED	0.62 (0.31, 1.23)
College	0.57 (0.31, 1.08)
Graduate school	Ref.
Missing	0.40 (0.18, 0.91)
Insurance status	
Private	Ref.
Government-assisted	2.04 (1.22, 3.41)
Other, including non-US insurance	0.88 (0.32, 2.39)
None	1.19 (0.26, 5.45)
Missing	2.53 (1.01, 6.34)

* Excludes 8 participants missing age-related variables. 95% CI = 95% confidence interval; CARRA = Childhood Arthritis and Rheumatology Research Alliance; GED = General Educational Development Test; NHPI = Native Hawaiian or Pacific Islander; OR = odds ratio; Ref. = reference; SDI = Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index; SLE = systemic lupus erythematosus; SLEDAI-2K = Systemic Lupus Erythematosus Disease Activity Index 2000
† Models run separately for each demographic or socioeconomic factor and adjusted for age (except age model), disease duration at enrollment (continuous), and time since first seen by a pediatric rheumatologist (continuous).
‡ Native American or Alaska Native, Middle Eastern, North African, race not previously mentioned, multiracial.

private health insurance (β = –0.07 [95% CI –1.66, 1.81]). Finally, in nonprobabilistic bias analyses, estimated SLEDAI-2K scores were slightly elevated when missing data were imputed to the most disadvantaged group (no health insurance β = 0.41 [95% CI –0.99, 1.81]) and slightly decreased when missing data were imputed to the most advantaged group (private health insurance β = –0.22 [95% CI –1.58, 1.14]). Conversely, we observed elevated SLEDAI-2K scores associated with having no health insurance, compared to private insurance, across all methods used to address missing data, ranging from 2.69-fold higher scores (95% CI 0.47, 4.92) when missing data were imputed to the most disadvantaged to 4.91-fold higher scores (95% CI 0.13, 9.70) in complete case analysis. Similar levels of variation in estimates, either in direction or magnitude, were observed across the 3 methods to address missing data when assessing the association between socioeconomic factors and SDI at enrollment (see Supplementary Table 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25136).

DISCUSSION

This is the first study to systematically assess the impact of missing data on the interpretation of findings in a pediatric rheumatology research setting. Although missing data were most highly correlated among variables of the same type (i.e., demographic, socioeconomic, or clinical factors), missingness of clinical data were socioeconomically patterned such that the odds of missing SLEDAI-2K or SDI scores at enrollment were noticeably higher among participants who reported government-assisted insurance compared to private insurance. Using multiple statistical methods to address missing data (i.e., complete case analysis, multiple imputation, and nonprobabilistic bias analyses) and to evaluate the association between socioeconomic factors and enrollment SLEDAI-2K scores, we observed estimates across methods that varied in both direction and magnitude and suggest conflicting conclusions: decreased disease activity, no difference in disease activity, or elevated disease activity at enrollment in relation to lower socioeconomic position. Notably, this variation in the results clearly demonstrates how missing data and the methodology used to address missingness can influence results and subsequent conclusions derived in observational studies. Differences in estimated effects would not have been captured or considered in the interpretation of results if a single analytic method had been selected. These results support the use of multiple methods to evaluate the role of missing data in quantitative analyses and further sensitivity analyses to better identify the underlying relationship between the exposures and outcomes of interest.

Similar to other registry studies, the CARRA Registry relies on longitudinal observational data collection, which makes it vulnerable to missing data due to loss to follow-up, missed study visits, and participant failure to complete visit questionnaires (32). Proactive measures that reduce the amount of missing data during the data collection phase can reduce the dependence on post hoc solutions in the analytic phase (33). For example, the prevalence of missing data in patient registries may be attributed to several factors, including training and staffing at registry sites. To ensure the completion of self-administered questionnaires, the medical staff may need to routinely check on and assist

Table 3. Associations between socioeconomic factors and SLEDAI-2K score at enrollment among participants with childhood-onset systemic lupus erythematosus enrolled in the CARRA Registry using different analytic methods to address missing data*

			Nonprobabilistic bias analysis (n = 676)	
	Complete case analysis (n = 381)	20 multiple imputations (n = 766)	Imputed most disadvantaged/ marginalized	Imputed least disadvantaged/ marginalized
Race and ethnicity				
Asian/NHPI	2.33 (–0.80, 5.45)	1.31 (–0.79, 2.90)	1.79 (–0.25, 3.83)	1.58 (–0.42, 3.59)
Hispanic	0.60 (–1.79, 2.99)	1.22 (–0.46, 2.90)	1.48 (–0.18, 3.13)	1.07 (–0.56, 2.69)
Non-Hispanic Black/ African American	–0.58 (–2.80, 1.65)	0.18 (–1.42, 1.78)	0.25 (–1.28, 1.79)	0.00 (–1.53, 1.54)
Non-Hispanic other†	0.51 (–2.56, 3.58)	2.00 (–0.31, 4.32)	1.87 (–0.36, 4.10)	1.91 (–0.31, 4.12)
Non-Hispanic White	Ref.	Ref.	Ref.	Ref.
Household income, $				
<25,000	1.89 (–1.79, 5.56)	–0.50 (–3.46, 2.46)	–1.36 (–3.60, 0.89)	0.88 (–0.93, 2.69)
25,000–49,999	1.06 (–2.24, 4.37)	–0.29 (–2.97, 2.40)	–0.25 (–2.78, 2.27)	1.08 (–0.62, 2.79)
50,000–74,999	0.66 (–2.50, 3.82)	–0.62 (–3.11, 1.87)	–0.60 (–3.18, 1.97)	0.58 (–1.31, 2.46)
75,000–99,999	1.46 (–1.71, 4.63)	0.14 (–2.46, 2.74)	0.39 (–2.29, 3.06)	1.37 (–0.73, 3.47)
100,000–150,000	0.12 (–3.03, 3.27)	0.06 (–2.49, 2.62)	0.11 (–2.52, 2.74)	1.13 (–0.93, 3.19)
>150,000	Ref.	Ref.	Ref.	Ref.
Parental educational attainment				
<High school	1.03 (–2.75, 4.81)	0.70 (–1.95, 3.34)	0.46 (–1.59, 2.51)	1.10 (–1.05, 3.25)
High school grad or GED	1.28 (–1.45, 4.01)	1.18 (–1.04, 3.40)	1.06 (–0.98, 3.10)	0.85 (–0.74, 2.45)
College	0.95 (–1.41, 3.30)	0.90 (–1.02, 2.82)	0.89 (–0.95, 2.73)	0.76 (–0.67, 2.18)
Graduate school	Ref.	Ref.	Ref.	Ref.
Insurance status				
Private	Ref.	Ref.	Ref.	Ref.
Government-assisted	–1.60 (–3.76, 0.56)	0.07 (–1.66, 1.81)	0.41 (–0.99, 1.81)	–0.22 (–1.58, 1.14)
Other, including non-US insurance	–1.09 (–5.81, 3.64)	–1.71 (–4.70, 1.27)	–1.59 (–4.49, 1.30)	–1.94 (–4.78, 0.90)
None	4.91 (0.13, 9.70)	3.81 (0.05, 7.57)	2.69 (0.47, 4.92)	3.58 (0.13, 7.04)

* Values are the β (95% confidence interval), except for multiple imputations (second column), which shows the mean β. Models were adjusted for age (continuous), sex assigned at birth (male, female, other), disease duration at enrollment (continuous), time since first seen by a pediatric rheumatologist (continuous), and study site (categorical). CARRA = Childhood Arthritis and Rheumatology Research Alliance; GED = General Educational Development Test; NHPI = Native Hawaiian or Pacific Islander; Ref. = reference; SLEDAI-2K = Systemic Lupus Erythematosus Disease Activity Index 2000.
† Native American or Alaska Native, Middle Eastern, North African, race not previously mentioned, multiracial.

participants during data collection. The identification of missing data after a questionnaire is submitted may require the medical or registry staff to take additional steps to complete the missing variables (34), which may not be feasible at understaffed sites due to lack of time, resources, or training to monitor missing data. However, individuals who are experienced with data collection associated with patient registries may be more inclined to assist participants while filling out questionnaires, assess for missing data after the study visit, and subsequently follow-up with participants for questionnaire completion (34,35).

To attempt to mitigate the impact of these concerns on data completeness in the CARRA Registry, the clinical and data coordinating center provides thorough training for registry site staff, including best practices for administering questionnaires and capturing critical registry data. This training is repeated periodically to account for ongoing site staff turnover. Additionally, the registry's database includes automated validation checks that flag fields with missing data for the site to review. These validation checks are especially useful for clinical data that may be readily available for extraction from the medical record. Despite these efforts, institutional barriers that facilitate incomplete data may persist or evolve over time, and therefore, regular evaluation of ongoing data collection efforts are needed to limit data missingness.

In addition to institutional barriers that may hinder data collection, missing data can also occur when participants either avoid questions that address sensitive topics or suffer from respondent fatigue and fail to complete the full questionnaire (36). Of the variables with missing data across the CARRA Registry, household income had the highest level of missingness. Participants may feel uncomfortable disclosing personal income information if they consider it sensitive information (34). Further, accurately describing income may be challenging for participants with unstable income or employment who may opt to defer or leave the item blank. Informing patients about how their identity will be protected when their data are used may be beneficial,

along with explaining the reasons why sensitive information, such as income, is important to the research question. In lieu of individually reported income data, area level data, such as the area deprivation index or the social vulnerability index, can be used to supplement missing socioeconomic data or corroborate reported values (37,38).

When missing data persist despite these preventative measures, strategies for dealing with missing data vary, and best practices depend on the study design, the distribution of missingness, and the underlying assumptions about the structure of the data and reasons contributing to data missingness (39,40). Being intentional about the presence of missing data and the use of multiple methods to assess missing data can help to elucidate the impact of selection bias, confounding, and exposure misclassification on the interpretation of results (41) and facilitate critical assessment of potential causal statements that would otherwise have the potential to under- or overestimate existing health inequities.

Although the use of multiple imputation to address missing data in medical research and clinical trials has gained traction (42,43), these methods have yet to become standard practice in epidemiologic studies in pediatric rheumatology. When looking at prior CARRA Registry research studies, we found that while many reported missing data, most used complete case analysis methods (7,8,44), and only a few CARRA Registry studies conducted sensitivity analyses or used multiple imputation to address missing data and to assess its impact on study results (11,20). These trends are similar to those observed in studies published in key rheumatology journals (17). However, when data are not missing at random, complete case analysis can bias study results, and when missing data are socially patterned, they can introduce systematic biases that can have subsequent health equity implications. Many pediatric rheumatology researchers may not be aware of how different patterns of missing data may impact study results and their interpretation.

We recommend, therefore, that 1) attempts be made to reduce the potential for missing data during the study planning and implementation phases, 2) trained statistical personnel be included in multidisciplinary research teams to help identify and address persisting issues related to missing data, 3) authors be transparent about their missing data by reporting missing values and the methods used to address the missingness, and 4) results be thoughtfully interpreted after considering the impact of any missing data and the methods used to mitigate their effect.

A primary limitation of this study relates to the creation of the analytic sample. We excluded participants classified as having childhood-onset SLE per SLICC diagnostic criteria who did not fulfill at least 4 ACR criteria for SLE or did not have renal involvement and fulfill 2 additional criteria (n = 159) in order to support generalizability among youth diagnosed with childhood-onset SLE. As a result, we did not capture potential missing data related to childhood-onset SLE diagnosis. Furthermore, whereas the CARRA Registry includes thousands of variables, as well as additional complications related to the collection of prospective longitudinal data, in this proof-of-concept study, we included a limited number of pertinent variables collected at enrollment to demonstrate the potential health equity implications of missing data. Additional analyses may be warranted to fully assess the impact of missing data across more variables. Further, the inclusion of additional variables that are correlated with missingness of our variables of interest could potentially improve the quality of the imputation models (45).

The methods presented in this study address missing data within the analytic sample, which is also limited by the sample of participants who consent to participate in the CARRA Registry. Our methods do not address the overall representativeness of youths with childhood-onset SLE who are not captured in the CARRA Registry because they are not approached for research, are not treated at a CARRA Registry site, or are not seen by a pediatric rheumatologist due to workforce shortages. Although there are methods to help address potential selection bias (i.e., probability weights), dedicated research is needed to elucidate what populations may be underrepresented from the underlying population of interest.

Finally, the reason for data missingness is unknown (e.g., due to missing at random, differences in research workflows between sites, participant refusal to answer specific questions, etc.). Further assessment of these reasons could help to inform reporting biases and subsequent health equity implications. Targeted interventions to improve reporting among groups where missing data are demonstrably higher may be warranted. Despite these limitations, this study sheds light on the potential impacts of missing data in pediatric rheumatology research and their role in how results may be interpreted.

Every health registry and corresponding database has missing data. Previous studies have relied heavily on complete case analysis to test study hypotheses; however, these practices can misrepresent existing biases. In this study, we demonstrate the wide variation in results that may occur when using different statistical methods to address missing data. Our findings highlight the need in pediatric rheumatology research for 1) careful review of study data to identify patterns of missingness and 2) selection of appropriate methodology to handle missing data to ensure that misleading findings do not exacerbate existent health inequities.

ACKNOWLEDGMENTS

The authors thank the following organizations: the NIH's National Institute of Arthritis and Musculoskeletal and Skin Diseases and the Arthritis Foundation. They also thank Drs. Paige Bommarito and Peter Grayson for review of this article and Yolanda L. Jones, NIH Library, for manuscript editing assistance. Finally, the authors thank all the Childhood Arthritis and Rheumatology Research Alliance Registry participants, without whom this work would not be possible.

AUTHOR CONTRIBUTIONS

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

Study conception and design. Woo, Simmonds, Lewandowski, Rubinstein.

Acquisition of data. Dennos, Son.

Analysis and interpretation of data. Woo, Dennos, Son, Lewandowski, Rubinstein.

REFERENCES

1. Hiraki LT, Benseler SM, Tyrrell PN, et al. Ethnic differences in pediatric systemic lupus erythematosus. J Rheumatol 2009;36:2539–46.
2. Carter EE, Barr SG, Clarke AE. The global burden of SLE: prevalence, health disparities and socioeconomic impact [review]. Nat Rev Rheumatol 2016;12:605–20.
3. Falasinnu T, Chaichian Y, Bass MB, et al. The representation of gender and race/ethnic groups in randomized clinical trials of individuals with systemic lupus erythematosus [review]. Curr Rheumatol Rep 2018;20:20.
4. AlE'ed A, Vega-Fernandez P, Muscal E, et al. Challenges of diagnosing cognitive dysfunction with neuropsychiatric systemic lupus erythematosus in childhood [review]. Arthritis Care Res (Hoboken) 2017;69:1449–59.
5. Ardoin SP, Daly RP, Merzoug L, et al. Research priorities in childhood-onset lupus: results of a multidisciplinary prioritization exercise. Pediatr Rheumatol Online J 2019;17:32.
6. Chang JC, Mandell DS, Knight AM. High health care utilization preceding diagnosis of systemic lupus erythematosus in youth. Arthritis Care Res (Hoboken) 2018;70:1303–11.
7. Driest KD, Sturm MS, O'Brien SH, et al. Factors associated with thrombosis in pediatric patients with systemic lupus erythematosus. Lupus 2016;25:749–53.
8. Hersh AO, Case SM, Son MB, et al. Predictors of disability in a childhood-onset systemic lupus erythematosus cohort: results from the CARRA Legacy Registry. Lupus 2018;27:494–500.
9. Knight AM, Trupin L, Katz P, et al. Depression risk in young adults with juvenile- and adult-onset lupus: twelve years of followup. Arthritis Care Res (Hoboken) 2018;70:475–80.
10. Knight AM, Vickery ME, Muscal E, et al. Identifying targets for improving mental healthcare of adolescents with systemic lupus erythematosus: perspectives from pediatric rheumatology clinicians in the United States and Canada. J Rheumatol 2016;43:1136–45.
11. Rubinstein TB, Mowrey WB, Ilowite NT, et al. Delays to care in pediatric lupus patients: data from the Childhood Arthritis and Rheumatology Research Alliance Legacy Registry. Arthritis Care Res (Hoboken) 2018;70:420–7.
12. Schanberg LE, Sandborg C, Barnhart HX, et al. Use of atorvastatin in systemic lupus erythematosus in children and adolescents. Arthritis Rheum 2012;64:285–96.
13. Wahezi DM, Ilowite NT, Wu XX, et al. Annexin A5 anticoagulant activity in children with systemic lupus erythematosus and the association with antibodies to domain I of β2-glycoprotein I. Lupus 2013;22:702–11.
14. Yen EY, Shaheen M, Woo JM, et al. 46-year trends in systemic lupus erythematosus mortality in the United States, 1968 to 2013: a nationwide population-based study. Ann Intern Med 2017;167:777–85.
15. Johnson CY, Howards PP, Strickland MJ, et al. Multiple bias analysis using logistic regression: an example from the National Birth Defects Prevention Study. Ann Epidemiol 2018;28:510–4.
16. Lash TL, Fox MP, Fink AK. Applying quantitative bias analysis to epidemiologic data. New York: Springer; 2009.
17. Lauper K, Kedra J, De Wit M, et al. Analysing and reporting of observational data: a systematic review informing the EULAR points to consider when analysing and reporting comparative effectiveness research with observational data in rheumatology. RMD Open 2021;7:e001818.
18. Balmuri N, Soulsby WD, Cooley V, et al. Community poverty level influences time to first pediatric rheumatology appointment in polyarticular juvenile idiopathic arthritis. Pediatr Rheumatol Online J 2021;19:122.
19. Ilowite NT, Prather K, Lokhnygina Y, et al. Randomized, double-blind, placebo-controlled trial of the efficacy and safety of rilonacept in the treatment of systemic juvenile idiopathic arthritis. Arthritis Rheumatol 2014;66:2570–9.
20. Kimura Y, Schanberg LE, Tomlinson GA, et al. Optimizing the start time of biologics in polyarticular juvenile idiopathic arthritis: a comparative effectiveness study of Childhood Arthritis and Rheumatology Research Alliance consensus treatment plans. Arthritis Rheumatol 2021;73:1898–909.
21. Liu K, Tomlinson G, Reed AM, et al. Pilot study of the juvenile dermatomyositis consensus treatment plans: a CARRA registry study. J Rheumatol 2021;48:114–22.
22. Ong MS, Ringold S, Kimura Y, et al. Improved disease course associated with early initiation of biologics in polyarticular juvenile idiopathic arthritis: trajectory analysis of a Childhood Arthritis and Rheumatology Research Alliance consensus treatment plans study. Arthritis Rheumatol 2021;73:1910–20.
23. Phillippi K, Hoeltzel M, Byun Robinson A, et al. Race, income, and disease outcomes in juvenile dermatomyositis. J Pediatr 2017;184:38–44.
24. Weitzman ER, Wisk LE, Salimian PK, et al. Adding patient-reported outcomes to a multisite registry to quantify quality of life and experiences of disease and treatment for youth with juvenile idiopathic arthritis. J Patient Rep Outcomes 2018;2:1.
25. Beukelman T, Kimura Y, Ilowite NT, et al. The new Childhood Arthritis and Rheumatology Research Alliance (CARRA) registry: design, rationale, and characteristics of patients enrolled in the first 12 months. Pediatr Rheumatol Online J 2017;15:30.
26. Tan EM, Cohen AS, Fries JF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1982;25:1271–7.
27. Adkins-Jackson PB, Chantarat T, Bailey ZD, et al. Measuring structural racism: a guide for epidemiologists and other health researchers. Am J Epidemiol 2022;191:539–47.
28. Bonilla-Silva E. Rethinking racism: toward a structural interpretation. Am Sociol Rev 1997;62:465–80.
29. Smedley A, Smedley BD. Race as biology is fiction, racism as a social problem is real: anthropological and historical perspectives on the social construction of race. Am Psychol 2005;60:16–26.
30. Williams DR, Priest N, Anderson NB. Understanding associations among race, socioeconomic status, and health: patterns and prospects. Health Psychol 2016;35:407–11.
31. Williams DR, Collins C. Racial residential segregation: a fundamental cause of racial disparities in health. Public Health Rep 2001;116:404–16.
32. Ayilara OF, Zhang L, Sajobi TT, et al. Impact of missing data on bias and precision when estimating change in patient-reported outcomes from a clinical registry. Health Qual Life Outcomes 2019;17:106.
33. Scharfstein DO, Hogan J, Herman A. On the prevention and analysis of missing data in randomized clinical trials: the state of the art. J Bone Joint Surg Am 2012;94 Suppl 1:80–4.

34. Kyte D, Ives J, Draper H, et al. Current practices in patient-reported outcome (PRO) data collection in clinical trials: a cross-sectional survey of UK trial staff and management. BMJ Open 2016;6:e012281.
35. Palmer MJ, Krupa T, Richardson H, et al. Clinical research associates experience with missing patient reported outcomes data in cancer randomized controlled trials. Cancer Med 2021;10:3026–34.
36. Olson K, Smyth JD, Ganshert A. The effects of respondent and question characteristics on respondent answering behaviors in telephone interviews. J Surv Stat Methodol 2019;7:275–308.
37. Kind AJ, Buckingham WR. Making neighborhood-disadvantage metrics accessible: the Neighborhood Atlas. N Engl J Med 2018;378: 2456–8.
38. Flanagan BE, Gregory EW, Hallisey EJ, et al. A social vulnerability index for disaster management. J Homel Secur Emer Manag 2011;8.
39. Graham JW. Missing data analysis: making it work in the real world [review]. Annu Rev Psychol 2009;60:549–76.
40. Raghunathan TE. What do we do with missing data? Some options for analysis of incomplete data. Annu Rev Public Health 2004;25: 99–117.
41. Van Smeden M, Penning de Vries BB, Nab L, et al. Approaches to addressing missing values, measurement error, and confounding in epidemiologic studies. J Clin Epidemiol 2021;131:89–100.
42. Mackinnon A. The use and reporting of multiple imputation in medical research: a review. J Intern Med 2010;268:586–93.
43. Alsaber A, Al-Herz A, Pan J, et al. Handling missing data in a rheumatoid arthritis registry using random forest approach. Int J Rheum Dis 2021;24:1282–93.
44. Vazzana KM, Daga A, Goilav B, et al. Principles of pediatric lupus nephritis in a prospective contemporary multi-center cohort [review]. Lupus 2021;30:1660–70.
45. Enders CK. Applied missing data analysis. New York: Guilford Press; 2010.

APPENDIX A: THE CHILDHOOD ARTHRITIS AND RHEUMATOLOGY RESEARCH ALLIANCE

Members of the Childhood Arthritis and Rheumatology Research Alliance Registry site principal investigators, subinvestigators, and research coordinators are as follows: N. Abel, K. Abulaban, A. Adams, M. Adams, R. Agbayani, J. Aiello, S. Akoghlanian, C. Alejandro, E. Allenspach, R. Alperin, M. Alpizar, G. Amarilyo, W. Ambler, E. Anderson, S. Ardoin, S. Armendariz, E. Baker, I. Balboni, S. Balevic, L. Ballenger, S. Ballinger, N. Balmuri, F. Barbar-Smiley, L. Barillas-Arias, M. Basiaga, K. Baszis, M. Becker, H. Bell-Brunson, E. Beltz, H. Benham, S. Benseler, W. Bernal, T. Beukelman, T. Bigley, B. Binstadt, C. Black, M. Blakley, J. Bohnsack, J. Boland, A. Boneparth, S. Bowman, C. Bracaglia, E. Brooks, M. Brothers, A. Brown, H. Brunner, M. Buckley, M. Buckley, H. Bukulmez, D. Bullock, B. Cameron, S. Canna, L. Cannon, P. Carper, V. Cartwright, E. Cassidy, L. Cerracchio, E. Chalom, J. Chang, A. Chang-Hoftman, V. Chauhan, P. Chira, T. Chinn, K. Chundru, H. Clairman, D. Co, A. Confair, H. Conlon, R. Connor, A. Cooper, J. Cooper, S. Cooper, C. Correll, R. Corvalan, D. Costanzo, R. Cron, L. Curiel-Duran, T. Curington, M. Curry, A. Dalrymple, A. Davis, C. Davis, C. Davis, T. Davis, F. De Benedetti, D. De Ranieri, J. Dean, F. Dedeoglu, M. DeGuzman, N. Delnay, V. Dempsey, E. DeSantis, T. Dickson, J. Dingle, B. Donaldson, E. Dorsey, S. Dover, J. Dowling, J. Drew, K. Driest, Q. Du, K. Duarte, D. Durkee, E. Duverger, J. Dvergsten, A. Eberhard, M. Eckert, K. Ede, B. Edelheit, C. Edens, C. Edens, Y. Edgerly, M. Elder, B. Ervin, S. Fadrhonc, C. Failing, D. Fair, M. Falcon, L. Favier, S. Federici, B. Feldman, J. Fennell, I. Ferguson, P. Ferguson, B. Ferreira, R. Ferrucho, K. Fields, T. Finkel, M. Fitzgerald, C. Fleming, O. Flynn, L. Fogel, E. Fox, M. Fox, L. Franco, M. Freeman, K. Fritz, S. Froese, R. Fuhlbrigge, J. Fuller, N. George, K. Gerhold, D. Gerstbacher, M. Gilbert, M. Gillispie-Taylor, E. Giverc, C. Godiwala, I. Goh, H. Goheer, D. Goldsmith, E. Gotschlich, A. Gotte, B. Gottlieb, C. Gracia, T. Graham, S. Grevich, T. Griffin, J. Griswold, A. Grom, M. Guevara, P. Guittar, M. Guzman, M. Hager, T. Hahn, O. Halyabar, E. Hammelev, M. Hance, A. Hanson, L. Harel, S. Haro, J. Harris, O. Harry, E. Hartigan, J. Hausmann, A. Hay, K. Hayward, J. Heiart, K. Hekl, L. Henderson, M. Henrickson, A. Hersh, K. Hickey, P. Hill, S. Hillyer, L. Hiraki, M. Hiskey, P. Hobday, C. Hoffart, M. Holland, M. Hollander, S. Hong, M. Horwitz, J. Hsu, A. Huber, J. Huggins, J. Hui-Yuen, C. Hung, J. Huntington, A. Huttenlocher, M. Ibarra, L. Imundo, C. Inman, A. Insalaco, A. Jackson, S. Jackson, K. James, G. Janow, J. Jaquith, S. Jared, N. Johnson, J. Jones, J. Jones, J. Jones, K. Jones, S. Jones, S. Joshi, L. Jung, C. Justice, A. Justiniano, N. Karan, K. Kaufman, A. Kemp, E. Kessler, U. Khalsa, B. Kienzle, S. Kim, Y. Kimura, D. Kingsbury, M. Kitcharoensakkul, T. Klausmeier, K. Klein, M. Klein-Gitelman, B. Kompelien, A. Kosikowski, L. Kovalick, J. Kracker, S. Kramer, C. Kremer, J. Lai, J. Lam, B. Lang, S. Lapidus, B. Lapin, A. Lasky, D. Latham, E. Lawson, R. Laxer, P. Lee, P. Lee, T. Lee, L. Lentini, M. Lerman, D. Levy, S. Li, S. Lieberman, L. Lim, C. Lin, N. Ling, M. Lingis, M. Lo, D. Lovell, D. Lowman, N. Luca, S. Lvovich, C. Madison, J. Madison, S. Magni Manzoni, B. Malla, J. Maller, M. Malloy, M. Mannion, C. Manos, L. Marques, A. Martyniuk, T. Mason, S. Mathus, L. McAllister, K. McCarthy, K. McConnell, E. McCormick, D. McCurdy, P. McCurdy Stokes, S. McGuire, I. McHale, A. McMonagle, C. McMullen-Jackson, E. Meidan, E. Mellins, E. Mendoza, R. Mercado, A. Merritt, L. Michalowski, P. Miettunen, M. Miller, D. Milojevic, E. Mirizio, E. Misajon, M. Mitchell, R. Modica, S. Mohan, K. Moore, L. Moorthy, S. Morgan, E. Morgan Dewitt, C. Moss, T. Moussa, V. Mruk, A. Murphy, E. Muscal, R. Nadler, B. Nahal, K. Nanda, N. Nasah, L. Nassi, S. Nativ, M. Natter, J. Neely, B. Nelson, L. Newhall, L. Ng, J. Nicholas, R. Nicolai, P. Nigrovic, J. Nocton, B. Nolan, E. Oberle, B. Obispo, B. O'Brien, T. O'Brien, O. Okeke, M. Oliver, J. Olson, K. O'Neil, K. Onel, A. Orandi, M. Orlando, S. Osei-Onomah, R. Oz, E. Pagano, A. Paller, N. Pan, S. Panupattanapong, M. Pardeo, J. Paredes, A. Parsons, J. Patel, K. Pentakota, P. Pepmueller, T. Pfeiffer, K. Phillippi, K. Phillippi, D. Pires Marafon, L. Ponder, R. Pooni, S. Prahalad, S. Pratt, S. Protopapas, B. Puplava, J. Quach, M. Quinlan-Waters, C. Rabinovich, S. Radhakrishna, J. Rafko, J. Raisian, A. Rakestraw, C. Ramirez, E. Ramsay, S. Ramsey, R. Randell, A. Reed, A. Reed, A. Reed, H. Reid, K. Remmel, A. Repp, A. Reyes, A. Richmond, M. Riebschleger, S. Ringold, M. Riordan, M. Riskalla, M. Ritter, R. Rivas-Chacon, A. Robinson, E. Rodela, M. Rodriquez, K. Rojas, T. Ronis, M. Rosenkranz, B. Rosolowski, H. Rothermel, D. Rothman, E. Roth-Wojcicki, K. Rouster-Stevens, T. Rubinstein, N. Ruth, N. Saad, S. Sabbagh, E. Sacco, R. Sadun, C. Sandborg, A. Sanni, L. Santiago, A. Sarkissian, S. Savani, L. Scalzi, L. Schanberg, S. Scharnhorst, K. Schikler, A. Schlefman, H. Schmeling, K. Schmidt, E. Schmitt, R. Schneider, K. Schollaert-Fitch, G. Schulert, T. Seay, C. Seper, J. Shalen, R. Sheets, A. Shelly, S. Shenoi, K. Shergill, J. Shirley, M. Shishov, C. Shivers, E. Silverman, N. Singer, V. Sivaraman, J. Sletten, A. Smith, C. Smith, J. Smith, J. Smith, E. Smitherman, J. Soep, M. Son, S. Spence, L. Spiegel, J. Spitznagle, R. Sran, H. Srinivasalu, H. Stapp, K. Steigerwald, Y. Sterba Rakovchik, S. Stern, A. Stevens, B. Stevens, R. Stevenson, K. Stewart, C. Stingl, J. Stokes, M. Stoll, E. Stringer, S. Sule, J. Sumner, R. Sundel, M. Sutter, R. Syed, G. Syverson, A. Szymanski, S. Taber, R. Tal, A. Tambralli, A. Taneja, T. Tanner, S. Tapani, G. Tarshish, S. Tarvin, L. Tate, A. Taxter, J. Taylor, M. Terry, M. Tesher, A. Thatayatikom, B. Thomas, K. Tiffany, T. Ting, A. Tipp, D. Toib, K. Torok, C. Toruner, H. Tory, M. Toth, S. Tse, V. Tubwell, M. Twilt, S. Uriguen, T. Valcarcel, H. Van Mater, L. Vannoy, C. Varghese, N. Vasquez, K. Vazzana, R. Vehe, K. Veiga, J. Velez, J. Verbsky, G. Vilar, N. Volpe, E. von Scheven, S. Vora, J. Wagner, L. Wagner-Weiner, D. Wahezi, H. Waite, J. Walker, H. Walters, T. Wampler Muskardin, L. Waqar, M. Waterfield, M. Watson, A. Watts, P. Weiser, J. Weiss, P. Weiss, E. Wershba, A. White, C. Williams, A. Wise, J. Woo, L. Woolnough, T. Wright, E. Wu, A. Yalcindag, M. Yee, E. Yen, R. Yeung, K. Yomogida, Q. Yu, R. Zapata, A. Zartoshti, A. Zeft, R. Zeft, Y. Zhang, Y. Zhao, A. Zhu, C. Zic.

Arthritis Care & Research
Vol. 75, No. 11, November 2023, pp 2295–2305
DOI 10.1002/acr.25143

American College of Rheumatology
Empowering Rheumatology Professionals

Development of American College of Rheumatology Quality Measures for Systemic Lupus Erythematosus: A Modified Delphi Process With Rheumatology Informatics System for Effectiveness (RISE) Registry Data Review

Christie M. Bartels,[1] April Jorge,[2] Candace H. Feldman,[3] JoAnn Zell,[4] Bonnie Bermas,[5] Claire E. H. Barber,[6] Alí Duarte-García,[7] Shivani Garg,[1] Leah Haseley,[8] Shraddha Jatwani,[9] Tracy Johansson,[10] Alex Limanni,[11] Wendy Rodgers,[12] Brad H. Rovin,[13] Yesenia Santiago-Casas,[14] Lisa G. Suter,[15] April Barnado,[16] Jennifer Ude,[10] Alfredo Aguirre,[17] Jing Li,[17] Gabriela Schmajuk,[17] and Jinoos Yazdany[17]

Objective. We aimed to develop readily measurable digital quality measure statements for clinical care in systemic lupus erythematosus (SLE) using a multistep process guided by consensus methods.

Methods. Using a modified Delphi process, an American College of Rheumatology (ACR) workgroup of SLE experts reviewed all North American and European guidelines from 2000 to 2020 on treatment, monitoring, and phenotyping of patients with lupus. Workgroup members extracted quality constructs from guidelines, rated these by importance and feasibility, and generated evidence-based quality measure statements. The ACR Rheumatology Informatics System for Effectiveness (RISE) Registry was queried for measurement data availability. In 3 consecutive Delphi sessions, a multidisciplinary Delphi panel voted on the importance and feasibility of each statement. Proposed measures with consensus on feasibility and importance were ranked to identify the top 3 measures.

Results. Review of guidelines and distillation of 57 quality constructs resulted in 15 quality measure statements. Among these, 5 met high consensus for importance and feasibility, including 2 on treatment and 3 on laboratory monitoring measures. The 3 highest-ranked statements were recommended for further measure specification as SLE digital quality measures: 1) hydroxychloroquine use, 2) limiting glucocorticoid use >7.5 mg/day to <6 months, and 3) end-organ monitoring of kidney function and urine protein excretion at least every 6 months.

Conclusion. The Delphi process selected 3 quality measures for SLE care on hydroxychloroquine, glucocorticoid reduction, and kidney monitoring. Next, measures will undergo specification and validity testing in RISE and US rheumatology practices as the foundation for national implementation and use in quality improvement programs.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a chronic autoimmune disease associated with significant morbidity and premature mortality. Studies have characterized numerous disparities in health care access and quality among people with SLE (1). Efforts to improve care for patients with SLE are needed. Digital quality measures leverage electronic health records

Contents are solely the responsibility of the authors and do not necessarily represent the official views of the Department of Health and Human Services or the Centers for Disease Control and Prevention (CDC).

Supported by the CDC Department of Health and Human Services (grant 1-NU-58DP006908-01-00-A), Developing and Disseminating Programs to Build Sustainable Lupus Awareness, Knowledge, Skills and Partnerships, and the American College of Rheumatology. Workgroup members received an honorarium as a member of the American College of Rheumatology Systemic Lupus Erythematosus Expert Workgroup as part of the CDC grant.

[1]Christie M. Bartels, MD, MS, Shivani Garg, MD, MS: University of Wisconsin School of Medicine and Public Health, Madison; [2]April Jorge, MD: Massachusetts General Hospital, Boston; [3]Candace H. Feldman, MD, MPH, ScD: Brigham and Women's Hospital, Boston, Massachusetts; [4]JoAnn Zell, MD: University of Colorado, Denver; [5]Bonnie Bermas, MD: UT Southwestern Medical Center, Dallas, Texas; [6]Claire E. H. Barber, MD, PhD, FRCPC: University of Calgary, Calgary, Alberta, Canada; [7]Alí Duarte-García, MD: Mayo Clinic, Rochester, Minnesota; [8]Leah Haseley, MD: University of Washington, Seattle; [9]Shraddha Jatwani, MD: Guthrie Robert Packer Hospital, Sayre, Pennsylvania; [10]Tracy Johansson, MS, Jennifer Ude, MS: American College of Rheumatology, Atlanta, Georgia; [11]Alex Limanni, MD: Arthritis Centers of Texas, Dallas; [12]Wendy Rodgers, MEd, MPH: Lupus Foundation Care and Support Services, Los Angeles, California; [13]Brad H. Rovin, MD: Ohio State University Wexner Medical Center, Columbus; [14]Yesenia Santiago-Casas, MD: Integral Rheumatology and Immunology Specialists, Plantation, Florida; [15]Lisa G. Suter, MD: Yale School of Medicine, New Haven, and Veterans Administration Medical Center, West Haven, Connecticut; [16]April Barnado, MD, MSCI: Vanderbilt University Medical Center, Nashville, Tennessee; [17]Alfredo Aguirre, MD, Jing Li, MPH, Gabriela Schmajuk, MD, MS, Jinoos Yazdany, MD, MPH: University of California San Francisco.

Drs. Bartels and Jorge contributed equally to this work.

SIGNIFICANCE & INNOVATIONS

- Despite significant morbidity and mortality among patients with lupus, none of the 25 American College of Rheumatology (ACR) quality measures specifically targets lupus.
- In collaboration with the ACR, multidisciplinary experts conducted a guideline review and modified Delphi process to generate and prioritize evidence-based quality measure statements for lupus.
- Emphasizing strong public health potential, panelists recommended 3 quality measures: 1) hydroxychloroquine use, 2) limiting glucocorticoid doses exceeding 7.5 mg daily to 6 months or less, and 3) measuring kidney function and urine protein at least as often as every 6 months.

(EHRs), claims, registries, and other digital data by facilitating timely monitoring and improvement of health care quality on a population level (2). Currently, several digital electronic clinical quality measures are tracked in the Rheumatology Informatics System for Effectiveness (RISE) Registry for rheumatoid arthritis and other conditions. Yet to date, among measures by the National Quality Forum, Centers for Medicare and Medicaid Services, and 25 American College of Rheumatology (ACR) measures, not one is specific to SLE (3).

As part of a collaboration between the Centers for Disease Control and Prevention and the ACR, we sought to develop candidate quality measures for SLE based on available guidelines that could leverage longitudinal EHR data and the ACR RISE registry. Additionally, we aimed to evaluate the importance and feasibility of potential measures, with the goal of prioritizing up to 3 measures for detailed testing. The ultimate goal is for eventual use of these SLE-specific digital quality measures in various national quality programs, including as part of the Centers for Medicare and Medicaid Services value-based care payment program known as the Quality Payment Program.

MATERIALS AND METHODS

We assembled an ACR workgroup of 10 SLE experts (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25143), including practicing rheumatologists and researchers from diverse geographic and rural-urban settings across the US and Canada. Members of the workgroup were selected based on a range of expertise in treating patients with SLE, health services research, research using longitudinal data from EHRs, and quality measure development. Proposed quality measure statements, evidence summaries, and feasibility data developed by the workgroup were subsequently reviewed by a 17-member invited multidisciplinary Delphi panel, including rheumatologists, nephrologists, and a patient representative. With oversight from the workgroup, candidate measure development included 5 phases (Figure 1): 1) literature review with identification of evidence-based SLE quality constructs, 2) evaluation of the importance and feasibility of these constructs, 3) development of IF/THEN statements for SLE quality measures, 4) assessment of data availability in the RISE registry, and 5) modified Delphi exercise (4), with evaluation and prioritization of final proposed measure statements. A priori, we planned to advance up to 3 statements with high consensus for importance and feasibility, for further development, testing, and eventual implementation as digital quality measures.

Phase I: literature review and identification of evidence-based SLE quality constructs. With assistance from a professional librarian, we conducted a literature search with PubMed, using medical subject heading (MeSH) terms for "systemic lupus erythematosus," "lupus," "lupus nephritis," and "practice guideline," and excluding the terms "child," "infant," or "adolescent," to identify all North American and European guidelines from 2000 to 2020 that focused on SLE or lupus nephritis (LN) management in adults. We similarly searched Ovid MEDLINE and Ovid Embase (see Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25143). We reviewed all peer-reviewed, published studies in English with full text available. When the same society published updated guidelines within the date range, the most recent version was included.

ACR workgroup members reviewed guidelines meeting inclusion criteria to develop a list of all potential quality constructs across 3 domains agreed upon a priori: SLE treatment, monitoring, and phenotyping. The workgroup selected these domains as being highly specific to SLE. Guideline recommendations in domains of preventive care (i.e., reproductive health, osteoporosis prevention) were excluded, since these concepts were not SLE specific.

Phase II: SLE quality construct importance and feasibility evaluation. The ACR workgroup rated the importance and feasibility of the preliminary constructs with an asynchronous web survey using a 9-point Likert scale (4,5). Importance was specified as important for high-quality SLE care on a population

Author disclosures and graphical abstract are available online at https://onlinelibrary.wiley.com/doi/10.1002/acr.25143.

Address correspondence via email to Christie M. Bartels, MD, MS, at cb4@medicine.wisc.edu.

Submitted for publication January 4, 2023; accepted in revised form May 4, 2023.

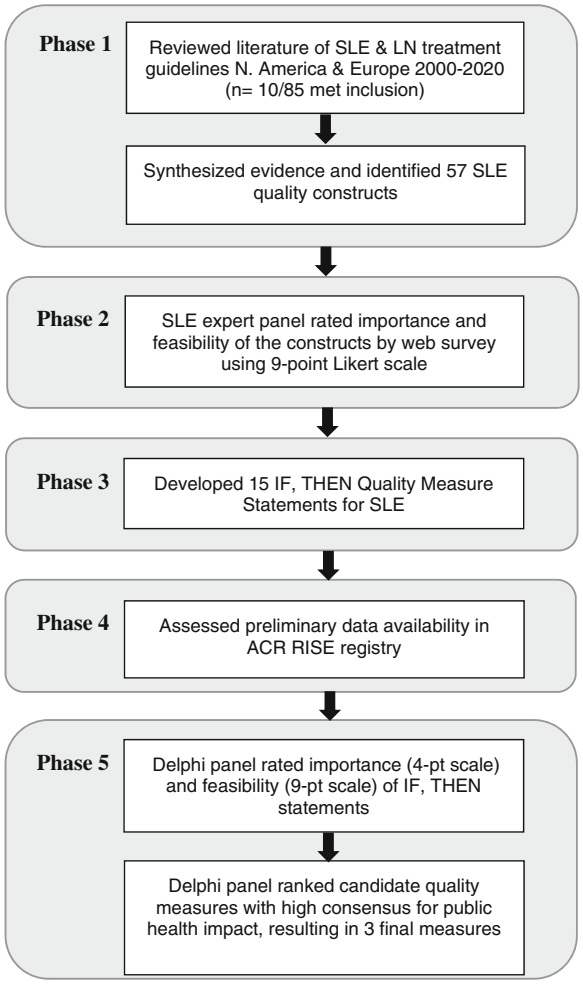

Figure 1. Overview of the process to develop digital quality measures for systemic lupus erythematosus (SLE). ACR = American College of Rheumatology; LN = lupus nephritis; RISE = Rheumatology Informatics System for Effectiveness.

level. The highest score was 9 (extremely important) and the lowest was 1 (not important). Feasibility specified whether an item would be feasible for implementation as a digital quality measure, using EHR or other electronic health information. The highest score was 9 (extremely feasible) with the lowest score of 1 (not feasible). Incorporating the RAND/University of California Los Angeles (UCLA) appropriateness method (6), consensus for high importance or high feasibility was defined a priori as ≥60% of ratings ≥7 and as ≤1 rating ≤3, respectively, after excluding 1 extreme low (i.e., 1–2) and 1 extreme high rating (i.e., 8–9) (4,5).

Phase III: SLE quality measure IF/THEN statement development. Informed by these ratings of the quality measure constructs, the workgroup then developed candidate IF/THEN statements for SLE quality measure constructs and accompanying evidence summaries. IF statements defined eligibility for the measures. All measures included adult populations age ≥18 years with SLE (7) or LN, and clinical exclusions by measure were proposed (8). THEN statements defined quality measure indicators as reflected in guideline literature.

Phase IV: data availability and preliminary gaps in candidate SLE quality measures. Next, we queried data from the ACR's RISE registry to inform the feasibility of implementing the candidate SLE quality measures using data derived from the EHR, as well as to assess potential gaps in meeting these candidate measures to help inform the potential public health impact of implementation. RISE is a national registry that collects EHR data from rheumatology practices across the US, including 1,000 US rheumatologists (3). Available RISE data included diagnostic codes, medications, and laboratory data captured in structured EHR fields but did not include unstructured fields (i.e., narrative text, clinical notes, pathology reports, and other scanned documents).

We identified all participating RISE practices and patients who met published definitions of SLE or LN (7,8). For each IF/THEN statement, we assessed the proportion of RISE practices with relevant data available and the preliminary proportion of patients who met the candidate measures in 2019. Failure to enter the numerator of a measure could reflect a lack of data availability or an actual gap in care. Findings were incorporated into a comprehensive evidence summary for the 15 candidate measures (see Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25143). This summary references relevant guideline recommendations, evidence, and RISE data availability that was shared with the panel in the project's next phase. The Western Institutional Review Board determined that RISE is a quality improvement registry deemed minimal risk with a waiver of individual informed consent.

Phase V: modified Delphi process and final prioritization. The Delphi panel convened 4 virtual video conference meetings in November 2021, December 2021, January 2022, and February 2022 and completed 2 rounds of ratings for each IF/THEN quality measure statement. A final Delphi round was conducted to rank measures and arrive at a final group of 3 recommended measures. Before each meeting, the panel members were instructed to review several IF/THEN statements and the evidence summaries document and to complete the first round of rankings via an anonymous premeeting online survey. The panelists asynchronously ranked each IF/THEN statement for importance for high quality SLE care and feasibility for implementation as a digital quality measure. Importance was rated on a 4-point Likert scale (A–D), with the highest score A (extremely important) and the lowest score D (not important). Feasibility was rated on a 9-point Likert scale (range 1–9), with the highest score 9 (extremely feasible) and the lowest score 1 (definitely not feasible). A priori, consensus for importance was defined as

≥60% of ratings A–B and ≤1 rating of D, after excluding 1 extreme low (i.e., D) and 1 extreme high rating (i.e., A). Consensus for feasibility was defined as ≥60% of ratings ≥7 and ≤2 rating ≤3, after excluding 1 extreme low (i.e., 1–2) and 1 extreme high rating (i.e., 8–9).

During the Delphi meetings, we presented the results of the premeeting surveys. After discussing each measure, we conducted a real-time second round survey of anonymous ratings for each measure using the same scales for importance and feasibility. The IF/THEN quality measure statements that reached high consensus for importance and feasibility on the second round Delphi surveys were identified. The mean ratings for each statement were calculated and normalized on a 100-point scale, and statements were ranked from highest to lowest mean importance followed by highest to lowest mean feasibility.

In the final Delphi meeting, we presented round-2 survey results and discussed the IF/THEN quality measure statements that had achieved high consensus for importance and feasibility. The Delphi panel then completed a real-time, anonymous survey to rank the highest rated statements in order of perceived public health benefit. The panel aimed a priori to endorse 2 to 3 quality measures with the highest public health impact for further specification and testing as an SLE quality measure.

RESULTS

Literature review and evidence-based SLE quality constructs. The literature review identified 85 relevant articles, and 10 met the inclusion criteria (9–18). The ACR workgroup distilled 57 quality measure constructs from these guidelines, including 15 in the treatment domain, 24 in the monitoring domain, and 18 in the phenotyping domain (see Supplementary Table 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25143). Fifteen quality constructs reached high consensus for importance ratings and were advanced to generate SLE quality measures posed as IF/THEN statements (Table 1). We did not exclude constructs that did not meet consensus for feasibility at this stage.

SLE quality measure IF/THEN statements. *Treatment domain.* The treatment domain included 7 quality measures. Nine guidelines supported the use of hydroxychloroquine by all people with SLE if there are no contraindications (9–17) including the 2019 EULAR guidelines for SLE, which gave an evidence grade of 1b/A (9). Evidence supporting the importance for SLE care included a systematic review including 4 small clinical trials and multiple observational studies indicating improvement in multiple outcomes, including lower flare rates, fewer renal relapses, reduced damage accumulation, improved overall survival, and possible prevention of thrombosis and atherosclerosis (14,19–32) (see Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25143).

The second quality measure statement was to limit hydroxychloroquine dosing in patients with SLE to ≤5 mg/kg/day to minimize the risk of toxic retinopathy. This recommendation had an evidence grade of 3b/C, per 2019 EULAR guidelines, with evidence linking this dose-threshold to retinopathy risk based on observational data (23,33), but it lacked evidence linking this dose threshold with efficacy for SLE treatment.

The third statement focused on limiting the prolonged use of glucocorticoids to doses to not exceed 7.5 mg/day for >6 months; this limiting was recommended by 7 guidelines (9–12,14–16) with an evidence grade of 1b/B (9) based on risks of long-term glucocorticoid toxicity and organ damage (9,11,12,14,15,34–43). Six months was designated as the maximal duration for higher glucocorticoid dosing based on recommended induction regimens for LN and other organ-threatening disease (10).

Next, 4 quality measures pertained to LN treatment, including the induction regimen for International Society of Nephrology/Renal Pathology Society class III/IV LN, maintenance treatment for class III/IV LN, maintenance treatment for class V LN, and the use of angiotensin-converting enzyme (ACE) inhibitors or angiotensin receptor blockers (ARBs) (10,11,13,16,17). Induction treatment for class III/IV was recommended by 9 guidelines (9–17), with evidence from a systematic review of randomized controlled trials (RCTs), and rated grade 1a/A from 2019 joint EULAR and European Renal Association–European Dialysis and Transplant Association (EULAR/ERA-EDTA) guidelines for LN (10) level A by 2012 ACR guidelines for LN (13) and rated grade 1b from 2012 consensus guidelines from the Systemic Autoimmune Disease Group of the Spanish Society of Internal Medicine and the Spanish Society of Nephrology (17). Voclosporin and belimumab were approved as adjunctive but not stand-alone therapies for LN at the time of this review, and were not yet incorporated into published treatment guidelines, so they were not included in the quality measure statement. Class III/IV maintenance had a similar level of evidence. The IF/THEN statement on class V maintenance regimen was rated grade 2b/B per 2019 EULAR/ERA-EDTA guidelines for LN based on evidence from small clinical trials and cohort studies. The use of ACE/ARBs by patients with LN with over 0.5 grams/day proteinuria and with no contraindications (e.g., pregnancy or low blood pressure) was recommended by 6 guidelines and largely based on RCT data in patients with diabetic nephropathy; data were extrapolated for LN, with 1 observational study in LN (10,11,13,14,16,17).

Monitoring domain. The monitoring domain included 2 quality measure statements for periodic laboratory monitoring, including SLE serologic testing and end-organ monitoring (i.e., nephritis, cytopenias), as well as 1 measure statement for disease activity monitoring using a validated instrument (e.g., SLE Disease Activity

Table 1. Candidate quality measures for adults with SLE*

Domain	IF/THEN quality measure statement
Treatment	HCQ use
IF	a patient has SLE,
THEN	they should have a prescription for HCQ in the measurement year unless a contraindication or adverse event is documented in the medical record
Treatment	HCQ dose
IF	a patient with SLE is receiving HCQ,
THEN	the most recent dose prescribed should be ≤5 mg/kg/day
Treatment	Limit glucocorticoid use
IF	a patient has SLE,
THEN	the glucocorticoid dose should not exceed 7.5 mg/day of prednisone (or equivalent) for more than 6 months
Treatment	Lupus nephritis class III/IV induction
IF	a patient with SLE has new class III or IV nephritis and is not pregnant,
THEN	induction therapy with mycophenolate or intravenous cyclophosphamide should be administered within 3 months of kidney biopsy or diagnosis
Treatment	Lupus nephritis class III/IV maintenance
IF	a patient with SLE has been diagnosed with class III or IV nephritis and is not pregnant,
THEN	they should be placed on therapy for at least 2 years with mycophenolate, azathioprine, or a calcineurin inhibitor
Treatment	Lupus nephritis class V maintenance
IF	a patient with SLE has been diagnosed with class V nephritis,
THEN	they should be placed on therapy for at least 2 years with either mycophenolate, a calcineurin inhibitor, or azathioprine
Treatment	ACE/ARB use in lupus nephritis
IF	a patient with lupus nephritis has proteinuria of >0.5 grams/24 hours on 2 occasions,
THEN	they should be treated with an ACE inhibitor or ARB in the absence of contraindications
Monitoring	End-organ laboratory monitoring in SLE
IF	a patient has SLE,
THEN	measurement of both kidney function and protein excretion (urinalysis and/or quantitative measurement) should be performed at least every 6 months
Monitoring	End-organ lab monitoring in lupus nephritis
IF	a patient has a history of lupus nephritis,
THEN	CBC, urinalysis, and quantitative measurement of kidney function and protein excretion should be performed every 3 months
Monitoring	SLE disease activity or damage
IF	a patient has SLE,
THEN	disease activity should be measured using a validated instrument at more than half of visits in the measurement year
Monitoring	SLE periodic serology results
IF	a patient has SLE,
THEN	the serum complements c3/c4 and anti-dsDNA antibody levels should be checked at least every 6 months
Phenotype	End-organ laboratory results at SLE diagnosis
IF	a patient has SLE,
THEN	CBC, creatinine, urinalysis, and a measure of urine protein should be performed within 6 months of diagnosis
Phenotype	SLE diagnosis serology results
IF	a patient has SLE,
THEN	ANA, anti-dsDNA antibody, anti-Smith antibody, c3, and c4 should be performed within 6 months of diagnosis
Phenotype	Kidney biopsy indications
IF	a patient with SLE has new persistent (e.g., ≥500 mg of proteinuria in 24 hours on 2 occasions), and/or worsening of serum creatinine (>30% elevation from baseline) and has not had prior lupus nephritis diagnosis or biopsy within 1 year,
THEN	a referral for a kidney biopsy should be placed
Phenotype	Antiphospholipid antibody laboratory testing at SLE diagnosis
IF	a patient has SLE,
THEN	antiphospholipid antibodies (anticardiolipin IgG and IgM, β_2 glycoprotein IgG and IgM, and lupus anticoagulant) should be checked within 1 year of SLE diagnosis

* ACE = angiotensin converting enzyme inhibitor; ANA = antinuclear antibody; ARB = angiotensin receptor blocker; CBC = complete blood count; dsDNA = double-stranded DNA; HCQ = hydroxychloroquine; SLE = systemic lupus erythematosus.

Index, British Isles Lupus Assessment Group). Multiple guidelines recommended periodic monitoring of anti–double-stranded DNA (anti-dsDNA) and complement c3 and c4 levels, although recommended frequencies varied or were not specified (10,11,13,15,17,18). Monitoring for LN with urine protein, serum creatinine kidney function, or both, with or without complete blood count monitoring, was recommended by multiple guidelines, with frequencies of at least every 6 months (10,11,13–15,17,18). Recommendations to monitor urine protein and creatinine were grade 1A and 2B, respectively, per 2019 EULAR/ERA-EDTA guidelines (10). Kidney monitoring is requisite for prompt treatment to improve kidney outcomes and prognosis

(44,45). Routine monitoring of disease activity was EULAR recommended, aiming at remission or low disease activity, but EULAR did not recommend specific validated instruments or frequency. Monitoring was recommended by Canadian and British Society of Rheumatology guidelines as grade B, with a low level of evidence for impacting SLE outcomes (14,18).

SLE phenotyping domain. The SLE phenotyping domain included 4 IF/THEN statements. Three included laboratory assessment at the time of SLE diagnosis, including antiphospholipid antibody (aPL) testing, SLE-specific serologic testing (e.g., antinuclear antibodies [ANAs], anti-dsDNA antibody, anti-Smith antibody, c3, and c4), and end-organ monitoring, including a complete blood count to identify cytopenias and urinalysis and kidney function to assess for LN (9,10,13–15,17,18). Testing for aPLs was recommended by multiple guidelines (9,10,14,17,18) with a grade of 1A per the 2019 EULAR SLE treatment guidelines (9). The fourth IF/THEN statement in the phenotyping domain pertained to indications for kidney biopsy, including the identification of new, persistent proteinuria and/or unexplained worsening kidney function as recommended by multiple guidelines, with a grade of B-C level evidence (9–17) (see Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25143).

Data availability and preliminary identification of measure gaps in the RISE Registry. Across 226 practices representing >1,000 rheumatologists in the RISE registry (3), we identified 35,859 patients with SLE and 4,826 patients with LN who had at least 2 rheumatology visits in 2019. Over 70% of patients were seen in single-specialty rheumatology practices; the mean number of annual visits was 4.2. Practice-level data availability assessments showed that at least 1 source of medication records (e.g., medication reconciliation tables) was available for all RISE practices. Medication dose, required to assess candidate measures of safe dosing, was most often available via e-prescriptions or orders (versus medication reconciliation lists); e-prescriptions or orders were available for 73% of practices for hydroxychloroquine and 56% for glucocorticoids. Laboratory monitoring of anti-dsDNA, complement, and urinalysis or quantitative urine protein were each available in >50% of practices. Only 6% of practices had structured data available containing kidney biopsy procedure codes or nephrology consult orders. LN class and dates of SLE/LN diagnosis were not reported in structured EHR fields.

Regarding preliminary measure-specific, patient-level data, 63% of patients with SLE had any documentation of hydroxychloroquine use in the assessment year (46). Among hydroxychloroquine users with dosing information available, 67% received hydroxychloroquine at doses ≤5 mg/kg (see Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25143). Hydroxychloroquine prescription dosing instructions or body weight values were missing for 29%. Few (0.3%) had a documented contraindication to hydroxychloroquine use according to International Classification of Diseases (ICD) codes for toxic retinopathy.

Glucocorticoids were used by 48% of patients with SLE in 2019. Over half (56%) of glucocorticoid users had prescription dosing instructions available, and 91% had pill size. The proportion using >7.5 mg/day for >6 months in 2019 was not readily available; as previously reported, 18.5% of glucocorticoid users with SLE in RISE used >7.5 mg/day for >90 days in 2018 (47).

Because the LN class was not available using structured data such as ICD codes, the proportions of patients with LN receiving recommended induction and maintenance therapy according to LN class are unknown. Just over one-third (36%) of patients with LN had documentation of ACE/ARB use. For end-organ monitoring, only 27% of patients with SLE and 32% with LN had ≥1 urinalysis or quantitative urine protein documented in 2019. For serologic monitoring, 51% with SLE had ≥1 anti-dsDNA test and 37% had ≥1 c3 or c4 test documented in 2019. Regarding phenotype, ANA was available in structured EHR fields for 59% of SLE patients, likely because historic data or outside testing were not captured. The proportion with serology results at SLE diagnosis is likewise unknown. Fewer than 1% of patients had an SLE-specific disease activity score (e.g., Systemic Lupus Erythematosus Disease Activity Index) documented, although 39% with SLE had a Routine Assessment of Patient Index Data score reported on ≥1 occasion, which is not a lupus-specific measure. This preliminary data assessment was presented to the Delphi panel to inform discussions regarding measure feasibility during the project's next phase.

Delphi panel discussion, ratings, and SLE quality measure endorsement. Of the 15 IF/THEN statements considered by the Delphi panel, 5 met high consensus for importance and feasibility: hydroxychloroquine use, limiting glucocorticoid doses exceeding 7.5 mg/day to ≤6 months, standardized screening for LN with end-organ monitoring for kidney function and urine protein excretion at least every 6 months, SLE serology results at diagnosis (e.g., ANA, anti-dsDNA antibody, anti-Smith antibody, c3, and c4), and end-organ laboratory evaluation at diagnosis (Table 2).

In the treatment domain, hydroxychloroquine was noted to have benefits on multiple outcomes, including SLE disease activity, damage accumulation, and overall survival, and measuring prevalent hydroxychloroquine use would be feasible using EHR data. However, discussions regarding a hydroxychloroquine dosing quality measure included the paucity of data regarding the impact of dose thresholds on SLE outcomes as well as the emerging role of hydroxychloroquine blood levels in guiding dosing.

Table 2. Delphi consensus results for quality measure statements*

	IF/THEN statements	Consensus for high importance	Mean importance	Consensus for high feasibility	Mean feasibility
1.	Treatment: HCQ use	yes	97.9	yes	83.3
2.	Treatment: HCQ dose	no	61.7	no	63.7
3.	Treatment: limit GC use	yes	87.5	yes	77.8
4.	Treatment: LN induction	yes	100.0	no	59.0
5.	Treatment: LN class III/IV maintenance	yes	92.7	no	60.8
6.	Treatment: LN class V maintenance	no	68.8	no	53.7
7.	Treatment: ACE/ARB use in LN	no	69.6	no	64.1
8.	Monitoring: SLE end-organ laboratory results	yes	92.9	yes	84.6
9.	Monitoring: LN end-organ laboratory results	yes	89.3	no	65.9
10.	Monitoring: SLE disease activity or damage	no	61.7	no	42.2
11	Monitoring: SLE serology results	no	48.2	yes	82.5
12.	Phenotype: end-organ tests at SLE diagnosis	yes	75.0	yes	85.5
13.	Phenotype: SLE diagnosis serology results	yes	73.1	yes	78.7
14	Phenotype: kidney biopsy indications	yes	85.7	no	55.6
15.	Phenotype: aPL testing at SLE diagnosis	yes	75.0	no	72.2

* n = 12–17 voters per measure. Importance was assessed on a 4-category ordinal scale (A = 4: extremely important; D = 1: not important); feasibility on a 9-point scale (1 = definitely not feasible; 9 = extremely feasible). Both were normalized to a 100-point scale. ACE = angiotensin converting enzyme inhibitor, aPL = antiphospholipid antibody; ARB = angiotensin receptor blockade; GC = glucocorticoid; HCQ = hydroxychloroquine; LN = lupus nephritis; SLE = systemic lupus erythematosus.

Glucocorticoid toxicity was noted as a major problem for patients with SLE and LN; reducing glucocorticoid exposure has the potential to reduce long-term harm as well as improve outcomes. Discussion of this measure included the challenges of assessing glucocorticoid dose from EHR data, from pill size and number dispensed, since patients may be instructed on increases or tapers that are not documented on the prescription. Panelists discussed ongoing work in the RISE registry to make steroid dose and duration interpretation possible. Discussion of measure thresholds for glucocorticoid dose and timing included evidence of harm over 7.5 mg daily and usual induction periods for severe manifestations of SLE, such as LN, of approximately 3–6 months.

In the laboratory monitoring domain, the panel noted a strong consensus regarding the evidence to recommend creatinine kidney function and urinary protein monitoring due to the morbidity associated with LN and the need for its prompt treatment. The panel concluded that either a quantitative or qualitative urine protein measurement could fulfill this measure but preferred quantification. Overall, a quality measure to screen for or monitor LN among all patients with SLE was considered of broader public health impact than more frequent end-organ kidney monitoring limited only to patients with established LN.

In the disease phenotyping domain, discussion included the major challenge of identifying incident SLE and LN in EHR data as well as gaps in historic data. Baseline serologic testing and screening for LN or flare reached consensus for importance and feasibility but had lower average rankings than the 3 measures that achieved consensus (Table 2).

The Delphi panel discussions of remaining candidate SLE quality measures that did not reach consensus for importance and feasibility are reported in Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25143. The final rankings for the top 3 recommended statements for patients with SLE were: 1) hydroxychloroquine use, 2) limiting glucocorticoids (to not exceed 7.5 mg/day for >6 months), and 3) end-organ monitoring of kidney function and urine protein excretion at least every 6 months (Table 3).

Future agenda. The Delphi panel endorsed a future infrastructure agenda to include building capacity to: 1) accurately capture SLE and LN diagnosis dates, 2) identify LN class, such as with new ICD codes for specific International Society of Nephrology/Renal Pathology Society LN

Table 3. Final rank of quality measure statements with high consensus by public health benefit for quality SLE care*

Final rank	IF/THEN statement	Ranked first, %	Recommend inclusion, %	Public health impact
First	IF a patient has SLE, THEN they should have a prescription for hydroxychloroquine on or after the date of the most recent rheumatology visit unless a contraindication or adverse event is documented in the medical record	55.6	100	Lower SLE flare rate Fewer kidney relapses Reduced damage accumulation Pregnancy safety and benefits Improved survival in observational studies Possible prevention of thrombosis and cardiovascular disease (refs. 14,20–33)
Second	IF a patient has SLE, THEN the glucocorticoid dose should not exceed 7.5 mg/day prednisone (or equivalent) for more than 6 months	33.3	100	Long-term glucocorticoid therapy can cause irreversible organ damage Doses ≥7.5 mg/day indicate patient does not meet lupus low disease activity state Prednisone dose <7.5 mg/day is associated with lower risk of cataracts, osteoporotic fractures, and cardiovascular disease versus higher dose (refs. 9,12,14,15,19,35–45)
Third	IF a patient has SLE, THEN measurement of both kidney function and protein excretion (urinalysis and/or quantitative measurement) should be performed at least every 6 months	11.1	100	Spot UPCr correlates with 24-hour protein in most studies in detecting nephritis Proteinuria can indicate lupus nephritis flare and can be used to monitor treatment response Proteinuria and creatinine at 6–12 months predict LN prognosis Low proteinuria at 1 year predicts better long-term kidney outcomes (refs. 45,46)

* Items recommended to be included for systemic lupus erythematosus (SLE) quality measure specification and testing in the Rheumatology Informatics System for Effectiveness registry (n = 9 voters). LN = lupus nephritis; SLE = systemic lupus erythematosus; UPCr = urine protein to creatinine ratio.

classification, and 3) improve interoperability to reliably retrieve laboratory and pathology results from outside the rheumatologist's EHR. These items were deemed important next steps in the feasibility of additional future digital quality measures for SLE and LN. The Delphi panel additionally endorsed a research agenda to include: 1) evidence for hydroxychloroquine dosing or blood levels and correlation with SLE outcomes and toxicity risks, 2) evidence for SLE serologic/biomarker monitoring frequency and correlation with outcomes, and 3) data on feasibility and impact of disease activity or damage monitoring in clinical practice. These items were deemed necessary to advance additional candidate SLE quality measures (Table 4).

DISCUSSION

Using a literature review and modified Delphi process, we developed evidence-based quality measures for the longitudinal care of patients with SLE. These measures are recommended for future testing and potential implementation in EHRs, including the RISE registry. We reached consensus agreement and Delphi-panel endorsement of the top 3 quality measures for

Table 4. Future agenda for systemic lupus erythematosus digital quality measures*

Agenda focus and goal	Agenda item
RISE data infrastructure	
SLE and LN diagnosis dates are retrievable from EHR data	Develop structured data fields across EHRs for SLE disease onset and LN diagnosis
LN class is retrievable from EHR data	Develop structured data fields across EHRs and new ICD codes for specific International Society of Nephrology/Renal Pathology Society LN classification
All laboratory tests are retrievable from EHR data, including tests performed outside the EHR	Incorporate outside laboratory tests into structured data fields across EHRs
Research agenda	
Consensus for the optimal hydroxychloroquine dosing strategy, balancing risks and benefits	Conduct research studies correlating hydroxychloroquine dose (or level) with SLE outcomes and risks
Optimal SLE serologic monitoring frequency is identified and linked with potential outcomes	Conduct research studies correlating SLE serologic monitoring frequency with outcomes
Evidence for feasibility and impact of disease activity or damage monitoring in clinical practice is established	Conduct research studies correlating disease activity or damage monitoring with outcomes and evaluating the feasibility of implementing these measures in clinical practice

* EHR = electronic health record; ICD = International Classification of Diseases; LN = lupus nephritis; RISE = Rheumatology Informatics System for Effectiveness; SLE = systemic lupus erythematosus.

SLE, focused on hydroxychloroquine use, limiting the dose and duration of glucocorticoids, and regular kidney monitoring every 6 months to screen for and monitor LN.

Despite documentation of gaps and disparities in SLE health care quality over the last decade, implementation of a national quality measure program to monitor and improve care has remained elusive (1,48). Multiple factors have posed challenges, including the low prevalence of SLE disease, heterogeneity in disease manifestations and severity, a lack of consensus on outcome measures that are feasible to assess in clinical practice, and lack of a platform that facilitates quality measurement nationally (1). Although our consensus panel agreed that many of these factors remain barriers, the robust platform of the RISE registry has increased the feasibility of advancing measures in several areas with the potential for significant public health impact in the care of patients with SLE. Ensuring appropriate use of hydroxychloroquine in eligible patients, reducing glucocorticoid exposure and associated morbidity, and early detection of LN through appropriate screening and monitoring all have the potential to reduce care gaps in SLE to ultimately improve patient outcomes. Moreover, our preliminary assessment of data in the RISE registry suggests that measurement is potentially feasible on a national scale.

Through this process, we also identified several important constructs for quality SLE care that were not deemed currently feasible for implementation as digital quality measures. Measures based on new-onset SLE or LN were limited by data availability, as the relevant dates of diagnosis and treatment initiation were not recorded in structured EHR fields. Information on the LN class was additionally lacking, as were relevant dates of kidney biopsies. Therefore, while quality constructs related to LN induction treatment and maintenance treatment regimens according to LN class were rated highly important with relatively high-quality evidence, measures based on LN class were rated poorly for feasibility of implementation as digital quality measures. Therefore, this work informs a future agenda of infrastructure changes to EHR data availability and documentation (e.g., specific ICD codes for LN class or hydroxychloroquine retinopathy) that would be needed to facilitate the implementation of digital quality measures pertaining to these important quality constructs.

We also identified areas where further evidence is needed, including hydroxychloroquine dosing or use of hydroxychloroquine blood levels to guide dosing, as well as the impact of serologic monitoring (e.g., anti-dsDNA and complement tests) and the optimal frequency. Finally, we identified an evidence gap regarding the feasibility and impact of monitoring disease activity or damage in clinical practice. Although tracking outcomes should be a long-term goal of a national SLE quality measurement program, lack of consensus on an SLE disease activity measure that is feasible and useful to implement in clinical practice remains a barrier. A separate ACR workgroup is currently working on advancing a quality measure relating to patient-reported outcomes as a first step in tracking standardized outcomes in SLE.

Strengths of this work include the use of a rigorous literature review and modified Delphi process, engaging a multidisciplinary panel of SLE experts representing various practice settings across the US and Canada to develop a set of quality measures for SLE care that are candidates for further development as digital quality measures and national implementation. A limitation of this work is that while these recommended SLE quality measures are based on SLE guidelines, our literature review was limited to articles published in English and did not consider work published outside of North America and Europe, before 2000, or after June 2021. In addition, we acknowledge that development of quality measures is just the first step in developing digital quality measures and implementing these nationally. Prior to implementation, detailed measure specifications and testing will need to be undertaken. Testing will include assessment of measure feasibility (e.g., data availability, data accuracy, data standards, and workflows), measure reliability (e.g., quantification of the proportion of provider performance variation explained by true quality differences), measure validity (e.g., ensuring agreement between data elements and performance scores obtained by automated EHR abstraction and manual abstraction of the same information). Measures that are feasible, reliable, and valid will then be implemented in the RISE registry as part of a comprehensive quality improvement effort in SLE.

In conclusion, we present the first ACR quality measures for SLE, based on a rigorous, modified Delphi process involving an expert panel, informed by systematic literature review and initial feasibility testing. Prioritizing the future public health impact, Delphi experts recommended 3 digital quality measures focused on hydroxychloroquine use, limiting glucocorticoid use, and kidney monitoring. Ultimately, these efforts aim to implement validated digital quality measures within US rheumatology practices to improve SLE outcomes.

ACKNOWLEDGMENTS

The authors thank Janet Waters, MLS, BSN, RN, Medical Librarian Consultant, for guideline search assistance, Dr. Marcy Bolster and the ACR Board for review, and Monica Messina, PhD, for editorial support.

AUTHOR CONTRIBUTIONS

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

Study conception and design. Bartels, Jorge, Feldman, Zell, Bermas, Barber, Duarte-García, Garg, Haseley, Jatwani, Johansson, Limanni, Rodgers, Rovin, Santiago-Casas, Suter, Barnado, Ude.

Acquisition of data. Bartels, Jorge, Feldman, Zell, Bermas, Barber, Duarte-García, Garg, Haseley, Jatwani, Johansson, Limanni, Rovin, Santiago-Casas, Suter, Barnado, Ude, Aguirre, Li, Schmajuk, Yazdany.

Analysis and interpretation of data. Bartels, Jorge, Feldman, Zell, Bermas, Barber, Duarte-García, Garg, Haseley, Jatwani, Johansson, Limanni, Rodgers, Rovin, Santiago-Casas, Suter, Barnado, Ude, Aguirre, Li, Schmajuk, Yazdany.

REFERENCES

1. Arora S, Yazdany J. Use of quality measures to identify disparities in health care for systemic lupus erythematosus [review]. Rheum Dis Clin North Am 2020;46:623–38.
2. Electronic Clinical Quality Improvement (eCQI) Resource Center. dQM. 2022. URL: https://ecqi.healthit.gov/glossary/dqm.
3. Rheumatology Informatics System for Effectiveness (RISE Qualified Clinical Data Registry). About Rise Registry. 2022. URL: https://rheumatology.org/about-rise-registry.
4. Feldman CH, Speyer C, Ashby R, et al. Development of a set of lupus-specific, ambulatory care-sensitive, potentially preventable adverse conditions: a Delphi consensus study [review]. Arthritis Care Res (Hoboken) 2021;73:146–57.
5. De Meyer D, Kottner J, Beele H, et al. Delphi procedure in core outcome set development: rating scale and consensus criteria determined outcome selection. J Clin Epidemiol 2019;111:23–31.
6. Fitch KA, Bernstein SJ, Aguilar MD, et al. The RAND/UCLA appropriateness method user's manual. Santa Monica (CA): RAND Corp.; 2001.
7. Hanly JG, Thompson K, Skedgel C. Identification of patients with systemic lupus erythematosus in administrative healthcare databases. Lupus 2014;23:1377–82.
8. Chibnik L, Massarotti E, Costenbader K. Identification and validation of lupus nephritis cases using administrative data. Lupus 2010;19:741–3.
9. Fanouriakis A, Kostopoulou M, Alunno A, et al. 2019 update of the EULAR recommendations for the management of systemic lupus erythematosus. Ann Rheum Dis 2019;78:736–45.
10. Fanouriakis A, Kostopoulou M, Cheema K, et al. 2019 update of the Joint European League Against Rheumatism and European Renal Association-European Dialysis and Transplant Association (EULAR/ERA-EDTA) recommendations for the management of lupus nephritis. Ann Rheum Dis 2020;79:713–23.
11. Xibillé-Friedmann D, Pérez-Rodríguez M, Carrillo-Vázquez S, et al. Clinical practice guidelines for the treatment of systemic lupus erythematosus by the Mexican College of Rheumatology. Reumatol Clin (Engl Ed) 2019;15:3–20.
12. Pons-Estel BA, Bonfa E, Soriano ER, et al. 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;77:1549–57.
13. Hahn BH, McMahon MA, Wilkinson A, et al. American College of Rheumatology guidelines for screening, treatment, and management of lupus nephritis. Arthritis Care Res (Hoboken) 2012;64:797–808.
14. Gordon C, Amissah-Arthur MB, Gayed M, et al. The British Society for Rheumatology guideline for the management of systemic lupus erythematosus in adults. Rheumatology (Oxford) 2018;57:e1–45.
15. Rovin BH, Caster DJ, Cattran DC, et al. Management and treatment of glomerular diseases (part 2): conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) controversies conference. Kidney Int 2019;95:281–95.
16. Van Tellingen A, Voskuyl AE, Vervloet MG, et al. Dutch guidelines for diagnosis and therapy of proliferative lupus nephritis [review]. Neth J Med 2012;70:199–207.
17. Ruiz-Irastorza G, Espinosa G, Frutos MA, et al. Diagnosis and treatment of lupus nephritis: consensus document from the systemic auto-immune disease group (GEAS) of the Spanish Society of Internal Medicine (SEMI) and Spanish Society of Nephrology (S.E.N.). Nefrología 2012;32 Suppl 1:1–35.
18. Keeling SO, Alabdurubalnabi Z, Avina-Zubieta A, et al. Canadian Rheumatology Association recommendations for the assessment and monitoring of systemic lupus erythematosus. J Rheumatol 2018;45:1426–39.
19. Wallace DJ, Gudsoorkar VS, Weisman MH, et al. New insights into mechanisms of therapeutic effects of antimalarial agents in SLE [review]. Nat Rev Rheumatol 2012;8:522–33.
20. Costedoat-Chalumeau N, Amoura Z, Hulot JS, et al. Very low blood hydroxychloroquine concentration as an objective marker of poor adherence to treatment of systemic lupus erythematosus. Ann Rheum Dis 2007;66:821–4.
21. Fessler BJ, Alarcón GS, McGwin G Jr, et al. Systemic lupus erythematosus in three ethnic groups: XVI. Association of hydroxychloroquine use with reduced risk of damage accrual. Arthritis Rheum 2005;52:1473–80.
22. Iudici M, Pantano I, Fasano S, et al. Health status and concomitant prescription of immunosuppressants are risk factors for hydroxychloroquine non-adherence in systemic lupus patients with prolonged inactive disease. Lupus 2018;27:265–72.
23. Jorge A, Ung C, Young LH, et al. Hydroxychloroquine retinopathy: implications of research advances for rheumatology care [review]. Nat Rev Rheumatol 2018;14:693–703.
24. Jung H, Bobba R, Su J, et al. The protective effect of antimalarial drugs on thrombovascular events in systemic lupus erythematosus. Arthritis Rheum 2010;62:863–8.
25. Mok CC, Penn HJ, Chan KL, et al. Hydroxychloroquine serum concentrations and flares of systemic lupus erythematosus: a longitudinal cohort analysis. Arthritis Care Res (Hoboken) 2016;68:1295–1302.
26. Molad Y, Gorshtein A, Wysenbeek AJ, et al. Protective effect of hydroxychloroquine in systemic lupus erythematosus: prospective long-term study of an Israeli cohort. Lupus 2002;11:356–61.
27. Pons-Estel GJ, Alarcón GS, Hachuel L, et al. Anti-malarials exert a protective effect while mestizo patients are at increased risk of developing SLE renal disease: data from a Latin-American cohort. Rheumatology (Oxford) 2012;51:1293–8.
28. Ruiz-Irastorza G, Egurbide MV, Pijoan JI, et al. Effect of antimalarials on thrombosis and survival in patients with systemic lupus erythematosus. Lupus 2006;15:577–83.
29. Ruiz-Irastorza G, Ramos-Casals M, Brito-Zeron P, et al. Clinical efficacy and side effects of antimalarials in systemic lupus erythematosus: a systematic review. Ann Rheum Dis 2010;69:20–8.
30. Wallace DJ. Does hydroxychloroquine sulfate prevent clot formation in systemic lupus erythematosus? Arthritis Rheum 1987;30:1435–6.
31. Jiménez Palop M. Antimalarials: an update in rheumatic diseases. Reumatol Clin 2006;2:190–201. In Spanish.
32. Canadian Hydroxychloroquine Study Group. A randomized study of the effect of withdrawing hydroxychloroquine sulfate in systemic lupus erythematosus. N Engl J Med 1991;324:150–4.
33. Melles RB, Marmor MF. The risk of toxic retinopathy in patients on long-term hydroxychloroquine therapy. JAMA Ophthalmol 2014;132:1453–60.
34. Al Sawah S, Zhang X, Zhu B, et al. Effect of corticosteroid use by dose on the risk of developing organ damage over time in systemic lupus erythematosus: the Hopkins Lupus Cohort. Lupus Sci Med 2015;2:e000066.
35. Bruce IN, O'Keeffe AG, Farewell V, et al. Factors associated with damage accrual in patients with systemic lupus erythematosus:

results from the Systemic Lupus International Collaborating Clinics (SLICC) inception cohort. Ann Rheum Dis 2015;74:1706–13.

36. Chen HL, Shen LJ, Hsu PN, et al. Cumulative burden of glucocorticoid-related adverse events in patients with systemic lupus erythematosus: findings from a 12-year longitudinal study. J Rheumatol 2018;45:83–9.
37. Lim LS, Pullenayegum E, Lim L, et al. From childhood to adulthood: the trajectory of damage in patients with juvenile-onset systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2017;69:1627–35.
38. Ruiz-Arruza I, Barbosa C, Ugarte A, et al. Comparison of high versus low-medium prednisone doses for the treatment of systemic lupus erythematosus patients with high activity at diagnosis [review]. Autoimmun Rev 2015;14:875–9.
39. Thamer M, Hernan MA, Zhang Y, et al. Prednisone, lupus activity, and permanent organ damage. J Rheumatol 2009;36:560–4.
40. Sutton EJ, Davidson JE, Bruce IN. The Systemic Lupus International Collaborating Clinics (SLICC) damage index: a systematic literature review. Semin Arthritis Rheum 2013;43:352–61.
41. Joo YB, Won S, Choi CB, et al. Lupus nephritis is associated with more corticosteroid-associated organ damage but less corticosteroid non-associated organ damage. Lupus 2017;26:598–605.
42. Hoes JN, Jacobs JW, Boers M, et al. EULAR evidence-based recommendations on the management of systemic glucocorticoid therapy in rheumatic diseases. Ann Rheum Dis 2007;66:1560–7.
43. Mosca M, Tani C, Carli L, et al. Glucocorticoids in systemic lupus erythematosus [review]. Clin Exp Rheumatol 2011;29 Suppl 68:S126–9.
44. Dall'Era M, Cisternas MG, Smilek DE, et al. Predictors of long-term renal outcome in lupus nephritis trials: lessons learned from the Euro-Lupus Nephritis cohort. Arthritis Rheumatol 2015;67:1305–13.
45. Tamirou F, Lauwerys BR, Dall'Era M, et al. A proteinuria cut-off level of 0.7 g/day after 12 months of treatment best predicts long-term renal outcome in lupus nephritis: data from the MAINTAIN Nephritis Trial. Lupus Sci Med 2015;2:e000123.
46. Bartels CM, Jorge A, Feldman C, et al. Development of ACR longitudinal digital quality measures for systemic lupus erythematosus: a literature review and modified Delphi consensus study [abstract]. Arthritis Rheumatol 2022;74 Suppl 9. URL: https://acrabstracts.org/abstract/development-of-acr-longitudinal-digital-quality-measures-for-systemic-lupus-erythematosus-a-literature-review-and-modified-delphi-consensus-study/.
47. Schmajuk G, Li J, Evans M, et al. Quality of care for patients with systemic lupus erythematosus: data from the American College of Rheumatology RISE Registry. Arthritis Care Res (Hoboken) 2022;74:179–86.
48. Golder V, Morand EF, Hoi AY. Quality of care for systemic lupus erythematosus: mind the knowledge gap. J Rheumatol 2017;44:271–8.

Arthritis Care & Research
Vol. 75, No. 11, November 2023, pp 2306–2315
DOI 10.1002/acr.25141

American College of Rheumatology
Empowering Rheumatology Professionals

Barriers and Enablers in the Use of Parenteral Methotrexate in Rheumatoid Arthritis Patients: A Scoping Review

Jiun Ming Tan,[1] Emily Reeve,[2] Lauren Fraser,[1] Susanna M. Proudman,[3] and Michael D. Wiese[1]

Objective. Methotrexate (MTX) is effective in controlling disease activity in rheumatoid arthritis (RA). Parenteral MTX may have benefits over oral MTX, but it is rarely used in practice. To better understand this low usage rate, it is necessary to explore the barriers and enablers of therapy from the perspective of RA patients. The objectives of this scoping review were to describe RA patients' perspectives on the barriers and enablers in the use of parenteral MTX and to identify the research gaps in this field.

Methods. The search was performed in Medline, Embase, Scopus, and Cochrane Library from inception to May 2021. Data synthesis was conducted using the Theoretical Framework of Acceptability. This scoping review included any type of study that explored the use of parenteral MTX by adult RA patients from the patients' perspective, written in English.

Results. Fifteen studies were included; findings related to the constructs "affective attitude," "burden," "intervention coherence," and "self-efficacy" were explored the most, while some were rarely ("opportunity cost" and "perceived effectiveness") or not ("ethicality") reported. RA patients were generally satisfied with MTX injections ("affective attitude"). From the burden construct, the requirement for dexterity for administering MTX by injection was considered a barrier, whereas the lack of significant pain from MTX injection was considered an enabler.

Conclusion. The findings suggested that patients generally preferred parenteral MTX formulations with attributes that facilitate self-administration. However, much of the identified research focused on prefilled pen devices, and significant gaps were identified, such as a lack of qualitative research.

INTRODUCTION

Methotrexate (MTX) remains the first-line disease-modifying antirheumatic drug (DMARD) for rheumatoid arthritis (RA) due to its favorable cost-effectiveness, manageable safety, and high efficacy in controlling RA disease activity (1,2). It is used as monotherapy or in combination with other DMARDs (2,3). To maximize the efficacy of MTX therapy, dosing is driven by a treat-to-target approach (4,5). Treat-to-target is a strategy whereby the dose of MTX is intensified to achieve a therapeutic goal in RA, such as a state of low disease activity or remission (6).

MTX, especially via the oral route, is not without issues regarding its pharmacokinetic and toxicity profile (7). First, oral MTX requires folate transporters in the gut to facilitate absorption, but its bioavailability plateaus at oral doses of ≥15 mg due to saturation of these folate transporters (8). Second, gastrointestinal side effects such as nausea, anorexia, and vomiting are frequent adverse effects of oral MTX, particularly at higher doses (9). Gastrointestinal intolerance is often the reason for the poor adherence to and discontinuation of MTX in RA patients (10). A common strategy to manage gastrointestinal side effects is the use of folic or folinic acid (11,12), and there is limited evidence for splitting the dose of MTX to twice or thrice weekly (13,14) and adding caffeine to the MTX treatment regimen (15). Another option is switching the route of MTX delivery from oral to parenteral (16,17).

Recent studies suggest that parenteral MTX is more effective than oral MTX and has a more acceptable gastrointestinal side-effect profile (16–18). It may also support the treat-to-target approach for which doses >15 mg/week may be required. RA patients who received a treat-to-target approach with MTX maintained remission for longer than patients receiving standard care with MTX, possibly because administration via the parenteral

Dr. Reeve's work was supported by the Australian National Health and Medical Research Council (Investigator grant GNT-1195460).

[1]Jiun Ming Tan, BPharm (Hons), Lauren Fraser, BPharm (Hons), Michael D. Wiese, BPharm, M Clin Pharm, PhD: University of South Australia, Adelaide, South Australia, Australia; [2]Emily Reeve, BPharm (Hons), PhD: Monash University, Parkville, Victoria, Australia, and University of South Australia, Adelaide, Australia; [3]Susanna M. Proudman, MBBS (Hons), FRACP: Royal Adelaide Hospital and University of Adelaide, Adelaide, South Australia, Australia.

Author disclosures graphical abstract are available online at https://onlinelibrary.wiley.com/doi/10.1002/acr.25141.

Address correspondence via email to Michael D. Wiese, BPharm, M Clin Pharm, PhD, at Michael.Wiese@unisa.edu.au.

Submitted for publication October 18, 2022; accepted in revised form April 27, 2023.

SIGNIFICANCE & INNOVATIONS

- Ease of using a self-injection device, limited pain, high perceived efficacy, a sense of independence when self-administering, and general positive perceptions were identified as enablers of parenteral methotrexate (MTX) use.
- Limited dexterity to handle the injection device (leading to difficulties removing the cap), concerns about side effects, loss of autonomy if not able to self-administer, and negative perception were barriers to parenteral MTX use.
- Further qualitative research is required to explore the barriers and enablers in the use of MTX injection in depth, as most of the included studies were quantitative in nature and focused on specific brands of MTX prefilled pens.

route ensures predictable bioavailability of MTX (7,19,20). Fewer gastrointestinal adverse effects could also improve the patient's adherence to MTX therapy. Optimizing effectiveness and tolerability of MTX could also delay the addition of other DMARDs, including biologic DMARDs (bDMARDs), thus improving cost-effectiveness (21–23).

Despite these advantages, parenteral MTX is underutilized in practice. Several studies have shown that MTX therapy in RA patients has been less intensive and of a shorter duration since the introduction of bDMARDs (21). Less than 30% of RA patients were switched to parenteral MTX after failure of oral MTX in the US, but the reason for this remains unknown (23,24–28).

Despite the widespread use of MTX, little is known about patients' attitudes towards MTX therapy for RA, particularly the barriers and enablers toward use in a parenteral form. Prior qualitative research has focused on the patients' perception toward injectable bDMARDs (29–32). Storage issues, needle problems, and difficulties in administration have been identified as barriers to the use of injectable bDMARDs in a recent qualitative study (30). Meanwhile, intravenous route of administration, better understanding of taking medications, and less frequent dosing regimens were facilitators (30). Similar barriers and enablers were identified in an integrative review on injectable medications for chronic diseases in general (33). Fear of injection and pain related to injection have been reported as major barriers to the use of parenteral MTX therapy for children with juvenile idiopathic arthritis (34,35). Whether these findings can be translated to the use of parenteral MTX in adult RA populations has not been determined. Therefore, the aim of this scoping review was to describe the reported barriers and enablers in the use of parenteral MTX by adult RA patients and to identify knowledge gaps to inform future research to support the use of parenteral MTX in practice.

MATERIALS AND METHODS

No current or ongoing systematic or scoping reviews examining the factors that facilitate or inhibit the use of parenteral MTX in adult RA patients were found. A preliminary search was conducted in Medline, the Cochrane Database of Systematic Reviews, and JBI Evidence Synthesis.

Study design and protocol. A scoping review was chosen for this study as we wanted to explore findings from studies with heterogeneous methods and disciplines, present a descriptive overview, and identify research gaps (36). The protocol was developed using the Joanna Briggs Institute Reviewers Manual (37), and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) extension for scoping reviews was used as the reporting guideline (38). The protocol was preregistered with Open Science Framework (https://archive.org/details/osf-registrations-upgd3-v1).

Information sources and search strategy. A preliminary search of relevant articles identified 4 dimensions for the search strategy: 'methotrexate'; 'parenteral'; 'rheumatoid arthritis'; and 'patient experience, perceptions, barriers and enablers' (the complete search strategy can be found in Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25141). The databases Embase, Medline, Scopus, and Cochrane were searched from their inception until May 31, 2021. Furthermore, reference lists of relevant articles were manually searched to identify additional literature for potential inclusion.

Inclusion criteria. Studies were included if they fulfilled the following criteria: 1) inclusion of at least 50% of participants with a diagnosis of RA who were ≥18 years of age; 2) exploration of satisfaction, experience, and/or attitude from the patient perspective about parenteral MTX; 3) any geographic and health care setting; and 4) original research studies with any methodology (qualitative, quantitative, or mixed), or systematic or scoping reviews of original research studies.

Exclusion criteria. Studies were excluded if they fulfilled the following criteria: 1) abstracts only, conference proceedings, narrative reviews, editorials, commentaries, letters, notes or retracted studies; 2) unable to access full text; 3) not published in English.

Study selection process. A 2-stage screening process using Covidence review article management software (Veritas Health Innovation) was followed to assess the relevance of studies. For the first stage of screening, titles and abstracts were independently screened by 2 reviewers (JMT and LF). This process was repeated in the second stage, where the full text of the

articles was assessed against the inclusion/exclusion criteria. Conflicts or uncertainties were resolved through discussion with 2 other team members (ER and MDW).

Extraction of results. A draft charting table was developed to record the details of the study and the reported findings. The draft charting table was pilot tested on the first 10 articles and revised as needed. The data extraction was performed independently by 2 reviewers (JMT and LF) and reviewed by 2 other team members (ER and MDW) if there was inconsistency. Methodologic quality was not appraised because the objective of this scoping review was to map the existing evidence and identify gaps in the literature to inform future studies.

Synthesis of results. The analytical stage was performed as described by Levac et al (39). Besides providing a descriptive summary, the qualitative analytical technique was used to apply meaning to the extracted findings. The quantitative data from the included studies were extracted and categorized into the 7 constructs of the Theoretical Framework of Acceptability (TFA). The TFA is a newly developed framework that conceptualizes the acceptability of health care interventions based on 7 constructs: affective attitude, burden, perceived effectiveness, ethicality, intervention coherence, opportunity costs, and self-efficacy (40). Qualitative data analysis was conducted using both the directed content analysis (deductive analysis) to the TFA and conventional content analysis (inductive analysis) as outlined by Hsieh and Shannon (41) where findings did not align with any of the TFA constructs. First, 2 researchers (JMT and LF) read the findings of the selected studies to capture the key concepts. After familiarizing themselves with the definition of all 7 constructs in TFA, the researchers began the coding process using qualitative data analysis software NVivo 12 (QSR International). For any meaning units (i.e., relevant pieces of text) that did not fit into the TFA framework, a new code was created and applied to those relevant meaning units. Throughout the coding process, the coding guideline was continuously reviewed to incorporate new codes as necessary, and the existing codes were refined.

RESULTS

Literature search. Database searches identified 3,841 results following the removal of duplicates. After title and abstract screening, 311 studies progressed to full-text review. Of the 311 studies, 299 were excluded, and their reasons for exclusion were reported (Figure 1). The remaining 12 studies were included, and a further 3 studies were included after hand-searching reference lists, which resulted in a total of 15 studies (the characteristics of the 15 studies are presented in Supplementary Table 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25141).

Study characteristics. Eleven quantitative studies, 3 qualitative studies, and 1 mixed quantitative and qualitative study were included in this scoping review. The included studies had a variety of purposes and methods used. The majority (8 studies) were focused on usability and patient preference and used quantitative methods (42–49), 1 was a quantitative survey study (50), 1 included a quantitative and qualitative survey (mixed-methods study) that investigated the experience of using subcutaneous MTX (51), 1 was a quantitative survey study regarding training on how to use MTX injection (52), 1 was a quantitative MTX adherence study (53), 2 were qualitative focus group studies discussing the barriers and enablers in the use of DMARDs in general (with specific findings presented about MTX) (54,55), and 1 was a qualitative in-depth interview study about medication use in early RA (56).

The included studies were conducted in 7 different countries (UK, US, France, Germany, Canada, Ireland, and Spain). The mean age of participants was between 40 and 60 years. Female participants outnumbered male participants by ≥25% in each of the included studies, except the study by Saraux et al (49), in which the number of male patients exceeded the number of female patients.

All 15 studies involved patients using the subcutaneous route to administer MTX. In terms of the experience of using injectable medication, only 2 studies (43,47) involved participants who had no prior experience of using parenteral medications in general, whereas the remaining studies included current or past users of parenteral MTX. The doses of MTX used in the included studies ranged from 7.5 to 25 mg once a week. Six studies involved participants using the prefilled pen device only, whereas 2 studies focused on prefilled syringes only. Two studies specifically compared the prefilled pen to a prefilled syringe, and the remaining 5 studies were not focused on a specific device. In all the studies, injections were mostly performed by the participants themselves, but 3 studies reported that some participants were assisted by health care professionals or carers/family members with MTX administration.

Review findings (the TFA). Results (barriers/enablers) of the included studies were coded to 6 of the 7 constructs of the TFA (Table 1; the full charting table can be found in Supplementary Table 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25141). No barriers and enablers that related to the TFA construct of ethicality were identified. Fourteen studies reported the participants' acceptability with using parenteral MTX during the study or in the past. Only 1 study reported the patient's expectation of parenteral MTX. No additional codes needed to be created.

Affective attitude. This construct includes participant feelings toward MTX injections overall and was often captured as satisfaction or through an expression of preferences. Within this construct, all 11 quantitative studies reported that a majority of

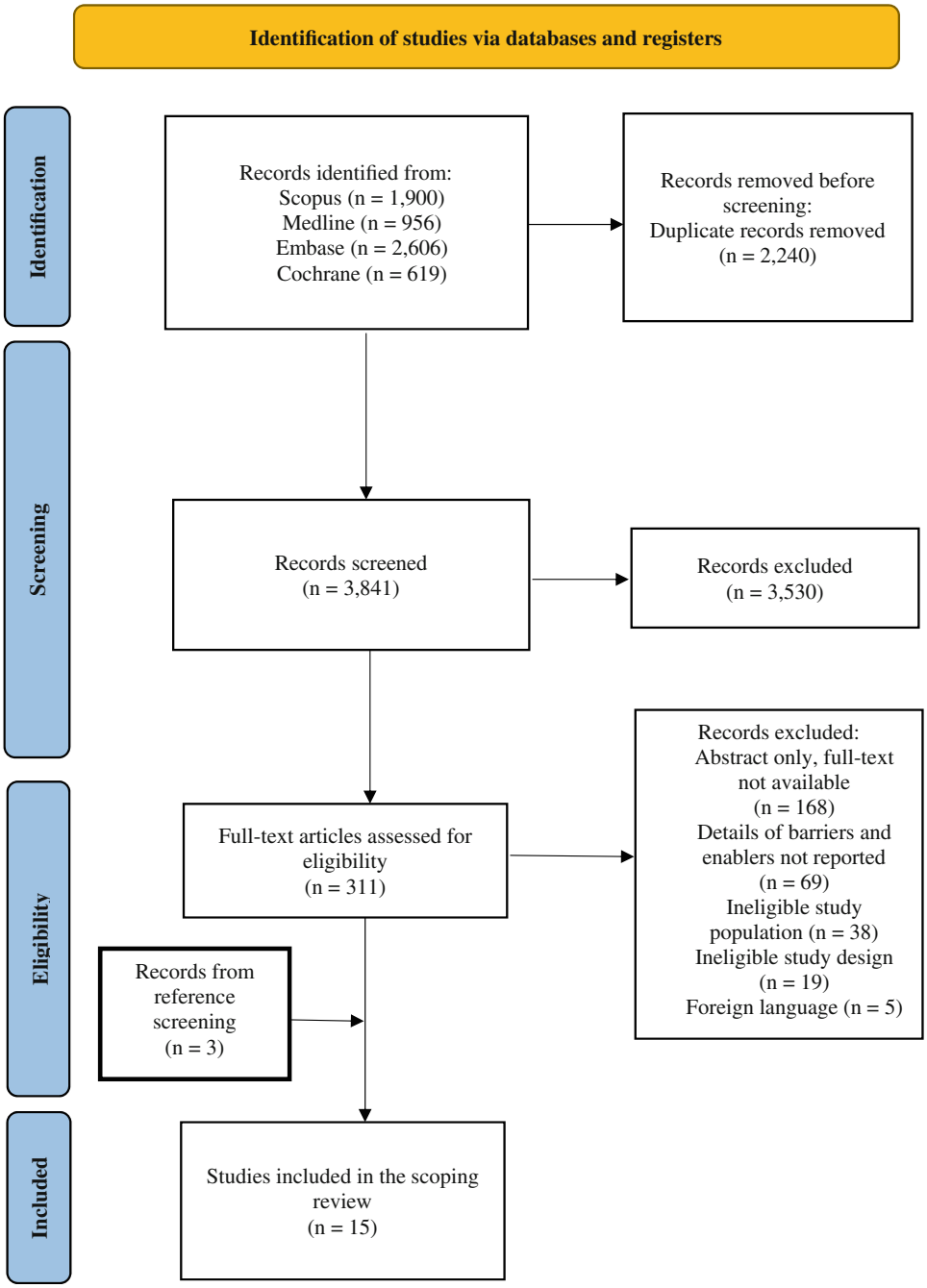

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) extension for scoping reviews flow diagram.

participants had positive responses to questions representing affective attitude (42–50,52,53). Even in the 2 studies in which participants had no prior experience with parenteral MTX treatment, participants reported high satisfaction upon using MTX injections for the first time (43,47). One study reported that up to 98% of participants had a positive overall impression of a prefilled pen, with more participants preferring the prefilled pen to oral MTX (48). Seven studies specifically assessed the usability of and preference for use of MTX prefilled pens, and in all 7 studies, the MTX pen had a higher satisfaction rating than MTX prefilled syringes (43–49).

In contrast to the positive findings of the quantitative studies, 2 qualitative studies reported negative feelings toward MTX injections, which could be barriers, while only 1 reported positive feelings. Negative feelings included the emotions of fear, hatred, and reluctancy to initiate and adhere to parenteral MTX therapy (54,56). One participant reported that injectable medications were inconvenient and intolerable, and there was a negative reaction to

Table 1. Coding frequency according to the Theoretical Framework of Acceptability (TFA) (40) with exemplary quotes

TFA construct	Definition	Coding frequency (no. of articles)	Exemplary quote from quantitative studies	Exemplary quote from qualitative studies
Affective attitude	How an individual feels about the intervention	33 (n = 13)	Overall patient preference for the methotrexate prefilled pen was 75% ($P < 0.0001$) (43).	...particularly the use of needles with the sentiment "...I hate needles..." (54)
Intervention coherence	The extent to which the participant understands the intervention and how it works	16 (n = 7)	The majority of patients (n = 18) (86%) agreed that they understood why the new type of syringe had been introduced (51).	–
Perceived effectiveness	The extent to which the intervention is perceived to be likely to achieve its purpose	3 (n = 2)	The most common reasons for a change to methotrexate self-administration were...improved usability (25.3%) and dislike of methotrexate tablets (13.6%) (57).	"...Well, methotrexate was also used to treat cancer so it's a very, very strong drug" (56)
Self-efficacy	The participant's confidence that they can perform the behavior required to participate in the intervention	25 (n = 10)	Patients reported that self-administration led to a feeling of more independence (89.1%) (57).	Another theme was patients' feelings that they could get on with their life: "I have found it a great help as I can get on with my life a bit easier" (51)
Burden	The perceived amount of effort that is required to participate in the intervention	47 (n = 13)	In total, 67% of the patients confirmed that it did not take much effort to overcome subcutaneous self-injection with the pen (43).	"...when my hands are stiff, I have problems pulling the stopper out" (51)
Opportunity costs	The extent to which benefits, profits, or values must be given up to engage in the intervention	4 (n = 4)	Reasons that patients missed ≥1 methotrexate doses in the last 4 weeks: 1 patient forgot; 2 had side effects; 1 had pharmacy shortages; 1 was doing well; 1 couldn't afford; and 2 had other reasons (50).	"...if someone is not at home now, I'm unable to get the tops off to inject myself...able to open stopper on the old ones on my own – now I have to have help" (51)
Ethicality	The extent to which the intervention has good fit with an individual's value system	0 (n = 0)	–	–

the knowledge that MTX is also used to treat cancer: "It was… injectable…it's a little bit more of a hassle to take…Methotrexate was also used to treat cancer so it's a very, very strong drug…" (56). However in another study, one participant claimed that MTX injection improved their quality of life: "…I have found it a great help as I can get on with my life a bit easier…" (51).

Intervention coherence. The intervention coherence construct represents participants' understanding of the practical aspects of MTX injections. Seven articles were coded to this construct, and all 7 were quantitative studies (44–46,48,49,51,52). The documented practical aspects that could act as barriers or enablers included the functionality of the MTX injector devices, the comprehension of self-injection training, and the handling of MTX injectors. More than 80% of participants demonstrated clarity in understanding the training, performing the self-injection, and safely disposing the MTX injectors (45,46,48,49,51,52).

Perceived effectiveness. This construct was considered as the perceived potential benefits of oral versus parenteral MTX where greater perceived effectiveness was considered an enabler. Only 2 articles were coded to this construct (56,57). One quantitative study reported that 13.6% of participants preferred parenteral MTX because of a disagreement with MTX tablets (no further explanation was provided) (57). One qualitative study reported that a participant expressed concern about the adverse effects of parenteral MTX, and the thought of risking unpleasant adverse effects outweighed the potential benefits that the MTX injections could bring (56).

Self-efficacy. The confidence and ability to use or adhere to parenteral MTX therapy were included in this construct. Ten articles were coded to this construct, 7 quantitative (42,43,46,48,49,52,53), 2 qualitative (54,55), and 1 mixed methods (51). In one study, there was an increase in adherence from 42.0% to 50.7% after switching from oral to subcutaneous MTX (53), and in another study, 89.1% of participants claimed that using MTX injection without assistance instilled a sense of independence (57). A third study reported that 94% of participants felt "empowered" or "very empowered" by being able to self-administer MTX injections (51). The higher confidence stemmed from the support and training provided by health care professionals. For example, self-injection performance improved from 80% to 100% after training was provided to participants (46). In terms of the training method, one study reported that

91% of participants agreed that they could confidently perform self-injection correctly without any assistance after viewing the self-training video (52). On the other hand, the qualitative findings from one study revealed that participants appreciated help from a nurse for dealing with MTX when required: "[I am] grateful that I can always speak to a nurse over the MTX if necessary…" (51).

Qualitative studies also reported strategies to enhance self-efficacy and hindrances that lower self-efficacy. Enablers involved cultivating a routine of using MTX injection by carefully scheduling the time of self-administration to facilitate the development of confidence and adherence (54). By spacing the self-injection of MTX away from daily tasks and activities, participants from one study claimed that they could reduce the impact of MTX injection, especially the general side effects on their daily life (54). Difficulty in removing the MTX injector packaging was identified as a hindrance to adherence and autonomy (55).

Burden. This construct concerns factors that the participant believes will impact their ability to use or adhere to treatment with MTX injections. Thirteen articles were coded to this construct (42–50,52,54–56), 4 of which used qualitative data (51,54–56). The pain associated with MTX injections, which could be a barrier, was mild to almost negligible in one study, as evidenced by a mean pain score of 3.6 on a 100-mm visual analog scale (0 = no pain and 100 mm = the worst pain imaginable) (45). Another study found that the pain score reported by participants was even lower than the pain caused by the injection of bDMARDs (50). Likewise, MTX was well tolerated postinjection, with redness, erythema, and itchiness reported as the most common self-resolving local reactions. No major side effects were recorded in any of the studies.

Difficulty with removing the cap, stopper, or packaging was identified as the most commonly reported barrier (46,48,51,55). Other injection-related issues were that the injection device was nonrefillable (48), potentially leaked (51), involved handling difficulties (51), and was associated with injection site reactions (45,47,54). Meanwhile, the favorable attributes of the MTX injectors included favorable ergonomics (44,52), decreased pain (48,52), high concentration with reduced volume (47,57), preattached needle (47,57), ease of storage (51), audible signal while injecting (44), button-free activation system (44,52), and the ability to self-administer (43,51,52,57). The anticipated burden identified by one qualitative study was related to the negative feelings or beliefs toward injectable medications, resulting in refusal to administer MTX injection (56).

Opportunity costs. This construct represents the trade-off process, in which an individual may have to lose one opportunity (or benefit) in order to gain another (or avoid something negative). Four articles were coded to this construct (50,51,54,55), and 3 of them were qualitative studies (50,54,55). In the quantitative study, 4% of patients agreed that adverse effects were the main reason for skipping the use of subcutaneous MTX (50). On the other hand, one of the qualitative studies described that participants had to allocate time in order to prepare themselves for the adverse effect from using MTX injections (54). The inability to self-administer an MTX injection was associated with a loss of autonomy and independence in some participants (51,55).

DISCUSSION

The aim of this scoping review was to explore the barriers and enablers in the use of parenteral MTX in patients with RA. Our systematic search identified 15 eligible studies, including 11 quantitative studies, 3 qualitative studies, and 1 mixed method (quantitative and qualitative) study. We were able to apply the TFA to understand the barriers and enablers in parenteral MTX use, and results from the identified studies were coded to 6 of the 7 TFA constructs. All the findings from included studies could be fitted to the constructs of the TFA, hence no new themes or subthemes were generated during the coding process. Most findings were primarily coded to 3 constructs (affective attitude, burden, and intervention coherence), while others were identified less frequently (opportunity costs and perceived effectiveness) or not at all (ethicality).

Through the TFA constructs, we identified a number of enablers in the use of parenteral MTX from the perspective of patients with RA. The participants from the included quantitative studies had positive experiences with MTX injections according to questions asked about satisfaction, albeit participants were more inclined to prefer a MTX prefilled pen compared to a prefilled syringe (affective attitude) (43–46,48,49,52). From the construct of burden, factors identified as enablers to acceptability were the use of prefilled pens with a preattached needle, high concentration/low volume of MTX solution, a cap/stopper that is easier to remove, and ergonomic injectors. The self-efficacy and intervention coherence constructs were closely associated with each other. From the construct of self-efficacy, it seemed that participants were generally inclined to learn/be taught to self-administer MTX. Likewise, the construct of intervention coherence suggested that participants had no difficulties being trained by health care professionals. Therefore, being trained to self-administer would act as an enabler. The studied populations commonly felt that the instruction provided by health care professionals was clear and understandable.

In our review, the most commonly reported barrier to MTX use was the difficulty using the injection device, especially removing the cap/stopper, which was identified in the construct of burden. Indeed, RA patients may have functional limitations in dexterity and reduced grip strength, which could be a significant barrier in using injectable medications (58). Increasing accessibility to the easy-to-use injection devices could potentially address this barrier and fear of needles (59). Besides functional limitations, participants could have preconceived negative feelings toward injectable MTX (affective attitude), which could increase avoidance behaviors towards MTX injection (56). It is important for health care professionals to counter any myth or misconception that MTX injection is only used for treatment of cancer (60) while

discussing with their patients that there is good rationale for using parenteral MTX, such as easing the burden on patients who have swallowing difficulties (61).

Our scoping review identified substantial gaps in knowledge about the barriers and enablers in the use MTX injection. The included 8 of 11 quantitative studies mostly aimed to investigate the device used to inject the MTX rather than the concept of MTX injection in general (42–49). Little attention was focused on the specific barriers and enablers in MTX injections outside of expressed preferences/satisfaction. Issues related to cost, cytotoxic and sharps handling, the involvement of health care professionals in the ongoing supply, administration, counselling, and support were rarely or not reported at all. Findings related to constructs of perceived effectiveness, opportunity costs, and ethicality also remained rarely discussed and reported. The scarcity of such findings indicates that it is not possible to establish the breadth of possible barriers and enablers with certainty (e.g., whether issues related to ethicality are potential barriers/enablers or not). Questions that require further exploration include how patients think the use of MTX via the parenteral route will affect their quality of life, what patients are willing to abnegate in order to engage in using MTX injection, and the attribute(s) of MTX injection that patients most value. Furthermore, one of the main gaps identified in this review is the lack of studies that explore the broad patient experience in using MTX injection, as most of the identified studies focused on particular products (e.g., studies that explored whether patients preferred a proprietary pen device or prefilled syringes). Additional research may be required to fully explore the barriers and enablers in parenteral use regardless of the specific device used.

In terms of the methodology, only 4 studies were identified that provided qualitative data in this scoping review. This is not ideal for addressing our research question, as qualitative methodology is more appropriate to produce a "thick description," which facilitates the understanding of certain behaviors and identifying the barriers and enablers (62). It will be interesting to see whether qualitative studies regarding MTX injection would yield any results that are relevant for the uncoded ethicality construct. Additionally, all 3 qualitative studies focused on injectable DMARDs in general (i.e., including subcutaneously and intravenously administered bDMARDs) and included participants with other inflammatory diseases, although we only extracted and included results where they were presented as specific to parenteral MTX when used for RA.

Tornero Molina et al (63) recently reviewed the available studies on the use of MTX prefilled pens and how they affected patients in terms of perceptions and treatment adherence. Our findings were consistent with those identified by Tornero Molina et al, in which the majority of the participants generally preferred prefilled pen devices over traditional injectors. Our review was of broader scope and included qualitative studies that revealed contrasting results to the quantitative data. For example, in the construct of affective attitude, qualitative data highlighted negative feelings toward MTX injections where the quantitative findings had been overwhelmingly positive. Again, this warrants further qualitative research to investigate the validity of the positive response identified in the quantitative studies.

Although pediatric populations were not the focus of this scoping review, the use of MTX injections in children with juvenile idiopathic arthritis has been investigated. Similar to our findings in adult populations (48,52), pediatric populations generally prefer prefilled MTX pens over prefilled syringes, citing less pain as the primary reason for such a preference (64). Autonomy in taking medicine was valued by young adults with inflammatory arthritis as highlighted by Hart et al (65), which was similar to our findings in older patients. This was clearly demonstrated in the construct of burden (participants appreciated being able to self-administer MTX injections) (43,51,52,57). In fact, having a user-friendly MTX injector that omitted the tedious process of preparing an MTX injection could potentially enhance autonomy in both children and adults, thus empowering patients to confidently manage their DMARD regimen. Anticipatory nausea associated with the use of parenteral MTX has been reported in pediatric populations and for injections more broadly. This was not identified in this review, which may be due to the generally limited findings rather than this not being an issue in the adult RA population.

During the rollout of the COVID-19 vaccination program, strategies and health measures have been established to reduce patients' fear and anxiety with injections (66,67). People who have past experience with the flu injection were more likely to receive the COVID-19 vaccine (68–70), and it is possible that patients' perceptions about needles/injections might have been changed by the education associated with the implementation of the COVID-19 vaccination program. More research is warranted to investigate the patients' acceptability towards MTX injection in the post-COVID-19 era.

This scoping review used a wide range of terms to search the literature and extract relevant data. The inclusion of quantitative and qualitative studies enabled us to provide an overview of the breadth of evidence, which is one of the aims of conducting a scoping review (71). To our knowledge, this is the first study that used the TFA in evaluating the acceptability of a medication delivered specifically via the parenteral route.

However, our review has some limitations. There was incomplete reporting of baseline clinical and demographic information such as RA disease duration, disease activity, and the type and number of DMARDs used. This limited our ability to consider how the characteristics of the populations influenced and interacted with our findings, for example, we could not determine the influence of use of other DMARDs (combination therapy) on the patients' attitudes toward parenteral MTX. Another limitation is that studies that were only available as abstracts or were published in languages other than English were excluded, and gray literature was not searched. It is possible that more enablers and barriers to the use of MTX injections could have been identified from these studies.

This review focused on the barriers and enablers in the use of parenteral MTX, however, there may be other strategies to optimize the effectiveness of MTX (and other DMARDs) and reduce side effects. For example, the feasibility of different parenteral MTX dosing regimens could be compared to split-dose oral MTX (13). Further work is required to explore the place of parenteral MTX among other treatment options in achieving the goals of RA therapy in practice.

Our findings provided an overview of barriers and enablers in the use of parenteral MTX from the perspective of RA patients. The 7 constructs of the TFA guided the data synthesis, which offered insights into the acceptability of parenteral MTX. Our findings will enable future researchers to focus on addressing the research gaps that presently exist.

Future research should employ qualitative study design to explore barriers and enablers in depth and to allow for positive and negative attitudes to be expressed. Additionally, we did not identify any studies that reported differences in geographic or health care settings as a barrier or enabler, as the included studies were conducted in different countries/settings and were too different to make any comparisons between studies. It may also be prudent for future research to focus on populations currently taking oral MTX that are most likely to benefit from parenteral MTX, such as those who are nonadherent, suffer significant gastrointestinal side effects, or have insufficient response to treatment. Due to the number of gaps/limitations of studies, we are not able to make specific recommendations for providers that aim to overcome barriers and leverage enablers such that uptake of parenteral MTX in patients with RA in clinical practice is enhanced, hence further research is needed. This research may help with the development of tailored interventions to enhance the uptake of parenteral MTX in patients with RA and therefore optimize benefits and minimize treatment harm in this population.

This scoping review offered insight into the barriers and enablers in the use of parenteral MTX in RA patients from the perspective of RA patients. It described the patients' experience and satisfaction with MTX injections, primarily from usability and preference studies. The findings suggested that patients generally preferred parenteral MTX formulations with attributes that facilitate the process of self-injection. In terms of the usability of MTX injectors, patients consistently demonstrated their ability to perform self-injection after instruction, although there were some barriers that hindered their confidence in using MTX injections. However, significant gaps exist, and more qualitative studies that focus on parenteral MTX users are warranted in order to explore patients' feelings and experiences with MTX injections.

ACKNOWLEDGMENTS

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

AUTHOR CONTRIBUTIONS

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

Study conception and design. Tan, Reeve, Proudman, Wiese.
Acquisition of data. Tan, Fraser.
Analysis and interpretation of data. Tan, Reeve, Wiese.

REFERENCES

1. Smolen JS, Landewé RB, Bijlsma JW, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. Ann Rheum Dis 2020;79:685–99.
2. Mysler E, Caubet M, Lizarraga A. Current and emerging DMARDs for the treatment of rheumatoid arthritis [review]. Open Access Rheumatol 2021;13:139–52.
3. Friedman B, Cronstein B. Methotrexate mechanism in treatment of rheumatoid arthritis [review]. Joint Bone Spine 2019;86:301–7.
4. Hughes CD, Scott DL, Ibrahim F, et al. Intensive therapy and remissions in rheumatoid arthritis: a systematic review. BMC Musculoskelet Disord 2018;19:389.
5. Schoels M, Knevel R, Aletaha D, et al. Evidence for treating rheumatoid arthritis to target: results of a systematic literature search [review]. Ann Rheum Dis 2010;69:638–43.
6. Solomon DH, Bitton A, Katz JN, et al. Review: treat to target in rheumatoid arthritis: fact, fiction, or hypothesis? [review] Arthritis Rheumatol 2014;66:775–82.
7. Taylor PC, Balsa Criado A, Mongey AB, et al. How to get the most from methotrexate (MTX) treatment for your rheumatoid arthritis patient? MTX in the treat-to-target strategy. J Clin Med 2019;8:515.
8. Goodman SM, Cronstein BN, Bykerk VP. Outcomes related to methotrexate dose and route of administration in patients with rheumatoid arthritis: a systematic literature review [review]. Clin Exp Rheumatol 2015;33:272–8.
9. Wang W, Zhou H, Liu L. Side effects of methotrexate therapy for rheumatoid arthritis: a systematic review [review]. Eur J Med Chem 2018;158:502–16.
10. Hope HF, Hyrich KL, Anderson J, et al. The predictors of and reasons for non-adherence in an observational cohort of patients with rheumatoid arthritis commencing methotrexate. Rheumatology (Oxford) 2019;59:213–23.
11. Shea B, Swinden MV, Tanjong Ghogomu E, et al. Folic acid and folinic acid for reducing side effects in patients receiving methotrexate for rheumatoid arthritis [review]. Cochrane Database Syst Rev 2013;2013:Cd000951.
12. Van Ede AE, Laan RF, Rood MJ, et al. Effect of folic or folinic acid supplementation on the toxicity and efficacy of methotrexate in rheumatoid arthritis: a forty-eight week, multicenter, randomized, double-blind, placebo-controlled study. Arthritis Rheum 2001;44:1515–24.
13. Dhaon P, Das SK, Srivastava R, et al. Oral methotrexate in split dose weekly versus oral or parenteral methotrexate once weekly in rheumatoid arthritis: a short-term study. Int J Rheum Dis 2018;21:1010–7.
14. Mohamedali N, Ravindran V. Splitting dose of methotrexate in the management of rheumatoid arthritis: making a mountain out of a molehill? [letter] Indian J Rheumatol 2017;12:114–5.
15. Malaviya A, Baghel S, Verma S, et al. Use of coffee for alleviating methotrexate intolerance in rheumatic diseases [letter]. Indian J Rheumatol 2019;14:79–80.

16. Bujor AM, Janjua S, LaValley MP, et al. Comparison of oral versus parenteral methotrexate in the treatment of rheumatoid arthritis: a meta-analysis. PloS One 2019;14:e0221823.
17. Li D, Yang Z, Kang P, et al. Subcutaneous administration of methotrexate at high doses makes a better performance in the treatment of rheumatoid arthritis compared with oral administration of methotrexate: a systematic review and meta-analysis [review]. Semin Arthritis Rheum 2016;45:656–62.
18. Bianchi G, Caporali R, Todoerti M, et al. Methotrexate and rheumatoid arthritis: current evidence regarding subcutaneous versus oral routes of administration [review]. Adv Ther 2016;33:369–78.
19. Verstappen SM, Jacobs JW, van der Veen MJ, et al. Intensive treatment with methotrexate in early rheumatoid arthritis: aiming for remission. Computer Assisted Management in Early Rheumatoid Arthritis (CAMERA, an open-label strategy trial). Ann Rheum Dis 2007;66:1443–9.
20. Mueller RB, Spaeth M, von Restorff C, et al. Superiority of a treat-to-target strategy over conventional treatment with fixed csDMARD and corticosteroids: a multi-center randomized controlled trial in RA patients with an inadequate response to conventional synthetic DMARDs, and new therapy with certolizumab pegol. J Clin Med 2019;8:302.
21. Bello AE, Perkins EL, Jay R, et al. Recommendations for optimizing methotrexate treatment for patients with rheumatoid arthritis [review]. Open Access Rheumatol 2017;9:67–79.
22. Fitzpatrick R, Scott DG, Keary I. Cost-minimisation analysis of subcutaneous methotrexate versus biologic therapy for the treatment of patients with rheumatoid arthritis who have had an insufficient response or intolerance to oral methotrexate. Clin Rheumatol 2013;32:1605–12.
23. Rohr MK, Mikuls TR, Cohen SB, et al. Underuse of methotrexate in the treatment of rheumatoid arthritis: a national analysis of prescribing practices in the US. Arthritis Care Res (Hoboken) 2017;69:794–800.
24. Curtis JR, Zhang J, Xie F, et al. Use of oral and subcutaneous methotrexate in rheumatoid arthritis patients in the United States. Arthritis Care Res (Hoboken) 2014;66:1604–11.
25. Mathew S, Sharma TS, Quintero O, et al. How are actual patterns of care for a health system in rheumatology in comparison with validated quality measures? [abstract] 2013 ACR/AHRP Annual Meeting. URL: https://acrabstracts.org/abstract/how-are-actual-patterns-of-care-for-a-health-system-in-rheumatology-in-comparison-with-validated-quality-measures/.
26. O'Dell JR, Cohen SB, Thorne JC, et al. Treatment of rheumatoid arthritis in the USA: premature use of tumor necrosis factor inhibition and underutilization of concomitant methotrexate. Open Access Rheumatol 2018;10:97–101.
27. O'Dell JR, Cohen S, Thorne C, et al. FRI0181 changing use of methotrexate (MTX) and biologics in the treatment of rheumatoid arthritis (RA) in the United States (US): results of a comprehensive pharmaceutical claims analysis [poster]. Ann Rheum Dis 2016;75 Suppl 2:495.
28. Tornero-Molina J, Andreu JL, Martín-Martínez MA, et al. Methotrexate in patients with rheumatoid arthritis in Spain: subanalysis of the AR Excellence Project. Reumatol Clin (Eng Ed) 2019;15:338–42.
29. Singh JA, Tornberg H, Goodman SM. Pop a pill or give myself a shot? Patient perspectives of disease-modifying anti-rheumatic drug choice for rheumatoid arthritis. Joint Bone Spine 2021;88:105053.
30. Voshaar M, Vriezekolk J, van Dulmen S, et al. Barriers and facilitators to disease-modifying antirheumatic drug use in patients with inflammatory rheumatic diseases: a qualitative theory-based study. BMC Musculoskelet Disord 2016;17:442.
31. Sarmiento-Monroy JC, Parada-Arias L, Rodríguez-López M, et al. Subcutaneous abatacept in rheumatoid arthritis: a real-life experience [review]. J Transl Autoimmun 2019;2:100016.
32. Durand C, Eldoma M, Marshall DA, et al. Patient preferences for disease-modifying antirheumatic drug treatment in rheumatoid arthritis: a systematic review. J Rheumatol 2020;47:176–87.
33. Lambrinou E, Kyriakou M, Lakatamitou I, et al. An integrative review on facilitators and barriers in delivering and managing injectable therapies in chronic conditions: a part of the ACNAP project 'injectable medicines among patients with cardiovascular conditions' [review]. Eur J Cardiovasc Nurs 2020;19:663–80.
34. Sørensen K, Skirbekk H, Kvarstein G, et al. Children's fear of needle injections: a qualitative study of training sessions for children with rheumatic diseases before home administration. Pediatr Rheumatol Online J 2020;18:13.
35. Jacobse J, Ten Voorde W, Rissmann R, et al. The effect of repeated methotrexate injections on the quality of life of children with rheumatic diseases [review]. Eur J Pediatr 2019;178:17–20.
36. Pham MT, Rajić A, Greig JD, et al. A scoping review of scoping reviews: advancing the approach and enhancing the consistency [review]. Res Synth Methods 2014;5:371–85.
37. Peters MD, Godfrey CM, Khalil H, et al. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015;13:141–6.
38. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169:467–73.
39. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5:69.
40. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017;17:88.
41. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005;15:1277–88.
42. Striesow F, Brandt A. Preference, satisfaction and usability of subcutaneously administered methotrexate for rheumatoid arthritis or psoriatic arthritis: results of a postmarketing surveillance study with a high-concentration formulation. Ther Adv Musculoskelet Dis 2012;4:3–9.
43. Demary W, Schwenke H, Rockwitz K, et al. Subcutaneously administered methotrexate for rheumatoid arthritis, by prefilled syringes versus prefilled pens: patient preference and comparison of the self-injection experience. Patient Prefer Adherence 2014;8:1061–71.
44. Zeitoun JD, Morvan Y. Patients' perceptions and preferences regarding two different forms of methotrexate autoinjectors for moderate to severe rheumatoid arthritis: a European crossover survey. Patient Prefer Adherence 2020;14:2177–85.
45. Freundlich B, Kivitz A, Jaffe JS. Nearly pain-free self-administration of subcutaneous methotrexate with an autoinjector: results of a phase 2 clinical trial in patients with rheumatoid arthritis who have functional limitations. J Clin Rheumatol 2014;20:256–60.
46. Hudry C, Lebrun A, Moura B, et al. Evaluation of usability and acceptance of a new autoinjector intended for methotrexate subcutaneous self-administration in the management of rheumatoid arthritis. Rheumatol Ther 2017;4:183–94.
47. Müller-Ladner U, Rockwitz K, Brandt-Jürgens J, et al. Tolerability and patient/physician satisfaction with subcutaneously administered methotrexate provided in two formulations of different drug concentrations in patients with rheumatoid arthritis. Open Rheumatol J 2010;4:15–22.
48. Pachon JA, Kivitz AJ, Heuer KU, et al. Assessing usability, label comprehension, pen robustness and pharmacokinetics of a self-administered prefilled autoinjector pen of methotrexate in patients with rheumatoid arthritis. SAGE Open Med 2014;2:2050312114564241.
49. Saraux A, Hudry C, Zinovieva E, et al. Use of auto-injector for methotrexate subcutaneous self-injections: high satisfaction level and good

compliance in SELF-I study, a randomized, open-label, parallel group study. Rheumatol Ther 2019;6:47–60.
50. Curtis JR, Xie F, Mackey D, et al. Patient's experience with subcutaneous and oral methotrexate for the treatment of rheumatoid arthritis. BMC Musculoskelet Disord 2016;17:405.
51. Hiley J, Homer D, Clifford C. Patient self-injection of methotrexate for inflammatory arthritis: a study evaluating the introduction of a new type of syringe and exploring patients' sense of empowerment. Musculoskeletal Care 2008;6:15–30.
52. Homer D. Using video-based training for button-free auto-injection of subcutaneous methotrexate: a pilot study. Musculoskeletal Care 2019;17:247–79.
53. Senbel E, Tropé S, Herman-Demars H et al. Benefits of switch from oral to subcutaneous route on adherence to methotrexate in patients with rheumatoid arthritis in real life setting. Patient Preference Adherence 2021;15:751–60.
54. Rai SK, Howren A, Wilcox ES, et al. Exploring strategies to support medication adherence in patients with inflammatory arthritis: a patient-oriented qualitative study using an interactive focus group activity. Patient Prefer Adherence 2018;12:2015–25.
55. Otón T, Carmona L, Andreu JL. Patient's experiences of the barriers and facilitators to methotrexate. Musculoskeletal Care 2022;20:158–66.
56. Townsend A, Backman CL, Adam P, et al. A qualitative interview study: patient accounts of medication use in early rheumatoid arthritis from symptom onset to early postdiagnosis. BMJ Open 2013;3:e002164.
57. Striesow F, Brandt A. Preference, satisfaction and usability of subcutaneously administered methotrexate for rheumatoid arthritis or psoriatic arthritis: results of a postmarketing surveillance study with a high-concentration formulation. Ther Adv Musculoskelet Dis 2012;4:3–9.
58. Palamar D, Er G, Terlemez R, et al. Disease activity, handgrip strengths, and hand dexterity in patients with rheumatoid arthritis. Clin Rheumatol 2017;36:2201–8.
59. Van den Bemt BJ, Gettings L, Domańska B, et al. A portfolio of biologic self-injection devices in rheumatology: how patient involvement in device design can improve treatment experience [review]. Drug Deliv 2019;26:384–92.
60. Arnold MH, Bleasel J, Haq I. Nocebo effects in practice: methotrexate myths and misconceptions. Med J Aust 2016;205:440–2.
61. Tornero Molina J, Calvo Alen J, Ballina J, et al. Recommendations for the use of parenteral methotrexate in rheumatic diseases. Rheumatol Clin (Engl Ed) 2018;14:142–9.
62. Jick TD. Mixing qualitative and quantitative methods: triangulation in action. Adm Sci Q 1979;24:602–11.
63. Tornero Molina J, López Robledillo JC, Casamira Ruiz N. Potential benefits of the self-administration of subcutaneous methotrexate with autoinjector devices for patients: a review [review]. Drug Healthc Patient Saf 2021;13:81–94.
64. Roszkiewicz J, Swacha Z, Smolewska E. Prefilled pen versus prefilled syringe: a pilot study evaluating two different methods of methotrexate subcutaneous injection in patients with JIA. Pediatr Rheumatol Online J 2020;18:64.
65. Hart RI, Foster HE, McDonagh JE, et al. Young people's decisions about biologic therapies: who influences them and how? Rheumatology (Oxford) 2015;54:1294–301.
66. Australian Government, Department of Health and Aged Care. Information for adults with a needle phobia. 2022. URL: https://www.health.gov.au/resources/publications/information-for-adults-with-a-needle-phobia.
67. Freeman D, Lambe S, Yu LM, et al. Injection fears and COVID-19 vaccine hesitancy. Psychol Med 2023;53:1185–95.
68. Dadras O, SeyedAlinaghi S, Karimi A, et al. Public acceptability of COVID-19 vaccines and its predictors in Middle Eastern/North African (MENA) countries: a systematic review. Hum Vaccin Immunother 2022;18:2043719.
69. Joshi A, Kaur M, Kaur R, et al. Predictors of COVID-19 vaccine acceptance, intention, and hesitancy: a scoping review [review]. Front Public Health 2021;9:698111.
70. Dietze PM, Hall C, Price O, et al. COVID-19 vaccine acceptability among people in Australia who inject drugs: implications for vaccine rollout. Drug Alcohol Rev 2022;41:484–7.
71. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8:19–32.

Arthritis Care & Research
Vol. 75, No. 11, November 2023, pp 2316–2327
DOI 10.1002/acr.25140

American College of Rheumatology
Empowering Rheumatology Professionals

How Does Exercise, With and Without Diet, Improve Pain and Function in Knee Osteoarthritis? A Secondary Analysis of a Randomized Controlled Trial Exploring Potential Mediators of Effects

Belinda J. Lawford,[1] Rana S. Hinman,[1] Fiona McManus,[1] Karen E. Lamb,[1] Thorlene Egerton,[1] Catherine Keating,[2] Courtney Brown,[2] Kathryn Oliver,[1] and Kim L. Bennell[1]

Objective. To explore the mediators of effects of two 6-month telehealth-delivered exercise programs, including exercise with and without weight-loss diet, on pain and function improvements in knee osteoarthritis (OA).

Methods. Secondary analysis of 345 participants from a 3-arm randomized controlled trial of exercise (Exercise program) and exercise plus diet (Diet + Exercise program) versus information (Control program) was conducted. Outcomes were changes in pain (11-point numeric rating scale) and function (Western Ontario and McMaster Universities Osteoarthritis Index [score range 0–68]) at 12 months. Potential mediators were change at 6 months in attitudes toward self-management, fear of movement, arthritis self-efficacy, weight, physical activity, and willingness for knee surgery. For the Diet + Exercise program versus the Exercise program, only change in weight was evaluated.

Results. Possible mediators of the Exercise program versus the Control program included reduced fear of movement (accounting for –1.11 units [95% confidence interval (95% CI) –2.15, –0.07] improvement in function) and increased arthritis self-efficacy (–0.40 units [95% CI –0.75, –0.06] reduction in pain, –1.66 units [95% CI –3.04, –0.28] improvement in function). The Diet + Exercise program versus the Control program mediators included reduced fear of movement (–1.13 units [95% CI –2.17, –0.08] improvement in function), increased arthritis self-efficacy (–0.77 units [95% CI –1.26, –0.28] reduction in pain, –5.15 units [95% CI –7.34, –2.96] improvement in function), and weight loss (–1.20 units [95% CI –1.73, –0.68] reduction in pain, –5.79 units [95% CI –7.96, –3.63] improvement in function). Weight loss mediated the Diet + Exercise program versus the Exercise program (–0.89 units [95% CI –1.31, –0.47] reduction in pain, –4.02 units [95% CI –5.77, –2.26] improvement in function).

Conclusion. Increased arthritis self-efficacy, reduced fear of movement, and weight loss may partially mediate telehealth-delivered exercise program effects, with and without diet, on pain and/or function in knee OA. Weight loss may partially mediate the effect of diet and exercise compared to exercise alone.

INTRODUCTION

All current clinical guidelines recommend education, exercise, and weight loss (if indicated) as first-line management approaches for knee osteoarthritis (OA) (1–4). In those individuals who are overweight or obese, there is evidence that combining a weight-loss diet with exercise is optimal, with benefits of the combination exceeding the effects of either treatment alone (5,6). A recent systematic review and meta-analysis found that diet-induced weight loss alone did not improve pain for people with knee OA who were overweight or obese, but there were moderate effects from interventions that combined diet and exercise

Supported by Medibank, the Medibank Better Health Foundation Research Fund, and the National Health and Medical Research Council Centre of Research Excellence (grant APP-1079078).

Dr. Hinman's work was supported by the National Health and Medical Research Council (Fellowship grant 1154217). Dr. Bennell's work was supported by the National Health and Medical Research Council (Investigator grant 1174431).

[1]Belinda J. Lawford, BMedSci, BHlthSci(Hons), PhD, Rana S. Hinman, BPhysio(Hons), PhD, Fiona McManus, BPhysio(Hons), MPH, MBiostat, Karen E. Lamb, BSc(Hons), PhD, Thorlene Egerton, BPhty, MPhil, PhD, Kathryn Oliver, BSc, DPT, Kim L. Bennell, BAppSci(Physio), PhD: The University of Melbourne, Melbourne, Victoria, Australia; [2]Catherine Keating, BOccThy, MPH, PhD, Courtney Brown, BHlthSc, BBus(Mktg): Medibank Private, Melbourne, Victoria, Australia.

Author disclosures and graphical abstracts are available online at https://onlinelibrary.wiley.com/doi/10.1002/acr.25140.

[Correction added on 16 August 2023, after first online publication: Fiona McManus' affiliation has been corrected in this version.]

Address correspondence via email to Belinda J. Lawford, BMedSci, BHlthSci(Hons), PhD, at belinda.lawford@unimelb.edu.au.

Submitted for publication September 20, 2022; accepted in revised form April 25, 2023.

SIGNIFICANCE & INNOVATIONS

- Our previous three-arm randomized controlled trial found that two 6-month telehealth-delivered programs, including exercise with and without weight-loss diet, led to improvements in pain and physical function in people with knee osteoarthritis compared to information-only control. In this study, we used causal mediation analyses to explore potential mediators of these effects.
- Reduced fear of movement and increased arthritis self-efficacy may mediate the effects of exercise, compared to control, on pain and/or function.
- Reduced fear of movement, increased arthritis self-efficacy, and weight loss may mediate the effects of exercise with weight-loss diet, compared to control, on pain and/or function.
- Weight loss may mediate the effects of exercise with weight-loss diet, compared to exercise alone, on pain and function.

(7). As such, scalable interventions that combine weight-loss diet with exercise may help maximize clinical outcomes and reduce the enormous individual and societal burden of the condition (8).

Despite clinical guidelines recommending diet and exercise for management of knee OA, the clinical benefits of these treatments for this population are only modest (1–4,7). This may be partly due to limited understanding about the mechanisms by which exercise and diet approaches work to improve pain and physical function. Identifying precise mechanisms of effect will help ensure future treatment programs are designed to target these mechanisms, potentially leading to enhanced effects on clinical outcomes such as pain and function. A robust method of identifying mechanisms of effect is through causal mediation analysis using data from a randomized controlled trial (RCT). Causal mediation analyses examine the causal links between an intermediate variable (mediator) and the effect of an intervention on outcomes (9). A recent scoping review of mediation analysis studies examining nonsurgical interventions for people with OA found that reduced inflammation, reduced body weight, increased muscle strength, and increased self-efficacy may mediate effects of nonsurgical interventions on pain and physical function (10). Specifically, increased knee extensor muscle strength (11) and knee flexor muscle perfusion (12) partially mediated the effects of exercise on changes in pain and physical function in adults with knee OA, but no previous studies had examined psychosocial putative mediators of exercise like self-efficacy or fear of movement. In combined diet and exercise programs, there was inconsistent evidence that weight loss (13), inflammatory biomarkers (14), self-efficacy (for walking duration [15] and for OA symptoms [16]), and pain control (perceived ability to exert control over one's pain) (16) mediate the effects of the programs on pain and physical function. Given the paucity and heterogeneity of existing evidence identified by authors of the review (10), more research is required to identify potential mediators of exercise interventions, including those with and without diet, for people with OA.

Recently, our RCT found that two 6-month telehealth-delivered programs, including exercise with and without weight-loss diet, led to improvements in pain and physical function at 6 months and 12 months, compared to a control group who received online information (6,17). Compared to the exercise-only group, the combined diet and exercise program led to modest additional improvements in pain and function. However, from the RCT alone, the mechanisms underpinning the effectiveness of both the diet and exercise and exercise-only programs on symptoms are not clear. Using data from our RCT (6), we used causal mediation analysis to explore potential mediators of the effects of our 2 exercise programs, one with and one without weight-loss diet, on improvements in pain and physical function in people with knee OA who are overweight/obese.

MATERIALS AND METHODS

This is a secondary analysis of data from an RCT comparing the effects of exercise, with and without a weight-loss diet, to an information-only control group in people with knee OA (Australian New Zealand Clinical Trials Registry ACTRN12618000930280) (6,17). All participants provided written informed consent and The University of Melbourne Human Research Ethics Committee approved the study.

Participants. Participants were recruited from members of Australia's largest private health insurer, Medibank Private. Medibank sent targeted invitations, predominantly via email, to members. Eligible participants including the those who: 1) held private health insurance with Medibank at a level that included cover for arthroplasty surgery; 2) met the National Institute for Health and Care Excellence OA clinical criteria (ages ≥45 years, activity-related joint pain, morning stiffness ≤30 minutes) (3); 3) had average knee pain ≥4 on 11-point numerical rating scale (NRS) in the past week (0 = no pain, 10 = worst pain possible); 4) had a history of knee pain on most days for at least 3 months; 5) were ages <81 years; and 6) had a body mass index ≥28 kg/m^2 and <41 kg/m^2. Detailed inclusion and exclusion criteria for the RCT are published (17).

Control group. Participants in the Control group were given access to a bespoke website that contained information about OA, treatment options, exercise and physical activity, weight loss, managing pain, sleep, and "success stories." The website also provided links to external websites for further information.

Interventions. The intervention protocol was previously published (17). All project clinicians underwent training prior to the start of the trial, including in best-practice management of

OA, motivational interviewing skills, specifics of the weight-loss diet, and study-specific protocols (17).

Exercise program. Participants in the Exercise group had 6 individual videoconferencing consultations (Zoom Video Communications Inc.) with a physical therapist over 6 months where they were prescribed a strengthening exercise and physical activity program. Initial consultations lasted ~45 minutes, with follow-up consultations lasting ~20 minutes. Physical therapists also provided individualized advice about treatment options and used motivational interviewing principles to support behavior change. Participants received hard copy information booklets (information about OA, exercise instructions, log book), exercise bands, and a Fitbit (Flex 2 model) to monitor physical activity.

Diet + Exercise program. Participants in the Diet + Exercise group received all components of the Exercise group, plus 6 individual videoconferencing consultations with a dietitian over 6 months to guide them through a ketogenic very low-calorie diet (VLCD). This diet involved consuming ~800 calories (or 3,280 kilojoules) per day (18) and replacing 2 meals per day with Optifast meal replacements (Nestlé Health Science) (or Optislim [Opti-Pharm Pty Ltd] if unavailable or if the participant was vegetarian). Participants were encouraged to lose at least 10% of their body weight on the diet (5) before transitioning off meal replacements to a healthy eating diet. Initial consultations lasted ~45 minutes, with follow-up consultations lasting ~20 minutes. Dietitians used motivational interviewing to help participants adhere to their weight management plan. Participants received additional weight management booklets ("how to" guide, recipe book, weight management activities), a plastic portion plate, and up to a 6-month supply of meal replacements.

Outcomes. Outcomes were self-reported via online questionnaires at baseline, 6 months, and 12 months. Outcomes relevant to this mediation analysis include the primary outcomes of change in knee pain and physical function at 12 months. Overall knee pain was measured using an 11-point NRS ranging from 0 (no pain) to 10 (worst pain possible). Physical function was measured using the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC OA Index) function subscale (19), with scores ranging from 0 (no dysfunction) to 68 (maximum dysfunction).

Mediator variables. Six potential mediator variables were measured at baseline and 6 months:

1. Attitudes towards self-management, measured using the Patient Activation Measure (20) (score range 13–42, with higher scores indicating greater patient activation);
2. Fear of movement, measured using the Brief Fear of Movement Scale for Osteoarthritis (21) (score range 6–24, with higher scores indicating greater fear);
3. Self-efficacy for managing arthritis symptoms, measured using the Arthritis Self-Efficacy Scale (22) (score range 3–30, with higher scores indicating greater self-efficacy);
4. Body weight in kilograms, which was self-reported;
5. Physical activity, measured using the Incidental and Planned Exercise Questionnaire, "past week" version (23) (score range 0–128, with higher scores indicating higher levels of activity); and
6. Willingness to have surgery, rated on a 5-point scale ranging from "definitely not willing" to "definitely willing," with those indicating "probably not willing" or "definitely not willing" classified as unwilling to have knee surgery in the near future and all other options classified as willing.

Statistical analysis. All statistical analyses were performed using Stata, version 16.1 (StataCorp LLC), on complete case data (i.e., excluding participants with any missing baseline data, missing outcome data [knee pain or physical function] at 12 months, and missing potential mediator data at 6 months). Complete case data were used in this exploratory study as characteristics of the complete case and omitted sample were comparable (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25140). Regression assumptions of linearity and homoscedasticity were assessed using standard diagnostic plots.

Initially, to explore the effect of the treatment group on each continuous and binary potential mediator, respectively, at 6 months, separate linear regression and logistic regression models were fitted (Figure 1, Pathway A). This was conducted separately for each pair of treatment groups, Diet + Exercise versus Control and Exercise versus Control. For Diet + Exercise versus Exercise, a linear regression model was fitted only on the potential mediator, change in weight. Results were calculated as the estimated mean (95% confidence interval [95% CI]) difference in change (6 months minus baseline) in each continuous potential mediator between groups. The estimated relative risk and risk difference (95% CI) in the binary mediator (unwillingness to have surgery at 6 months) between groups were calculated.

Next, to explore if mediation was present, full causal mediation analyses (based on the potential outcomes framework and the counterfactual framework [24,25]) were conducted where 2 regression models were simultaneously fitted for each outcome (change in knee pain and change in physical function), considering each potential mediator and each relevant treatment group comparison separately. Potential mediators were investigated separately as this was an exploratory study, and there is a paucity of evidence regarding psychosocial mediators of exercise and diet. The first of the 2 regression models estimated the direct

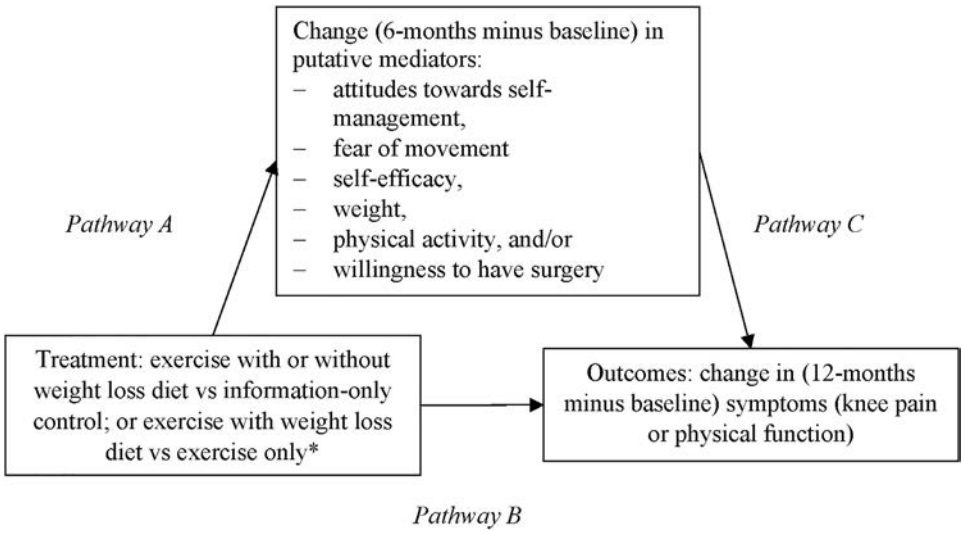

Figure 1. The effect of treatment on potential mediators (Pathway A), the direct effect of treatment on outcomes (Pathway B), and the effect of potential mediators on outcomes (Pathway C). Total effect is the sum of direct effect (Pathway B) and indirect effect (Pathway A multiplied by Pathway C). * For the Diet + Exercise versus Exercise comparison, only change (6 months minus baseline) in weight was considered.

effect of the first named group in the pairwise comparison (e.g., Diet + Exercise in Diet + Exercise versus Control comparison) and the effect of the potential mediator on the outcome (Figure 1, Pathways B and C). The second model estimated the effect of the first named group in the pairwise comparison on the potential mediator (Figure 1, Pathway A). These 2 models permitted the total effect of the first named group in the pairwise comparison on the outcome to be decomposed into the direct effect (Figure 1, Pathway B) and the indirect effect (Figure 1, Pathway A multiplied by C). The direct effect refers to the effect of the first named group in the pairwise comparison on the outcome that does not occur through the potential mediator. The indirect effect is the effect of the first named group that does occur through (i.e., is mediated by) the potential mediator.

Estimating direct and indirect effects using causal mediation analysis assumes that: 1) there are no unmeasured treatment–outcome confounders; 2) there are no unmeasured mediator–outcome confounders; 3) there are no unmeasured treatment–mediator confounders; and 4) there is no effect of treatment that confounds the mediator–outcome relationship (26,27). As treatment was randomly allocated and it appears reasonable that missing data were missing completely at random, assumptions 1 and 3 are satisfied. History of knee surgery (the stratification variable in the original trial), baseline mediator and baseline outcome scores (for the specific mediator and outcome considered in that model) were included as covariates in both regression models, as these were assumed to be the only potential confounders of the potential mediator–outcome relationship (Figure 1, Pathway C). There were no effects of treatment known to the authors a priori that could confound each mediator–outcome relationship (assumption 4). These causal mediation analyses were each conducted using the "paramed" function (28). The "medeff" function in Stata (29) was used to calculate the proportion of the total effect mediated through the potential mediator, estimated as the ratio of the indirect effect to the total effect, and presented as a percentage (95% CI). Percentages can exceed 100% if the sum of indirect effects exceeds total effects, which occurs if mediators affect one another or if there are interactions between mediators (30).

Finally, to estimate the magnitude of the association between each potential mediator (6 months minus baseline or at 6 months) and change in each symptom (12 months minus baseline), separate linear regression models were used (Figure 1, Pathway C). In these models, change in symptoms was entered as the outcome, the potential mediator was the independent variable and relevant baseline mediator scores, relevant baseline outcome scores, the stratifying variable, and each relevant pair of treatment groups were entered as covariates. Results were calculated as the estimated mean (95% CI) effect of the potential mediator on change in symptoms (12 months minus baseline). Analyses for Pathways A and C in Figure 1 are provided in Supplementary Table 6, (available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25140). However, these do not influence the causal mediation analyses. Therefore, the results are not described further (30).

We did not adjust for multiplicity as this was an exploratory study, where all results are hypothesis-generating and not confirmatory. We have reported all effects, confidence intervals, and *P* values to let readers use their own judgment about the relative weight of the conclusions. This approach aligns with the usage of *P* values favored by the American Statistical Association (31).

Sensitivity analyses (Supplementary Table 2, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25140) for mediation effects were conducted to investigate how robust the full

causal mediation analyses results were to violation of the sequential ignorability assumption (i.e., that there is an unmeasured confounder related to both the mediator and the outcome) (32). Violation was assessed using a sensitivity parameter, rho, which represents the correlation between the error terms of the mediator and outcome models, a measure of the degree of unmeasured mediator–outcome confounding (29). This parameter was allowed to vary to determine the impact on the resulting estimated indirect (mediation) effect, then the value of rho at which this effect is zero was examined.

RESULTS

Of the 415 participants enrolled in the trial, 345 (83%) had complete case data and were analyzed in this mediation analysis. The control group had a higher proportion of female participants, otherwise baseline characteristics of participants were similar between groups (Table 1).

Effect of treatment group on potential mediators. Results for Pathway A, the effect of treatment on potential mediators, are shown in Tables 2 and 3. Compared to the Control program, the Exercise program treatment led to improvements in fear of movement, self-efficacy, reduction in weight, physical activity, and increased unwillingness to have surgery. Compared to the Control program, the Diet + Exercise program led to improvements in attitudes toward self-management, fear of movement, self-efficacy, reduction in weight, and increased unwillingness to have surgery. The Diet + Exercise program led to a reduction in weight compared to the Exercise program.

Mediators of effects of Exercise compared to Control. The full causal mediation analysis for the effects of the Exercise program, compared to the Control program, on outcomes is shown in Table 4 and Supplementary Table 7 (available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25140). For change in knee pain, the effect of the Exercise program versus the Control program was only mediated by change in self-efficacy. The total effect of the Exercise program versus the Control program was a 0.70-unit reduction on the NRS (95% CI –1.45, 0.05). An estimated 0.40 units (–0.75, –0.06) of that reduction was through an increase in self-efficacy (corresponding to 54% of the total effect).

For change in physical function, the effect of the Exercise program versus the Control program was mediated by changes in both fear of movement and self-efficacy. The total effect of the Exercise program versus the Control program was a 4.99-unit reduction on the WOMAC (–8.13, –1.85). An estimated 1.11 units (–2.15, –0.07) of that reduction was through a reduction in fear of movement, (corresponding to 23% of the total effect), and 1.66 units (–3.04, –0.28) was through an increase in self-efficacy (corresponding to 33% of the total effect).

Mediators of effects of Diet + Exercise compared to Control. The full causal mediation analysis for effects of the Diet + Exercise program compared to the Control program on outcomes is shown in Table 5 and Supplementary Table 7 (available on the Arthritis Care & Research website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25140). The effect of the Diet + Exercise program versus the Control program on knee pain was mediated by an increase in self-efficacy and reduction in weight. The total effect of the Diet + Exercise program versus

Table 1. Participant baseline characteristics*

	Control (n = 45)	Exercise (n = 137)	Diet + Exercise (n = 163)	Missing data (n = 69)†
Age, mean ± SD years	64.9 ±8.7	66.0 ± 8.2	64.2 ± 8.2	64.0 ± 8.1
Female, no. (%)	31 (68.9)	77 (56.2)	85 (52.1)	34 (49.3)
Body mass index (kg/m^2)	33.2 (30.8–36.5)	32.0 (29.9–34.6)	32.4 (30.4–35.7)	34.6 (31.6–37.6)
Knee pain (NRS)	6 (5–7)	6 (5–7)	6 (5–7)	6 (5–6)
Physical function (WOMAC), mean ± SD	21.2 ± 9.9	22.2 ± 10.7	24.1 ± 9.0	24.7 ± 10.0
Attitudes towards self-management (PAM-13)	46 (40–49)	44 (40–49)	44 (40–48)	44 (40–47)
Fear of movement (BFMS)	11 (10–14)	12 (9–14)	12 (10–15)	13 (10–14)
Self-efficacy (ASES), mean ± SD	21.4 ± 3.5	20.7 ± 4.1	20.6 ± 3.7	19.4 ± 3.7
Weight (kg), mean ± SD	96.6 ± 13.6	94.1 ± 13.0	95.4 ± 13.6	99.5 ± 15.3
Physical activity (IPEQ)	17.1 (11.9–35.8)	22.8 (12.9–34.0)	21.6 (11.9–32.3)	16.6 (11.0–29.4)
Unwilling to have surgery, no. (%)‡	12 (26.7)	42 (30.7)	45 (27.6)	16 (23.2)
History of knee surgery (arthroscopy or contralateral arthroplasty), no. (%)	26 (57.8)	80 (58.4)	93 (57.1)	35 (50.7)

* Values are the median (interquartile range [IQR]) unless indicated otherwise. ASES = Arthritis Self-Efficacy Scale (scored 3–30, with higher scores indicating greater self-efficacy); BFMS = Brief Fear of Movement Scale for osteoarthritis (scored 6–24, with higher scores indicating greater fear); IPEQ = Incidental and Planned Exercise Questionnaire (version W; scored 0–128, with higher scores indicating higher levels of activity); NRS = numerical rating scale (rated 0–10, with higher scores indicating worse pain); PAM-13 = Patient Activation Measure (scored 13–52, with higher scores indicating greater patient activation); WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index (physical function subscale; rated 0–68, with higher scores indicating worse function).
† One participant requested withdrawal of all data from the study (i.e., of 415 participants enrolled, data for 414 participants are shown).
‡ Rated using a 5-point scale with terminal descriptors of "definitely not willing" to "definitely willing," with those indicating "probably not willing" or "definitely not willing" classified as unwilling to have knee surgery in the near future and all other options classified as willing.

Table 2. Mean ± SD change by group and mean (95% CI) difference in change for each potential mediator between groups*

Potential mediators	Control (n = 45)†	Exercise (n = 137)†	Diet + Exercise (n = 163)†	Exercise vs. Control (n = 182)‡	*P*	Diet + Exercise vs. Control (n = 208)‡	*P*	Diet + Exercise vs. Exercise (n = 300)‡	*P*
Attitudes toward self-management (PAM-13)¶	0.16 ± 6.53	1.44 ± 6.62	3.25 ± 6.29	1.33 (–0.50, 3.17)	0.154	2.88 (1.00, 4.75)	0.003	–	–
Fear of movement (BFMS)§	–0.13 ± 3.42	–1.67 ± 3.30	–2.06 ± 4.04	–1.61 (–2.56, –0.66)	<0.001	–1.61 (–2.73, –0.49)	0.005	–	–
Self-efficacy (ASES)¶	–0.51 ± 3.59	2.95 ± 5.73	4.81 ± 4.66	3.01 (1.43, 4.59)	<0.001	4.77 (3.51, 6.03)	<0.001	–	–
Weight (kg)§	–0.52 ± 4.12	–1.71 ± 3.72	–10.29 ± 6.05	–1.36 (–2.62, –0.09)	0.035	–9.95 (–11.69, –8.21)	<0.001	–8.39 (–9.49, –7.30)	<0.001
Physical activity (IPEQ-W)¶	0.43 ± 13.45	6.49 ± 18.62	5.25 ± 16.71	5.87 (0.54, 11.19)	0.031	4.40 (–0.36, 9.16)	0.070	–	–

* Mean (95% CI) difference in change (6 months minus baseline) in each potential mediator between treatment groups, adjusted for baseline mediator scores and the stratifying variable, history of knee surgery (arthroscopy or contralateral arthroplasty), estimated using separate regression models for each treatment group comparison. 95% CI = 95% confidence interval; ASES = Arthritis Self-Efficacy Scale (scored 3–30, with higher scores indicating greater self-efficacy); BFMS = Brief Fear of Movement Scale for osteoarthritis (scored 6–24, with higher scores indicating greater fear); IPEQ-W = Incidental and Planned Exercise Questionnaire ("past week" version; scored 0–128, with higher scores indicating higher levels of activity); PAM-13 = Patient Activation Measure (scored 13–52, with higher scores indicating greater patient activation).
† Values are the mean ± SD change within groups (6 months minus baseline) (positive changes indicate improvement).
‡ Values are the mean (95% confidence interval [95% CI]) difference in change (6 months minus baseline) between groups (positive differences favor the first named group in the pairwise comparison).
§ For change within groups, negative changes indicate improvement. For difference in change between groups, negative differences favor the first named group in the pairwise comparison.
¶ For change within groups, positive changes indicate improvement. For difference in change between groups, positive differences favor the first named group in the pairwise comparison.

Table 3. Counts (proportions) at 6 months by group and relative risks and risk differences for each potential mediator between groups*

Counts (proportions) at 6 months	Control (n = 45	Exercise (n = 137)	Diet + Exercise (n = 163)	Exercise vs. Control (n = 182)	*P*	Diet + Exercise vs. Control (n = 208)	*P*	Diet + Exercise vs. Exercise (n = 300)	*P*
Unwilling to have surgery, values†									
Mean ± SD	19 ± 42.2	84 ± 61.3	116 ± 71.2	–	–	–	–	–	–
Relative risk (95% CI)‡	–	–	–	1.40 (1.00, 1.97)	0.047	1.66 (1.22, 2.26)	0.001	–	–
Risk difference (95% CI)§	–	–	–	0.18 (0.02, 0.33)	0.027	0.28 (0.14, 0.42)	<0.001	–	–

* Relative risks and risk differences (95% confidence interval [95% CI]) estimated using separate logistic regression models for each treatment group comparison, adjusted for baseline mediator scores and the stratifying variable, history of knee surgery (arthroscopy or contralateral arthroplasty).
† Rated using a 5-point scale with terminal descriptors of "definitely not willing" to "definitely willing," with those indicating "probably not willing" or "definitely not willing" classified as unwilling to have knee surgery in the near future, and all other options classified as willing.
‡ Relative risks >1 favor the first named group in the pairwise comparison.
§ Risk differences >0 favor the first named group in the pairwise comparison.

the Control program was a 1.25-point reduction on the NRS (–1.95, –0.54). An estimated 0.77 units (–1.26, –0.28) of that reduction was through an increase in self-efficacy (63% of the total effect) and 1.20 units (–1.73, –0.68) was through a reduction in weight (96% of the total effect).

For physical function, the effect of the Diet + Exercise program versus the Control program was mediated by change in fear of movement, self-efficacy, and weight. The total effect of the Diet + Exercise program versus the Control program was a 7.54-point reduction on the WOMAC (–10.42, –4.67). An estimated 1.13 units (–2.17, –0.08) of that reduction was through a reduction in fear of movement (14% of the total effect), 5.15 units (–7.34, –2.96) was through an increase in self-efficacy (69% of the total effect), and 5.79 units (–7.96, –3.63) was through a reduction in weight (79% of the total effect).

Mediators of effects of Diet + Exercise compared to Exercise. The full causal mediation analysis for effects of the Diet + Exercise program on outcomes, compared to the Exercise program, is shown in Table 6. For knee pain and physical function, the effects of the Diet + Exercise program versus the Exercise program were both mediated by change in weight. The total effect of the Diet + Exercise program versus the Exercise program was a 0.56-point reduction in pain on the NRS (–1.05, –0.07), of which an estimated 0.89 units (–1.31, –0.47) was through a

Table 4. Estimated mean (95% CI) total, direct, and indirect effects of the Exercise program on change in symptoms (12 months minus baseline) compared to the Control program (n = 182)*

Potential mediator†	Total effect, mean (95% CI)	*P*	Direct effect, mean (95% CI)	*P*	Indirect effect, mean (95% CI)	*P*	% mediated (95% CI)‡
Knee pain (NRS)							
Attitudes towards self-management (PAM-13)	–0.69 (–1.44, 0.05)	0.067	–0.61 (–1.35, 0.13)	0.108	–0.09 (–0.28, 0.10)	0.367	12 (–52, 79)
Fear of movement (BFMS)	–0.68 (–1.43, 0.06)	0.070	–0.49 (–1.27, 0.29)	0.218	–0.20 (–0.42, 0.03)	0.089	28 (–94, 171)
Self-efficacy (ASES)	–0.70 (–1.45, 0.05)	0.068	–0.29 (–1.04, 0.46)	0.443	–0.40 (–0.75, –0.06)	0.023	54 (–183, 496)
Weight (kg)	–0.68 (–1.42, 0.05)	0.069	–0.58 (–1.31, 0.16)	0.124	–0.11 (–0.28, 0.07)	0.228	15 (–86, 96)
Physical activity (IPEQ-W)	–0.70 (–1.45, 0.05)	0.069	–0.61 (–1.37, 0.15)	0.116	–0.09 (–0.25, 0.07)	0.268	11 (–45, 77)
Unwilling to have surgery§	–0.70 (–1.45, 0.05)	0.066	–0.64 (–1.38, 0.10)	0.091	–0.06 (–0.23, 0.10)	0.461	6 (–22, 39)
Physical function (WOMAC)							
Attitudes toward self-management (PAM-13)	–4.88 (–8.01, –1.75)	0.002	–4.71 (–7.81, –1.62)	0.003	–0.17 (–0.80, 0.47)	0.603	3 (2, 9)
Fear of movement (BFMS)	–4.79 (–7.92, –1.65)	0.003	–3.68 (–6.78, –0.58)	0.020	–1.11 (–2.15, –0.07)	0.037	23 (14, 60)
Self–efficacy (ASES)	–4.99 (–8.13, –1.85)	0.002	–3.33 (–6.42, –0.23)	0.035	–1.66 (–3.04, –0.28)	0.018	33 (20, 92)
Weight (kg)	–4.68 (–7.85, –1.51)	0.004	–4.37 (–7.48, –1.26)	0.006	–0.31 (–1.02, 0.40)	0.392	7 (4, 19)
Physical activity (IPEQ-W)	–4.90 (–8.08, –1.72)	0.003	–4.66 (–7.78, –1.55)	0.003	–0.24 (–0.80, 0.32)	0.407	5 (3, 13)
Unwilling to have surgery§	–4.96 (–8.16, –1.76)	0.002	–4.85 (–7.98, –1.73)	0.002	–0.11 (–0.73, 0.52)	0.738	1 (1, 3)

* Adjusted for baseline mediator scores, baseline outcome scores, and the stratifying variable, history of knee surgery (arthroscopy or contralateral arthroplasty); negative effects favor the Exercise group. 95% CI = 95% confidence interval; ASES = Arthritis Self-Efficacy Scale (scored 3–30, with higher scores indicating greater self-efficacy); BFMS = Brief Fear of Movement Scale for osteoarthritis (scored 6–24, with higher scores indicating greater fear); IPEQ-W = Incidental and Planned Exercise Questionnaire ("past week" version; scored 0–128, with higher scores indicating higher levels of activity); NRS = numerical rating scale (rated 0–10, with higher scores indicating worse pain); PAM-13 = Patient Activation Measure (scored 13–52, with higher scores indicating greater patient activation); WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index (physical function subscale; rated 0–68, with higher scores indicating worse function).
† Potential mediator is change in (6 months minus baseline) except for the binary mediator, unwilling to have surgery, which is at 6 months.
‡ If the sum of the proportion mediated exceeds 100%, then it is likely the mediators affect one another, or there are interactions between the mediators (30).
§ Rated using a 5-point scale with terminal descriptors of "definitely not willing" to "definitely willing," with those indicating "probably not willing" or "definitely not willing" classified as unwilling to have knee surgery in the near future, and all other options classified as willing.

Table 5. Estimated mean (95% CI) total, direct, and indirect effects of the Diet + Exercise program on change in symptoms (12 months minus baseline) compared to Control program (n = 208)*

Potential mediator†	Total effect, mean (95% CI)	*P*	Direct effect, mean (95% CI)	*P*	Indirect effect, mean (95% CI)	*P*	% mediated (95% CI)‡
Knee pain (NRS)							
Attitudes towards self-management (PAM-13)	−1.29 (−2.01, −0.57)	<0.001	−1.19 (−1.90, −0.48)	0.001	−0.10 (−0.31, 0.11)	0.365	8 (5, 15)
Fear of movement (BFMS)	−1.31 (−2.02, −0.61)	<0.001	−1.23 (−1.95, −0.51)	<0.001	−0.08 (−0.24, 0.08)	0.301	7 (4, 13)
Self-efficacy (ASES)	−1.25 (−1.95, −0.54)	<0.001	−0.48 (−1.25, 0.29)	0.226	−0.77 (−1.26, −0.28)	0.002	63 (40, 123)
Weight (kg)	−1.27 (−1.97, −0.57)	<0.001	−0.06 (−0.89, 0.76)	0.878	−1.20 (−1.73, −0.68)	<0.001	96 (63, 190)
Physical activity (IPEQ-W)	−1.29 (−2.00, −0.57)	<0.001	−1.23 (−1.93, −0.53)	<0.001	−0.05 (−0.15, 0.04)	0.269	4 (3, 8)
Unwilling to have surgery§	−1.31 (−2.03, −0.60)	<0.001	−1.07 (−1.76, −0.38)	0.002	−0.24 (−0.51, 0.02)	0.070	13 (8, 27)
Physical function (WOMAC)							
Attitudes towards self-management (PAM-13)	−7.57 (−10.44, −4.70)	<0.001	−6.81 (−9.63, −3.99)	<0.001	−0.76 (−1.97, 0.45)	0.216	10 (7, 16)
Fear of movement (BFMS)	−7.61 (−10.44, −4.78)	<0.001	−6.48 (−9.37, −3.60)	<0.001	−1.13 (−2.17, −0.08)	0.035	14 (10, 21)
Self-efficacy (ASES)	−7.54 (−10.42, −4.67)	<0.001	−2.39 (−5.35, 0.56)	0.113	−5.15 (−7.34, −2.96)	<0.001	69 (49, 109)
Weight (kg)	−7.41 (−10.19, −4.62)	<0.001	−1.61 (−4.96, 1.74)	0.345	−5.79 (−7.96, −3.63)	<0.001	79 (57, 125)
Physical activity (IPEQ-W)	−7.58 (−10.49, −4.67)	<0.001	−7.38 (−10.28, −4.47)	<0.001	−0.20 (−0.65, 0.25)	0.378	3 (2, 4)
Unwilling to have surgery§	−7.67 (−10.66, −4.68)	<0.001	−6.55 (−9.46, −3.64)	<0.001	−1.12 (−2.30, 0.05)	0.061	10 (7, 16)

* Estimated mean adjusted for baseline mediator scores, baseline outcome scores and the stratifying variable, history of knee surgery (arthroscopy or contralateral arthroplasty); for total, direct, and indirect effects. Negative effects favor the Diet + Exercise group. ASES = Arthritis Self-Efficacy Scale; scored 3–30, with higher scores indicating greater self-efficacy. BFMS = Brief Fear of Movement Scale for osteoarthritis (scored 6–24, with higher scores indicating greater fear); IPEQ-W = Incidental and Planned Exercise Questionnaire, "past week" version (scored 0–128, with higher scores indicating higher levels of activity); NRS = numerical rating scale; rated 0–10, with higher scores indicating worse pain; PAM-13 = Patient Activation Measure (scored 13–52, with higher scores indicating greater patient activation); WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index (physical function subscale; rated 0–68, with higher scores indicating worse function).
† Potential mediator is change in (6 months minus baseline) except for the binary mediator, unwilling to have surgery, which is at 6 months.
‡ If the sum of the proportion mediated exceeds 100%, then one of the following must be true: 1) There are other mediators with a negative proportion mediated, 2) the mediators affect one another, or 3) there are interactions between the mediators (30).
§ Rated using a 5-point scale with terminal descriptors of "definitely not willing" to "definitely willing," with those indicating "probably not willing" or "definitely not willing" classified as unwilling to have knee surgery in the near future, and all other options classified as willing.

change in weight. The total effect of the Diet + Exercise program versus the Exercise program was a 3.09-point reduction in physical dysfunction on the WOMAC (–5.18, –0.99), of which an estimated 4.02 units (–5.77, –2.26) was through a change in weight.

Associations between changes in potential mediators and outcomes (Pathway C in Figure 1), where the change in the mediator may or may not be attributed to group allocation (i.e., intervention received), are shown in Supplementary Table 5, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25140. A comparison of the results of the analyses for Pathways A and C in Figure 1 with the results from the causal mediation analyses are provided in Supplementary Table 6 (available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25140). However, only the causal mediation analyses results were used to investigate mediation (33).

Sensitivity analyses. Sensitivity analyses are presented in Supplementary Table 2. Supplementary Table 3 and Supplementary Figures 1–6 (available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25140) show the value of the sensitivity parameter (rho) required to change the direction of the indirect effect for each treatment comparison (mediator and outcome). Where the causal mediation analysis results suggested mediation may be present, the sensitivity analyses suggested the indirect effect changes to 0 and then reverses direction at values of rho of at least 0.2. Although

Table 6. Estimated mean (95% CI) total, direct, and indirect effects of the Diet + Exercise program on change in symptoms (12 months minus baseline) compared to Exercise program (n = 300)*

Potential mediator†	Total effect, mean (95% CI)	*P*	Direct effect, mean (95% CI)	*P*	Indirect effect, mean (95% CI)	*P*	% mediated, (95% CI)‡
Knee pain (NRS)							
Weight (kg)	−0.56 (−1.05, −0.07)	0.024	0.33 (−0.32, 0.99)	0.317	−0.89 (−1.31, −0.47)	<0.001	–
Physical function (WOMAC)							
Weight (kg)	−3.09 (−5.18, −0.99)	0.004	0.93 (−1.84, 3.71)	0.510	−4.02 (−5.77, −2.26)	<0.001	–

* Estimated mean adjusted for baseline mediator scores, baseline outcome scores and the stratifying variable, history of knee surgery (arthroscopy or contralateral arthroplasty). Negative effects favor the Diet + Exercise group. 95% CI = 95% confidence interval; NRS = numerical rating scale (rated 0–10, with higher scores indicating worse pain); WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index (physical function subscale; rated 0–68, with higher scores indicating worse function).
† Potential mediator is change in (6 months minus baseline).
‡ Percent mediated not presented as the direct and indirect effects are in opposite directions (30).

these appear large, there is no cutoff value for rho to judge the robustness of the results to the violation of the ignorability assumption (32). However, omitting an observed confounder (the relevant mediator at baseline) reduced rho by at most 0.13, suggesting 0.2 is a large critical value so results appear robust to the violation of the ignorability assumption.

DISCUSSION

This study aimed to evaluate mediators of the effects of 2 telehealth-delivered exercise programs, including 1 program with and 1 program without a weight-loss diet, on pain and physical function in people with knee OA. We found that reduced fear of movement and increased arthritis self-efficacy may mediate the effects of exercise, and diet and exercise, on improvements in pain and physical function. In addition, weight loss may mediate the effects of diet and exercise on pain and physical function, compared to exercise only and to information only.

Our findings add to the limited similar research available. To our knowledge, only 3 previous studies have evaluated mediators of exercise for people with OA (11,12,34), and only 1 of those evaluated psychosocial potential mediators (34). This 1 study, by Rejeski and colleagues, found that self-efficacy did not mediate the effects of exercise on physical function (34), which differs from our findings. However, Rejeski et al used a measure of self-efficacy for stair climbing rather than for arthritis symptom management as was used in our study. In combined diet and exercise interventions compared to information alone, 2 previous studies found that self-efficacy mediated effects on physical function (16) and that weight loss mediated effects on pain and physical function (13), both of which concur with our findings. We also found that weight loss may mediate the effects of our Diet + Exercise program on pain and function when compared to exercise only. To our knowledge, no previous studies have examined weight loss as a potential mediator of the effects of diet and exercise compared to exercise alone. Only 1 previous study (15) that examined mediators of a diet and exercise program compared to exercise only found that knee extensor strength and self-efficacy for walking duration (neither of which were measured in our study) mediated effects on physical function and the 6-minute walk test. Finally, we found that reduced fear of movement may mediate the effects of both exercise and diet and exercise on physical function. To our knowledge, no previous studies examined fear of movement as a potential mediator, so further research is required to confirm our findings.

We found no evidence of mediation by attitudes toward self-management, physical activity, or willingness to have knee surgery, for either diet and exercise or exercise alone. Attitudes toward self-management or willingness to have surgery have not previously been examined as potential mediators of exercise or diet and exercise programs. One previous study (35) examined physical activity as a potential mediator of a diet and exercise program for people with knee OA, also finding no evidence of mediation on changes in pain. This suggests that these variables may not contribute toward the mechanism of effect on symptoms of pain and function, or the tools used to measure these domains are problematic. Further research is needed to confirm our findings and to evaluate whether they have an important role in other outcomes such as quality of life, mental health, or health care costs. Future interventions could explore inclusion of components targeting "activation."

Our findings have implications for the design of future exercise and diet programs for people with knee OA. We found that fear of movement and/or arthritis self-efficacy have important roles in the mechanism of effect of exercise, and diet and exercise, programs. Intuitively, it makes sense that greater belief in one's capability to manage their OA, and reduced fear of engaging in exercising or physical activity, could contribute to greater improvements in self-reported pain or physical function. Indeed, arthritis self-efficacy has been found to have significant overall associations with pain severity and function (36,37). There is also evidence that greater arthritis-related self-efficacy is associated with better overall health status and lower health care costs (38), suggesting that there are potentially additional benefits. Similarly, reduced fear of movement has been associated with lower pain and better physical function (39–41). Collectively, this suggests that future exercise, and diet and exercise, programs should include elements that target self-efficacy and fear of movement. For example, self-efficacy has been shown to be enhanced by various strategies like motivational interviewing (42,43), pain-coping skills training (44), education about self-management (43,45,46), or a home-based exercise program (45,46), all of which were key components of both of our programs. In addition, education about OA using an empowerment and participatory discourse (rather than a disease and impairment discourse) has recently been shown to improve self-efficacy and fear of movement and could be incorporated into the information provided in our interventions (47). Other behavioral and psychological interventions targeting fear of movement (48) may also be helpful additional components and contribute to further improved physical function.

We found that weight loss may mediate the effects of the diet and exercise program on outcomes when compared to control and to exercise alone. Our findings suggest that a 10 kg loss of weight (approximately proportional to a 10% reduction in weight, based on the mean weight of the cohort at baseline) corresponded to a 1.2-unit (compared to control) and a 1.1-unit (compared to exercise alone) improvement in pain (on the 0–10 NRS), and a 5.8-unit and 4.8-unit improvement in physical function (on the 0–68 WOMAC scale), respectively. These changes do not quite reach the minimal clinically important difference (MCID) for pain and function (1.8 units and 6.0 units, respectively [49]), though MCIDs vary depending on the characteristics of the

population and the treatment they receive (49). Indeed, a recent systematic review that aimed to identify the most credible MCIDs for outcomes in those with chronic knee pain stated that their best estimates of MCIDs were ~10% of the instrument's total range (50), suggesting a change of 1 NRS unit in pain may be considered worthwhile. In addition, if combined with other mechanisms such as improvement in self-efficacy and fear of movement, weight loss may contribute to clinically significant improvements in outcomes. Weight loss is thought to alter mechanical pathways and loading within the joint, contributing to reduced pain (5,51). Weight loss has also been shown to lower joint inflammation and change levels of joint biomarkers (52–54), improving quality of life (including subdomains of physical functioning, vitality, stress, and mental health) (55). There are numerous diets which could facilitate weight loss in people with OA. Our study used a ketogenic VLCD, which has been perceived by participants to be easy to use, convenient, and effective (56). This is supported by other research suggesting that ketogenic VLCDs are associated with significant (10–16 kg) weight loss in people who are overweight or obese, which is maintained at 2 years follow-up (57). Importantly, our diet program also involved a suite of information booklets and behavior change support that likely helped support weight loss (56), suggesting that future exercise and weight-loss programs for this population may consider including similar components.

Our study has limitations. Our findings should be interpreted with caution, as other mediating factors, like joint loading, muscle strength, and inflammatory biomarkers (none of which were measured in our trial), may also account for part of the treatment effects (11,14). As potential mediators were examined separately in this study, our analyses do not account for potential interactions between mediators. For example, it is not clear whether reduced fear of movement led to improved arthritis self-efficacy, rather than the program itself being responsible for the change in self-efficacy. In addition, there may be other confounders that were not accounted for; however, our sensitivity analyses suggested findings were insensitive to unmeasured mediator–outcome confounding. Our analysis was based on participants with complete data, and attrition rates differed across the 3 trial groups (6). This may have introduced bias if missing data were related to mediators or outcomes; however, the assumption that data were missing completely at random appeared reasonable since baseline characteristics were similar between those with missing data versus those with complete data. Our findings cannot necessarily be generalized to those without private health insurance or to those residing outside of Australia. Finally, our findings may also not be generalizable to exercise or diet programs that do not include the level of behavior change support included in our programs, such as use of motivational interviewing by clinicians, and provision of bespoke booklets with behavior change activities.

In conclusion, increased arthritis self-efficacy, reduced fear of movement, and weight loss may partially mediate telehealth-delivered exercise program effects, including those programs with and without diet, on pain and/or function in knee OA. Weight loss may partially mediate the effect of diet and exercise compared to exercise alone.

ACKNOWLEDGMENTS

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

AUTHOR CONTRIBUTIONS

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

Study conception and design. Lawford, Hinman, Egerton, Keating, Brown, Oliver, Bennell.

Acquisition of data. Lawford.

Analysis and interpretation of data. Lawford, Hinman, McManus, Lamb, Bennell.

REFERENCES

1. Kolasinski SL, Neogi T, Hochberg MC, et al. 2019 American College of Rheumatology/Arthritis Foundation guideline for the management of osteoarthritis of the hand, hip, and knee. Arthritis Rheumatol 2020;72:220–33.
2. Bannuru RR, Osani MC, Vaysbrot EE, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis [review]. Osteoarthritis Cartilage 2019;27:1578–89.
3. National Institute for Health and Care Excellence (UK). Osteoarthritis: care and management, clinical guideline no. 177. London: NICE; 2014. URL: https://www.nice.org.uk/guidance/cg177.
4. Royal Australian College of General Practitioners. Guideline for the management of knee and hip osteoarthritis, 2nd edition. East Melbourne (Australia): RACGP; 2018. URL: https://www.racgp.org.au/getattachment/71ab5b77-afdf-4b01-90c3-04f61a910be6/Guideline-for-the-management-of-knee-and-hip-osteoarthritis.aspx.
5. Messier SP, Mihalko SL, Legault C, et al. Effects of intensive diet and exercise on knee joint loads, inflammation, and clinical outcomes among overweight and obese adults with knee osteoarthritis: the IDEA randomized clinical trial. JAMA 2013;310:1263–73.
6. Bennell KL, Lawford BJ, Keating C, et al. Comparing video-based, telehealth–delivered exercise and weight loss programs with online education on outcomes of knee osteoarthritis: a randomized trial. Ann Intern Med 2022;175:198–209.
7. Hall M, Castelein B, Wittoek R, et al. Diet-induced weight loss alone or combined with exercise in overweight or obese people with knee osteoarthritis: a systematic review and meta-analysis. Semin Arthritis Rheum 2019;48:765–77.
8. Allen KD, Choong PF, David AM, et al. Osteoarthritis: models for appropriate care across the disease continuum [review]. Best Pract Res Clin Rheumatol 2016;30:503–35.
9. Lee H, Herbert RD, McAuley JH. Mediation analysis. JAMA 2019;321:697–8.
10. Lima YL, Lee H, Klyne DM, et al. How do non-surgical interventions improve pain and physical function in people with osteoarthritis? A scoping review of mediation analysis studies [review]. Arthritis Care Res (Hoboken) 2023;75:467–81.
11. Hall M, Hinman RS, Wrigley TV, et al. Knee extensor strength gains mediate symptom improvement in knee osteoarthritis: secondary

analysis of a randomised controlled trial. Osteoarthritis Cartilage 2018;26:495–500.

12. Bandak E, Boesen M, Bliddal H, et al. Exercise-induced pain changes associate with changes in muscle perfusion in knee osteoarthritis: exploratory outcome analyses of a randomised controlled trial. BMC Musculoskelet Disord 2019;20:491.
13. Foy CG, Lewis CE, Hairston KG, et al. Intensive lifestyle intervention improves physical function among obese adults with knee pain: findings from the Look AHEAD trial. Obesity (Silver Spring) 2011;19: 83–93.
14. Runhaar J, Beavers DP, Miller GD, et al. Inflammatory cytokines mediate the effects of diet and exercise on pain and function in knee osteoarthritis independent of BMI. Osteoarthritis Cartilage 2019;27: 1118–23.
15. Mihalko SL, Cox P, Beavers DP, et al. Effect of intensive diet and exercise on self-efficacy in overweight and obese adults with knee osteoarthritis: the IDEA randomized clinical trial. Transl Behav Med 2019; 9:227–35.
16. Taylor SS, Oddone EZ, Coffman CJ, et al. Cognitive mediators of change in physical functioning in response to a multifaceted intervention for managing osteoarthritis. Int J Behav Med 2018;25:162–70.
17. Bennell KL, Keating C, Lawford BJ, et al. Better Knee, Better Me™: effectiveness of two scalable health care interventions supporting self-management for knee osteoarthritis-protocol for a randomized controlled trial. BMC Musculoskelet Disord 2020;21:160.
18. Sumithran P, Prendergast LA, Delbridge E, et al. Ketosis and appetite-mediating nutrients and hormones after weight loss. Eur J Clin Nutr 2013;67:759–64.
19. Bellamy N, Buchanan WW, Goldsmith CH, et al. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833–40.
20. Hibbard JH, Mahoney ER, Stockard J, et al. Development and testing of a short form of the patient activation measure. Health Serv Res 2005;40:1918–30.
21. Shelby RA, Somers TJ, Keefe FJ, et al. Brief fear of movement scale for osteoarthritis. Arthritis Care Res (Hoboken) 2012; 64:862–71.
22. Lorig K, Chastain RL, Ung E, et al. Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis. Arthritis Rheum 1989;32:37–44.
23. Delbaere K, Hauer K, Lord SR. Evaluation of the incidental and planned activity questionnaire (IPEQ) for older people. Br J Sports Med 2010;44:1029–34.
24. Holland PW. Statistics and causal inference. J Am Stat Assoc 1986; 81:945–60.
25. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. J Ed Psychol 1974;66:688.
26. Valeri L, VanderWeele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychol Methods 2013;18:137–50.
27. VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. Stat Interface 2009;2: 457–68.
28. Emsley R, Liu H. PARAMED: stata module to perform causal mediation analysis using parametric regression models. Stat Softw Compon 2013.
29. Hicks R, Tingley D. Causal mediation analysis. Stata J 2011;11: 605–19.
30. VanderWeele TJ. Explanation in causal inference: methods for mediation and interaction. Oxford (UK): Oxford University Press; 2015.
31. Wasserstein RL, Schirm AL, Lazar NA. Moving to a world beyond "p< 0.05". Am Stat 2019;73:1–19.
32. Zhang Z, Zheng C, Kim C, et al. Causal mediation analysis in the context of clinical research. Ann Transl Med 2016;4:425.
33. VanderWeele TJ. Mediation analysis: a practitioner's guide. Annu Rev Public Health 2016;37:17–32.
34. Rejeski WJ, Ettinger WH Jr, Martin K, et al. Treating disability in knee osteoarthritis with exercise therapy: a central role for self-efficacy and pain. Arthritis Care Res 1998;11:94–101.
35. Williams A, Lee H, Kamper SJ, et al. Causal mechanisms of a healthy lifestyle intervention for patients with musculoskeletal pain who are overweight or obese. Clin Rehabil 2019;33:1088–97.
36. Degerstedt Å, Alinaghizadeh H, Thorstensson CA, et al. High self-efficacy – a predictor of reduced pain and higher levels of physical activity among patients with osteoarthritis: an observational study. BMC Musculoskelet Disord 2020;21:380.
37. Jackson T, Xu T, Jia X. Arthritis self-efficacy beliefs and functioning among osteoarthritis and rheumatoid arthritis patients: a meta-analytic review [review]. Rheumatology (Oxford) 2020;59:948–58.
38. Cross M, March LM, Lapsley HM, et al. Patient self-efficacy and health locus of control: relationships with health status and arthritis-related expenditure. Rheumatology (Oxford) 2006;45:92–6.
39. Somers TJ, Keefe FJ, Pells JJ, et al. Pain catastrophizing and pain-related fear in osteoarthritis patients: relationships to pain and disability. J Pain Symptom Manag 2009;37:863–72.
40. Alshahrani MS, Reddy RS, Tedla JS, et al. Association between kinesiophobia and knee pain intensity, joint position sense, and functional performance in individuals with bilateral knee osteoarthritis. Healthcare (Basel) 2022;10:120.
41. Aykut Selçuk, M, Karakoyun A. Is there a relationship between kinesiophobia and physical activity level in patients with knee osteoarthritis? Pain Med 2020;21:3458–69.
42. Marks R. Self-efficacy and arthritis disability: an updated synthesis of the evidence base and its relevance to optimal patient care [review]. Health Psychol Open 2014;1:2055102914564582.
43. Marks R. Self-Efficacy-an important arthritis pain correlate and possible treatment factor general review and commentary. Nov Tech Arthritis Bone Res 2017;1:1–10.
44. Somers TJ, Blumenthal JA, Guilak F, et al. Pain coping skills training and lifestyle behavioral weight management in patients with knee osteoarthritis: a randomized controlled study. Pain 2012;153:1199–209.
45. Oroh CT, Suprapti F, Susilo WH. Improving self-efficacy and life activities of arthritis patients: a quasy-experiment study. Enfermería Clínica 2020;30:30–3.
46. Uritani D, Koda H, Sugita S. Effects of self-management education programmes on self-efficacy for osteoarthritis of the knee: a systematic review of randomised controlled trials. BMC Musculoskelet Disord 2021;22:515.
47. Egerton T, Bennell KL, McManus F, et al. Comparative effect of two educational videos on self-efficacy and kinesiophobia in people with knee osteoarthritis: an online randomised controlled trial. Osteoarthritis Cartilage 2022;30:1398–410.
48. Gunn AH, Schwartz TA, Arbeeva LS, et al. Fear of movement and associated factors among adults with symptomatic knee osteoarthritis. Arthritis Care Res (Hoboken) 2017;69:1826–33.
49. Tubach F, Ravaud P, Baron G, et al. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. Ann Rheum Dis 2005; 64:29–33.
50. Devji T, Guyatt GH, Lytvyn L, et al. Application of minimal important differences in degenerative knee disease outcomes: a systematic review and case study to inform BMJ Rapid Recommendations [review]. BMJ Open 2017;7:e015587.

51. Messier SP, Legault C, Loeser RF, et al. Does high weight loss in older adults with knee osteoarthritis affect bone-on-bone joint loads and muscle forces during walking? Osteoarthritis Cartilage 2011;19:272–80.

52. Richette P, Poitou C, Garnero P, et al. Benefits of massive weight loss on symptoms, systemic inflammation and cartilage turnover in obese patients with knee osteoarthritis. Ann Rheum Dis 2011;70:139–44.

53. Bliddal H, Leeds AR, Christensen R. Osteoarthritis, obesity and weight loss: evidence, hypotheses and horizons – a scoping review [review]. Obes Reviews 2014;15:578–86.

54. Schrepf A, Harte SE, Miller N, et al. Improvement in the spatial distribution of pain, somatic symptoms, and depression after a weight loss intervention. J Pain 2017;18:1542–50.

55. Imayama I, Alfano CM, Kong A, et al. Dietary weight loss and exercise interventions effects on quality of life in overweight/obese postmenopausal women: a randomized controlled trial. Int J Behav Nutr Phys Act 2011;8:118.

56. Lawford BJ, Bennell KL, Jones SE, et al. "It's the single best thing I've done in the last 10 years": a qualitative study exploring patient and dietitian experiences with, and perceptions of, a multi-component dietary weight loss program for knee osteoarthritis. Osteoarthritis Cartilage 2021;29:507–17.

57. Castellana M, Conte E, Cignarelli, et al. Efficacy and safety of very low calorie ketogenic diet (VLCKD) in patients with overweight and obesity: a systematic review and meta-analysis. Rev Endocr Metab Disord 2020;21:5–16.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 75, No. 11, November 2023, pp 2328–2335
DOI 10.1002/acr.25160

Hip Abductor Weakness and Its Association With New or Worsened Knee Pain: Data From the Multicenter Osteoarthritis Study

Cara L. Lewis,[1] Neil A. Segal,[2] Gabriela V. Rabasa,[1] Michael P. LaValley,[1] Glenn N. Williams,[3] Michael C. Nevitt,[4] Cora E. Lewis,[5] David T. Felson,[1] and Joshua J. Stefanik[6]

Objective. Hip abductors, important for controlling pelvic and femoral orientation during gait, may affect knee pain. Our objective was to evaluate the relation of hip abductor strength to worsened or new-onset frequent knee pain. Given previously noted associations of knee extensor strength with osteoarthritis in women, we performed sex-specific analyses.

Methods. We used data from the Multicenter Osteoarthritis study. Hip abductor and knee extensor strength was measured. Knee pain was assessed using the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) questionnaire and a question about frequent knee pain at baseline (144-month visit), and 8, 16, and 24 months thereafter. Knee pain outcomes were worsened knee pain (2-point increase in WOMAC pain) and incident frequent knee pain (answering yes to the frequent knee pain question among those without frequent knee pain at baseline). Leg-specific analyses tested hip abductor strength as a risk factor for worsened and new frequent knee pain, adjusting for potential covariates. Additionally, we stratified by knee extensor strength (high versus low).

Results. Among women, compared to the highest quartile of hip abductor strength, the lowest quartile had 1.7 (95% confidence interval [95% CI] 1.1–2.6) times the odds of worsened knee pain; significant associations were limited to women with high knee extensor strength (odds ratio 2.0 [95% CI 1.1–3.5]). We found no relation of abductor strength to worsening knee pain in men or with incident frequent knee pain in men or women.

Conclusion. Hip abductor weakness was associated with worsening knee pain in women with strong knee extensors, but not with incident frequent knee pain in men or women. Knee extensor strength may be necessary, but not sufficient, to prevent pain worsening.

INTRODUCTION

Knee osteoarthritis (OA) and knee pain are highly prevalent in older adults and lead to poor function and increased disability. A need exists to identify modifiable risk factors that can be targeted in interventions in this population. Clinical guidelines recommend exercise and strength training as a first line of treatment for knee OA. Knee extensor weakness has been extensively studied as a risk factor for knee OA and knee pain (1–6). Other muscles in the lower extremity have also been implicated in the knee OA disease process, but limited data exist on their relation to knee pain.

Hip abductor weakness and decreased muscle volume are present in adults with knee OA (7–10). Hip abductors are important for controlling the orientation of the pelvis on top of the femur and the alignment of the femur relative to the tibia when weightbearing, both of which affect knee mechanics. Weakness of these muscles may increase pelvic drop, a movement thought to contribute to knee joint loading and pain (11,12), although direct measurements of the knee adduction moment, a primary variable of joint loading, in individuals with hip abductor weakness do not support this assertion (13–15). Nonetheless, relationships between hip abductor weakness and knee pain and function have

Drs. C. E. Lewis, Nevitt, and Felson's work was supported by the NIH/Department of Health and Human Services (grants AG-18947, AG-19069, and AG-18820, respectively). Drs. C. L. Lewis, Felson, and Stefanik's work was supported by the NIH (grants K23-AR-063235, P30-AR-072571, and K23-AR-070913, respectively).

[1]Cara L. Lewis, PT, PhD, Gabriela V. Rabasa, BS, Michael P. LaValley, PhD, David T. Felson, MD, MPH: Boston University, Boston, Massachusetts; [2]Neil A. Segal, MD, MS: University of Kansas Medical Center, Kansas City; [3]Glenn N. Williams, PT, PhD, ATC: Drexel University, Philadelphia, Pennsylvania; [4]Michael C. Nevitt, PhD, MPH: University of California San Francisco, San Francisco; [5]Cora E. Lewis, MD, MSPH: University of Alabama at Birmingham, Birmingham; [6]Joshua J. Stefanik, MSPT, PhD: Northeastern University, Boston, Massachusetts.

Drs. Felson and Stefanik contributed equally to this work.

Author disclosures are available online at https://onlinelibrary.wiley.com/doi/10.1002/acr.25160.

Address correspondence via email to Cara L. Lewis, PT, PhD, at lewisc@bu.edu.

Submitted for publication December 7, 2022; accepted in revised form May 18, 2023.

SIGNIFICANCE & INNOVATIONS

- Hip abductor strength affects pelvic orientation during gait and may alter moments across the knee.
- Hip abductor weakness increased knee pain in women, especially those with strong knee extensors.
- There was no association of hip abductor strength with knee pain in men.
- Few studies have examined the longitudinal association of hip abductor strength with knee pain, even though it may be a target of rehabilitation strategies.

been reported in individuals preceding (16) or following (17) total knee arthroplasty. In individuals with knee OA, studies demonstrate that strengthening hip abductors leads to reduced knee pain and improved function (18–21).

Hip abductor weakness could be instrumental in the development or worsening of knee pain. Cross-sectional and interventional studies cannot distinguish whether the weakness is a result of or a contributor to the development of knee pain. Longitudinal studies are necessary to elucidate relationships between this potentially modifiable factor and the risk for pain. To date, no studies have investigated the relation of hip abductor weakness to new-onset or worsening knee pain. Chang et al reported that greater baseline hip abductor strength was associated with reduced risks of cartilage worsening and function (repeated chair stands and self-reported function) (22); however, they did not assess relations with knee pain.

While evaluations often focus on weakness in individual muscles, such as the hip abductors, the observed weakness may indicate overall decreased muscle strength due to deconditioning and thus may not be muscle specific. Therapeutic interventions are frequently designed to address weakness in individual muscles noted during evaluation, emphasizing the importance of a thorough evaluation. Currently, no studies have investigated the contribution of hip abductor weakness to knee pain while considering the presence or absence of knee extensor muscle weakness.

Thus, the purpose of this study was to evaluate the longitudinal relation of hip abductor strength to worsened or new-onset knee pain. As knee extensor strength is known to have a sex-specific effect on knee pain, we analyzed this relationship in the entire cohort as well as separately for women and men. We also evaluated this relationship with and without stratifying by knee extensor strength, to investigate whether any relationship noted is specific to the hip abductor muscles or more generally a reflection of overall muscle weakness.

PATIENTS AND METHODS

Study sample. We used data from the Multicenter Osteoarthritis (MOST) study. The MOST study is a prospective, observational, National Institutes of Health–funded cohort study of risk factors for the incidence and progression of knee OA (23). The original cohort was enrolled between 2003 and 2005 at the Iowa City, Iowa and Birmingham, Alabama study sites. Eligible participants were age 50–79 years at the initial study visit and were at high risk of symptomatic knee OA, defined as either having knee pain, being overweight, or having a history of knee injury or surgery. At 144 months, participants from the original cohort without end-stage knee OA were invited to return for a study visit. This 144-month study visit for the original cohort was the baseline visit for this analysis. At this same time, a new cohort was recruited. Eligibility for the new cohort included individuals ages 45–69 years, without severe or constant knee pain, and with at most mild radiographic OA (all Kellgren/Lawrence [K/L] grades ≤2). Additional exclusion criteria for hip and knee strength testing were high blood pressure on the day of the examination, any history of a brain aneurysm, cerebral hemorrhage in the past 6 months, knee or hip replacement or back surgery in the previous 3 months, heart attack or cataract surgery in the past 6 weeks, or groin hernia that has not been operated on. This initial study visit for the new cohort was the baseline visit for this analysis. There was considerable overlap between the 2 cohorts, as many of the original cohort still did not have OA and many in the new cohort had mild OA. The MOST study was approved by the local institutional review board at each site, and all participants gave informed consent.

Hip abductor and knee extensor strength assessment. At the baseline visit (144-month visit for the original cohort, initial visit for the new cohort), participants had hip abductor and knee extensor muscle strength measured in each leg. Briefly, hip abductor and knee extensor strength were measured with the participant seated in the chair of a HUMAC NORM Testing and Rehabilitation System (Computer Sports Medicine). For assessment of isometric hip abductor strength, the chair was positioned with the seat back at approximately 15 degrees from horizontal. An additional leg rest extension was provided to position the knees in neutral extension. A custom-built device to measure force with a load cell (MLP-150, Transducer Techniques) was positioned at the level of the lateral femoral epicondyle. The distance between the greater trochanter and the location of the load cell was measured as an estimation of lever arm distance. Force was multiplied by this distance to obtain the torque measurement of hip abductor strength. A wide velcro strap was placed at the level of the anterior superior iliac spines and tightened snugly to stabilize the pelvis during testing. Legs were positioned shoulder width apart, with the toes pointing upward toward the ceiling. Contralateral to the hip being tested, a large padded block was secured to maintain the position of the contralateral leg. Individuals were instructed to push as hard as they could against a stationary padded bar positioned at the lateral femoral epicondyle of the knee. They maintained the contraction

for approximately 3 seconds before resting. The measurement was repeated 3 times, with 10 seconds between each trial. Participants were provided with the opportunity to practice at approximately 50% effort to become accustomed to the procedures before data were recorded at full effort. Verbal encouragement was provided during each trial. Test–retest reliability was assessed in 60 participants tested 2 times, approximately 7 days apart. The intraclass correlation coefficient (ICC) for hip abduction strength was 0.80 with a 95% confidence interval (95% CI) of ±0.09.

As a measure of knee extensor strength, we used the 1-repetition maximum isotonic knee strength obtained as part of the larger study using the torque motor of the HUMAC NORM testing system. Individuals were positioned within the HUMAC NORM with the seat back positioned at 85 degrees relative to horizontal with the thigh strapped to the seat. Participants were instructed to fully extend the knee from a position of approximately 90 degrees of flexion by pushing as hard and as fast as they could. Resistance was increased until the participant was unable to complete the full range of motion. The highest load participants were able to move through the full range of motion was considered their knee extensor strength. Participants were given 30 seconds of rest between repetitions to minimize muscle fatigue. The ICC for this measurement was 0.80 with an approximate 95% CI width of ±0.09.

Assessment of knee pain. Knee pain was assessed using the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) questionnaire (24). WOMAC was administered at the baseline visit for this analysis and then 8, 16, and 24 months later. Frequent knee pain was assessed by asking the question, "Have you had knee pain, aching, or stiffness on most of the last 30 days?" We then characterized 2 knee pain outcomes. First, using the WOMAC pain subscale, we defined worsened as an increase in pain score by at least 2 on the 0–20 scale (25). We characterized each knee as having worsened pain if reported pain was worse than baseline for at least 2 of the 3 follow-ups. Among those who did not report frequent knee pain at baseline, we characterized the knee as having new frequent knee pain if they responded yes to the frequent knee pain question for at least 2 of the 3 follow-ups. Knees with WOMAC pain >18 or with frequent knee pain at baseline were also excluded from the individual analyses.

Statistical analysis. We carried out leg-specific analyses (i.e., 1 participant could contribute 2 observations) using logistic regression to assess hip abductor weakness (sex-specific quartiles) as a risk factor for ipsilateral worsened (WOMAC) and new-onset frequent knee pain, accounting for the correlation between limbs with generalized estimating equations and adjusting for age, sex, body mass index, race (White versus non-White), depressive symptoms (using the Center for Epidemiological Studies Depression Scale score >15) (26,27), and radiographic tibiofemoral OA (K/L grade ≥2) in that knee. The baseline WOMAC pain score was also accounted for as a continuous variable in the WOMAC pain worsening analyses. Analyses were first conducted in the entire cohort and then separately for women and men since sex-specific differences in the effect of muscle weakness on OA outcomes have been reported (1–4). Participants were excluded from analyses if they were missing outcome measurements or covariates. All analyses were conducted in SAS 9.4, with an alpha of 0.05, 2-sided.

We were also interested in the relation of hip abductor strength to knee pain while accounting for knee extensor strength. Because hip abductor strength and knee extensor strength are highly correlated in our sample (r = 0.7), including both in the same model has limitations. Instead, we stratified our analysis by knee extensor strength (high versus low using the median value, which was 37 Nm for women and 60 Nm for men). This approach allowed us to evaluate the relation of hip abductor strength to pain in those with stronger knee extensors (greater than the median) and those with weaker knee extensors (less than the median).

Analyses of lower-extremity muscle strength often divide the strength values by body mass and may divide by a measure of height as well. Normalizing data by dividing by body mass and/or height introduces an interaction without separately assessing the variables in the interaction (28–30). Therefore, we choose not to normalize strength values.

RESULTS

Of the 3,447 participants in the MOST study, participants or knees were removed from analysis for multiple reasons. A total of 1,058 participants were removed for not having hip abductor force; these were due to exclusion factors or strength testing equipment not being available. In total, 153 participants had missing or inconsistent lever arm (i.e., thigh length) measures. After removing ineligible knees, the remaining knees with missing covariates were removed; the majority of these were missing knee extensor strength (Figure 1). A total of 2,167 and 2,028 participants contributed 4,142 and 3,993 knees, respectively, to the WOMAC pain and frequent knee pain analyses. Characteristics of the participants are in Table 1.

In the full cohort, compared to the highest (strongest) quartile of hip abductor strength, the lowest quartile (weakest) was associated with 1.5 (95% CI 1.1–2.0) times greater odds of worsened knee pain (Table 2). There was also a significant linear trend across quartiles of hip abductor strength (P = 0.01). There was no relation between hip abductor strength and incident frequent knee pain in the full cohort.

In sex-specific analyses, among women compared to the highest (strongest) quartile of hip abductor strength, the lowest

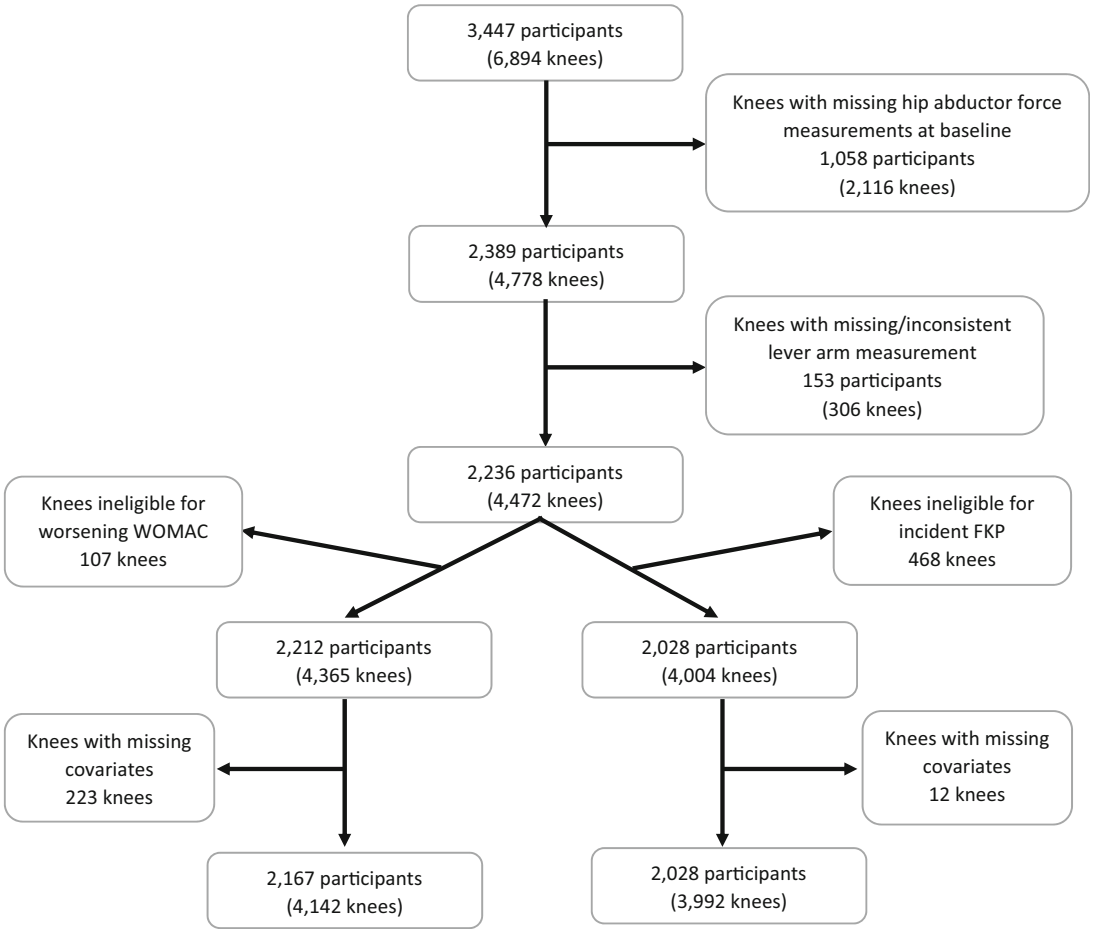

Figure 1. Flow chart depicting derivation of participants included in the analysis for worsened Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) knee pain and for new-onset frequent knee pain (FKP).

quartile (weakest) was associated with 1.7 (95% CI 1.1–2.6) times greater odds of worsened knee pain (Table 3). There was also a significant linear trend across quartiles of hip abductor strength (P = 0.02). There was no relation between hip abductor strength and incident frequent knee pain in women. In men, there was no relation to either pain outcome (Table 4).

In participants with greater than the median knee extensor strength, compared to the highest (strongest) quartile of hip abductor strength, the lowest quartile (weakest) was associated with 1.5 (95% CI 1.0–2.4) times greater odds of worsened knee pain (Table 5) (P for linear trend = 0.04). In women with high knee extensor strength, compared to the highest (strongest) quartile of

Table 1. Cohort characteristics by sex*

	WOMAC sample		Frequent knee pain sample	
	Women (n = 2,332 knees)	Men (n = 1,810 knees)	Women (n = 2,221 knees)	Men (n = 1,771 knees)
Age, years	62.3 ± 9.9	62.0 ± 10.0	61.8 ± 9.8	61.6 ± 9.8
Body mass index, kg/m^2	28.6 ± 5.8	29.7 ± 4.9	28.5 ± 5.7	29.7 ± 4.8
Depressive symptoms, %	12.4	10	11.8	9.7
White, %	77.2	82.1	77.6	82
Radiographic tibiofemoral OA, %	28.1	21.8	30.2	23.1
Hip abductor strength, Nm	73.26 ± 24.21	114.51 ± 34.50	74.20 ± 24.03	114.80 ± 34.49
Knee extensor strength, Nm	41.47 ± 13.59	67.10 ± 22.75	41.94 ± 13.61	67.64 ± 22.8
WOMAC pain at baseline	2.2 ± 2.9	1.8 ± 2.6	2.1 ± 2.9	1.7 ± 2.5
WOMAC pain worsened, %	13.8	13.5	12.8	13
New frequent knee pain, %	6.5	6.2	6.5	6.2

* Values are the mean ± SD unless indicated otherwise. OA = osteoarthritis; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index.

Table 2. Relation of quartiles of hip abductor strength to knee pain*

	Incident frequent knee pain (n = 3,992; 255 events)		WOMAC pain worsening (n = 4,142; 567 events)	
Hip abductor strength quartiles	Frequency of outcome	OR (95% CI)†	Frequency of outcome	OR (95% CI)†
Q1 (weakest) Women: 18.45–55.55 Men: 27.24–91.00	82/997 (8.2)	1.2 (0.8–1.8)	189/1,035 (18.3)	1.5 (1.1–2.0)
Q2 Women: 55.56–72.36 Men: 91.01–110.95	59/998 (5.9)	1.0 (0.7–1.4)	133/1,036 (12.8)	1.1 (0.8–1.5)
Q3 Women: 72.37–88.23 Men: 110.96–135.87	59/999 (5.9)	1.0 (0.7–1.4)	126/1,037 (12.2)	1.1 (0.8–1.4)
Q4 (strongest) Women: 88.24–164.20 Men: 135.88–266.50	55/998 (5.5)	1.0 (Ref.)	119/1,035 (11.5)	1.0 (Ref.)
P for trend	–	0.5	–	0.01

* Values are the number/total number (%) unless indicated otherwise. 95% CI = 95% confidence interval; OR = odds ratio; Ref. = reference; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index.
† OR adjusted for age, sex, race, body mass index, depressive symptoms, radiographic tibiofemoral osteoarthritis, and baseline WOMAC pain score (for WOMAC analysis only).

hip abductor strength, the lowest quartile (weakest) was associated with 2.0 (95% CI 1.1–3.5) times greater odds of worsened knee pain (Table 5) (*P* for linear trend = 0.03). There was no relation in women with either high or low knee extensor strength to incident frequent knee pain. In men, there was no relation in either the high or low knee extensor strength group to either pain outcome.

DISCUSSION

This study evaluated the relation of hip abductor strength to worsening or new onset of frequent knee pain, with and without stratifying by knee extensor strength. In women, hip abductor weakness was strongly associated with WOMAC knee pain worsening. When stratifying by knee extensor strength, this effect persisted in women with stronger, but not in those with weaker, knee extensor muscles. In men, however, there was no association between hip abductor strength and WOMAC knee pain worsening. In neither women nor men was hip abductor strength associated with incidence of frequent knee pain. Overall, we found differences in effect between women and men, highlighting the importance of using sex-stratified analyses, especially in musculoskeletal conditions with known sex differences.

Our hip abductor strength values and distributions were generally consistent with those in the literature (22,31). Our values for knee extensor strength were lower than in the literature and lower than our measure of hip abductor strength. This finding is likely attributable to differences in methods. We used an isotonic 1-repetition maximum measurement for knee extensor strength that may result in slightly lower values than would have been obtained with an isometric measurement. As we do not compare between hip abductor and knee extensor strength, the measurement differences are unlikely to impact our findings.

Hip abductor weakness has been hypothesized to increase contralateral pelvic drop, a movement thought to contribute to knee joint loading and pain (11,12). This presumed mechanism whereby abductor weakness increases the knee adduction moment during walking, a primary variable of joint loading, has not been supported (13–15). While hip abductor weakness has been correlated with peak pelvic drop (32), strength alone does

Table 3. Relation of hip abductor strength to knee pain in women*

	Incident frequent knee pain (n = 2,221; 145 events)		WOMAC pain worsening (n = 2,332; 322 events)	
Hip abductor strength quartiles	Frequency of outcome	OR (95% CI)†	Frequency of outcome	OR (95% CI)†
Q1 (weakest): 18.45–55.55	49/555 (8.8)	1.2 (0.7–2.1)	111/583 (19.0)	1.7 (1.1–2.6)
Q2: 55.56–72.36	39/555 (7.0)	1.1 (0.6–1.8)	83/583 (14.2)	1.4 (0.9–2.1)
Q3: 72.37–88.23	27/556 (4.9)	0.9 (0.5–1.5)	71/584 (12.2)	1.3 (0.8–1.9)
Q4 (strongest): 88.24–164.20	30/555 (5.4)	1.0 (Ref.)	57/583 (9.8)	1.0 (Ref.)
P for trend	–	0.5	–	0.02

* Values are the number/total number (%) unless indicated otherwise. 95% CI = 95% confidence interval; OR = odds ratio; Ref. = reference; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index.
† OR adjusted for age, sex, race, body mass index, depressive symptoms, radiographic tibiofemoral osteoarthritis, and baseline WOMAC pain score (for WOMAC analysis only).

Table 4. Relation of hip abductor strength to knee pain in men*

	Incident frequent knee pain (n = 1,771; 110 events)		WOMAC pain worsening (n = 1,810; 245 events)	
Hip abductor strength quartiles	Frequency of outcome	OR (95% CI)†	Frequency of outcome	OR (95% CI)†
Q1 (weakest): 27.24–91.00	33/442 (7.5)	1.1 (0.6–2.1)	78/452 (17.3)	1.3 (0.8–1.0)
Q2: 91.01–110.95	20/443 (4.5)	0.8 (0.4–1.4)	50/453 (11.0)	0.9 (0.6–1.3)
Q3: 110.96–135.87	32/443 (7.2)	1.0 (0.6–1.8)	55/453 (12.1)	0.8 (0.6–1.3)
Q4 (strongest): 135.88–266.50	25/443 (5.6)	1.0 (Ref.)	62/452 (13.7)	1.0 (Ref.)
P for trend	–	0.99	–	0.3

* Values are the number/total number (%) unless indicated otherwise. 95% CI = 95% confidence interval; OR = odds ratio; Ref. = reference; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index.
† OR adjusted for age, sex, race, body mass index, depressive symptoms, radiographic tibiofemoral osteoarthritis, and baseline WOMAC pain score (for WOMAC analysis only).

not determine movement patterns. With our data, we were unable to determine how hip abductor weakness affected knee adduction moment or movement patterns in our cohort.

Hip abductor strength is associated with WOMAC physical function (33) and does contribute uniquely to the explained variance in performance-based measures of physical function, even after accounting for knee extensor strength (17,34). Hip abductor strength has also been related to turning speed (35), a metric not typically captured during straight walking, providing a potential explanation for why hip abductor strength affects function without changing knee adduction moment during walking. Similarly, interventions that target hip abductor strength have been effective in reducing symptoms and improving function (16,18,19).

While hip abductor weakness was strongly associated with WOMAC knee pain worsening in women, there was no relation to incident frequent knee pain, suggesting that these 2 measures capture different aspects of the pain experience. The WOMAC pain subscale focuses on the difficulty caused by pain during different activities and at rest; frequent knee pain focuses on frequency of knee pain. An individual could have increased difficulty with activities due to pain (worsened WOMAC pain), without having pain more often (incident frequent knee pain). In our sample, we had twice as many individuals experience worsened WOMAC pain than individuals who had incident frequent knee pain.

The sex difference in the effect of hip abductor strength is consistent with findings of Chang et al (22). Hip abductor strength may be more important for women due to their structure, movement patterns, and/or an increased need for muscle to stabilize joints. For example, the female pelvis is wider and broader than the male pelvis (36–38); women typically have more hip anteversion and knee valgus than men (39). During walking, women tend to maintain the pelvis in a more anterior pelvic tilt (40) and have greater excursion of the pelvis in the frontal (41–43) and transverse planes (42–44). Women may be more flexible than men (45), suggesting more reliance on muscle strength for stability, although data on this possibility are not consistent (46). The combined effect of these differences may indicate both a need for greater muscle strength due to structural differences and a greater reliance on muscle strength for joint stability in women than in men.

The relationship of hip abductor weakness to worsening knee pain was affected by knee extensor strength in women. In the group with low knee extensor strength, there was no relation between hip abductor strength and pain; however, in the group with high knee extensor strength, individuals with hip abductor weakness were at an increased risk of worsened knee pain. Chang et al noted a similar effect of hip abductor strength on structural damage and function (22). While our findings suggest that hip abductor strength may be uniquely important in those with strong knee extensors, risk factors for disease are often better detected in those who have few other risk factors, a concept underlying the rich contributions of groups with particular dietary

Table 5. Relation of hip abductor strength quartiles to knee pain in those with knee extensors strength greater than median: WOMAC pain worsening*

	Full cohort (n = 2,284; 277 events)		Women only (n = 1,361; 155 events)	
Hip abductor strength quartiles	Frequency of outcome	OR (95% CI)†	Frequency of outcome	OR (95% CI)†
Q1 (weakest)	45/235 (19.2)	1.5 (1.0–2.4)	28/140 (20.0)	2.0 (1.1–3.5)
Q2	62/490 (12.7)	1.2 (0.8–1.7)	40/304 (13.2)	1.3 (0.8–2.2)
Q3	77/681 (11.3)	1.0 (0.7–1.5)	42/402 (10.5)	1.1 (0.7–1.9)
Q4 (strongest)	93/878 (10.6)	1.0 (Ref.)	45/515 (8.7)	1.0 (Ref.)
P for trend	–	0.04	–	0.03

* Values are the number/total number (%) unless indicated otherwise. 95% CI = 95% confidence interval; OR = odds ratio; Ref. = reference; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index.
† OR adjusted for age, sex, race, body mass index, depressive symptoms, radiographic tibiofemoral osteoarthritis, and baseline WOMAC pain score. Median knee extensor values were 37 Nm for women and 60 Nm for men.

habits (e.g., Seventh-day Adventists) to our understanding of risk factors for heart disease (47,48). Those with strong knee extensors constitute those without a major risk factor for knee pain. Even so, our findings highlight the fact that, while knee extensor strength is necessary, it may not be sufficient. Intervention programs targeted at reducing the risk of knee pain should address both hip abduction and knee extension.

The study findings conducted in individuals with or at risk of knee OA are generalizable to similar populations, but may not generalize to a younger cohort. Similarly, the use of study-specific quartiles instead of predefined cut points may limit generalizability to other study populations. Other limitations exist. The impact of hip abductor strength on knee pain may depend on the location of knee pain; subclassifying the individuals based on knee pain location may demonstrate different relationships. The use of isometric hip abductor strength measured in a single hip position may not fully capture hip strength as used in functional activities, especially our selected position of neutral hip rotation. Knee extensor strength was measured isotonically, not isometrically, potentially influencing our results and interpretation; the motion captured also may not reflect strength during function, given the seated position. The WOMAC threshold used to define worsened may not capture all worsening experiences. The limited number of cases, especially for incident frequent knee pain, may have reduced our ability to detect small-to-moderate associations.

Hip abductor weakness is associated with worsening knee pain in women, but not men, in this cohort. Hip abductor weakness was not associated with incident frequent knee pain in either women or men. A focus on strength beyond that of knee extensors alone may provide a comprehensive understanding of how lower extremity muscle strength affects knee pain.

AUTHOR CONTRIBUTIONS

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

Study conception and design. C. L. Lewis, Segal, LaValley, Williams, Nevitt, C. E. Lewis, Felson, Stefanik.

Acquisition of data. C. L. Lewis, Segal, Williams, Nevitt, C. E. Lewis, Felson, Stefanik.

Analysis and interpretation of data. C. L. Lewis, Segal, Rabasa, LaValley, Williams, Nevitt, C. E. Lewis, Felson, Stefanik.

REFERENCES

1. Slemenda C, Heilman DK, Brandt KD, et al. Reduced quadriceps strength relative to body weight: a risk factor for knee osteoarthritis in women? Arthritis Rheum 1998;41:1951–9.
2. Bacon KL, Segal NA, Øiestad BE, et al. Thresholds in the relationship of quadriceps strength with functional limitations in women with knee osteoarthritis. Arthritis Care Res (Hoboken) 2019;71:1186–93.
3. Culvenor AG, Segal NA, Guermazi A, et al. Sex-specific influence of quadriceps weakness on worsening patellofemoral and tibiofemoral cartilage damage: a prospective cohort study. Arthritis Care Res (Hoboken) 2019;71:1360–5.
4. Segal NA, Glass NA, Torner J, et al. Quadriceps weakness predicts risk for knee joint space narrowing in women in the MOST cohort. Osteoarthritis Cartilage 2010;18:769–75.
5. Culvenor AG, Ruhdorfer A, Juhl C, et al. Knee extensor strength and risk of structural, symptomatic, and functional decline in knee osteoarthritis: a systematic review and meta-analysis. Arthritis Care Res (Hoboken) 2017;69:649–58.
6. Øiestad BE, Juhl CB, Culvenor AG, et al. Knee extensor muscle weakness is a risk factor for the development of knee osteoarthritis: an updated systematic review and meta-analysis including 46,819 men and women [review]. Br J Sports Med 2022;56:349–55.
7. Hinman RS, Hunt MA, Creaby MW, et al. Hip muscle weakness in individuals with medial knee osteoarthritis. Arthritis Care Res (Hoboken) 2010;62:1190–3.
8. Deasy M, Leahy E, Semciw AI. Hip strength deficits in people with symptomatic knee osteoarthritis: a systematic review with meta-analysis. J Orthop Sports Phys Ther 2016;46:629–39.
9. Costa RA, de Oliveira LM, Watanabe SH, et al. Isokinetic assessment of the hip muscles in patients with osteoarthritis of the knee. Clinics (Sao Paulo) 2010;65:1253–9.
10. Ackland DC, Denton M, Schache AG, et al. Hip abductor muscle volumes are smaller in individuals affected by patellofemoral joint osteoarthritis. Osteoarthritis Cartilage 2019;27:266–72.
11. Mundermann A, Dyrby CO, Andriacchi TP. Secondary gait changes in patients with medial compartment knee osteoarthritis: increased load at the ankle, knee, and hip during walking. Arthritis Rheum 2005;52:2835–44.
12. Chang A, Hayes K, Dunlop D, et al. Hip abduction moment and protection against medial tibiofemoral osteoarthritis progression. Arthritis Rheum 2005;52:3515–9.
13. Rutherford DJ, Hubley-Kozey C, Stanish W. Hip abductor function in individuals with medial knee osteoarthritis: implications for medial compartment loading during gait. Clin Biomech (Bristol, Avon) 2014;29:545–50.
14. Lewinson RT, Worobets JT, Stefanyshyn DJ. The relationship between maximal hip abductor strength and resultant loading at the knee during walking. Proc Inst Mech Eng H 2014;228:1258–63.
15. Kean CO, Bennell KL, Wrigley TV, et al. Relationship between hip abductor strength and external hip and knee adduction moments in medial knee osteoarthritis. Clin Biomech (Bristol, Avon) 2015;30:226–30.
16. Loyd BJ, Jennings JM, Judd DL, et al. Influence of hip abductor strength on functional outcomes before and after total knee arthroplasty: post hoc analysis of a randomized controlled trial. Phys Ther 2017;97:896–903.
17. Alnahdi AH, Zeni JA, Snyder-Mackler L. Hip abductor strength reliability and association with physical function after unilateral total knee arthroplasty: a cross-sectional study. Phys Ther 2014;94:1154–62.
18. Sled EA, Khoja L, Deluzio KJ, et al. Effect of a home program of hip abductor exercises on knee joint loading, strength, function, and pain in people with knee osteoarthritis: a clinical trial. Phys Ther 2010;90:895–904.
19. Bennell KL, Hunt MA, Wrigley TV, et al. Hip strengthening reduces symptoms but not knee load in people with medial knee osteoarthritis and varus malalignment: a randomised controlled trial. Osteoarthritis Cartilage 2010;18:621–8.
20. Yuenyongviwat V, Duangmanee S, Iamthanaporn K, et al. Effect of hip abductor strengthening exercises in knee osteoarthritis: a randomized controlled trial. BMC Musculoskelet Disord 2020;21:284.

21. Thomas DT, Shruthi R, Prabhakar AJ, et al. Hip abductor strengthening in patients diagnosed with knee osteoarthritis: a systematic review and meta-analysis. BMC Musculoskelet Disord 2022;23:622.
22. Chang AH, Chmiel JS, Almagor O, et al. Hip muscle strength and protection against structural worsening and poor function and disability outcomes in knee osteoarthritis. Osteoarthritis Cartilage 2019;27: 885–94.
23. Segal NA, Nevitt MC, Gross KD, et al. The Multicenter Osteoarthritis Study: opportunities for rehabilitation research. PM R 2013;5:647–54.
24. Bellamy N, Buchanan WW, Goldsmith CH, et al. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833–40.
25. Angst F, Benz T, Lehmann S, et al. Multidimensional minimal clinically important differences in knee osteoarthritis after comprehensive rehabilitation: a prospective evaluation from the Bad Zurzach Osteoarthritis Study. RMD Open 2018;4:e000685.
26. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977;1: 385–401.
27. Smarr KL, Keefer AL. Measures of depression and depressive symptoms: Beck Depression Inventory-II (BDI-II), Center for Epidemiologic Studies Depression Scale (CES-D), Geriatric Depression Scale (GDS), Hospital Anxiety And Depression Scale (HADS), and Patient Health Questionnaire-9 (PHQ-9) [review]. Arthritis Care Res (Hoboken) 2011;63 Suppl 11:S454–66.
28. Curran-Everett D. Explorations in statistics: the analysis of ratios and normalized data. Adv Physiol Educ 2013;37:213–9.
29. Packard GC, Boardman TJ. The use of percentages and size-specific indices to normalize physiological data for variation in body size: wasted time, wasted effort? Comp Biochem Physiol A Physiol 1999; 122:37–44.
30. Kronmal RA. Spurious correlation and the fallacy of the ratio standard revisited. J R Statist Soc A 1993;156:379–92.
31. Tevald MA, Murray A, Luc BA, et al. Hip abductor strength in people with knee osteoarthritis: a cross-sectional study of reliability and association with function. Knee 2016;23:57–62.
32. Park SK, Kobsar D, Ferber R. Relationship between lower limb muscle strength, self-reported pain and function, and frontal plane gait kinematics in knee osteoarthritis. Clin Biomech (Bristol, Avon) 2016; 38:68–74.
33. Selistre LF, Goncalves GH, Nakagawa TH, et al. The role of hip abductor strength on the frontal plane of gait in subjects with medial knee osteoarthritis. Physiother Res Int 2019;24:e1779.
34. Piva SR, Teixeira PE, Almeida GJ, et al. Contribution of hip abductor strength to physical function in patients with total knee arthroplasty. Phys Ther 2011;91:225–33.
35. Iijima H, Yorozu A, Suzuki Y, et al. Hip abductor muscle weakness and slowed turning motion in people with knee osteoarthritis. J Biomech 2020;101:109652.
36. Lewis CL, Laudicina NM, Khuu A, et al. The human pelvis: variation in structure and function during gait [review]. Anat Rec (Hoboken) 2017; 300:633–42.
37. LaVelle M. Natural selection and developmental sexual variation in the human pelvis. Am J Phys Anthropol 1995;98:59–72.
38. Tague RG. Variation in pelvic size between males and females. Am J Phys Anthropol 1989;80:59–71.
39. Nguyen A, Shultz SJ. Sex differences in clinical measures of lower extremity alignment. J Orthop Sports Phys Ther 2007;37:389–98.
40. Cho SH, Park JM, Kwon OY. Gender differences in three dimensional gait analysis data from 98 healthy Korean adults. Clin Biomech (Bristol, Avon) 2004;19:145–52.
41. Smith LK, Lelas JL, Kerrigan DC. Gender differences in pelvic motions and center of mass displacement during walking: stereotypes quantified. J Womens Health Gend Based Med 2002;11:453–8.
42. Bruening DA, Frimenko RE, Goodyear CD, et al. Sex differences in whole body gait kinematics at preferred speeds. Gait Posture 2015; 41:540–5.
43. Chumanov ES, Wall-Scheffler C, Heiderscheit BC. Gender differences in walking and running on level and inclined surfaces. Clin Biomech (Bristol, Avon) 2008;23:1260–8.
44. Crosbie J, Vachalathiti R, Smith R. Age, gender and speed effects on spinal kinematics during walking. Gait Posture 1997;5:13–20.
45. Soucie JM, Wang C, Forsyth A, et al. Range of motion measurements: reference values and a database for comparison studies. Haemophilia 2011;17:500–7.
46. McKay MJ, Baldwin JN, Ferreira P, et al. Normative reference values for strength and flexibility of 1,000 children and adults. Neurology 2017;88:36–43.
47. Phillips RL, Lemon FR, Beeson WL, et al. Coronary heart disease mortality among Seventh-day Adventists with differing dietary habits: a preliminary report. Am J Clin Nutr 1978;31 Suppl:S191–8.
48. Fraser GE. Diet as primordial prevention in Seventh-day Adventists [review]. Prev Med 1999;29:S18–23.

Arthritis Care & Research
Vol. 75, No. 11, November 2023, pp 2336–2344
DOI 10.1002/acr.25163

American College of Rheumatology
Empowering Rheumatology Professionals

Examination of the Increased Risk for Falls Among Individuals With Knee Osteoarthritis: A Canadian Longitudinal Study on Aging Population-Based Study

Jessica M. Wilfong, Anthony V. Perruccio, and Elizabeth M. Badley

Objective. To characterize the profile of individuals with and without knee osteoarthritis (OA) who fell, and to identify factors contributing to an individual with knee OA experiencing 1 or multiple injurious falls.

Methods. Data are from the baseline and 3-year follow-up questionnaires of the Canadian Longitudinal Study on Aging, a population-based study of people ages 45–85 years at baseline. Analyses were limited to individuals either reporting knee OA or no arthritis at baseline (n = 21,710). Differences between falling patterns among those with and without knee OA were tested using chi-square tests and multivariable-adjusted logistic regression models. An ordinal logistic regression model examined predictors of experiencing 1 or more injurious falls among individuals with knee OA.

Results. Among individuals reporting knee OA, 10% reported 1 or more injurious falls; 6% reported 1 fall, and 4% reported 2+ falls. Having knee OA significantly contributed to the risk of falling (odds ratio [OR] 1.33 [95% confidence interval (95% CI) 1.14–1.56]), and individuals with knee OA were more likely to report having a fall indoors while standing or walking. Among individuals with knee OA, reporting a previous fall (OR 1.75 [95% CI 1.22–2.52]), previous fracture (OR 1.42 [95% CI 1.12–1.80]), and having urinary incontinence (OR 1.38 [95% CI 1.01–1.88]) were significant predictors of falling.

Conclusion. Our findings support the idea that knee OA is an independent risk factor for falls. The circumstances in which falls occur differ from those for individuals without knee OA. The risk factors and environments that are associated with falling may provide opportunities for clinical intervention and fall prevention strategies.

INTRODUCTION

Falls are a leading cause of morbidity and mortality in older adults and pose a major public health concern. Estimates in the 2017 Global Burden of Diseases Study ranked falls as the eighteenth leading cause of disability-adjusted life years and the second leading cause of death due to unintentional injuries (1,2). Recurrent fallers, those who fall 2 or more times per year, experience greater morbidity than those who are not recurrent fallers (3).

Knee osteoarthritis (OA) is a leading cause of disability and a known risk factor for falls (4–10). Data from a representative national sample in the US showed that older adults with arthritis are at an increased risk of fall-related injuries, and are more than twice as likely to have recurrent falls compared to people without any arthritis (7). Another study found that older adults with knee OA had an increased risk of recurrent falls regardless of the severity of their OA (10). The incidence of both falls and knee OA are predicted to increase with the aging population (11). Despite this prediction, factors that contribute to the risk of falling among people with knee OA are not yet well understood. A recent systematic review examined the risk factors for falls in adults with knee OA and found that only limited to moderate evidence is available (12). Within the current body of literature examining risk factors for falling among people with knee OA, evidence is lacking for many of the known risk factors among the general population,

The opinions expressed in this article are the authors' own and do not reflect the views of the Canadian Longitudinal Study on Aging.

Funding for the Canadian Longitudinal Study on Aging is provided by the Government of Canada through the Canadian Institutes of Health Research (grant LSA 94473) and the Canada Foundation for Innovation, as well as the following provinces: Newfoundland, Nova Scotia, Quebec, Ontario, Manitoba, Alberta, and British Columbia. Ms. Wilfong and Dr. Perruccio's work was supported by the Arthritis Society Canada (collaborative service agreement and grant STAR-20-0000000012, respectively).

Jessica M. Wilfong, MPH, Anthony V. Perruccio, PhD, Elizabeth M. Badley, DPhil: Schroeder Arthritis Institute, University Health Network, Dalla Lana School of Public Health, University of Toronto, and Arthritis Community Research and Epidemiology Unit, Toronto, Ontario, Canada.

Author disclosures are available online at https://onlinelibrary.wiley.com/doi/10.1002/acr.25163.

Address correspondence via email to Anthony V. Perruccio, PhD, at anthony.perruccio@uhnresearch.ca.

Submitted for publication December 23, 2022; accepted in revised form May 18, 2023.

SIGNIFICANCE & INNOVATIONS

- We confirmed knee osteoarthritis (OA) as an independent risk factor for experiencing 1 or more injurious falls compared to those without knee OA.
- It is important to consider environmental factors when assessing the risk of falling among individuals with knee OA. Those with knee OA were much more likely to report having an injurious fall indoors versus outdoors, or while standing or walking indoors compared to those without knee OA.
- Urinary incontinence was found to be a unique risk factor that was associated with a risk of falling for those with knee OA. That 20% of fallers with knee OA reported this condition suggests the risk of falling is likely to be increased by the limitations in mobility, namely poor balance, associated with knee OA.

as well as the identification of risk factor profiles, such as those who are recurrent fallers versus those who are not.

The etiology of falls is multifactorial. They can be caused by intrinsic factors (person-related), extrinsic factors (environment-related), or behavioral factors (activity-related). Many falls result from a complex interaction of risk factors, so that identifying those most at risk is difficult (13). For recurrent fallers, intrinsic risk factors play a greater role than extrinsic or behavioral factors (14). The most common intrinsic factors reported are advanced age, impaired balance and gait, visual impairment, and the presence of certain diseases (13,15,16).

The purpose of the current study was to characterize the profile of risk factors for falling among those with knee OA and those without knee OA, and to examine where and how a fall occurred using a representative sample of the Canadian population. Additionally, we sought to identify intrinsic factors that contributed to an individual with knee OA experiencing 1 or multiple injurious falls.

MATERIALS AND METHODS

Data were obtained from the baseline and 3-year follow-up of the Canadian Longitudinal Study on Aging (CLSA), collected between September 2011 and December 2015. The CLSA is a longitudinal study that will follow approximately 50,000 community-dwelling Canadians ages 45–85 years over a period of 20 years or until death to gain insight into the development of disease and disability throughout the aging process. The design and recruitment of the 2 cohorts comprising this study has been described fully elsewhere (17). The comprehensive cohort, used for the present study, consists of a community-based sample of 30,097 individuals who reside within a 50 km radius of 11 data collection sites (Victoria, Vancouver, Surrey, Calgary, Winnipeg, Hamilton, Ottawa, Montreal, Sherbrooke, Halifax, and St. John's) across 7 Canadian provinces. Participants provided information through questionnaires and through physical examinations performed at the data collection sites. Participants were excluded if they could not communicate in English or French, had a cognitive impairment at the time of contact, were a full-time member of the Canadian Armed Forces, were a resident in a long-term care institution, or were living on reserves or other Aboriginal settlements. Additionally, no proxy responses were allowed at baseline, but were available as an option for those who screened positive for a cognitive impairment at follow-up. None of the participants in the present study used a proxy respondent. The present study received ethics approval from the University Health Network's Research Ethics Board.

Analytic sample. The Comprehensive cohort of the CLSA included 30,097 individuals at baseline. Individuals were asked at baseline if they have knee OA (yes/no). They were also asked if they have hand OA (yes/no), hip OA (yes/no), rheumatoid arthritis (yes/no), or other arthritis (yes/no). Those indicating any of the additional types of arthritis and no knee OA were excluded from further consideration in analyses. This exclusion left an analytic sample of 21,710 individuals with and without knee OA.

Outcome. The outcome of interest was whether an individual reported an injurious fall in the year leading up to the follow-up questionnaire. Falls were determined from questions asking participants whether they had any injuries that were serious enough to limit some of their normal activities in the last 12 months and asking whether any of their injuries were caused by a fall. Those who responded yes to both were then asked how many times they had fallen in the past 12 months. We categorized the number of falls as none (0), 1, or multiple (2+).

Where and how the fall occurred. Participants who reported experiencing a fall in the 12 months leading up to the follow-up questionnaire were asked where the fall that resulted in the most serious injury or problem occurred. The response options were: 1) inside of your home, 2) outside of your home, but inside a building, and 3) outdoors. We collapsed the former 2 groups to compare indoors versus outdoors. Those who responded that their fall occurred indoors were then asked how their fall occurred. For analysis we categorized the response options into 4 groups: fell while standing or walking, fell on stairs or steps, fell while exercising (except walking), and other (which included fell from height of >1 meter or 3 feet, fell from furniture, fell while getting in or out of the bathtub, or fell while getting in or out of the shower). For those who responded that their fall occurred outdoors the response options were similar, except the "other" category excluded response options such as "fell

while getting in or out of the bathtub" and included additional options such as "fell on snow or ice."

Baseline sociodemographic and health variables. The occurrence of an injurious fall in the year preceding baseline was determined in the same way as stated above for the outcome. Knee OA was determined from a positive response to the question "Has a doctor ever told you that you have osteoarthritis in the knee?" Age was categorized into 10-year intervals (45–54, 55–64, 65–74, and 75–85). Body mass index (BMI; weight/height [kg/m^2]) was calculated from measured height and weight and then categorized as under/normal weight (<24.9), overweight (25–29.9), or obese (≥30). Respondents were also asked how often in the past 12 months they drank alcohol, which we categorized as never, ≤1/week, 2–5/week, and 6+/week. Respondents were considered to have knee symptoms if they responded yes to having had knee pain on most days, or knee pain while climbing down stairs or walking down slopes, or knee swelling in the past 4 weeks. Respondents were determined to have a lower-extremity fracture if they reported ever having suffered a break or fracture in their leg, ankle, foot/toes, hip, or knee. Participants were asked to self-rate their vision, and responses were dichotomized as those with vision problems (fair or poor) and those without vision problems (excellent or very good or good).

Chronic conditions. Specific chronic conditions and groups of conditions that are known to be associated with falls were assessed as separate predictors. These included respiratory conditions (chronic obstructive pulmonary disease, asthma), cardiovascular disease (CVD; heart disease, stroke, transient ischemic attack), urinary incontinence, neurologic conditions (Parkinson's disease, multiple sclerosis, epilepsy), diabetes mellitus or taking medication for diabetes mellitus, high blood pressure or taking medication for high blood pressure, and depression or taking medication for depression. All chronic conditions were self-reported as diagnosed by a doctor.

Performance-based tests. Three performance-based tasks were considered in this study, the 1-leg standing balance test, the timed-up-and-go (TUG), and the chair rise test. For the standing balance test, participants began by standing 1 meter from a wall, facing the wall. The time (in seconds) that they were able to balance on 1 foot before their foot touches the ground or they lose balance and touch the wall was recorded with a maximum time of 60 seconds. A cutoff based on previous work published with this data set was chosen to be 4.5 seconds (18); a time of 4.5 seconds or less indicates impaired balance. The TUG was used to assess mobility. Participants began by sitting back in a standard chair with armrests. The time (in seconds) was recorded for how long it takes the participant to stand up from the chair, walk 3 meters, turn around, walk back to the chair and sit down again at their normal pace. A cutoff of 14.2 seconds was used (18); a time of 14.2 seconds or more indicates impaired mobility. The chair-rise test assesses balance and coordination. Participants began by sitting back in a chair without armrests with their arms crossed over their chest. The time (in seconds) was recorded for how long it takes the participant to stand up and sit back down 5 times as quickly as possible with no rest in between. A cutoff of 15.9 seconds was used (18); a time of 15.9 seconds or more indicates impaired balance and coordination.

Statistical analysis. In the full analytic sample (n = 21,710), chi-square tests and observation of nonoverlapping 95% confidence intervals (95% CIs) were used to compare the number of falls between those with and without knee OA as well as baseline characteristics between those who did and did not fall within those with knee OA and those without knee OA. Logistic regression models were used to calculate odds ratios (ORs) with 95% CIs to determine whether knee OA was independently associated with falling, and then separately among those with knee OA and those without knee OA to identify the pattern of risk factors of falling. The outcome for all models was reporting an injurious fall, with not having experienced a fall as the referent group.

Investigating only those individuals who experienced a fall (n = 1,399), descriptive analysis was used to examine differences between the location and circumstance of the experienced fall among those with and without knee OA. Differences between groups were assessed using chi-square tests.

Finally, investigating only among those individuals with knee OA (n = 4,112), descriptive analysis with chi-square tests was used to compare risk factors across groups based on the number of injurious falls (0, 1, or multiple). An ordinal logistic regression was used to profile risk factors for the number of reported falls (0, 1, or multiple falls [cumulated over lower values]). Finding that the proportional odds assumption was violated, based on a significant score test, a partial proportional odds model was used instead.

For all regression models, we controlled for the baseline characteristics, including sociodemographic and health variables (age, sex, BMI, alcohol use, baseline fall, knee symptoms, previous lower-extremity fracture, and vision problems), chronic conditions (respiratory conditions, CVD, urinary incontinence, neurologic conditions, diabetes mellitus, high blood pressure, and depression), and performance-based tests (standing balance, chair rise, and TUG). Individuals with missing covariate data were dropped from the multivariable analyses. Each covariate had 5% or fewer missing values, with most being 2% or fewer. All analyses were performed using SAS, version 9.4.

RESULTS

Knee OA was reported by 4,112 individuals (19%) in the full analytic sample. Of those who reported knee OA, 10% reported having a least 1 fall during the year leading up to the follow-up

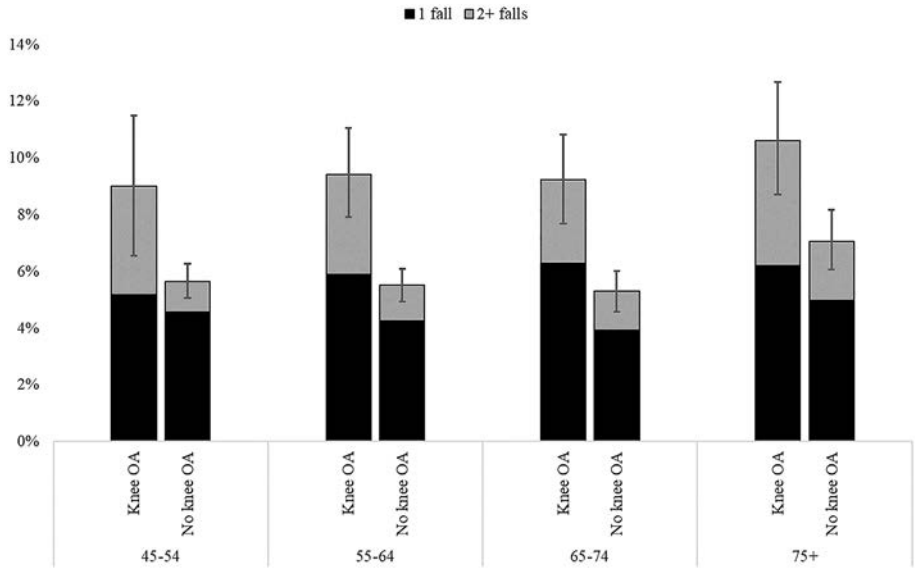

Figure 1. Proportion of individuals, by age, reporting 1 or multiple falls, by knee osteoarthritis (OA) status. Error bars show 95% Wald confidence intervals for the proportion reporting 1 or more falls at follow-up. Significant differences between groups were determined where error bars do not overlap.

Table 1. Characteristics of individuals in the full analytic sample stratified by knee OA status, with comparisons between those who did and did not fall*

	Knee OA		No knee OA	
Baseline characteristics	Fallers (n = 394)	Nonfallers (n = 3,718)	Fallers (n = 1,005)	Nonfallers (n = 16,593)
Age group, years				
45–54	11.9	12.8	31.5	31.9
55–64	32.1	32.6	32.6	33.9
65–74	30.8	32.2	19.7	21.3
75–85	25.3	22.5	16.1	12.8
Female	65.0	60.0	56.4	45.2
Body mass index categories				
Underweight/normal	18.5	18.8	34.4	33.0
Overweight	36.2	35.9	40.0	41.8
Obese	45.4	45.3	25.6	25.2
Alcohol use				
6+ times/week	15.1	15.8	18.0	16.7
2–5 times/week	28.6	29.2	31.6	33.2
0–1 times/week	56.4	55.1	50.4	50.1
Injurious fall, yes	13.2	6.4†	8.3	4.1†
Knee symptoms, yes	75.8	70.5†	22.2	18.7†
Lower limb fracture, yes	41.4	30.9†	31.5	24.2†
Vision problem, yes	8.9	8.4	8.2	6.4
Chronic conditions				
Respiratory, yes	29.4	22.7†	17.0	13.8†
Cardiovascular disease, yes	20.7	18.2	14.2	12.0†
Urinary incontinence, yes	20.9	13.1†	9.1	5.7†
Neurologic, yes	4.9	1.9†	3.4	2.0†
Diabetes mellitus, yes	26.9	22.3†	14.0	13.9
High blood pressure, yes	48.8	47.9	32.9	32.1
Depression, yes	24.6	20.1†	21.1	13.8†
Impaired performance				
One-leg balance ≤4.5 seconds	27.4	21.1†	14.2	9.2†
TUG time ≥14.2 seconds	10.8	7.0†	3.9	1.8†
Chair rise test ≥15.9 seconds	36.3	29.5†	21.4	18.2†

* Values are the percentage. OA = osteoarthritis; TUG = timed-up-and-go.
† Significant chi-square test comparing individuals who did and did not fall ($P < 0.05$).

Table 2. Examination of the predictors of reporting an injurious fall, overall and among those with knee OA and without knee OA from the full analytic sample (outcome: fallers versus nonfallers)*

	Model 1 Full sample	Model 2 Knee OA	Model 3 No knee OA
Knee OA, yes (ref. no)	1.33 (1.14–1.56)†	–	–
Age, years (ref. 75–85)			
45–54	0.93 (0.76–1.13)	0.96 (0.62–1.48)	0.90 (0.71–1.14)
55–64	0.87 (0.72–1.06)	0.98 (0.69–1.37)	0.83 (0.66–1.05)
65–74	0.85 (0.70–1.03)	0.92 (0.66–1.26)	0.81 (0.64–1.03)
Female (ref. male)	1.44 (1.27–1.63)†	1.14 (0.89–1.46)	1.55 (1.34–1.78)†
Body mass index (ref. underweight/normal)			
Overweight	0.96 (0.83–1.11)	1.04 (0.75–1.43)	0.95 (0.81–1.11)
Obese	0.88 (0.74–1.03)	0.86 (0.62–1.21)	0.90 (0.75–1.09)
Alcohol use (ref. 0–1 times/week)			
6+ times/week	1.18 (1.00–1.39)†	1.14 (0.81–1.59)	1.19 (0.98–1.44)
2–5 times/week	1.10 (0.96–1.25)	1.14 (0.87–1.48)	1.08 (0.93–1.26)
Baseline fall, yes (ref. no)	1.85 (1.50–2.26)†	1.75 (1.22–2.52)†	1.89 (1.47–2.42)†
Knee symptoms, yes (ref. no)	1.13 (0.98–1.30)	1.10 (0.85–1.43)	1.14 (0.97–1.35)
Lower fracture, yes (ref. no)	1.39 (1.23–1.57)†	1.42 (1.12–1.80)†	1.38 (1.20–1.60)†
Vision problems, yes (ref. no)	1.10 (0.89–1.36)	0.95 (0.63–1.43)	1.18 (0.92–1.52)
Respiratory, yes (ref. no)	1.17 (1.01–1.35)†	1.22 (0.94–1.59)	1.16 (0.97–1.39)
Cardiovascular disease, yes (ref. no)	1.15 (0.97–1.36)	1.05 (0.77–1.42)	1.19 (0.97–1.47)
Urinary incontinence, yes (ref. no)	1.29 (1.06–1.56)†	1.38 (1.01–1.88)†	1.26 (0.98–1.62)
Neurologic, yes (ref. no)	1.60 (1.15–2.22)†	2.13 (1.18–3.87)†	1.38 (0.92–2.06)
Diabetes mellitus, yes (ref. no)	1.05 (0.89–1.24)	1.26 (0.96–1.67)	0.94 (0.76–1.16)
High blood pressure, yes (ref. no)	0.97 (0.85–1.11)	0.97 (0.76–1.24)	0.96 (0.82–1.13)
Depression, yes (ref. no)	1.40 (1.21–1.62)†	1.10 (0.83–1.46)	1.55 (1.31–1.83)†
One leg balance ≤4.5 seconds (ref. >4.5 seconds)	1.34 (1.12–1.60)†	1.18 (0.87–1.59)	1.47 (1.18–1.83)†
TUG time ≥14.2 seconds (ref. <14.2 seconds)	1.34 (0.96–1.86)	1.25 (0.78–2.01)	1.54 (0.97–2.44)
Chair rise test ≥15.9 seconds (ref. <15.9 seconds)	1.09 (0.94–1.26)	1.11 (0.85–1.44)	1.09 (0.92–1.30)

* Values are the odds ratio (95% confidence interval). OA = osteoarthritis; ref. = reference; TUG = timed-up-and-go.
† $P < 0.05$.

questionnaire: 6% reported 1 fall, while 4% reported 2 or more falls. Comparable estimates were 4% and 1%, respectively, among individuals without knee OA. Individuals with knee OA were more likely to report 1 or multiple falls within each age group than individuals without knee OA (Figure 1).

The characteristics of the sample are shown in Table 1. The table is stratified by the presence of knee OA, and comparisons were made between those who did and did not experience an injurious fall. Among those with knee OA, there was no difference in age or sex between those who did and did not fall; this lack of difference was not the case among those without knee OA. In both those with and without knee OA, fallers were more likely to report having a previous fall, knee symptoms, and previously having a lower-extremity fracture. With reference to chronic conditions and the performance-based measures, the profile of differences between fallers and non-fallers was similar in those with knee OA as in those without.

Findings from logistic regression analysis showed that the risk of falling was significantly higher for individuals with knee OA compared to those without, with an OR of 1.33 (95% CI 1.14–1.56), adjusting for baseline characteristics (Table 2). Table 2 further shows the risk factor profiles for falling among those with knee OA and those without knee OA. Overall, the pattern of risk factors was similar. For both groups, those who fell were more likely to have reported a baseline fall or lower-extremity fracture. Among those with knee OA, those who fell were more likely to report urinary incontinence (OR 1.38 [95% CI 1.01–1.88]) or neurologic conditions (OR 2.13 [95% CI 1.18–3.87]) than those who did not report a fall at follow-up. Among individuals without knee OA, those who fell were more likely to be female

Table 3. Where and how falls occurred among those who experienced a fall, by OA status*

Where and how fall occurred	Knee OA	No knee OA
Indoors†	46.8	38.7
Standing or walking‡	48.4	35.7
On the stairs or steps	17.9	27.0
Exercising	6.0	9.0
Other	27.7	28.3
Outdoors	53.2	61.3
Standing or walking§	41.6	32.0
On the stairs or steps	12.9	9.3
Exercising	8.6	12.4
Other	36.8	46.3

* Values are the percentage. OA = osteoarthritis.
† Significant chi-square test comparing the location of a fall between those with and without knee OA ($P < 0.05$).
‡ Significant chi-square test comparing how an indoor fall occurred between those with and without knee OA ($P < 0.05$).
§ Significant chi-square test comparing how an outdoor fall occurred between those with and without knee OA ($P < 0.05$).

(OR 1.55 [95% CI 1.34–1.78]) and to report depression (OR 1.55 [95% CI 1.31–1.83]) and worse balance (OR 1.47 [95% CI 1.18–1.83]) than their peers who did not fall.

When limiting examination to those individuals who experienced a fall, differences were found in where and how individuals fell, depending on the presence of knee OA (Table 3). Individuals with knee OA were significantly more likely to report falling indoors than individuals without knee OA, any (46.8% versus 38.7%, respectively), and significantly less likely to report falling outdoors (53.2% versus 61.3%, respectively). Among those who reported falling indoors, individuals with knee OA had a different profile of how their fall occurred than did individuals without knee OA. For example, among those who fell indoors, 48.4% of those with knee OA reported falling while standing or walking, while this type of fall was the case for 35.7% of those without knee OA.

Focusing only on individuals with knee OA, Table 4 shows the characteristics of individuals by the number of falls reported. Individuals with knee OA who reported multiple falls were more likely to report a previous injurious fall, knee symptoms, a lower-extremity fracture, and impairment on the performance tests, compared to those with knee OA who reported either none or 1 fall. Overall, individuals with knee OA reporting multiple falls were also more likely to report having co-occurring chronic conditions.

Table 5 shows the results from the partial proportional odds model that examined contributors to the risk of 1 and 2 or more falls among individuals with knee OA. Reporting a previous fall at baseline ($OR_{1\ fall}$ 1.75 [95% CI 1.21–2.51]; $OR_{2+\ falls}$ 2.48 [95% CI 1.53–4.04]), a previous lower-extremity fracture (cumulative OR 1.42 [95% CI 1.12–1.80]), urinary incontinence (cumulative OR 1.38 [95% CI 1.01–1.88]), or a neurologic condition ($OR_{1\ fall}$ 2.08 [95% CI 1.14–3.78]; $OR_{2+\ falls}$ 3.67 [95% CI 1.77–7.59]) were all significant predictors of reporting 1 or multiple falls at follow-up compared to no falls in knee OA. Individuals with knee OA who reported a respiratory condition ($OR_{2+\ falls}$ 1.55 [95% CI 1.03–2.32]) and those who performed poorly on the balance test ($OR_{2+\ falls}$ 1.82 [95% CI 1.20–2.77]) had higher odds of reporting multiple falls compared to just 1 or none at all.

DISCUSSION

In the current prospective study, we sought to identify the contribution of knee OA to the risk of falling and to determine what

Table 4. Characteristics among individuals with knee osteoarthritis by number of falls reported*

Baseline characteristic	0 falls (n = 3,718)	1 fall (n = 247)	2+ falls (n = 147)
Sociodemographic, health, and lifestyle variables			
Age group			
45–54	12.8	10.9	13.6
55–64	32.6	32.0	32.0
65–74	32.2	33.6	26.5
75–85	22.5	23.5	27.9
Female	60.0	64.8	65.3
Body mass index categories			
Underweight/normal	18.8	19.4	16.8
Overweight	35.9	38.1	32.9
Obese	45.3	42.5	50.4
Alcohol use			
6+ times/week	15.8	16.9	12.0
2–5 times/week	29.2	27.2	31.0
0–1 times/week	55.1	56.0	57.0
Injurious fall, yes†	6.4	8.5	21.2
Knee symptoms, yes†	70.5	71.5	83.0
Lower limb fracture, yes†	30.9	36.4	49.7
Vision problem, yes	8.4	8.5	9.5
Chronic conditions			
Respiratory, yes†	22.7	26.4	34.5
Cardiovascular disease, yes	18.2	19.0	23.6
Urinary incontinence, yes†	13.1	17.5	26.7
Neurologic, yes†	1.9	2.4	9.0
Diabetes mellitus, yes	22.3	27.1	26.5
High blood pressure, yes	47.9	47.5	51.0
Depression, yes†	20.1	21.5	29.9
Impaired performance			
One leg balance ≤4.5 seconds†	21.1	20.1	40.6
TUG time ≥14.2 seconds†	7.0	5.7	19.4
Chair rise test ≥15.9 seconds†	29.5	32.4	43.2

* Values are the percentage. TUG = timed-up-and-go.
† Significant chi-square test comparing groups ($P < 0.05$).

Table 5. Partial proportional odds model identifying contributors to the risk of falling among individuals with knee osteoarthritis*

Baseline characteristics	Proportional OR	Partial proportional OR	
		2+ or 1 vs. 0 falls	2+ vs. 1 or 0 falls
Age, years (ref. 75–85)			
45–54	0.99 (0.64–1.52)	–	–
55–64	0.96 (0.68–1.35)	–	–
65–74	0.91 (0.66–1.25)	–	–
Female (ref. male)	1.14 (0.89–1.46)	–	–
Body mass index (ref. underweight/normal)			
Overweight	1.03 (0.75–1.43)	–	–
Obese	0.87 (0.62–1.22)	–	–
Alcohol use (ref. 0–1 times/week)			
6+ times/week	1.11 (0.79–1.56)	–	–
2–5 times/week	1.14 (0.87–1.49)	–	–
Baseline fall, yes (ref. no)	–	1.75 (1.21–2.51)†	2.48 (1.53–4.04)†
Knee symptoms, yes (ref. no)	–	1.10 (0.84–1.42)	1.50 (0.93–2.40)
Lower fracture, yes (ref. no)	1.42 (1.12–1.80)†	–	–
Vision problems, yes (ref. no)	0.93 (0.62–1.41)	–	–
Respiratory, yes (ref. no)	–	1.22 (0.94–1.59)	1.55 (1.03–2.32)†
Cardiovascular disease, yes (ref. no)	1.05 (0.77–1.42)	–	–
Urinary incontinence, yes (ref. no)	1.38 (1.01–1.88)†	–	–
Neurologic, yes (ref. no)	–	2.08 (1.14–3.78)†	3.67 (1.77–7.59)†
Diabetes mellitus, yes (ref. no)	1.25 (0.95–1.65)	–	–
High blood pressure, yes (ref. no)	0.98 (0.76–1.25)	–	–
Depression, yes (ref. no)	1.12 (0.85–1.48)	–	–
One leg balance ≤4.5 seconds (ref. >4.5 seconds)	–	1.17 (0.87–1.57)	1.82 (1.20–2.77)†
TUG time ≥14.2 seconds(ref. <14.2 seconds)	1.29 (0.81–2.07)	–	–
Chair rise test ≥15.9 seconds (ref. <15.9 seconds)	1.10 (0.85–1.44)	–	–

* Values are the odds ratio (OR) (95% confidence interval). ref. = reference; TUG = timed-up-and-go.
† $P < 0.05$.

predicts experiencing an injurious fall among individuals with knee OA. We confirmed in our analyses that knee OA is a predictor of falling, as has been previously reported (4–10) and that having knee OA significantly contributed to the risk of falling compared to those without knee OA. We also found differences in the patterns of where and how a fall occurred for those with and without knee OA. Generally, individuals with knee OA had similar risk factors for having an injurious fall to those without knee OA. However, for those with knee OA, the risk was higher for those with urinary incontinence or a neurologic condition.

We found differences in where and how those with and without knee OA reported that their fall occurred. Individuals with knee OA were significantly more likely to report falling indoors compared to individuals without knee OA. The most common circumstance of falling reported by individuals with knee OA indoors was while standing or walking, while individuals without knee OA were much more likely to report falling while on the stairs or steps or while exercising. Possibly individuals with knee OA, especially those with symptomatic knee OA, use more caution while walking down stairs or while exercising or avoid these activities altogether (19,20). While outdoors, individuals with knee OA were more likely to report falling while standing or walking or while on the stairs or steps, while individuals without knee OA were more likely to report that their fall occurred while exercising or another reason. Possibly while outdoors, avoiding stairs or steps is not as easy or convenient, causing more falls among individuals with knee OA. Our findings also support the hypothesis that individuals with lower-extremity OA are more likely to become unstable and are less able to perform compensatory stepping responses to avoid falling compared to individuals without lower-extremity OA, exacerbating existing risk factors for falling (6).

Painful knee symptoms did not appear to contribute to the risk of falling among individuals with knee OA, likely because the majority of individuals with knee OA reported knee symptoms (71%), although there was some indication of a contribution to the risk of 2 or more falls. Conflicting evidence has been found in the literature for the association between pain and falling among individuals with knee OA. While 4 studies from the review by Manlapaz et al (12) identified an association, 2 studies found no association (19, 20). The latter studies suggest that the presence of severe pain deters participants from doing too much activity, especially activity associated with the risk of increased pain or loss of balance, therefore reducing the risk of falling (19,20). On the other hand, those who do not reduce their activity despite pain may be more at risk due to increased postural sway and lower range of motion in their knees (21,22). In a more recent study, Barbour et al found that while having radiographic knee OA without pain symptoms was not associated with having an injurious fall, having radiographic knee OA with pain was, but only among men (9). Perhaps additional factors are involved, such as sex, which influence whether a person continues to participate in an

activity despite their pain, therefore increasing or decreasing their risk of falling.

We also found that although older age is a well-established risk factor for falling among the general population, older age did not increase the risk of falling in this study. Although most previous studies only looked at older adults (age 65+ years) for the risk of falls, a few studies have looked at the risk factors of falling in younger adults in the general population (23,24), and among younger individuals with OA as well (10,25). The age-related increase in the prevalence of OA has led to the belief that OA is a disease that only impacts older adults; the condition is common, however, in younger and middle-aged people as well. Younger individuals with OA experience similar, if not worse, outcomes compared to older adults with OA (26). This similarity across age groups is also likely to be true for the risk of falls among individuals with knee OA, as we have demonstrated in the current study. One recent study found that middle-aged adults with more advanced knee OA do not have an increased likelihood of falls, unlike older-aged adults, compared to individuals in their respective age groups without OA (10). The authors speculated that differences in balance, compared to older adults, might reduce the likelihood of a middle-aged adult experiencing a fall. Possibly controlling for performance on the standing balance test in our model explained age differences, resulting in a lack of association between age and the risk of falling.

Urinary incontinence is recognized as a risk factor for falls (27), although this condition has not been specifically studied in knee OA. Knee OA can be associated with impairments of balance and limitations in mobility, including transfers from sitting or lying to standing (28). These impairments and limitations could contribute to slowness, leading to an increased risk of falls in response to an urgency to get to the bathroom in a timely manner. Although we did not find that any of our performance-based measures were associated with experiencing a fall among individuals with knee OA, we did find that a lack of balance was associated with individuals with knee OA reporting multiple falls.

Strengths of this study include the use of a large longitudinal population-based sample and the inclusion of adults with knee OA across a wide-range of ages from middle aged to elderly adults. Another strength was the use of a very specific outcome, falls which resulted in an injury stratified by the number of falls (single versus multiple), whereas most previous studies use more general definitions of falls that may mask particular subgroups of fallers. However, this definition may also be a limitation, as it excludes falls not considered the cause of an injury associated with limitation of activity. Other potential limitations of our study were the self-reported nature of the questions about falls and knee OA. Individuals with poorer cognitive function are less likely to recall falling in the previous 12 months (29). Participants in the CLSA were screened for mild cognitive impairment at baseline and follow-up, and those who screened positive at follow-up had the option of having their designated proxy respondent help them or respond on their behalf. None of the respondents in this study used a proxy, and therefore there is minimal likelihood for recall bias due to cognitive impairment to be an issue.

Additionally, although self-report of knee OA has the potential of introducing recall bias into the study, self-reported OA has been found to be a valid measure of prevalence in population-based surveillance studies (30,31). A further limitation of the current study includes the relatively short follow-up period. Additionally, the available survey data did not allow us to assess the contribution of extrinsic or behavioral factors that may contribute to the increased risk of falls among individuals with knee OA, factors that may play an important role in understanding this increased risk.

In the current study, we found that the presence of knee OA predicts experiencing an injurious fall. Our findings support the idea that knee OA is an independent risk factor for falls, though more research is needed to understand the complex interactions between intrinsic and extrinsic factors that contribute to this increased risk. We also found differences in where and how falls occurred between individuals with and without knee OA. A novel finding was the increased risk of falls in knee OA associated with urinary incontinence. Fall prevention is an important clinical target, especially among individuals with knee OA, as falling can cause further damage to the joint and other injury leading to decreased physical activity and social participation, both of which are important factors for outcomes among individuals with knee OA. The modifiable intrinsic risk factors identified in the current study along with specific target environments associated with falling among individuals with knee OA may provide opportunities for clinical intervention and fall prevention strategies.

ACKNOWLEDGMENTS

This research was made possible using the data collected by the Canadian Longitudinal Study on Aging (CLSA). This research has been conducted using the CLSA datasets Baseline Comprehensive Dataset version 7.0 and the Follow-up 1 Comprehensive Dataset version 3.2, under Application Number 2010009. The CLSA is led by Drs. Parminder Raina, Christina Wolfson, and Susan Kirkland.

AUTHOR CONTRIBUTIONS

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

Study conception and design. Wilfong, Perruccio, Badley.
Acquisition of data. Perruccio, Badley.
Analysis and interpretation of data. Wilfong, Perruccio, Badley.

REFERENCES

1. Roth GA, Abate D, Abate KH, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and

territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392:1736–88.

2. Kyu HH, Abate D, Abate KH, et al. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392:1859–922.
3. Jehu D, Davis J, Falck R, et al. Risk factors for recurrent falls in older adults: a systematic review with meta-analysis. Maturitas 2021;144: 23–8.
4. Brand C, Aw J, Lowe A, et al. Prevalence, outcome and risk for falling in 155 ambulatory patients with rheumatic disease. Int J Rheum Dis 2005;8:99–105.
5. Nevitt MC, Cummings SR, Kidd S, et al. Risk factors for recurrent nonsyncopal falls: a prospective study. JAMA 1989;261:2663–8.
6. Hoops ML, Rosenblatt NJ, Hurt CP, et al. Does lower extremity osteoarthritis exacerbate risk factors for falls in older adults? Womens Health (Lond) 2012;8:685–98.
7. Barbour KE, Stevens JA, Helmick CG, et al. Falls and fall injuries among adults with arthritis: United States, 2012. MMWR Morb Mortal Wkly Rep 2014;63:379–83.
8. Doré AL, Golightly YM, Mercer VS, et al. Lower-extremity osteoarthritis and the risk of falls in a community-based longitudinal study of adults with and without osteoarthritis. Arthritis Care Res (Hoboken) 2015;67:633–9.
9. Barbour KE, Sagawa N, Boudreau RM, et al. Knee osteoarthritis and the risk of medically treated injurious falls among older adults: a community-based US cohort study. Arthritis Care Res (Hoboken) 2019;71:865–74.
10. Harris R, Strotmeyer ES, Sharma L, et al. The association between severity of radiographic knee OA and recurrent falls in middle and older aged adults: the Osteoarthritis Initiative. J Gerontol A Biol Sci Med Sci 2023;78:97–103.
11. Do M, Chang V, Kuran N, et al. Fall-related injuries among Canadian seniors, 2005–2013: an analysis of the Canadian Community Health Survey. Health Promot Chronic Dis Prev Can 2015;35:99–108.
12. Manlapaz DG, Sole G, Jayakaran P, et al. Risk factors for falls in adults with knee osteoarthritis: a systematic review. PM R 2019;11:745–57.
13. Deandrea S, Lucenteforte E, Bravi F, et al. Risk factors for falls in community-dwelling older people: a systematic review and meta-analysis [review]. Epidemiology 2010:658–68.
14. Stalenhoef P, Diederiks J, Knottnerus J, et al. A risk model for the prediction of recurrent falls in community-dwelling elderly: a prospective cohort study. J Clin Epidemiol 2002;55:1088–94.
15. Ambrose AF, Paul G, Hausdorff JM. Risk factors for falls among older adults: a review of the literature. Maturitas 2013;75:51–61.
16. Paliwal Y, Slattum PW, Ratliff SM. Chronic health conditions as a risk factor for falls among the community-dwelling US older adults: a zero-inflated regression modeling approach. Biomed Res Int 2017; 2017:5146378.
17. Raina PS, Wolfson C, Kirkland SA, et al. The Canadian longitudinal study on aging (CLSA). Can J Aging 2009;28:221–9.
18. Beauchamp MK, Kuspinar A, Sohel N, et al. Mobility screening for fall prediction in the Canadian Longitudinal Study on Aging (CLSA): implications for fall prevention in the decade of healthy ageing. Age Ageing 2022;51:afac095.
19. Alencar M, Arantes P, Dias J, et al. Muscular function and functional mobility of faller and non-faller elderly women with osteoarthritis of the knee. Braz J Med Biol Res 2007;40:277–83.
20. De Zwart AH, van der Esch M, Pijnappels MA, et al. Falls associated with muscle strength in patients with knee osteoarthritis and self-reported knee instability. J Rheumatol 2015;42:1218–23.
21. Foley S, Lord SR, Srikanth V, et al. Falls risk is associated with pain and dysfunction but not radiographic osteoarthritis in older adults: Tasmanian Older Adult Cohort study. Osteoarthritis Cartilage 2006; 14:533–9.
22. Sturnieks DL, Tiedemann A, Chapman K, et al. Physiological risk factors for falls in older people with lower limb arthritis. J Rheumatol 2004;31:2272–9.
23. Peeters G, van Schoor NM, Cooper R, et al. Should prevention of falls start earlier? Co-ordinated analyses of harmonised data on falls in middle-aged adults across four population-based cohort studies. PLoS One 2018;13:e0201989.
24. White AM, Tooth LR, Peeters GG. Fall risk factors in mid-age women: the Australian Longitudinal Study on Women's Health. Am J Prev Med 2018;54:51–63.
25. Ofori-Asenso R, Ackerman IN, Soh SE. Prevalence and correlates of falls in a middle-aged population with osteoarthritis: data from the Osteoarthritis Initiative. Health Soc Care Community 2021;29: 436–44.
26. O'Donnell S, Rusu C, Hawker GA, et al. Arthritis has an impact on the daily lives of Canadians young and old: results from a population-based survey. BMC Musculoskelet Disord 2015;16:230.
27. Chiarelli PE, Mackenzie LA, Osmotherly PG. Urinary incontinence is associated with an increase in falls: a systematic review. Aust J Physiother 2009;55:89–95.
28. Davis M, Ettinger W, Neuhaus J, et al. Knee osteoarthritis and physical functioning: evidence from the NHANES I epidemiologic followup study. J Rheumatol 1991;18:591–8.
29. Ganz DA, Higashi T, Rubenstein LZ. Monitoring falls in cohort studies of community-dwelling older people: effect of the recall interval [review]. J Am Geriatr Soc 2005;53:2190–4.
30. Ratzlaff C, Koehoorn M, Cibere J, et al. Clinical validation of an internet-based questionnaire for ascertaining hip and knee osteoarthritis. Osteoarthritis Cartilage 2012;20:1568–73.
31. March LM, Schwarz JM, Carfrae BH, et al. Clinical validation of self-reported osteoarthritis. Osteoarthritis Cartilage 1998;6:87–93.

Arthritis Care & Research
Vol. 75, No. 11, November 2023, pp 2345–2358
DOI 10.1002/acr.25157

American College of Rheumatology
Empowering Rheumatology Professionals

Prevalence of Nonsteroidal Antiinflammatory Drugs Prescribed for Osteoarthritis: A Systematic Review and Meta-Analysis of Observational Studies

Zijing Yang,[1] Stephanie Mathieson,[2] Sarah Kobayashi,[3] Christina Abdel Shaheed,[2] Leandro Alberto Calazans Nogueira,[4] Milena Simic,[2] Gustavo Machado,[2] and Andrew J. McLachlan[2]

Objective. Our systematic review aimed to investigate the proportion of participants with osteoarthritis who were prescribed nonsteroidal antiinflammatory drugs (NSAIDs) by their health care provider.

Methods. Electronic databases were searched for observational studies reporting NSAID prescribing to participants with diagnosed osteoarthritis of any region. Risk of bias was assessed using a tool designed for observational studies measuring prevalence. Random and fixed-effects meta-analysis was used. Meta-regression investigated study-level factors associated with prescribing. The overall evidence quality was assessed using Grading of Recommendations Assessment, Development, and Evaluation criteria.

Results. Fifty-one studies were included, published between 1989 and 2022, with 6,494,509 participants. The mean age of participants was 64.7 years (95% confidence interval [95% CI] 62.4, 67.0; n = 34 studies). Most studies were from Europe and Central Asia (n = 23 studies), and North America (n = 12 studies). Most studies were judged to be at low risk of bias (75%). Heterogeneity was eliminated when removing studies with a high risk of bias, to give a pooled estimate of NSAIDs prescribing to participants with osteoarthritis of 43.8% (95% CI 36.8, 51.1; moderate quality of evidence). Meta-regression determined that prescribing was associated with year (decreased prescribing over time; $P = 0.05$) and geographic region ($P = 0.03$; higher in Europe and Central Asia and in South Asia than in North America) but not with clinical setting.

Conclusion. Data from over 6.4 million participants with osteoarthritis between 1989 and 2022 indicate that NSAID prescribing has decreased over time and that prescribing differs between geographic locations.

INTRODUCTION

Osteoarthritis is the most common type of arthritis (1). Clinical guidelines for the management of osteoarthritis recommend nonpharmacologic treatments, such as educational, psychosocial, and physical interventions, as well as pharmacologic management such as topical and oral nonsteroidal antiinflammatory drugs (NSAIDs) (2,3). NSAIDs have been shown, through meta-analyses, to be effective in achieving clinical improvements in pain and function (4,5) in people with osteoarthritis symptoms and are recommended as an effective symptomatic treatment for early arthritis in some guidelines (2,6). Guidelines frequently recommend NSAIDs to be prescribed at the smallest effective dose for the shortest possible time (2,6). Although NSAIDs can be a less costly management strategy than conservative care (e.g., ongoing physical therapy) they are not without risk of harm (4,5). Caution should be taken in prescribing NSAIDs for use in people with a high risk of diabetes mellitus, hypertension, renal impairment, and heart disease (7,8), with consideration that cyclooxygenase 2 (COX-2) selective NSAIDs are associated with fewer gastrointestinal ulcers and complications than nonselective NSAIDs (9,10).

The incidence of NSAIDs use for the management of osteoarthritis is common as evidenced by numerous individual studies (11,12). However, the extent to which NSAIDs are prescribed for osteoarthritis globally and what factors may be associated with prescribing are unclear. Previous systematic reviews related to osteoarthritis have focused on clinical outcomes such as efficacy

[1]Zijing Yang, MS: King's College, London, UK; [2]Stephanie Mathieson, PhD, Christina Abdel Shaheed, PhD, Milena Simic, PhD, Gustavo Machado, PhD, Andrew J. McLachlan, PhD: University of Sydney, Sydney, New South Wales, Australia; [3]Sarah Kobayashi, PhD: University of Sydney and Australian Catholic University, Sydney, New South Wales, Australia; [4]Leandro Alberto Calazans Nogueira, PhD: Augusto Motta University Centre and Federal Institute of Rio de Janeiro, Rio de Janeiro, Brazil.

Author disclosures and graphical abstracts are available online at https://onlinelibrary.wiley.com/doi/10.1002/acr.25157.

Address correspondence via email to Stephanie Mathieson, PhD, at stephanie.mathieson@sydney.edu.au.

Submitted for publication December 13, 2022; accepted in revised form May 11, 2023.

SIGNIFICANCE & INNOVATIONS

- This is the first review to assess changes in, and factors associated with, nonsteroidal antiinflammatory drugs (NSAIDs) prescribing for osteoarthritis.
- This large review analyzed data from observational studies of 6,494,509 participants between 1989 and 2022.
- NSAID prescribing for osteoarthritis decreased over time and was associated with geographic region but not with clinical setting.

and safety of NSAIDs (13–16). Previous studies have suggested that both oral and topical NSAIDs exhibit pain relief among people with osteoarthritis (4), but topical NSAIDs had a lower risk of toxicity (13), while there is no difference in efficacy between selective and nonselective NSAIDs in reducing pain and improving function (17). However, the prevalence of NSAID prescribing for the clinical management of osteoarthritis is unclear, and little is known about prescribing practices across countries and any differences in the management of regional types of osteoarthritis. Understanding to what extent NSAIDs are prescribed for osteoarthritis will determine any differences in prescribing and provide a benchmark for future studies. Therefore, this systematic review aimed to investigate the proportion of participants with osteoarthritis who were prescribed an NSAID by their health care provider, factors associated with prescribing, and geographic differences in prescribing.

MATERIALS AND METHODS

Eligibility criteria. The protocol for this review was devised in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement guidelines (18) and was registered on PROSPERO (CRD42021238699; www.crd.york.ac.uk). We included observational studies (cross-sectional, prospective, or retrospective cohort or case–control studies) of adults (age ≥18 years) with clinician-diagnosed osteoarthritis at any site, and who were prescribed an NSAID to manage their osteoarthritis symptoms. We included pharmacy dispensing data provided that the data were specific for clinician-diagnosed osteoarthritis and for which NSAIDs were prescribed. We excluded studies that did not include the representative population sample (e.g., not consecutive cases or randomly sampled), studies of self-reported NSAID use, over-the-counter supply of NSAIDs, and those with self-reported osteoarthritis diagnoses.

Search strategy. We searched the following electronic databases: PubMed (National Library of Medicine database), MEDLINE, EMBASE, and International Pharmaceutical Abstracts (the latter 3 from OvidSP), and Web of Science (Thomson Reuters) on April 23, 2022. We conducted backward and forward author and reference citation tracking of included articles and communicated with content experts to identify any missing studies. Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25157, contains the details of the search strategy.

Screening. Two authors from a panel (ZY, SM, or SK) independently screened records against the eligibility criteria. Duplicate studies were removed manually and using the automated function in Endnote. Disagreements were resolved first by discussion, then by arbitration with an independent third review author if needed. For articles written in languages that the review authors could not read, we asked colleagues to assist with reading and appraising the article.

Data extraction and management. Two review authors independently extracted data from eligible studies using a piloted, standardized extraction form in Excel (ZY and SM). Disagreements were resolved first by discussion, then by arbitration with an independent third review author if needed (CAS and AJM). We contacted the authors of studies for clarification and additional data if relevant data were missing. Information extracted included bibliometric data (authors, title, year of publication, language, funding sources), study characteristics (study design, data source, sample size, sampling dates and methods, country), participants (age, sex, site of diagnosis, symptom duration, first or ongoing presentation of index visit), pain intensity (e.g., numerical pain rating scale), interventions (profession of prescribing clinician, the number of NSAIDs prescribed or dispensed on prescription, dose, mode of delivery, frequency, duration; the proportion of other medicines and nonpharmacologic therapies coprescribed with the NSAIDs), and data completeness (i.e., the percentage of missing data, how missing data were handled).

Medicines were categorized using the Anatomical Therapeutic Chemical classification system (19), and NSAIDs were classed as nonselective or COX-2 selective, followed by the mode of delivery. A list of nonselective and COX-2 selective NSAIDs is in Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25157. Combination medicines were initially classified by the NSAIDs. Data on co-administered therapy were retrieved if the therapy was prescribed to alleviate osteoarthritis and coprescribed with an NSAID. Nonpharmacologic treatments were categorized based on the therapies (e.g., physical therapy).

Countries were grouped according to World Health Organization (WHO) regions (East Asia and Pacific, Europe and Central Asia, Latin America and Caribbean, Middle East and North Africa, North America, South Asia) (20) and income status (low-, middle- and high-income) as per the World Bank (21).

Risk-of-bias assessment. Risk of bias was assessed using the tool developed by Hoy et al (22) to assess the risk of bias

in observational studies that measure prevalence. A study's overall risk of bias was low if further research was very unlikely to change our confidence in the estimate, moderate if further research was likely to have an important impact on our confidence in the estimate and may change the estimate, or high if further research was very likely to have an important impact on our confidence in the estimate and was likely to change the estimate (22). The criteria for the risk-of-bias assessment are shown in Supplementary Table 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25157.

Data synthesis. Study characteristics and study participants are descriptively reported. Random and fixed-effects meta-analyses were used to pool the main prevalence estimate and random effects were used for the subgroup analyses. Statistical heterogeneity among the studies was assessed using a visual inspection of the forest plot and I^2 statistics following the recommended guide for interpretation of I^2 as 0–40% = might not be important, 30–60% = may represent moderate heterogeneity, 50–90% = may represent substantial heterogeneity, and 75–100% = considerable heterogeneity (23). Meta-regression analyses were performed to explore sources of heterogeneity across the included studies and to determine possible study-related factors associated with prescribing. Factors included the WHO region (compared to North America), sampling year (continuous; defined as the year associated with the midpoint of the prevalence sampling period), setting (primary care, tertiary care, multiple clinical settings, population based, compared database [e.g., prescribing database, dispensing claims database]), the duration of the prevalence period (continuous in months), and whether funding was reported (compared to none). Analyses were conducted in Comprehensive Meta-Analysis, version 3.3.070. The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach (24) was used to assess the quality of the evidence. Supplementary Appendix B, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25157 contains the details of GRADE criteria.

Subgroup and sensitivity analyses. We conducted 4 planned subgroup analyses to 1) investigate differences in pooled prescribing estimates per osteoarthritis site, including participants with spinal-related osteoarthritis, 2) compare the pooled prescribing estimates per WHO geographic region and country income status, 3) determine the proportion of participants using different types of NSAIDs and dose, including grouped per nonselective and COX-2 selective NSAIDs, and 4) determine the differences in the proportion of participants prescribed NSAIDs per mode of delivery (i.e., topical) and action. Sensitivity analysis was performed as there was an adequate number of studies (>10 studies) by excluding studies assessed to have high risk of bias and then repeating the analyses.

RESULTS

A total of 9,220 records were identified by searching electronic databases, plus 10 additional articles were identified through citation tracking. Fifty-one studies met the inclusion criteria and were included in this review. The flow of studies is shown in Figure 1.

The 51 studies provided data on a total of 6,494,509 participants with a mean age of 64.7 years (95% confidence interval [95% CI] 62.4, 67.0; n = 34 studies) (11,12,25–55). The included studies were published between 1989 and 2022 and were all in English except 1 study published in Croatian (56). Studies were from 31 countries across the globe, including South Asia (n = 4 studies) (37,47,49,57), Middle East and North Africa (n = 1 study) (33), East Asia and Pacific (n = 10 studies) (12,25,32,38,40,51,53,58–60), Europe and Central Asia (n = 23 studies) (26–29,39,42,43,45,46,48,50,52,54–56,61–68), Latin America and Caribbean (n = 1 study) (44), and North America (n = 12 studies) (11,30,31,34–36,41,69–73). Most studies (90.2%) were from high-income countries with 1 study from an upper-middle income country (44), and 4 studies were from lower-middle income countries (37,47,49,57). Half the studies (52.9%) were from clinical settings, with 20 studies from primary care (26,28,29,34,39,42,46,48–50,52,54,55,58,60,64,65–68), 7 studies from tertiary care clinics (37,43,44,47,56,57,72), and 5 studies from multiple care (30,31,41,45,62); 18 studies (35%) provided prevalence data from a database (11,12,25,27,32,33,35,36,38,40,51,53,59,63,69–71,73), and 1 was a population-based study (61). Characteristics of included studies are shown in Table 1. No study reported the coprescribing of analgesic drugs or nonpharmacologic therapies specifically occurring at the same time of NSAID prescribing. However, 26 studies reported that participants used other medicines (12,25,27,28,31,33–36,42–45,48,51–53,59,60,62,66–68,70,72) or physical therapy (32,45,58,60,70) at some time during the sampling period.

Risk of bias. The majority of studies (75%) were judged to be at low risk of bias. Eight studies (30,37,43,56,57,61,66,72) were classified as having a moderate risk of bias (16%), while 5 studies (46–48,67,69) were scored as having a high risk of bias (9%). The domain that most frequently scored poorly was related to using validated outcome measures, as most studies evaluated clinical records. Only 5% of studies collected data using validated measures. The risk-of-bias scores are shown in Table 2.

Proportion of patients with osteoarthritis who were prescribed NSAIDs. High heterogeneity was present when pooling NSAID prescribing estimates across all studies (I^2 = 99.9%). A forest plot of individual studies is shown in Figure 2. We conducted a sensitivity analysis to explore heterogeneity related to risk of bias. When studies scored as having a high

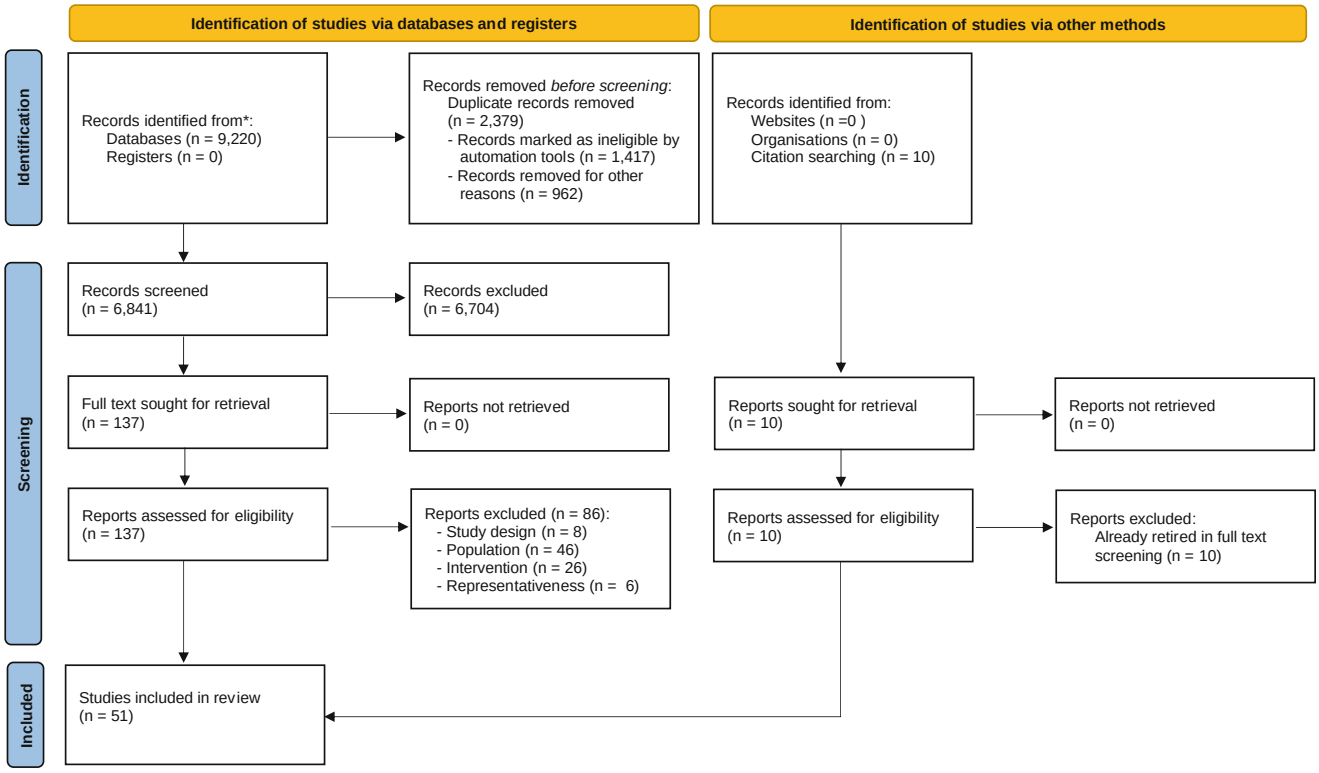

Figure 1. Study flow diagram.

risk of bias were removed (n = 5 studies) (46–48,67,69), the pooled prescribing estimate remained similar (43.8% [95% CI 36.8, 51.1], n = 46 studies, high quality of evidence I^2 = 5.1%) (11,12,25,26,28–46,49–56,58–61,58–64,65,66,68,70–73) compared to the original estimate with high heterogeneity (43.1% [95% CI 36.3, 50.1], n = 51 studies, I^2 = 99.9%, low quality of evidence). A post hoc sensitivity analysis was conducted to explore the primary analyses using an alternative statistical approach (see Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25157), which resulted in less conservative estimates than our original model.

Factors associated with prescribing of NSAIDs. Meta-regression was used to explore potential sources of heterogeneity and to determine potential factors associated with prescribing. Meta-regression analyses of study-related factors explained 42% of heterogeneity (R^2 = 0.42). Prescribing was associated with the WHO region (*P* = 0.026), with increased prescribing in the regions of East Asia and Pacific (coefficient 0.86 [95% CI –0.098, 1.81]; *P* = 0.078), Europe and Central Asia (coefficient 1.26 [95% CI 0.23, 2.28]; *P* = 0.02), Latin America and Caribbean (coefficient 2.02 [95% CI –0.62, 4.65]; *P* = 0.13), Middle East and North Africa (coefficient 0.26 [95% CI –2.11, 2.63]; *P* = 0.83), and South Asia (coefficient 3.02 [95% CI 1.27, 4.76]; *P* = 0.001), compared to North America (US and Canada). There was a decrease in NSAID prescribing over time (coefficient –0.04 [95% CI –0.08, 0.00]; *P* = 0.05) and longer sampling duration (coefficient –0.006 [95% CI –0.009, –0.002]; *P* = 0.001). Reporting of funding (*P* = 0.59) and clinical setting (*P* = 0.20) did not influence prescribing. A summary of the meta-regression analysis is shown in Supplementary Figure 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25157.

Subgroup analyses. *The proportion of NSAIDs prescribed to participants per osteoarthritis site.* The pooled estimate of NSAIDs prescribed to patients with hip osteoarthritis (27,32,60) was 34.9% (95% CI 23.8, 47.9; n = 3 studies, I^2 = 0%, high quality of evidence). In contrast, NSAID prescribing to patients with knee osteoarthritis was 46.3% (95% CI 36.9, 55.9; n = 11 studies, I^2 = 28.8%, moderate quality of evidence) (27,29,32,45,47,48,57,59,60,66,67) and for spine osteoarthritis was 66.9% (95% CI 66.6, 67.2; n = 1 study, I^2 = 0%, high quality of evidence) (27). The stratified analyses results are summarized in Table 3 and the forest plot shown in Supplementary Figure 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25157.

Prescribing estimates across WHO regions. The pooled prevalence of NSAIDs varied widely across geographical locations. The pooled estimate of NSAIDs prescribing was highest in South Asia at 83.4% (95% CI 74.8, 89.4; n = 4 studies, I^2 = 3.0%, moderate quality of evidence) (37,47,49,57), followed by Latin America and

Table 1. Characteristics of included studies*

Author, year (ref.)	Prevalence type	Setting	Country	OA sample size	OA site	Radiologic diagnosis	Age, mean ± SD years
Akazawa et al, 2019 (25)	Retrospective	Database	Japan	118,996	All regions	No	68.8 ± 13.1
Alacqua et al, 2008 (68)	Retrospective	Primary	Italy	142,346	All regions	No	NR
Arboleya et al, 2003 (26)	Retrospective	Primary	Spain	897	All regions	Yes	66.0 ± 9.0
Barcella et al, 2019 (27)	Retrospective	Database	Denmark	533,502	All regions	No	62.2 ± 14.3
Bennell et al, 2021 (60)	Retrospective	Primary	Australia	9,812	Hip/knee	No	NR
Castaño Carou et al, 2015 (28)	Prospective	Primary	Spain	1,258	Hip, knee, and hand	Yes	68 ± 9.5
Chandan et al, 2021 (55)	Retrospective	Primary	UK	25,659	All regions	No	68.53 ± 11.0
Colombo et al, 2021 (54)	Retrospective	Primary	Italy	71,467	All regions	No	71.36 ± 12.2
Cunnington et al, 2008 (73)	Retrospective	Database	US	80,826	All regions	No	NR
Denoeud et al, 2005 (29)	Prospective	Primary	France	2,430	Knee	Yes	66.8 ± 10.6
Dominick et al, 2003 (30)	Retrospective	Multiple	US	2,473	All regions	No	61.1 ± 14.0
Dominick et al, 2003 (31)	Retrospective	Multiple	US	11,298	All regions	No	80.2 ± 6.9
Ebata-Kogure et al, 2020 (32)	Retrospective	Database	Japan	328,631	Hip/knee	No	69.7 ± 11.5
Fallach et al, 2021 (33)	Retrospective	Database	Israel	180,126	All regions	No	58.5 ± 11.9
Gore et al, 2011 (35)	Retrospective	Database	US	207,010	All regions	Yes	53.2 ± 9.8
Gore et al, 2011 (36)	Retrospective	Database	US	112,951	All regions	Yes	56.9 ± 9.5
Gore et al, 2012 (34)	Retrospective	Primary	UK	18,184	All regions	No	70.6 ± 11.0
Gupta et al, 2018 (37)	Prospective	Tertiary	India	188	All regions	No	61.7 ± 6.9
Barbero et al, 2017 (67)	Prospective	Primary	Spain	646	Knee	No	NR
Hsu et al, 2017 (38)	Retrospective	Database	China (Taiwan)	43,635	All regions	No	60 ± 14.1
Jackson et al, 2017 (39)	Prospective	Primary	UK	1,724	All regions	No	66.1 ± 11.9
Kanneppady et al, 2017 (72)	Retrospective	Tertiary	US	296	All regions	No	47.5 ± NR
Kikuchi et al, 2021 (40)	Retrospective	Database	Japan	180,371	All regions	No	49.3 ± 11.8
Lanas et al, 2011 (62)	Prospective	Multiple	Spain	17,105	All regions	No	NR
Li et al, 2022 (71)	Retrospective	Database	Canada	100,358	All regions	No	68 ± NR
McDonald and Walsh, 2012 (41)	Retrospective	Multiple	US	128	All regions	No	74.1 ± 8.3
Patel et al, 2020 (70)	Retrospective	Database	US	44,990	All regions	No	75.9 ± NR
Paterson et al, 2018 (58)	Retrospective	Primary	Australia	621	Foot/ankle	No	NR
Pontes et al, 2018 (42)	Retrospective	Primary	Spain	22,652	All regions	No	75.6 ± 9.82
Rajamäki et al, 2019 (43)	Retrospective	Tertiary	Finland	13,739	All regions	No	68.7 ± 10.1
Reginato et al, 2015 (41)	Prospective	Tertiary	13 Latin American countries	3,040	All regions	Yes	62.5 ± 10.5
Reijman et al, 2005 (61)	Prospective	Population	Netherlands	3,585	Hip/knee	Yes	66 ± 6.9
Richette et al, 2011 (45)	Prospective	Multiple	France	1,821	Knee	Yes	67.3 ± 9.7
Russo et al, 2003 (65)	Retrospective	Primary	Italy	3,090	All regions	No	NR
Sakai et al, 2019 (59)	Retrospective	Database	Korea/Japan	1,143,636	Knee	No	NR
Shelbaya et al, 2018 (11)	Retrospective	Database	US	1,610,375	All regions	No	61 ± 12.2
Spitaels et al, 2020 (66)	Prospective	Primary	Belgium	1,595	Knee	No	55.3 ± NR
Spitaels et al, 2020 (66)	Prospective	Primary	Belgium	5,049	Knee	No	56.9 ± NR
Stambuk et al, 1989 (56)	Retrospective	Tertiary	Croatia	50	Hip	No	NR
Subramanian et al, 2020 (57)	Prospective	Tertiary	India	256	Knee	Yes	NR
Summanen et al, 2021 (46)	Retrospective	Primary	Finland	51,608	Hip/knee	No	56.6 ± 10.1
Togo et al, 2022 (53)	Retrospective	Database	Japan	114,078	All regions	No	70.9 ± 12.1
Tomeczkowski et al, 2014 (63)	Retrospective	Database	Germany	163,800	All regions	No	NR
Ullal et al, 2010 (47)	Retrospective	Tertiary	US	154	Knee	No	62.3 ± 7.8
Milano et al, 2016 (48)	Prospective	Primary	Spain	1,152	Knee	No	67.9 ± 6.8
Wang et al, 2019 (49)	Retrospective	Primary	China	212,546	All regions	No	65.5 ± 8.1
Wilson et al, 2015 (50)	Retrospective	Primary	Spain	238,536	All regions	No	67 ± 12.0
Wu et al, 2012 (69)	Retrospective	Database	US	96,666	All regions	No	65.2 ± NR
Xue et al, 2018 (12)	Retrospective	Database	China (Taiwan)	3,4338	All regions	No	61.9 ± 8.2
Yeh et al, 2021 (51)	Retrospective	Database	China (Taiwan)	13,520	All regions	No	50.1 ± 12.7
Yu et al, 2017 (64)	Retrospective	Primary	UK	432,343	All regions	Yes	67.2 ± NR
Zeng et al, 2019 (52)	Retrospective	Primary	UK	88,902	Knee, hip, and hand	No	70.1 ± 9.5

* NR = not reported; OA = osteoarthritis; ref. = reference.

Caribbean at 68.5% (95% CI 66.8, 70.1; n = 1 study, I^2 = 0%, high quality of evidence) (33), East Asia and Pacific at 46.8% (95% CI 35.0, 58.9; n = 10 studies, I^2 = 31.7%, high quality of evidence) (12,25,32,38,40,51,53,58–60), Europe and Central Asia at 40.2% (95% CI 31.8, 49.3; n = 23 studies, I^2 = 12.2%, moderate quality of evidence) (26,27,28,29,39,42,43,45,46,48,50,52,54–56,61–68), Middle East and North Africa at 34.1% (95% CI 33.9, 34.3; n = 1 study, I^2 = 0%, high quality of evidence) (44), and North America at 32.6% (95% CI 16.9, 53.6; n = 12 studies, I^2 = 11.0%, moderate quality of evidence) (11,30,31,34–36,41,69–73). The stratified

Table 2. Risk of bias scores*

Author, year (ref.)	1. Target population a close representation of the national population?	2. The sampling frame a representation of the target population?	3. Was random selection or census used to select the sample?	4. Was the likelihood of nonresponse bias minimal?	5. Were data collected directly from the subjects?	6. Was an acceptable case definition used?	7. Was a valid/ reliable instrument used to measure the parameter of interest?	8. Was the same mode of data collection used?	9. Was the length of the shortest prevalence period for the parameter of interest appropriate?	10. Appropriate numerators and denominators used?	11. Overall score
Akazawa et al, 2019 (25)	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Alacqua et al, 2008 (68)	Low	High	Low	Low	Low	Low	High	Low	Low	Low	Low
Arboleya et al, 2003 (26)	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Barcella et al, 2019 (27)	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Bennell et al, 2021 (60)	Low	High	Low	Low	Low	Low	Low	Low	Low	Low	Low
Castaño Carou et al, 2015 (28)	High	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Chandan et al, 2021 (55)	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Colombo et al, 2021 (54)	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Cunnington et al, 2008 (73)	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Denoeud et al, 2005 (29)	High	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Dominick et al, 2003 (30)	High	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Dominick et al, 2003 (31)	Low	High	Low	Low	Low	Low	High	Low	High	Low	Moderate
Ebata-Kogure et al, 2020 (32)	High	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Fallach et al, 2021 (33)	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Gore et al, 2011 (35)	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Gore et al, 2011 (36)	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Gore et al, 2012 (34)	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Gupta et al, 2018 (37)	Low	High	Low	Low	Low	High	High	Low	Low	Low	Moderate
Barbero et al, 2017 (67)	High	High	High	High	High	High	High	Low	Low	Low	High

(Continued)

Table 2. (*Cont'd*)

Author, year (ref.)	1. Target population a close representation of the national population?	2. The sampling frame a representation of the target population?	3. Was random selection or census used to select the sample?	4. Was the likelihood of nonresponse bias minimal?	5. Were data collected directly from the subjects?	6. Was an acceptable case definition used?	7. Was a valid/ reliable instrument used to measure the parameter of interest?	8. Was the same mode of data collection used?	9. Was the length of the shortest prevalence period for the parameter of interest appropriate?	10. Appropriate numerators and denominators used?	11. Overall score
Hsu et al, 2017 (38)	Low	Low	Low	Low	Low	High	High	Low	Low	Low	Low
Jackson et al, 2017 (39)	Low	High	Low	Low	Low	Low	High	Low	Low	Low	Low
Kanneppady et al, 2017 (72)	Low	High	Low	Low	Low	High	High	Low	Low	Low	Moderate
Kikuchi et al, 2021 (40)	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Lanas et al, 2011 (62)	Low	Low	Low	Low	Low	High	High	Low	Low	Low	Low
Li et al, 2022 (71)	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
McDonald and Walsh, 2012 (41)	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Patel et al, 2020 (70)	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Paterson et al, 2018 (58)	High	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
Pontes et al, 2018 (42)	Low	High	Low	Low	Low	Low	Low	Low	Low	Low	Low
Rajamäki et al, 2019 (43)	Low	High	Low	Low	Low	High	High	Low	Low	Low	Moderate
Reginato et al, 2015 (41)	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Reijman et al, 2005 (61)	High	High	Low	Low	Low	Low	High	Low	Low	Low	Moderate
Richette et al, 2011 (45)	High	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Russo et al, 2003 (65)	Low	High	Low	Low	Low	Low	High	Low	Low	Low	Low
Sakai et al, 2019 (59)	High	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Shelbaya et al, 2018 (11)	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Spitaels et al, 2020 (66)	High	High	Low	Low	Low	Low	High	Low	Low	Low	Moderate
Stambuk et al, 1989 (56)	Low	High	Low	Low	Low	High	High	Low	Low	Low	Moderate
Subramanian et al, 2020 (57)	High	High	Low	Low	Low	Low	High	Low	Low	Low	Moderate

(Continued)

Table 2. (*Cont'd*)

Author, year (ref.)	1. Target population a close representation of the national population?	2. The sampling frame a representation of the target population?	3. Was random selection or census used to select the sample?	4. Was the likelihood of nonresponse bias minimal?	5. Were data collected directly from the subjects?	6. Was an acceptable case definition used?	7. Was a valid/ reliable instrument used to measure the parameter of interest?	8. Was the same mode of data collection used?	9. Was the length of the shortest prevalence period for the parameter of interest appropriate?	10. Appropriate numerators and denominators used?	11. Overall score
Summanen et al, 2021 (46)	High	Low	Low	High	Low	Low	High	High	Low	Low	High
Togo et al, 2022 (53)	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Tomeczkowski et al, 2014 (63)	Low	Low	Low	High	Low	Low	High	Low	Low	Low	Low
Ullal et al, 2010 (47)	High	High	Low	Low	Low	High	High	Low	Low	Low	High
Milano et al, 2016 (48)	High	High	High	High	High	High	High	High	Low	Low	High
Wang et al, 2019 (49)	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Wilson et al, 2015 (50)	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Wu et al, 2012 (69)	Low	Low	Low	High	Low	High	High	High	Low	Low	High
Xue et al, 2018 (12)	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Yeh et al, 2021 (51)	High	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Yu et al, 2017 (64)	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low
Zeng et al, 2019 (52)	High	Low	Low	Low	Low	Low	High	Low	Low	Low	Low

* Ref. = reference.

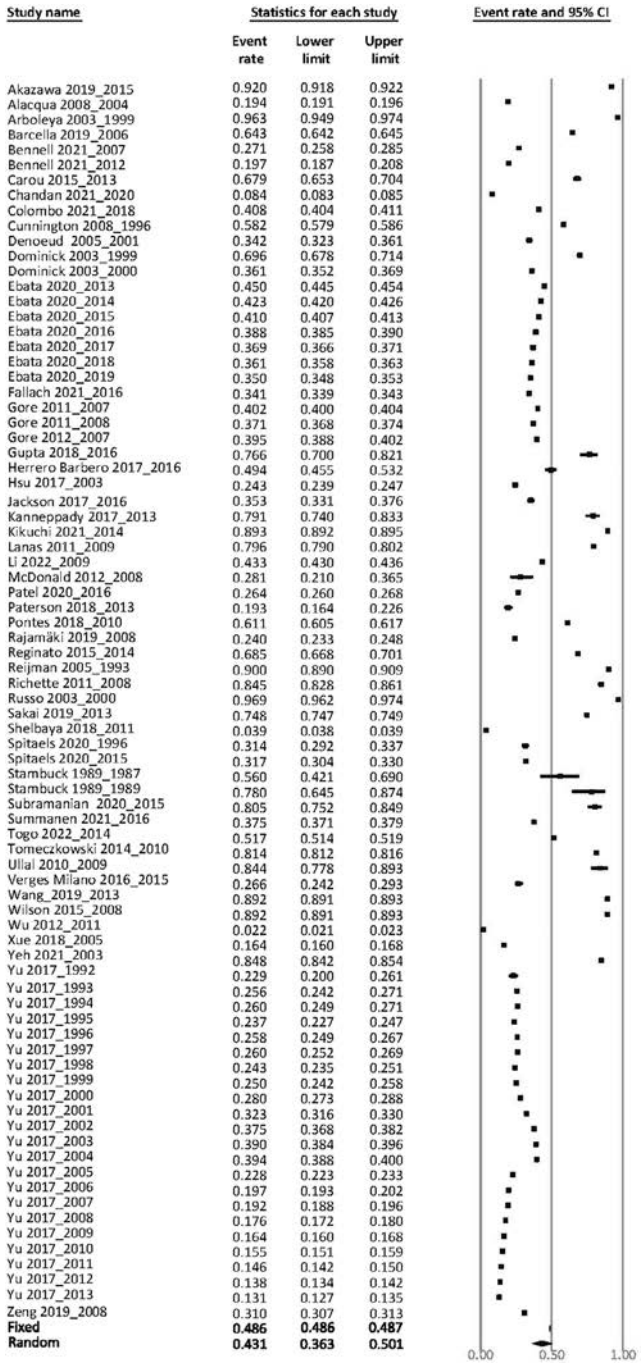

Figure 2. Proportion of participants with osteoarthritis prescribed a nonsteroidal antiinflammatory drug. The study name reports the name of the first author and publication year, followed by the associated data year.

analyses results are summarized in Table 3 and the forest plot shown in Supplementary Figure 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25157.

Prescribing estimates across country income status. Based on WHO income status, the pooled estimate of NSAIDs prescribing in high-income countries was 40.3% (95% CI 33.6, 47.4; n = 46 studies, I^2 = 8.5%, moderate quality of evidence) (11,12,25–36,38–43,45,46,48,50–56,58–64,65–73), greater in middle-income, including, respectively, lower-middle and upper-middle income countries, 83.4% (95% CI 74.8, 89.4; n = 4 studies, I^2 = 0%, moderate quality of evidence) (37,47,49,57) and 68.5% (95% CI 66.8, 70.1; n = 1 study, I^2 = 0%, high quality of evidence) (44). There were no studies from low-income countries. The stratified analyses results are summarized in Table 3 and the forest plot shown in Supplementary Figure 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25157.

The proportion of participants using different types of NSAIDs and dose. Fourteen studies (12,27,28,30,37,42,46,48,52,57,59,61,68,71) reported specific types of NSAIDs prescribed. A summary of the types of NSAIDs reported is shown in Supplementary Figure 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25157. Individual NSAIDs reported included aceclofenac, celecoxib, dexibuprofen, dexketoprofen, diclofenac, etodolac, etoricoxib, flurbiprofen, ibuprofen, indomethacin, ketoprofen, ketorolac, lornoxicam, meloxicam, nabumetone, naproxen, nimesulide, oxaprozin, piroxicam, rofecoxib, and tenoxicam. The most frequently reported prescribed NSAIDs in our sample was diclofenac, ibuprofen, and naproxen. High heterogeneity prevented pooling. Four studies (26–28,67) reported dosages. A summary of reported doses is detailed in Supplementary Table 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25157.

The proportion of participants using nonselective and COX-2 selective NSAIDs. Twenty-five studies provided data on the types of NSAIDs prescribed to patients with osteoarthritis. The pooled estimate of COX-2 selective NSAIDs was 11.0% (95% CI 8.0, 14.8; n = 23 studies, I^2 = 51.8%, moderate quality of evidence) (11,12,27,28,30,31,34–38,46,52,57,59,62,61,65–68,71,73) compared to nonselective NSAIDs at 34.5% (95% CI 27.0, 42.8; n = 23 studies, I^2 = 48.8%, moderate quality of evidence) (12,27,28,30,31,34,36–38,46,48,52,57,59,61,62,63,65–68,71,73). The stratified analyses results are summarized in Table 3 and the forest plot shown in Supplementary Figure 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25157.

Prescribing estimates per mode of delivery and mode of action. Ten studies (25,28,33,39,40,42,45,49,50,66) provided data on how NSAIDs were delivered. A summary is shown in Supplementary Figure 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25157, grouping NSAIDs as either oral, topical, transdermal patch, or suppository, and grouping them as systemic and topical. High heterogeneity prevented pooling.

DISCUSSION

Our review established that 4 in every 10 participants diagnosed with osteoarthritis seeking health care were

Table 3. Summary of estimates from subgroup analyses*

	Studies, no.	I^2 value, %	Event rate (95%CI)
Osteoarthritis site			
Hip	3	0	0.349 (0.238, 0.479)
Knee	11	28.8	0.463 (0.369, 0.559)
Spine	1	0	0.669 (0.666, 0.672)
WHO regions			
East Asia and Pacific	10	31.7	0.468 (0.350, 0.589)
Europe and Central Asia	23	12.2	0.402 (0.318, 0.493)
Latin America and Caribbean	1	0	0.685 (0.668, 0.701)
Middle East and North Africa	1	0	0.341 (0.339, 0.343)
North America	12	11.0	0.326 (0.169, 0.536)
South Asia	4	3.0	0.834 (0.748, 0.894)
Income status			
High income	46	8.5	0.403 (0.336, 0.474)
Lower to middle income	4	0	0.834 (0.748, 0.894)
Upper to middle income	1	0	0.685 (0.668, 0.701)
NSAID type†			
Aceclofenac	6	–	0.143 (0.044, 0.376)
Celecoxib	7	–	0.033 (0.019, 0.055)
Dexibuprofen	1	–	0.000 (0.000, 0.000)
Dexketoprofen	2	–	0.055 (0.004, 0.470)
Diclofenac	13	–	0.133 (0.080, 0.213)
Etodolac	1	–	0.121 (0.086, 0.167)
Etoricoxib	5	–	0.023 (0.006, 0.078)
Flurbiprofen	1	–	0.001 (0.001, 0.001)
Ibuprofen	10	–	0.106 (0.046, 0.226)
Indomethacin	2	–	0.009 (0.005, 0.015)
Ketoprofem	1	–	0.042 (0.041, 0.043)
Ketorolac	1	–	0.010 (0.005, 0.020)
Lornoxicam	2	–	0.041 (0.003, 0.392)
Meloxicam	4	–	0.041 (0.006, 0.227)
Nabumetone	1	–	0.072 (0.068, 0.077)
Naproxen	10	–	0.047 (0.027, 0.078)
Nimesulide	1	–	0.111 (0.109, 0.112)
Oxaprozin	1	–	0.043 (0.040, 0.047)
Piroxicam	5	–	0.022 (0.010, 0.050)
Rofecoxib	3	–	0.022 (0.012, 0.042)
Rofecoxib/etoricoxib/valdecoxib	1	–	0.176 (0.128, 0.237)
Tenoxicam	1	–	0.003 (0.003, 0.003)
Selective versus nonselective			
Selective	23	51.8	0.110 (0.080, 0.148)
Nonselective to selective	23	48.8	0.345 (0.270, 0.428)
Delivery mode†			
Oral	10	–	0.387 (0.233, 0.568)
Patch	1	–	0.068 (0.066, 0.069)
Suppository	1	–	0.002 (0.002, 0.002)
Topical	1	–	0.212 (0.118, 0.350)
Mode of action†			
Systemic	10	–	0.400 (0.253, 0.568)
Topical	4	–	0.212 (0.118, 0.350)

* 95% CI = 95% confidence interval; NSAID = nonsteroidal antiinflammatory drug; WHO = World Health Organization.
† High heterogeneity present, except when 1 study was present.

prescribed a type of NSAID over 30 years. Prescribing was greater in middle-income countries, but there was no evidence available from low-income countries. NSAID prescribing was influenced by geographic region, and there has been a decrease in prescribing over time. Half of the included studies reported details on the types of NSAIDs prescribed, in which prescribing of nonselective NSAIDs was more prevalent than selective NSAID prescribing. Data were limited on prescribing for spine-related osteoarthritis, but NSAID prescribing was prevalent in approximately one-third of participants with hip-related osteoarthritis and nearly half in those with knee osteoarthritis.

Our review with a large sample is the first to examine the extent of NSAID prescribing for the clinical management of osteoarthritis and the potential factors associated with prescribing. Our thorough and sensitive search was conducted without

restrictions and used backward and forward reference and author citation tracking. The limitations of this study include some reporting bias, as most studies did not use a validated measurement instrument, and the use of observational studies, which is unavoidable in prevalence-based studies. We acknowledge that osteoarthritis can affect any joint, and clinical management can vary, and we conducted meta-regression to explore factors associated with NSAID prescribing, However, other factors than what we were able to include in the analysis, such as patient-related factors, were unlikely to contribute to prescribing, as only 42% of the variance was explained with the included study-related factors. We noted that data were limited on prescribing for spine-related osteoarthritis and on specific dosing regimens (regular or "when needed" use patterns), dose form, and duration. Our estimates are likely be an underestimate of actual NSAID prescribing, as some NSAIDs are available over-the-counter and do not always need a prescription. Only 1 study (39) reported the inclusion of NSAIDs prescribed as over-the-counter, and there was no difference in the estimates from clinical records of prescribing versus dispensing claims records. Our estimates could also be an underestimate. Our post hoc sensitivity analysis explored meta-analysis robustness, as there can be variance from studies contributing proportional data when close to 0 and 1. The analysis revealed higher pooled estimates.

The prevalence of NSAID prescribing to participants in primary and tertiary care with osteoarthritis was greater than in many reports of NSAID prescribing in the general population (74,75), for example, 16% in 2015 in the US (76), 22% in tertiary care in Nigeria (77), and 36% in Malaysian primary care (78). Half of the studies included in this review were from Europe. Included European studies as well as studies from high-income countries saw a rate of NSAID prescribing for osteoarthritis similar to what the literature indicates, as the general NSAID prescribing rate in the general population is lower than 40% (74,75,79). NSAID prescribing can differ between countries but also between populations, such as in older populations, where NSAID prescribing has been reported to be as high as 55% (80). Geographic differences of NSAID prescribing may be related to variance in the under- or overuse of medicines and variances in medical systems between different countries, including differences in reimbursement policies, national education campaigns for clinicians to promote the judicious use of NSAIDs (81,82) and marketing practices (83).

The majority of included studies were from high-income countries. Previous studies (84,85) determining prescribing patterns and use of NSAIDs in the general population have observed similar findings. The number of studies of prescribing patterns from middle-income countries continues to be limited. The few studies from middle-income countries suggested that NSAID prescribing is greater than in high-income countries. There could be several reasons to explain these differences, such as the availability and low cost of NSAIDs, and a greater number of NSAIDs may require prescription rather than being available over-the-counter compared to high-income countries. However, the extent of NSAID prescribing for osteoarthritis in low-income countries, and whether this prescribing has changed over time, is uncertain. The decrease in NSAID prescribing noted over time in our review coincides with the increase in opioid prescribing (84) for chronic noncancer pain over the last 2 decades, although recent opioid mitigation strategies following rises in opioid-related harms have begun to take effect. Previous studies have found that NSAID prescribing in the general population from high-income countries has also decreased over time (83,86,87).

The focus of this review was to determine NSAID prescribing among patients diagnosed with osteoarthritis. Therefore, we are still unclear about the prevalence of NSAID prescribing and use among people who self-reported nonclinically diagnosed osteoarthritis. We noticed that most studies (85%) did not require radiographic evidence for confirmation of osteoarthritis in their inclusion criteria. The use of NSAIDs may be higher than our pooled estimates and future research could explore differences between NSAIDs use and prescribing rates to understand adherence to clinical recommendations. Understanding the differences between NSAID prescribing and utilization can identify scenarios where overprescribing occurs.

ACKNOWLEDGMENTS

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

AUTHOR CONTRIBUTIONS

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

Study conception and design. Mathieson, Kobayashi, Abdel Shaheed, Simic, Machado, McLachlan.

Acquisition of data. Yang, Mathieson, Kobayashi, Abdel Shaheed, Nogueira, Simic, Machado, McLachlan.

Analysis and interpretation of data. Yang, Mathieson, Kobayashi, Abdel Shaheed, Nogueira, Simic, Machado, McLachlan.

REFERENCES

1. Litwic A, Edwards MH, Dennison EM, et al. Epidemiology and burden of osteoarthritis. Br Med Bull 2013;105:185–99.
2. Kolasinski SL, Neogi T, Hochberg MC, et al. 2019 American College of Rheumatology/Arthritis Foundation guideline for the management of osteoarthritis of the hand, hip, and knee. Arthritis Care Res (Hoboken) 2020;72:149–62.
3. Bannuru RR, Osani MC, Vaysbrot EE, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis [review]. Osteoarthritis Cartilage 2019;27:1578–89.
4. Zeng C, Wei J, Persson MS, et al. Relative efficacy and safety of topical non-steroidal anti-inflammatory drugs for osteoarthritis: a

systematic review and network meta-analysis of randomised controlled trials and observational studies [review]. Br J Sports Med 2018;52:642–50.
5. Da Costa BR, Reichenbach S, Keller N, et al. Effectiveness of non-steroidal anti-inflammatory drugs for the treatment of pain in knee and hip osteoarthritis: a network meta-analysis. Lancet 2017;390: e21–33.
6. Zhang W, Moskowitz RW, Nuki G, et al. OARSI recommendations for the management of hip and knee osteoarthritis, part II: OARSI evidence-based, expert consensus guidelines [review]. Osteoarthritis Cartilage 2008;16:137–62.
7. Al-Shidhani A, Al-Rawahi N, Al-Rawahi A, et al. Non-steroidal anti-inflammatory drugs (NSAIDs) use in primary health care centers in A'Seeb, Muscat: a clinical audit. Oman Med J 2015;30:366–71.
8. Bijlsma JW, Berenbaum F, Lafeber FP. Osteoarthritis: an update with relevance for clinical practice [review]. Lancet 2011;377:2115–26.
9. Yeomans ND, Graham DY, Husni ME, et al. Randomised clinical trial: gastrointestinal events in arthritis patients treated with celecoxib, ibuprofen or naproxen in the PRECISION trial. Aliment Pharmacol Ther 2018;47:1453–63.
10. Nørgård B, Pedersen L, Johnsen SP, et al. COX-2-selective inhibitors and the risk of upper gastrointestinal bleeding in high-risk patients with previous gastrointestinal diseases: a population-based case-control study. Aliment Pharmacol Ther 2004;19:817–25.
11. Shelbaya A, Solem CT, Walker C, et al. The economic and clinical burden of early versus late initiation of celecoxib among patients with osteoarthritis. Clinicoecon Outcomes Res 2018;10:213–22.
12. Xue YH, Peng YS, Ting HF, et al. Etoricoxib and diclofenac might reduce the risk of dementia in patients with osteoarthritis: a nationwide, population-based retrospective cohort study. Dement Geriatr Cogn Disord 2018;45:262–71.
13. McPherson ML, Cimino NM. Topical NSAID formulations [review]. Pain Med 2013;14 Suppl 1:S35–9.
14. Rannou F, Pelletier JP, Martel-Pelletier J. Efficacy and safety of topical NSAIDs in the management of osteoarthritis: evidence from real-life setting trials and surveys. Semin Arthritis Rheum 2016;4 Suppl: S18–21.
15. Machado GC, Maher CG, Ferreira PH, et al. Non-steroidal anti-inflammatory drugs for spinal pain: a systematic review and meta-analysis [review]. Ann Rheum Dis 2017;76:1269–78.
16. Puljak L, Marin, A, Vrdoljak, D, et al. Celecoxib for osteoarthritis [review]. Cochrane Database Syst Rev 2017;5:CD009865.
17. American College of Rheumatology Ad Hoc Group on use of selective and nonselective nonsteroidal antiinflammatory drugs. Recommendations for use of selective and nonselective nonsteroidal antiinflammatory drugs: an American College of Rheumatology white paper. Arthritis Care Res (Hoboken) 2008;59:1058–73.
18. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev 2021;10:89.
19. WHO Collaborating Centre for Drug Statistics Methodology. Guidelines for ATC classification and DDD assignment. Oslo (Norway): Norwegian Institute of Public Health; 2020. p. 1–251.
20. World Health Organization. Countries. URL: https://www.who.int/countries.
21. World Bank. World bank country and lending groups. 2019. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups.
22. Hoy D, Brooks P, Woolf A, et al. Assessing risk of bias in prevalence studies: modification of an existing tool and evidence of interrater agreement. J Clin Epidemiol 2012;65:934–9.
23. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses [review]. BMJ 2003;327:557–60.
24. Guyatt GH, Oxman AD, Kunz R, et al. What is "quality of evidence" and why is it important to clinicians? [review] BMJ 2008;336:995–8.
25. Akazawa M, Mimura W, Togo K, et al. Patterns of drug treatment in patients with osteoarthritis and chronic low back pain in Japan: a retrospective database study. J Pain Res 2019;12:1631–48.
26. Arboleya LR, de la Figuera E, García MS, et al. Management pattern for patients with osteoarthritis treated with traditional non-steroidal anti-inflammatory drugs in Spain prior to introduction of coxibs. Curr Med Res Opin 2003;19:278–87.
27. Barcella CA, Lamberts M, McGettigan P, et al. Differences in cardiovascular safety with non-steroidal anti-inflammatory drug therapy: a nationwide study in patients with osteoarthritis. Basic Clin Pharmacol Toxicol 2019;124:629–41.
28. Castaño Carou A, Pita Fernández S, Pértega Díaz S, et al. Clinical profile, level of affection and therapeutic management of patients with osteoarthritis in primary care: the Spanish multicenter study EVALÚA. Reumatol Clin 2015;11:353–60.
29. Denoeud L, Mazieres B, Payen-Champenois C, et al. First line treatment of knee osteoarthritis in outpatients in France: adherence to the EULAR 2000 recommendations and factors influencing adherence. Ann Rheum Dis 2005;64:70–4.
30. Dominick KL, Ahern FM, Gold CH, et al. Gender differences in NSAID use among older adults with osteoarthritis. Ann Pharmacother 2003; 37:1566–71.
31. Dominick KL, Dudley TK, Grambow SC, et al. Racial differences in health care utilization among patients with osteoarthritis. J Rheumatol 2003;30:2201–6.
32. Ebata-Kogure N, Murakami A, Nozawa K, et al. Treatment and healthcare cost among patients with hip or knee osteoarthritis: a cross-sectional study using a real-world claims database in Japan between 2013 and 2019. Clin Drug Investig 2020;40:1071–84.
33. Fallach N, Chodick G, Tirosh M, et al. Pain pharmacotherapy in a large cohort of patients with osteoarthritis: a real-world data analysis. Rheumatol Ther 2021;8:1129–41.
34. Gore M, Sadosky AB, Leslie DL, et al. Therapy switching, augmentation, and discontinuation in patients with osteoarthritis and chronic low back pain. Pain Pract 2012;12:457–68.
35. Gore M, Sadosky A, Leslie D, et al. Patterns of therapy switching, augmentation, and discontinuation after initiation of treatment with select medications in patients with osteoarthritis. Clin Ther 2011;33: 1914–31.
36. Gore M, Tai KS, Sadosky A, et al. Clinical comorbidities, treatment patterns, and direct medical costs of patients with osteoarthritis in usual care: a retrospective claims database analysis. J Med Econ 2011;14:497–507.
37. Gupta R, Malhotra A, Malhotra P. Study of prescription pattern of drugs used in the treatment of osteoarthritis in a tertiary care teaching hospital: an observational study. Int J Res Med Sci 2018;6:985.
38. Hsu PS, Lin HH, Li CR, et al. Increased risk of stroke in patients with osteoarthritis: a population-based cohort study. Osteoarthritis Cartilage 2017;25:1026–31.
39. Jackson H, Barnett LA, Jordan KP, et al. Patterns of routine primary care for osteoarthritis in the UK: a cross-sectional electronic health records study. BMJ Open 2017;7:e019694.
40. Kikuchi S, Togo K, Ebata N, et al. Database analysis on the relationships between nonsteroidal anti-inflammatory drug treatment variables and incidence of acute myocardial infarction in Japanese patients with osteoarthritis and chronic low back pain. Adv Ther 2021;38:1601–13.
41. McDonald DD, Walsh S. Older adult osteoarthritis pain management: results from the 2008 National Ambulatory Medical Care Survey. J Am Acad Nurse Pract 2012;24:107–12.

42. Pontes C, Marsal JR, Elorza JM, et al. Analgesic use and risk for acute coronary events in patients with osteoarthritis: a population-based, nested case-control study. Clin Ther 2018;40:270–83.
43. Rajamäki TJ, Puolakka PA, Hietaharju A, et al. Use of prescription analgesic drugs before and after hip or knee replacement in patients with osteoarthritis. BMC Musculoskelet Disord 2019;20:427.
44. Reginato AM, Riera H, Vera M, et al. Osteoarthritis in Latin America: study of demographic and clinical characteristics in 3040 patients. J Clin Rheumatol 2015;21:391–7.
45. Richette P, Hilliquin P, Bertin P, et al. Comparison of general practitioners and rheumatologists' prescription patterns for patients with knee osteoarthritis. BMC Musculoskelet Disord 2011;12:72.
46. Summanen M, Ukkola-Vuoti L, Kurki S, et al. The burden of hip and knee osteoarthritis in Finnish occupational healthcare. BMC Musculoskelet Disord 2021;22:501.
47. Ullal SD, Narendranath S, Kamath RK, et al. Prescribing pattern for osteoarthritis in a tertiary care hospital. J Clin Diagn Res 2010;4: 2421–6.
48. Verges Milano JV, Barbero MH, Basallote SG, et al. EMARTRO: study to assess comorbidities in patients with symptomatic knee osteoarthritis [abstract]. Osteoarthritis Cartilage 2016;24:S194–5.
49. Wang F, Wang Y, Xing H, et al. Drug prescription patterns in osteoarthritis patients in a tertiary care hospital in China. Indian J Pharm Sci 2019;81:555–61.
50. Wilson N, Sanchez-Riera L, Morros R, et al. Drug utilization in patients with OA: a population-based study. Rheumatology (Oxford) 2015;54: 860–7.
51. Yeh CH, Chang WL, Chan PC, et al. Women with osteoarthritis are at increased risk of ischemic stroke: a population-based cohort study. J Epidemiol 2021;31:628–34.
52. Zeng C, Dubreuil M, LaRochelle MR, et al. Association of tramadol with all-cause mortality among patients with osteoarthritis. JAMA 2019;321:969–82.
53. Togo K, Ebata N, Yonemoto N, et al. Safety risk associated with use of nonsteroidal anti-inflammatory drugs in Japanese elderly compared with younger patients with osteoarthritis and/or chronic low back pain: a retrospective database study. Pain Pract 2022;22:200–9.
54. Colombo GL, Heiman F, Peduto I. Utilization of healthcare resources in osteoarthritis: a cost of illness analysis based on real-world data in Italy. Ther Clin Risk Manag 2021;17:345–56.
55. Chandan JS, Zemedikun DT, Thayakaran R, et al. Nonsteroidal antiinflammatory drugs and susceptibility to COVID-19. Arthritis Rheumatol 2021;73:731–9.
56. Stambuk B, Dürrigl T, Zenić N, et al. Use of non-steroidal antiinflammatory drugs in specialty polyclinic practice. Reumatizam 1989;36: 45–8. In Croatian.
57. Subramanian A, Adhimoolam M, Gnanasegaran S, et al. Study to assess the prescription pattern and quality of life in osteoarthritis patients at a tertiary care hospital. J Clin Diagn Res 2020;14:1–5.
58. Paterson KL, Harrison C, Britt H, et al. Management of foot/ankle osteoarthritis by Australian general practitioners: an analysis of national patient-encounter records. Osteoarthritis Cartilage 2018;26: 888–94.
59. Sakai R, Cho S, Jang EJ, et al. International descriptive study for comparison of treatment patterns in patients with knee osteoarthritis between Korea and Japan using claims data. Int J Rheum Dis 2019; 22:2052–8.
60. Bennell KL, Bayram C, Harrison C, et al. Trends in management of hip and knee osteoarthritis in general practice in Australia over an 11-year window: a nationwide cross-sectional survey. Lancet Reg Health West Pac 2021;12:100187.
61. Reijman M, Bierma-Zeinstra SM, Pols HA, et al. Is there an association between the use of different types of nonsteroidal antiinflammatory drugs and radiologic progression of osteoarthritis? The Rotterdam study. Arthritis Rheum 2005;52:3137–42.
62. Lanas A, Garcia-Tell G, Armada B, et al. Prescription patterns and appropriateness of NSAID therapy according to gastrointestinal risk and cardiovascular history in patients with diagnoses of osteoarthritis. BMC Med 2011;9:38.
63. Tomeczkowski J, Wimmer A, Pappert K. Pain therapy for osteoarthritis in Germany: analysis of sickness fund claims data. Value Health 2014;17:A375.
64. Yu D, Jordan KP, Bedson J, et al. Population trends in the incidence and initial management of osteoarthritis: age-period-cohort analysis of the Clinical Practice Research Datalink, 1992–2013. Rheumatology (Oxford) 2017;56:1902–17.
65. Russo P, Capone A, Attanasio E, et al. Pharmacoutilization and costs of osteoarthritis: changes induced by the introduction of a cyclooxygenase-2 inhibitor into clinical practice. Rheumatology (Oxford) 2003;42:879–87.
66. Spitaels D, Mamouris P, Vaes B, et al. Epidemiology of knee osteoarthritis in general practice: a registry-based study. BMJ Open 2020;10: e031734.
67. Herrero Barbero MH, Gimenez S, Vergara J, et al. Pharmacological approach of knee osteoarthritis treatment in primary care in Spain [abstract]. Ann Rheum Dis 2017;76:1456.
68. Alacqua M, Trifirò G, Cavagna L, et al. Prescribing pattern of drugs in the treatment of osteoarthritis in Italian general practice: the effect of rofecoxib withdrawal. Arthritis Rheum 2008;59:568–74.
69. Wu N, Chen SY, Andrews JS, et al. Factors predicting pain medication selection among patients with osteoarthritis. Value Health 2012; 15:A48.
70. Patel J, Ladani A, Sambamoorthi N, et al. A machine learning approach to identify predictors of potentially inappropriate nonsteroidal anti-inflammatory drugs (NSAIDs) use in older adults with osteoarthritis. Int J Environ Res Public Health 2020;18:155.
71. Li L, Marozoff S, Lu N, et al. Association of tramadol with all-cause mortality, cardiovascular diseases, venous thromboembolism, and hip fractures among patients with osteoarthritis: a population-based study. Arthritis Res Ther 2022;24:85.
72. Kanneppady SS, Kanneppady SK, Raghavan V, et al. Prescription pattern of primary osteoarthritis in tertiary medical centre. J Health Allied Sci NU 2017;07:037–42.
73. Cunnington M, Webb D, Qizilbash N, et al. Risk of ischaemic cardiovascular events from selective cyclooxygenase-2 inhibitors in osteoarthritis. Pharmacoepidemiol Drug Saf 2008;17:601–8.
74. Bedene A, Strmljan A, van Dorp EL, et al. Comparison of two different analgesic prescription strategies and healthcare systems: Slovenia vs. The Netherlands. Front Pain Res (Lausanne) 2021;2:723797.
75. Brusselaers N, Lagergren J. Maintenance use of non-steroidal anti-inflammatory drugs and risk of gastrointestinal cancer in a nationwide population-based cohort study in Sweden. BMJ Open 2018;8: e021869.
76. Fassio V, Aspinall SL, Zhao X, et al. Trends in opioid and nonsteroidal anti-inflammatory use and adverse events. Am J Manag Care 2018;3: e61–72.
77. Awodele O, Fadipe A, Adekoya M, et al. Prescribing pattern of nonsteroidal anti-inflammatory drugs at the outpatient pharmacy department of a university teaching hospital in Nigeria. Ghana Med J 2015; 49:25–9.
78. Dhabali AA, Awang R, Hamdan Z, et al. Associations between prescribing nonsteroidal anti-inflammatory drugs and the potential prescription-related problems in a primary care setting. Int J Clin Pharmacol Ther 2012;50:851–61.
79. Tielemans MM, van Rossum LG, Eikendal T, et al. Gastrointestinal symptoms in NSAID users in an 'average risk population': results of

a large population-based study in randomly selected Dutch inhabitants. Int J Clin Pract 2014;68:512–9.

80. Abdu N, Mosazghi A, Teweldemedhin S, et al. Non-steroidal anti-inflammatory drugs (NSAIDs): usage and co-prescription with other potentially interacting drugs in elderly. A cross-sectional study. PLoS One 2020;15:e0238868.
81. Bruyndonckx R, Verhoeven V, Anthierens S, et al. The implementation of academic detailing and its effectiveness on appropriate prescribing of pain relief medication: a real-world cluster randomized trial in Belgian general practices. Implement Sci 2018;13:6.
82. Langaas HC, Hurley E, Dyrkorn R, et al. Effectiveness of an academic detailing intervention in primary care on the prescribing of non-steroidal anti-inflammatory drugs. Eur J Clin Pharmacol 2019; 75:577–86.
83. Barozzi N, Sketris I, Cooke C, et al. Comparison of nonsteroidal anti-inflammatory drugs and cyclooxygenase-2 (COX-2) inhibitors use in Australia and Nova Scotia (Canada). Br J Clin Pharmacol 2009;68: 106–15.
84. Mathieson S, Wertheimer G, Maher CG, et al. What proportion of patients with chronic noncancer pain are prescribed an opioid medicine? Systematic review and meta-regression of observational studies. J Intern Med 2020;287:458–74.
85. Wertheimer G, Mathieson S, Maher CG, et al. The prevalence of opioid analgesic use in people with chronic noncancer pain: systematic review and meta-analysis of observational studies. Pain Med 2021; 22:506–17.
86. Chen Y, Bedson J, Hayward RA, et al. Trends in prescribing of non-steroidal anti-inflammatory drugs in patients with cardiovascular disease: influence of national guidelines in UK primary care. Fam Pract 2018;35:426–32.
87. Davis JS, Lee HY, Kim J, et al. Use of non-steroidal anti-inflammatory drugs in US adults: changes over time and by demographic. Open Heart 2017;4:e000550.

Arthritis Care & Research
Vol. 75, No. 11, November 2023, pp 2359–2368
DOI 10.1002/acr.25142

Efficacy and Safety of Sublingual Cyclobenzaprine for the Treatment of Fibromyalgia: Results From a Randomized, Double-Blind, Placebo-Controlled Trial

Seth Lederman,[1] Lesley M. Arnold,[2] Ben Vaughn,[3] Mary Kelley,[1] and Gregory M. Sullivan[1]

Objective. To evaluate the efficacy and safety of TNX-102 SL, a once-nightly sublingual formulation of cyclobenzaprine, in reducing pain in patients with fibromyalgia (FM).

Methods. RELIEF was a double-blind, randomized, placebo-controlled trial. Overall, 503 patients received TNX-102 SL 2.8 mg for 2 weeks, followed by 5.6 mg for 12 weeks (248 patients), or matching placebo (255 patients). The primary end point was change from baseline at week 14 in the weekly average of daily pain scores. Secondary end points included Patient Global Impression of Change (PGIC) scores, Fibromyalgia Impact Questionnaire Revised (FIQR) scores, Patient-Reported Outcomes Measurement Information System (PROMIS) Sleep Disturbance and Fatigue scores, and daily sleep quality. Safety was assessed by adverse event (AE) reporting.

Results. Reduction in daily pain from baseline at week 14 was significantly greater with TNX-102 SL (least squares [LS] mean change –1.9 [95% confidence interval (95% CI) –2.1, –1.7]) versus placebo (LS mean change –1.5 [95% CI –1.7, –1.3]; $P = 0.01$). TNX-102 SL was not associated with significant improvement in PGIC at week 14 but was associated with improvements in FIQR scores, PROMIS scores, and daily sleep quality. Overall, 59.7% of patients receiving TNX-102 SL and 46.3% receiving placebo reported treatment-emergent AEs; the most common were oral hypoesthesia (17.3% with TNX-102 SL versus 0.4% with placebo), oral paresthesia (5.6% versus 0.4%, respectively), and product taste abnormal (4.4% versus 0.4%, respectively).

Conclusion. In this phase III, randomized, controlled trial of patients with FM, treatment with TNX-102 SL was associated with significant reductions in daily pain and was safe and well tolerated.

INTRODUCTION

Patients with fibromyalgia (FM) experience chronic multisite pain, disturbed sleep, and chronic fatigue; other common features of FM include tenderness, cognitive impairment (difficulty concentrating, forgetfulness, and disorganized thinking), musculoskeletal stiffness, and environmental sensitivity (intolerance to bright lights, loud noises, perfumes, and cold) (1–3). FM affects an estimated 2.0–6.4% of people in the US, and prevalence tends to be higher in women than in men (4–6).

Chronic pain has historically been seen as the defining feature of FM; however, recent diagnostic criteria from the Analgesic, Anesthetic, and Addiction Clinical Trial Translations Innovations Opportunities and Networks (ACTTION)–American Pain Society Pain Taxonomy identified fatigue and sleep problems as key associated symptoms occurring in most patients (3). Patients with FM experience nonrestorative, poor-quality sleep with longer duration of wakefulness and shorter sleep duration; they additionally spend more time in light sleep and have greater sleep latency than healthy controls (7,8).

The US Food and Drug Administration (FDA) has approved 3 medications to treat FM: pregabalin, duloxetine, and milnacipran (9). In a survey of 800 patients with FM, 70% reported using pain medications prescribed by their physician, but only 19% reported being very satisfied with their current treatment, whereas 28% were not very satisfied or not at all satisfied (10). Furthermore, 35% of patients reported that chronic widespread pain was not well managed by their current treatment, and 22% reported that fatigue, joint pain, and concentration difficulties were also not well managed (10), highlighting an unmet need for additional treatments that can adequately treat multiple symptoms of FM.

ClinicalTrials.gov identifier: NCT04172831.
Supported by Tonix Pharmaceuticals, Inc.
[1]Seth Lederman, MD, Mary Kelley, MPH, Gregory M. Sullivan, MD: Tonix Pharmaceuticals, Inc., Chatham, New Jersey; [2]Lesley M. Arnold, MD: University of Cincinnati College of Medicine, Cincinnati, Ohio; [3]Ben Vaughn, MS: Rho Inc., Chapel Hill, North Carolina.

Author disclosures are available online at https://onlinelibrary.wiley.com/doi/10.1002/acr.25142.
Address correspondence via email to Gregory M. Sullivan, MD, at Greg.Sullivan@TonixPharma.com.
Submitted for publication October 10, 2022; accepted in revised form May 2, 2023.

SIGNIFICANCE & INNOVATIONS

- Fibromyalgia (FM) affects an estimated 2–6% of people in the US, with symptoms including chronic multisite pain and disturbed sleep.
- TNX-102 SL is a novel, low-dose, sublingual formulation of cyclobenzaprine designed to be taken at bedtime.
- In this phase III, double-blind, multicenter, placebo-controlled trial, treatment with TNX-102 SL was associated with significant reductions in daily pain relative to placebo and was well tolerated by patients with FM.
- Secondary results suggested that treatment with TNX-102 SL improved sleep, reduced fatigue, and positively impacted a broad array of associated symptoms and dysfunction in patients with FM.

Oral cyclobenzaprine, although not approved to treat FM, was initially studied as a possible treatment owing to pharmacologic similarities with tricyclic antidepressants (11). A meta-analysis of 5 randomized controlled trials found that treatment with oral cyclobenzaprine was associated with significant improvements in sleep after 4, 8, and 12 weeks of treatment, with significant improvements in pain at the 4-week time point (12). On the basis of this evidence, EULAR recommends oral cyclobenzaprine to treat sleep disturbance associated with FM, while cautioning that it has not been demonstrated to improve pain after 12 weeks of treatment (13). Treatment with oral cyclobenzaprine was also associated with adverse reactions of dry mouth, somnolence, dizziness, drowsiness/fatigue, and weight gain (14,15).

TNX-102 SL (sublingual cyclobenzaprine) is a low-dose sublingual formulation of cyclobenzaprine designed for transmucosal absorption to produce diurnal variation in peak-to-trough drug levels. Peak cyclobenzaprine levels (mean ± SD steady state maximum concentration 11,206 ± 5,659 pg/ml) are achieved a median of 5 hours after dosing of TNX-102 SL 5.6 mg, near the middle of the sleep phase, with daytime concentrations falling to a mean ± SD minimum concentration of 4,910 ± 3,531 pg/ml (16). Compared with oral (immediate release) cyclobenzaprine, treatment with TNX-102 SL resulted in more rapid cyclobenzaprine absorption, 54% higher bioavailability, and reduced 24-hour plasma exposure to the cyclobenzaprine metabolite norcyclobenzaprine (16).

Cyclobenzaprine antagonizes serotonin 2A (5-HT_{2A}), 5-HT_{2B}, and 5-HT_{2C}; histamine 1 (H_1); α_{1A}-, α_{1B}-, α_{2B}-, and α_{2C}-adrenergic; and muscarinic 1 (M_1) acetylcholine receptors, as well as relatively weakly inhibits activity at the norepinephrine transporter (NET) and serotonin transporter (SERT) (17). A single 5-mg dose of oral cyclobenzaprine results in accumulation of norcyclobenzaprine, which generally antagonizes the same receptors as cyclobenzaprine but with lower potency, although it is a more potent inhibitor of NET (17,18). Treatment with TNX-102 SL bypasses first-pass hepatic metabolism and results in lower norcyclobenzaprine exposure relative to the parent compared with treatment with oral cyclobenzaprine (16).

Unlike other cyclobenzaprine formulations, TNX-102 SL is intended to be taken once daily at bedtime to enhance nocturnal treatment effects, while limiting daytime side effects such as drowsiness, dry mouth, and dizziness. The use of TNX-102 SL 2.8 mg administered daily at bedtime to reduce pain and improve sleep quality in patients with FM was supported by the results of a proof-of-concept trial and more recently by the results of phase IIb (ClinicalTrials.gov identifier: NCT01903265) and phase III (ClinicalTrials.gov identifier: NCT02436096) trials (19–21). This phase III, double-blind, multicenter, placebo-controlled trial (RELIEF) (ClinicalTrials.gov identifier: NCT04172831) evaluated the efficacy and safety of once-daily TNX-102 SL 5.6 mg taken at bedtime for the treatment of patients with FM.

PATIENTS AND METHODS

Patients. Patients were 18–65 years old with a diagnosis of primary FM according to the 2016 revision to the 2010/2011 FM diagnostic criteria (2), including generalized pain, symptoms present at a similar level for ≥3 months, widespread pain index ≥7 and symptom severity score ≥5 *or* widespread pain index between 4 and 6 and symptom severity score ≥9, and with no other disorder that could explain the pain. All patients additionally had to meet diary-based criteria for FM-related pain, reporting a 7-day average daily pain score of ≥4 and ≤9 on an 11-point (scale of 0–10) numeric rating scale (NRS) in the 7 days before randomization.

Patients were excluded if they had medical or psychiatric conditions (e.g., infection, systemic autoimmune disease, other pain syndromes, bipolar mood disorders, psychotic disorders, increased suicide risk, substance use disorders, pregnancy, or nursing) that could affect their ability to participate in the trial or their well-being during the trial. Patients were additionally excluded if they were unwilling or unable to withdraw from prohibited medications, including duloxetine, milnacipran, pregabalin, gabapentin, tricyclic antidepressants, trazodone, opioids, naltrexone, benzodiazepines, and other formulations of cyclobenzaprine. Allowed concomitant treatments included acetaminophen and nonsteroidal antiinflammatory drugs (NSAIDs) for pain, certain antidepressants such as selective serotonin reuptake inhibitors, nonhypnotic insomnia treatments such as sedating antihistamines (other than during baseline data collection), and any other nonexcluded medications taken at a stable dose. Patients receiving permitted treatment for depression had to be clinically stable for 3 months before randomization. Additionally, patients were excluded if they had a day with a pain score of 10, >2 days with a pain score of <4, or <5 days with any recorded pain score in the 7 days before randomization.

Study design. This was a phase III, double-blind, multicenter, randomized, placebo-controlled 14-week trial of TNX-102 SL 5.6 mg daily at bedtime for the treatment of FM. The study consisted of an up to 35-day screening period, which included up to 28 days to wash out excluded medications, followed by a 7-day run-in baseline period prior to randomization, and a 14-week treatment period. The washout of excluded medications could be extended up to 49 days with medical monitor approval, and follow-up could be extended if an adverse event (AE) was unresolved at the end of the treatment period. Patients visited the study center at screening; at baseline/randomization; and at weeks 2, 6, 10, and 14. The study was conducted from November 2019 to October 2020; telephone visits were available if needed because of the COVID-19 pandemic.

The study protocol was written in accordance with the Declaration of Helsinki and approved by a central institutional review board (Advarra), which also conducted site-level reviews and approval. One academic site additionally received an internal institutional review board approval. The study was conducted in accordance with FDA regulations and International Council of Harmonisation Guidelines for Good Clinical Practice. All patients provided written informed consent.

At the screening visit, patients provided in-clinic 7-day recall NRS pain intensity scores. Patients were also trained to record average daily pain intensity and sleep quality using the electronic diary each evening and, if necessary, developed a plan to stop use of excluded medications >21 days before randomization. Baseline data were collected using the daily diary during the 7-day run-in period immediately prior to the randomization visit, and 14 days after stopping any excluded medications.

During the randomization visit, patients were randomly assigned in a 1:1 ratio to receive TNX-102 SL 5.6 mg or matching placebo tablets using interactive response technology. Randomization was stratified by study center, and all patients and study personnel were blinded with regard to treatment assignment. Investigators could unblind patients with serious AEs if necessary for patient safety. Patients were instructed to take 1 tablet (TNX-102 SL 2.8 mg or placebo) daily at bedtime for the first 2 weeks; at the week 2 visit, patients were instructed to take 2 tablets (TNX-102 SL 5.6 mg or placebo) through the duration of the trial. If a patient reported AEs that were intolerable, the dose could be reduced to 1 tablet at the discretion of the investigator.

Assessments and end points. *Efficacy assessments.* Patients completed a daily electronic diary from screening through the week 14 end-of-study visit; diary entries were to be completed in the evening before dosing. In the diary entries, patients reported their average pain severity over the previous 24 hours using an 11-point NRS, with 0 indicating no pain and 10 indicating worst possible pain. They also reported the quality of sleep for the prior night on an 11-point NRS, with 0 indicating the best possible sleep and 10 indicating the worst possible sleep quality.

At the week 2, 6, 10, and 14 visits, patients completed the Patient Global Impression of Change (PGIC) questionnaire, the revised Fibromyalgia Impact Questionnaire (FIQR), and the Patient-Reported Outcomes Measurements Information System (PROMIS) questionnaires for sleep disturbance and fatigue; the FIQR and PROMIS questionnaires were also completed on day 1. The PGIC is a validated instrument that gauges patient assessment of change in their overall condition on a 7-point scale, with 1 indicating "very much improved," 4 indicating "no change," and 7 indicating "very much worse" (22). The FIQR is a validated questionnaire comprising 21 questions, all of which are framed in the context of the last 7 days and use a 0–10 scale NRS (with 10 indicating worst) to assess 3 domains: functional (9 questions), symptoms (10 questions), and overall impact (2 questions). PROMIS is an initiative developed by the National Institutes of Health to assess patient-reported outcomes across chronic conditions (23). The PROMIS Sleep Disturbance instrument focuses on perceptions of sleep quality, restorative sleep, difficulty getting to sleep, and satisfaction with sleep, whereas the Fatigue instrument focuses on the experience of fatigue (including intensity, frequency, and duration) and its impact on physical, mental, and social activities (23).

Safety assessments. Patients were monitored for AEs throughout the study after informed consent was obtained. FM symptoms were reported as AEs if they worsened or became more frequent and were outside the normal experience in the opinion of the participant. An examination of the oral cavity was conducted at the screening visit and at each in-clinic study visit; patients with any AE involving the oral cavity that was possibly related to study drug were encouraged to contact the investigative site as soon as possible and come in for an unscheduled visit to conduct an oral cavity examination.

End points. The primary efficacy end point was change from baseline to week 14 in the weekly average of daily pain NRS severity scores from the daily diary. Predefined secondary end points were the proportion of patients considered PGIC responders (reporting a rating of "much improved" or "very much improved") at week 14 and change from baseline at week 14 in the FIQR symptoms domain score, FIQR function domain score, PROMIS score for sleep disturbance, PROMIS score for fatigue, and weekly average of daily sleep quality NRS scores (from the daily diary).

Safety end points included treatment-emergent AEs (TEAEs) and the Beck Depression Inventory-II (BDI-II). TEAEs were defined as either a new-onset AE or an AE (or medical history) present prior to randomization that increased in severity or frequency and were summarized by severity and relationship to study drug.

Statistical analysis. Planned enrollment was 470 patients (235 per arm), providing a power of ≥90% to detect an effect size

of 0.3 using a 2-sided *t*-test with an alpha level of 0.05. A prespecified interim analysis of efficacy was performed by a separate unblinded team after randomization of 236 patients; the primary outcome was tested at a 1-sided alpha level of 0.005, and the corresponding final 2-sided alpha level to account for the first stage alpha was set to 0.0452.

Efficacy was assessed in the intent-to-treat population, which included all randomized patients; safety was assessed in the safety population, which included all patients who took investigational product. The primary end point was analyzed using a restricted maximum likelihood–based repeated-measures approach with data imputed using multiple imputation with the assumption that patients who discontinued owing to AEs and lack of efficacy would revert to the distribution of baseline values, whereas other dropouts and intermittent missing data would follow the distribution of values in the assigned treatment group. The models included the fixed, categorical effects of treatment, site, study week, and treatment-by–study week interaction, as well as the fixed covariates of baseline value and baseline value score-by–study week interaction. A 2-sided *P* value of <0.0452 was set as significant, owing to the alpha spend from the interim analysis.

Secondary end points were tested sequentially to adjust for multiplicity and to control for type I error; if any secondary end point did not produce a significant result (i.e., $P \geq 0.0452$), the remaining secondary end points were considered descriptive and reported with nominal *P* values. For the first secondary end point, a categorical analysis of PGIC was performed using a logistic regression model for each visit with effects for treatment and investigative site; patients with missing data were considered nonresponders. A post hoc analysis was conducted to compare the proportion of patients with any improvement in PGIC in the TNX-102 SL and placebo arms using the same methodology. An approach identical to the primary analysis was used for the remaining (continuous) secondary end points. AE end points were summarized descriptively. BDI-II scores were analyzed using a mixed model for repeated measures approach.

RESULTS

Patients. A total of 503 patients were randomized to receive either TNX-102 SL (n = 248) or placebo (n = 255) (Figure 1). Among randomized patients, 417 patients (82.9%) completed the trial; 17.7% of patients receiving TNX-102 SL discontinued the trial compared with 16.5% of patients receiving placebo. Baseline characteristics were well balanced between treatment groups with no notable differences between groups for demographic variables (Table 1). Patients experienced FM symptoms for a mean of 9.1 years and were broadly representative of the diagnosed clinical FM population. Overall, 95% of patients were female, and the mean age was 49.6 years. Most patients were White (87%) and college educated (82%). The most commonly used classes of concomitant medications were NSAIDs (45.5%), other analgesics and antipyretics (25.0%), treatments for ulcer or acid reflux (21.7%), protocol-allowed antidepressants (19.5%), and nonsedating antihistamines (16.3%) (Table 1). Sedating antihistamines were used by 5.4% of patients.

Efficacy. *Primary end point.* At week 14, the least squares (LS) mean change from baseline (95% CI) in the weekly average of daily diary pain scores was significantly greater in the TNX-102 SL group (LS mean –1.91 [95% CI –2.15, –1.68]) than in the placebo group (LS mean –1.51 [95% CI –1.74, –1.28]; $P = 0.01$) (Figure 2A). In an exploratory analysis, a greater proportion of patients receiving TNX-102 SL (46.8%) experienced a ≥30% reduction in daily pain at week 14 relative to placebo (34.9%) (odds ratio [OR] 1.67 [95% CI 1.16, 2.40]; $P = 0.006$) (Figure 2B).

Secondary end points. At week 14, the percentage of patients classified as responders (reporting a rating of "much improved" or higher) on the PGIC was numerically higher in the TNX-102 SL group than in the placebo group; however, the OR (95% CI) did not achieve statistical significance (OR 1.44 [95% CI 0.99, 2.10]; $P = 0.058$) (Figure 3A). Because the first key secondary end point did not reach significance, analyses of remaining secondary end points are considered descriptive and are reported with nominal *P* values. The percentage of PGIC responders was greater in the TNX-102 SL group than in the placebo group at week 10 (nominal $P = 0.044$). In a post hoc analysis, the percentage of patients reporting any improvement in PGIC (ratings of "minimally improved," "much improved," or "very much improved") was greater in the TNX-102 SL group than in the placebo group at week 14 (OR 1.62 [95% CI 1.13, 2.33]; nominal $P = 0.009$).

The LS mean change from baseline in FIQR symptom domain scores at week 14 was greater in the TNX-102 SL group than in the placebo group (–18.38 and –14.05, respectively; LS mean difference [LSMD] –4.3 [95% CI –7.4, –1.2]; nominal $P = 0.007$) (Figure 3B). Similarly, the LS mean change from baseline in FIQR function domain scores at week 14 was greater in the TNX-102 SL group than in the placebo group (–13.63 and –9.25, respectively; LSMD –4.4 [95% CI –7.7, –1.0]; nominal $P = 0.009$) (Figure 3C). In addition, the LS mean change from baseline in the FIQR impact domain scores at week 14 was greater in the TNX-102 SL group than in the placebo group (–4.73 and –3.34, respectively; LSMD –1.4 [95% CI –2.3, –0.5]; nominal $P = 0.002$). Nominally significant improvements relative to placebo were observed across a broad range of individual FIQR items, including level of pain, energy, stiffness, quality of sleep, depression, memory problems, tenderness to touch, and sensory (environmental) sensitivity (Supplementary Table 1, available on the *Arthritis Care & Research* website at https://onlinelibrary.wiley.com/doi/10.1002/acr.25142/abstract).

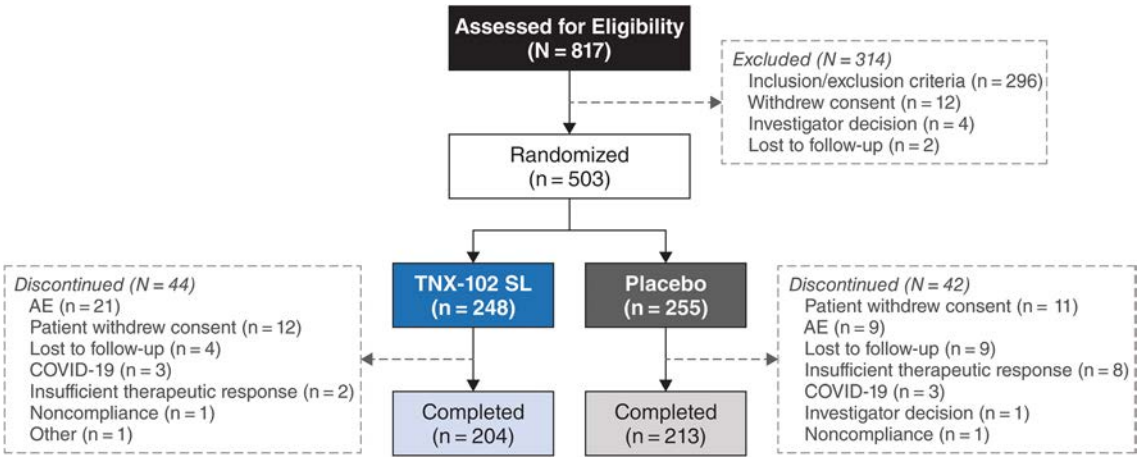

Figure 1. Patient disposition. A total of 503 patients were randomized. All patients received ≥1 dose of study drug, and 417 patients (82.9%) completed the study. AE = adverse event; SL = sublingual.

At week 14, the LS mean change from baseline in PROMIS Sleep Disturbance scores was greater in the TNX-102 SL group than in the placebo group (–9.47 and –6.55, respectively; LSMD –2.9 [95% CI –4.5, –1.3]; nominal $P < 0.001$) (Figure 4A). The LS mean change from baseline in PROMIS Fatigue scores at week 14 was greater in the TNX-102 SL group than in the

Table 1. Patient demographic and baseline clinical characteristics*

Characteristic	TNX-102 SL (n = 248)	Placebo (n = 255)	Total (N = 503)
Age, years	50.0 ± 9.4	49.3 ± 10.2	49.6 ± 9.8
Female sex, no. (%)	232 (93.5)	247 (96.9)	479 (95.2)
Hispanic/Latino ethnicity, no. (%)	43 (17.3)	42 (16.5)	85 (16.9)
Race, no. (%)			
White	222 (89.5)	216 (84.7)	438 (87.1)
Black/African American	19 (7.7)	20 (7.8)	39 (7.8)
Asian	2 (0.8)	5 (2.0)	7 (1.4)
American Indian/Alaska Native	1 (0.4)	2 (0.8)	3 (0.6)
Native Hawaiian/other Pacific Islander	0 (0.0)	1 (0.4)	1 (0.2)
Multiple†	3 (1.2)	9 (3.5)	12 (2.4)
Body mass index, kg/m^2	32.4 ± 6.6	31.6 ± 6.3	32.0 ± 6.4
Some college or greater education, no. (%)	205 (82.7)	212 (83.1)	417 (82.9)
Duration of fibromyalgia, years	9.2 ± 8.4	9.0 ± 8.1	9.1 ± 8.2
Diary pain score	6.1 ± 1.1	6.0 ± 1.1	6.1 ± 1.1
PROMIS Sleep Disturbance score	58.3 ± 6.4	60.0 ± 6.5	59.2 ± 6.5
PROMIS Fatigue score	62.5 ± 6.0	63.5 ± 6.3	63.0 ± 6.1
FIQR Symptoms domain score	52.3 ± 14.2	54.2 ± 14.5	53.3 ± 14.3
FIQR Function domain score	35.6 ± 20.5	37.1 ± 20.0	36.4 ± 20.2
FIQR Impact domain score	9.7 ± 5.2	9.7 ± 5.0	9.7 ± 5.1
BDI-II total score	8.3 ± 6.4	9.2 ± 6.2	8.8 ± 6.3
C-SSRS items, no. (%)			
Any suicidal ideation, lifetime	42 (16.5)	39 (15.7)	81 (16.1)
Any suicidal ideation, past 6 months	3 (1.2)	3 (1.2)	6 (1.2)
Actual suicide attempt, lifetime	11 (4.3)	5 (2.0)	16 (3.2)
Concomitant medication use, no. (%)‡			
NSAIDs	110 (44.4)	119 (46.7)	229 (45.5)
Other analgesics and antipyretics	62 (25.0)	64 (25.1)	126 (25.0)
Drugs for peptic ulcer and GERD	55 (22.2)	54 (21.2)	109 (21.7)
Antidepressants	42 (16.9)	56 (22.0)	98 (19.5)
Nonsedating antihistamines	44 (17.7)	38 (14.9)	82 (16.3)

* Except where indicated otherwise, values are the mean ± SD. BDI-II = Beck Depression Inventory-II; C-SSRS = Columbia Suicide Severity Rating Scale; FIQR = Fibromyalgia Impact Questionnaire Revised; GERD = gastroesophageal reflux disease; NSAIDs = nonsteroidal antiinflammatory drugs; PROMIS = Patient-Reported Outcomes Measurement Information System; SL = sublingual.
† Patients who selected >1 race.
‡ Concomitant medications taken by ≥15% of patients in the overall population.

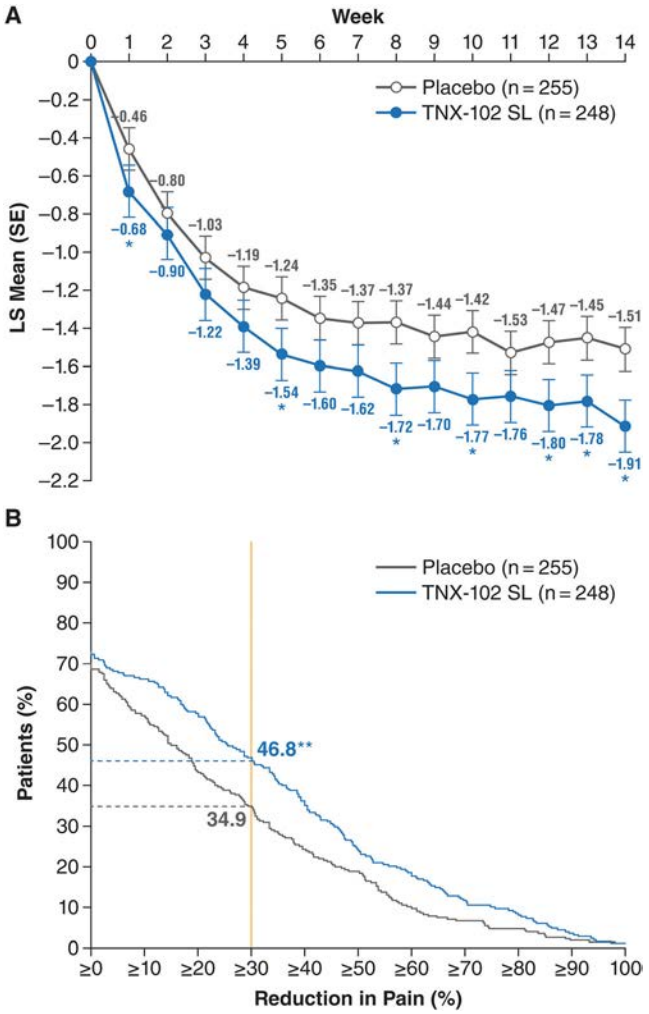

Figure 2. **A,** Least squares (LS) mean (SE) change from baseline in the weekly average of the daily pain numeric rating scale severity scores from the daily diary. **B,** Patients achieving various levels of improvement in pain intensity. The yellow line indicates the percentage of patients who experienced a ≥30% reduction in daily pain. SL = sublingual. * = $P < 0.0452$; ** = $P = 0.006$.

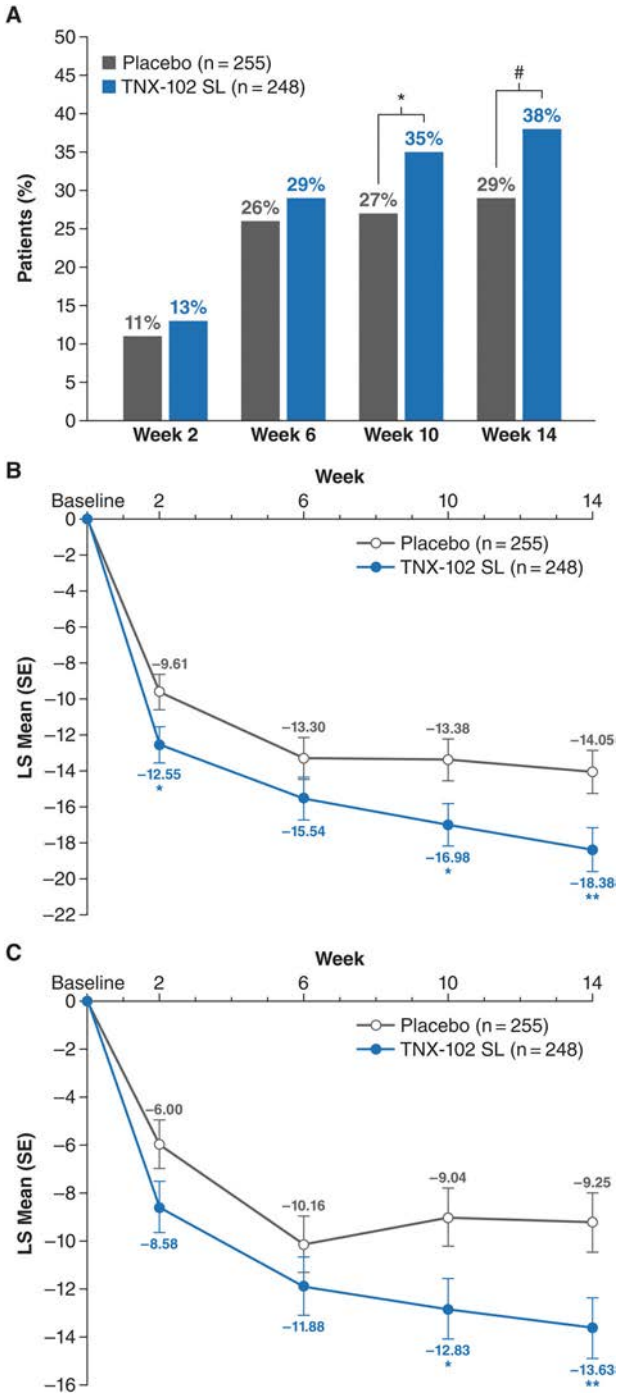

Figure 3. **A,** Percentage of patients indicating "much improved" or "very much improved" on the Patient Global Impression of Change. **B,** Least squares (LS) mean (SE) change from baseline on the Fibromyalgia Impact Questionnaire Revised (FIQR) Symptoms domain score. **C,** LS mean (SE) change from baseline in FIQR Function domain score. SL = sublingual. # = $P = 0.058$; * = nominal $P < 0.0452$; ** = nominal $P < 0.01$.

placebo group (–7.99 and –6.21, respectively; LSMD –1.8 [95% CI –3.3, –0.3]; nominal $P = 0.018$) (Figure 4B). Similarly, the LS mean change from baseline in daily diary sleep scores at week 14 was greater in the TNX-102 SL group than in the placebo group (–2.04 and –1.45, respectively; LSMD –0.6 [95% CI –0.9, –0.3]; nominal $P < 0.001$) (Figure 4C).

Safety. Overall, 8.9% of patients receiving TNX-102 SL discontinued study drug owing to a TEAE compared with 3.9% of patients receiving placebo. Of patients receiving TNX-102 SL, 6.0% underwent a dose reduction compared with 3.1% of patients receiving placebo (Supplementary Table 2, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25142/abstract). There was no clear pattern in the AEs that led to dose reductions or discontinuations. TEAEs were reported in 59.7% of patients in the TNX-102 SL group and 46.3% of patients in the placebo group; most TEAEs with TNX-102 SL were mild or moderate in

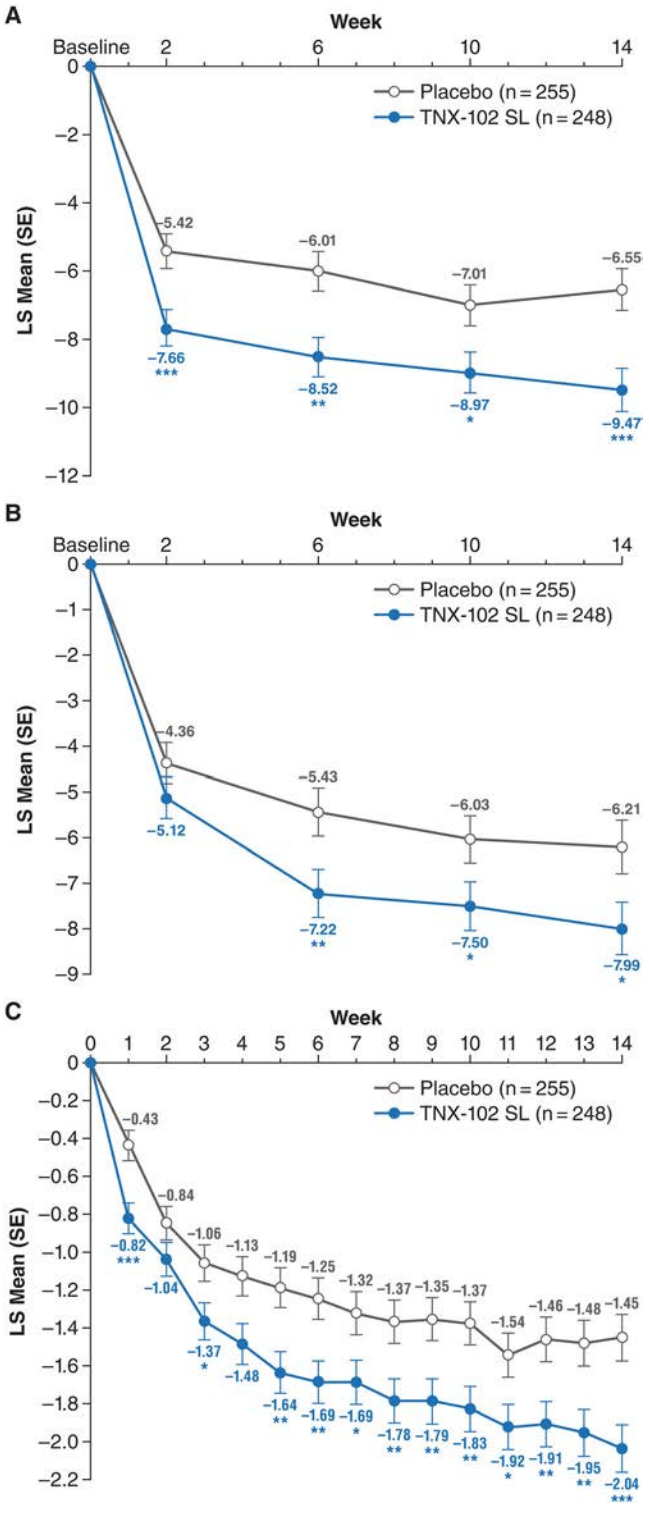

Figure 4. **A,** Least squares (LS) mean (SE) change from baseline in Patient-Reported Outcomes Measurement Information System (PROMIS) scores for sleep disturbance. **B,** LS mean (SE) change from baseline in PROMIS scores for fatigue. **C,** Weekly average of diary-reported sleep quality. SL = sublingual. * = nominal $P < 0.0452$; ** = nominal $P < 0.01$; *** = nominal $P < 0.001$.

severity (Table 2). The most commonly reported TEAEs with TNX-102 SL were oral hypoesthesia (17.3% versus 0.4% with placebo), oral paresthesia (5.6% versus 0.4%, respectively), and product taste abnormal (4.4% versus 0.4%, respectively). Most oral cavity TEAEs (72.8%) in the TNX-102 SL group were temporally related to dosing and lasted <60 minutes (71.9%). In the placebo group, 32.3% of oral cavity TEAEs were temporally related to dosing, and 60% lasted <60 minutes. Overall, 32.2% of patients reported ≥1 TEAE considered possibly related to study drug.

BDI-II total scores improved from baseline to week 14 in both groups, with greater improvements in the TNX-102 SL group than in the placebo group (LSMD −1.2 [95% CI −2.1, −0.3]; nominal $P = 0.012$). Improvements in BDI-II total scores reflect improvements in symptoms including crying, indecisiveness, pessimism, and self-dislike (Supplementary Table 3, http://onlinelibrary.wiley.com/doi/10.1002/acr.25142/abstract).

DISCUSSION

Among patients with FM, treatment with TNX-102 SL was associated with significantly reduced daily pain compared with placebo at week 14, meeting the primary end point of the RELIEF trial. Nominally significant reductions in pain were apparent at week 1 and were seen at weeks 5, 8, 10, 12, and 13, in addition to the week 14 end point, suggesting rapid and sustained pain relief. A total of 46.8% of patients achieved ≥30% pain reduction with TNX-102 SL, significantly more than with placebo, and a higher percentage of patients responded to TNX-102 SL compared with placebo across the range of possible pain reduction thresholds (Figure 2B).

Although the first key secondary efficacy end point of PGIC responders did not meet the prespecified threshold for statistical significance, nominally significant improvements in PGIC in the TNX-102 SL arm were apparent at week 10 and in a post hoc analysis of any improvement in PGIC at week 14, suggesting that TNX-102 SL treatment may be associated with clinically meaningful improvements in this patient-reported outcome. TNX-102 SL treatment was also associated with clinically meaningful and nominally significant improvements in the FIQR function, impact, and symptom domains, with improvement in the symptom domain reflecting improvement in a range of symptoms including pain, energy level, quality of sleep, depression, and memory and sensory issues. Improvements in sleep disturbance, fatigue, and sleep quality with TNX-102 SL relative to placebo were supported by results from PROMIS assessments and diary sleep quality scores. In addition, consistent with improvements in symptoms of depression reported in the FIQR, changes from baseline in BDI-II total and individual item scores suggested that some depressive symptoms improved more in patients receiving TNX-102 SL than in patients receiving placebo.

Table 2. Summary of safety*

Adverse event	TNX-102 SL (n = 248)	Placebo (n = 255)	Total (N = 503)
Treatment duration, mean ± SD days	88.9 ± 26.2	88.7 ± 24.9	88.8 ± 25.5
≥1 TEAE	148 (59.7)	118 (46.3)	266 (52.9)
Possibly related to treatment	110 (44.4)	52 (20.4)	162 (32.2)
Severe	11 (4.4)	9 (3.5)	20 (4.0)
Serious†	2 (0.8)	5 (2.0)	7 (1.4)
Oral	101 (40.7)	23 (9.0)	124 (24.7)
Discontinued study drug owing to TEAE	22 (8.9)	10 (3.9)	32 (6.4)
Deaths	0 (0.0)	0 (0.0)	0 (0.0)
TEAEs occurring in ≥3% of patients in the TNX-102 SL group			
Oral cavity AE			
Oral hypoesthesia	43 (17.3)	1 (0.4)	44 (8.7)
Oral paresthesia	14 (5.6)	1 (0.4)	15 (3.0)
Product taste abnormal	11 (4.4)	1 (0.4)	12 (2.4)
Glossodynia	9 (3.6)	2 (0.8)	11 (2.2)
Dry mouth	8 (3.2)	7 (2.7)	15 (3.0)
Systemic AE			
Fatigue	9 (3.6)	4 (1.6)	13 (2.6)
Sedation	9 (3.6)	1 (0.4)	10 (2.0)

* Except where indicated otherwise, values are the no. (%) of patients. AE = adverse event; SL = sublingual; TEAE = treatment-emergent AE.
† No serious TEAEs were considered by the investigator to be related to study drug.

Low baseline scores indicated that most patients enrolled in the study did not have scores above threshold for a diagnosable depressive disorder, although patients could continue permitted antidepressants during the treatment period, which may have contributed to the lower scores. Although further studies of patient-reported outcomes and quality of life may be warranted to better define the clinical impact of TNX-102 SL treatment, these results suggest that treatment with TNX-102 SL can provide clinically meaningful improvements in global symptoms and functioning in patients with FM including pain, sleep, fatigue, affective symptoms, and cognitive symptoms central to the FM diagnosis.

Although 17.7% of patients receiving TNX-102 SL discontinued the trial compared with 16.5% of patients receiving placebo, few patients (8.9%) receiving TNX-102 SL discontinued owing to a TEAE compared with 3.9% in the placebo arm. The most commonly reported TEAEs with TNX-102 SL were orally related. These TEAEs were generally transient, related temporally to dosing, and most likely were a result of sublingual dosing (i.e., administration site reactions). Assessing any impact of oral AEs on patient-reported treatment tolerability and adherence may be an important topic for future studies.

Interestingly, TEAEs of fatigue and sedation were each only reported by 3.6% of patients receiving TNX-102 SL. In comparison, in a published study of patients receiving 5 mg of oral cyclobenzaprine, 29% experienced drowsiness and 6% experienced fatigue (24). The apparent reduction in daytime AEs observed with TNX-102 SL could be explained by the dynamic diurnal peak-to-trough changes in cyclobenzaprine levels with bedtime sublingual transmucosal dosing and possibly reduced exposure to norcyclobenzaprine. Finally, the incidence of TEAEs of dry mouth was similar between the TNX-102 SL (3.2%) and placebo (2.7%) groups. The rate of dry mouth with oral cyclobenzaprine 5 mg was reported to be 21% compared with 7% for placebo (24), and this may also be related to higher NET inhibition from greater accumulation of norcyclobenzaprine relative to the parent when cyclobenzaprine is administered orally rather than sublingually with transmucosal absorption (17).

Although the pathogenesis of FM is not well understood, similarities between the symptoms of patients with FM and healthy controls who were deprived of stage 4 sleep initially led to the hypothesis that nonrestorative sleep may contribute to chronic pain and other FM symptoms (25). Consistent with this model, higher baseline levels of sleep disturbance have been shown to be associated with higher pain levels 1 year later in patients with FM (26). It was also hypothesized that FM-related sleep abnormalities might be mediated by an abnormality in central serotonergic neurotransmission (25).

Current FM therapies have limited efficacy in alleviating symptoms of sleep disturbance and fatigue; pregabalin improves sleep quality but is associated with an AE of fatigue (27,28), whereas duloxetine (29) and milnacipran (30) have shown reduced fatigue in some trials but did not improve sleep quality (31). Although additional studies are needed to confirm the positive effects of TNX-102 SL treatment on sleep, the descriptive results of the current trial suggest that treatment with TNX-102 SL may improve sleep quality and fatigue in patients with FM.

Treatment with oral cyclobenzaprine has not previously been associated with meaningful reductions in FM-associated pain (12,13). However, in the current trial, patients who received treatment with TNX-102 SL reported both significant improvements in daily pain and descriptive improvements in sleep quality and fatigue, demonstrating that the sublingual formulation of cyclobenzaprine and dosing regimen studied here can meaningfully

improve core FM symptoms. Since treatment with TNX-102 SL at bedtime achieves peak cyclobenzaprine levels near the middle of the sleep phase, we hypothesize that improved sleep quality in patients receiving TNX-102 SL may contribute to reduced daytime pain, fatigue, and other FM symptoms.

We hypothesize that TNX-102 SL improves sleep quality primarily by modulating 5-HT_{2A}, α_1-adrenergic, and M_1-muscarinic acetylcholine receptor activity, rather than H_1 receptor activity (17). Thus, short-term use of sedating antihistamines was allowed for intolerable insomnia and was managed by the investigators. Use of benzodiazepines and nonbenzodiazepine hypnotics for intolerable insomnia was excluded because these medications interfere with sleep architecture, which could confound the hypothesized mechanism of TNX-102 SL.

Limitations of the present trial include the 14-week duration, limited sample size, and descriptive nature of the secondary analyses, per the prespecified analysis plan. A total of 95% of the study population was female, although this is consistent with previous reports of the demographic characteristics of patients with diagnosed FM (5). Male, Hispanic/Latino, and non-White patients who meet diagnostic criteria for FM may be underrepresented relative to the real-world population, and future trials should include expanded efforts to recruit patients from these groups. FM symptoms could have been affected by use of allowed analgesics (NSAIDs), sleep aids (antihistamines), or antidepressants in some patients. The trial occurred during the emergence of the COVID-19 pandemic, which may have impacted the trial, such as by reducing patient access to regular medical care. The promising post hoc analysis of PGIC and the nominally significant change in PGIC at week 10 suggest that improvements in PGIC may have reached significance given a larger sample size, and additional expanded studies of TNX-102 SL are warranted.

In conclusion, in this phase III, randomized, controlled trial, treatment with TNX-102 SL was associated with significant reductions in daily pain and was generally safe and well tolerated in patients with FM. Secondary results also suggest that treatment with TNX-102 SL can improve sleep and reduce fatigue, which together with pain are recognized as key FM-associated symptoms.

ACKNOWLEDGMENTS

Medical writing and editorial support were provided by Joseph Kruempel, PhD, CMPP, and Tania R. Iqbal, PhD, CMPP, of The Curry Rockefeller Group, LLC (Tarrytown, NY), and were funded by Tonix Pharmaceuticals, Inc.

AUTHOR CONTRIBUTIONS

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

Study conception and design. Lederman, Arnold, Kelley, Sullivan.

Acquisition of data. Lederman, Arnold, Kelley, Sullivan.

Analysis and interpretation of data. Lederman, Arnold, Vaughn, Sullivan.

ROLE OF THE STUDY SPONSOR

Publication of this article was not contingent upon approval by Tonix Pharmaceuticals, Inc. Tonix Pharmaceuticals, Inc. had a role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

REFERENCES

1. Wolfe F, Clauw DJ, Fitzcharles MA, et al. The American College of Rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity. Arthritis Care Res (Hoboken) 2010;62:600–10.
2. Wolfe F, Clauw DJ, Fitzcharles MA, et al. 2016 revisions to the 2010/2011 fibromyalgia diagnostic criteria. Semin Arthritis Rheum 2016;46:319–29.
3. Arnold LM, Bennett RM, Crofford LJ, et al. AAPT diagnostic criteria for fibromyalgia. J Pain 2019;20:611–28.
4. Queiroz LP. Worldwide epidemiology of fibromyalgia [review]. Curr Pain Headache Rep 2013;17:356.
5. Vincent A, Lahr BD, Wolfe F, et al. Prevalence of fibromyalgia: a population-based study in Olmsted County, Minnesota, utilizing the Rochester Epidemiology Project. Arthritis Care Res (Hoboken) 2013; 65:786–92.
6. Wolfe F, Ross K, Anderson J, et al. The prevalence and characteristics of fibromyalgia in the general population. Arthritis Rheum 1995;38: 19–28.
7. Jacobson SA, Simpson RG, Lubahn C, et al. Characterization of fibromyalgia symptoms in patients 55–95 years old: a longitudinal study showing symptom persistence with suboptimal treatment. Aging Clin Exp Res 2015;27:75–82.
8. Wu YL, Chang LY, Lee HC, et al. Sleep disturbances in fibromyalgia: a meta-analysis of case-control studies. J Psychosom Res 2017;96: 89–97.
9. Forte ML, Butler M, Andrade KE, et al. Treatments for fibromyalgia in adult subgroups [review]. Rockville, MD: Agency for Healthcare Research and Quality (US);2015. URL: https://www.ncbi.nlm.nih.gov/books/NBK274474/.
10. Choy E, Perrot S, Leon T, et al. A patient survey of the impact of fibromyalgia and the journey to diagnosis. BMC Health Serv Res 2010; 10:102.
11. Arnold LM, Keck PE Jr, Welge JA. Antidepressant treatment of fibromyalgia. A meta-analysis and review. Psychosomatics 2000;41: 104–13.
12. Tofferi JK, Jackson JL, O'Malley PG. Treatment of fibromyalgia with cyclobenzaprine: a meta-analysis. Arthritis Rheum 2004;51:9–13.
13. Macfarlane GJ, Kronisch C, Dean LE, et al. EULAR revised recommendations for the management of fibromyalgia. Ann Rheum Dis 2017;76:318–28.
14. Carette S, Bell MJ, Reynolds WJ, et al. Comparison of amitriptyline, cyclobenzaprine, and placebo in the treatment of fibromyalgia. A randomized, double-blind clinical trial. Arthritis Rheum 1994;37:32–40.

15. Bennett RM, Gatter RA, Campbell SM, et al. A comparison of cyclobenzaprine and placebo in the management of fibrositis. A double-blind controlled study. Arthritis Rheum 1988;31:1535–42.
16. Sullivan G, Kiu R, Stillwell H, et al. Steady-state pharmacokinetic properties of a sublingual formulation of cyclobenzaprine (CBP) HCl (TNX-102 SL*): comparison to simulations of oral immediate release CBP [poster]. Presented at the American Society of Clinical Psychopharmacology; 2019 May 28–31; Scottsdale, Arizona.
17. Daugherty B, Gershell L, Lederman S. Cyclobenzaprine (CBP) and its major metabolite norcyclobenzaprine (nCBP) are potent antagonists of human serotonin receptor 2a (5HT2a), histamine receptor H-1 and alpha-adrenergic receptors: mechanistic and safety implications for treating fibromyalgia syndrome by improving sleep quality [abstract no. 960]. Presented at the ACR/ARHP Annual Meeting; 2012 November 9–14; Washington, DC. URL: https://acrabstracts.org/abstract/cyclobenzaprine-cbp-and-its-major-metabolite-norcyclobenzaprine-ncbp-are-potent-antagonists-of-human-serotonin-receptor-2a-5ht2a-histamine-receptor-h-1-and-a-adrenergic-receptors-mechanistic-a/.
18. Daugherty B, Sullivan G, Gershell L, et al. Serotonin receptor profiles of bedtime pharmacotherapies targeting posttraumatic stress disorder (PTSD) [abstract no. 728]. Presented at the Society of Biological Psychiatry 70th Annual Scientific Convention; 2015 May 14–16; Toronto, Ontario, Canada.
19. Lederman S, Gendreau M, Clauw D, et al. Bedtime, rapidly absorbed sublingual cyclobenzaprine (TNX-102 SL) for the treatment of fibromyalgia: results of a phase 2b randomized, double-blind, placebo-controlled study [abstract no. 2309]. Arthritis Rheumatol 2015;67 Suppl 10:1–4046.
20. Moldofsky H, Harris HW, Archambault WT, et al. Effects of bedtime very low dose cyclobenzaprine on symptoms and sleep physiology in patients with fibromyalgia syndrome: a double-blind randomized placebo-controlled study. J Rheumatol 2011;38:2653–63.
21. Sullivan G, Gendreau RM, Gendreau J, et al. A phase 3 randomized, double-blind, placebo-controlled trial of bedtime sublingual cyclobenzaprine (TNX-102 SL) for the treatment of fibromyalgia (FM): evidence for a broad spectrum of activity on the FM syndrome. Arthritis Rheumatol 2019;71 Suppl 10:1–5420.
22. Rampakakis E, Ste-Marie PA, Sampalis JS, et al. Real-life assessment of the validity of patient global impression of change in fibromyalgia. RMD Open 2015;1:e000146.
23. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. J Clin Epidemiol 2010;63:1179–94.
24. Merck & Co, Inc. Flexeril (cyclobenzaprine HCl). Full prescribing information. Merck; 2003. URL: https://www.accessdata.fda.gov/drugsatfda_docs/label/2003/017821s045lbl.pdf
25. Moldofsky H, Scarisbrick P, England R, et al. Musculosketal symptoms and non-REM sleep disturbance in patients with "fibrositis syndrome" and healthy subjects. Psychosom Med 1975;37:341–51.
26. Bigatti SM, Hernandez AM, Cronan TA, et al. Sleep disturbances in fibromyalgia syndrome: relationship to pain and depression. Arthritis Rheum 2008;59:961–7.
27. Arnold LM, Emir B, Pauer L, et al. Time to improvement of pain and sleep quality in clinical trials of pregabalin for the treatment of fibromyalgia. Pain Med 2015;16:176–85.
28. Pfizer. Lyrica (pregabalin). Full prescribing information. New York NY: Pfizer; 2004. URL: https://www.accessdata.fda.gov/drugsatfda_docs/label/2018/021446s035,022488s013lbl.pdf
29. Arnold LM, Wang F, Ahl J, et al. Improvement in multiple dimensions of fatigue in patients with fibromyalgia treated with duloxetine: secondary analysis of a randomized, placebo-controlled trial. Arthritis Res Ther 2011;13:R86.
30. Arnold LM, Gendreau RM, Palmer RH, et al. Efficacy and safety of milnacipran 100 mg/day in patients with fibromyalgia: results of a randomized, double-blind, placebo-controlled trial. Arthritis Rheum 2010;62:2745–56.
31. Welsch P, Üçeyler N, Klose P, et al. Serotonin and noradrenaline reuptake inhibitors (SNRIs) for fibromyalgia [review]. Cochrane Database Syst Rev 2018;2:Cd010292.

Arthritis Care & Research
Vol. 75, No. 11, November 2023, pp 2369–2378
DOI 10.1002/acr.25139

American College of Rheumatology
Empowering Rheumatology Professionals

Validity, Reliability, and Differential Item Functioning of English and French Versions of the 10-Item Connor-Davidson Resilience Scale in Systemic Sclerosis: A Scleroderma Patient-Centered Intervention Network Cohort Study

Marieke A. Neyer,[1] Richard S. Henry,[1] Marie-Eve Carrier,[2] Linda Kwakkenbos,[3] Robyn K. Wojeck,[4] Amy Gietzen,[5] Karen Gottesman,[6] Geneviève Guillot,[7] Amanda Lawrie-Jones,[8] Maureen D. Mayes,[9] Luc Mouthon,[10] Warren R. Nielson,[11] Michelle Richard,[12] Maureen Worron-Sauvé,[13] Daphna Harel,[14] Vanessa L. Malcarne,[15] Susan J. Bartlett,[16] and Brett D. Thombs,[17] on behalf of the SPIN Investigators

Objective. Some individuals with systemic sclerosis (SSc) report positive mental health, despite severe disease manifestations, which may be associated with resilience, but no resilience measure has been validated in SSc. This study was undertaken to assess the validity, reliability, and differential item functioning (DIF) between English- and French-language versions of the 10-item Connor-Davidson Resilience Scale (CD-RISC-10) in SSc.

Methods. Eligible participants were enrolled in the Scleroderma Patient-centered Intervention Network Cohort and completed the CD-RISC-10 between August 2022 and January 2023. We used confirmatory factor analysis (CFA) to evaluate the CD-RISC-10 factor structure and conducted DIF analysis across languages with Multiple Indicators Multiple Causes models. We tested convergent validity with another measure of resilience and measures of self-esteem and depression and anxiety symptoms. We assessed internal consistency and test–retest reliability using Cronbach's alpha and intraclass correlation coefficient (ICC).

Results. A total of 962 participants were included in this analysis. CFA supported a single-factor structure (Tucker–Lewis index = 0.99, comparative fit index = 0.99, root mean square error of approximation = 0.08 [90% confidence interval (90% CI) 0.07, 0.09]). We found no meaningful DIF. Internal consistency was high (α = 0.93 [95% CI 0.92, 0.94]), and we found that correlations with other measures of psychological functioning were moderate to large (|r| = 0.57–0.78) and confirmed study hypotheses. The scale showed good 1–2-week test–retest reliability (ICC 0.80 [95% CI 0.75, 0.85]) in a subsample of 230 participants.

Conclusion. The CD-RISC-10 is a valid and reliable measure of resilience in SSc, with score comparability across English and French versions.

INTRODUCTION

Systemic sclerosis (SSc; scleroderma) is a rare, chronic autoimmune disorder characterized by vascular abnormalities and fibrosis of the skin and internal organs, including the gastrointestinal (GI) tract, lungs, heart, and kidneys (1,2). Disease manifestation is heterogeneous, and the disease course is unpredictable (1,3). Researchers have estimated the standardized mortality rate

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

Supported by the Canadian Institutes of Health Research, the Lady Davis Institute for Medical Research of the Jewish General Hospital, the Jewish General Hospital Foundation, McGill University, the Scleroderma Society of Ontario, Scleroderma Canada, Sclérodermie Québec, Scleroderma Manitoba, Scleroderma Atlantic, the Scleroderma Association of BC, Scleroderma SASK, Scleroderma Australia, Scleroderma New South Wales, and Scleroderma Victoria.

[1]Marieke A. Neyer, BSc, Richard S. Henry, PhD: Lady Davis Institute of the Jewish General Hospital and McGill University, Montreal, Quebec, Canada; [2]Marie-Eve Carrier, MSc: Lady Davis Institute of the Jewish General Hospital, Montreal, Quebec, Canada; [3]Linda Kwakkenbos, PhD: Radboud University Medical Center, Nijmegen, The Netherlands; [4]Robyn K. Wojeck, PhD: University of Rhode Island, Kingston, Rhode Island; [5]Amy Gietzen: National Scleroderma Foundation, Tri-State Chapter, Buffalo, New York; [6]Karen Gottesman, BA: National Scleroderma Foundation, Los Angeles, California; [7]Geneviève Guillot, PDt: Sclérodermie Québec, Longueuil, Quebec, Canada; [8]Amanda Lawrie-Jones: Scleroderma Australia and Scleroderma Victoria, Melbourne, Victoria, Australia; [9]Maureen D. Mayes, MD: University of Texas McGovern School of Medicine, Houston; [10]Luc Mouthon, MD: Hôpital Cochin, Assistance Publique - Hôpitaux de Paris (AP-HP), Paris, France; [11]Warren R. Nielson, PhD: Western University and Lawson Research Institute, London, Ontario, Canada; [12]Michelle Richard, DSW: Scleroderma Atlantic, Halifax, Nova Scotia, Canada; [13]Maureen Worron-Sauvé, BA: Scleroderma Society of Ontario and Scleroderma Canada, Hamilton, Ontario, Canada; [14]Daphna Harel, PhD: New York University, New York, New York; [15]Vanessa L. Malcarne, PhD: San Diego State University and University of California, San Diego;

SIGNIFICANCE & INNOVATIONS

- Some individuals with severe systemic sclerosis (SSc) burden and high levels of pain, fatigue, and sleep disturbance report positive mental health, which may be associated with resilience.
- This is the first study to validate a resilience scale in SSc and the first to compare measurements for English and French versions of the 10-item Connor-Davidson Resilience Scale (CD-RISC-10).
- The CD-RISC-10 had good reliability and validity, and measurement properties were comparable for English- and French-language participants.
- The CD-RISC-10 can be used to evaluate resilience in individuals with SSc, including in international studies with English- and French-language participants.

to be almost 3 times as high as sex- and age-matched peers (4), and individuals with SSc report substantially lower quality of life compared to those with other rheumatic diseases (5) and the general population (6). Symptoms often include impaired function and mobility, breathing problems, GI symptoms, fatigue, pain, pruritus, sleep disturbances, body image distress from disfigurement (e.g., skin tightening, pigment changes, hand contractures, telangiectasias), and reduced mental health (3,7–10).

A recent cross-sectional study (Wojeck et al, unpublished observations) of >2,000 participants in the Scleroderma Patient-centered Intervention Network (SPIN) Cohort found that 5 latent classes characterized patterns of patient-reported outcomes, including fatigue, sleep, pain, anxiety symptoms, and depression symptoms (members of the SPIN investigators are shown in Appendix A). Participants were separated into four classes: low, normal, high, and very high symptom severity, and levels of patient-reported symptoms in these classes closely correlated with the severity or presence of specific disease manifestations. The fifth class, however, identified individuals with high fatigue, sleep, and pain symptoms but low mental health problems, even though members of this class had underlying disease burdens similar to the high class. The difference between individuals in this class and others with similarly severe SSc might be associated with resilience (11,12).

Research has defined resilience as positive adjustment or the ability to preserve or restore mental health despite adverse circumstances (13,14). Psychological factors associated with resilience include self-efficacy, self-esteem, optimism, hardiness, determination, an internal locus of control, and a sense of self-empowerment and mastery (11,12). Individuals with chronic medical conditions who score higher on resilience measures report lower anxiety and depression symptoms and better quality of life (11,12). In addition, researchers have found that intervention strategies that enhance resilience and adaptive coping improve psychological adaptation and reduce symptom burden (15).

No resilience measure has been validated in scleroderma, and there are no studies of resilience in individuals with SSc. A methodologic review (16) of tools to measure resilience reported that >15 scales had been developed and that, based on a set of predefined criteria to assess overall quality and usability, the 25-item Connor-Davidson Resilience Scale (CD-RISC) (17) was among 3 measures with the strongest ratings for measurement properties. It was the only measure that researchers had successfully used to evaluate change in response to an intervention. Researchers originally developed the CD-RISC in English and simultaneously validated it in a general population sample, primary care outpatients, mixed psychiatry outpatients, anxiety patients, and individuals with post-traumatic stress disorder (17). The 10-item short version of the scale, the CD-RISC-10, which researchers initially validated in English-speaking undergraduate students (18), reduces burden on study participants and has similar measurement properties as the CD-RISC (16,19). Additionally, compared to the original CD-RISC, the factor structure of the 10-item version may be more stable across studies and different cultural groups (20). The CD-RISC-10 has been validated in multiple languages (21,22), including French (21), and is therefore well-suited for use in international cohorts.

The objectives of the present study were to evaluate the validity and reliability of the 10-item CD-RISC-10 for use in SSc by 1) testing its unidimensional structure; 2) performing a differential item functioning (DIF) analysis to identify possible differences in measurement properties between English- and French-language respondents and assess the magnitude of any DIF; 3) evaluating internal consistency and test–retest reliability; and 4) evaluating convergent validity by comparing scores to another measure of resilience: the 14-item Resilience Scale (RS14) (23), a measure of self-esteem: the Rosenberg Self-Esteem Scale (24), and measures of depression and anxiety symptoms: Patient Reported Outcomes Measurement Information System (PROMIS) Anxiety 4a version 2.0 and PROMIS Depression 4a version 2.0 scales (25). For convergent validity, we hypothesized that the CD-RISC-10 would moderately to highly correlate with all other measures and that the magnitude of correlation with the RS14, another measure of resilience, would be the largest.

PATIENTS AND METHODS

We evaluated cross-sectional data collected from the regular SPIN Cohort assessments to evaluate English- (18) and French-

[16]Susan J. Bartlett, PhD: McGill University, Montreal, Quebec, Canada; [17]Brett D. Thombs, PhD: Jewish General Hospital and McGill University, Montreal, Quebec, Canada.

Author disclosures and graphical abstracts are available online at https://onlinelibrary.wiley.com/doi/10.1002/acr.25139.

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

Submitted for publication March 20, 2023; accepted in revised form April 25, 2023.

language (21) versions of the CD-RISC-10 for factor structure, language-based DIF, internal consistency reliability, and convergent validity. We administered the CD-RISC-10 a second time to a subset of participants 1–2 weeks after their first assessment to assess test–retest validity. A protocol was published online prior to study initiation (https://osf.io/dx3b6/). We reported the study consistent with the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) reporting guideline for studies on properties of patient-reported outcome measures (26).

Participants and procedure. The SPIN Cohort (27,28) is a convenience sample of participants recruited from 47 sites in 7 countries (Australia, Canada, France, Mexico, Spain, the UK, and the US). To be eligible for the SPIN Cohort, participants must be ≥18 years old, fluent in English, French, or Spanish, have access to and be able to respond to questionnaires via the internet, and meet the 2013 American College of Rheumatology/EULAR criteria for SSc (29) verified by a physician at a SPIN site. Participants are invited to participate in the SPIN Cohort by attending physicians or nurse coordinators at recruiting sites. Site personnel obtain written informed consent, including consent to be contacted by the SPIN team about additional studies, and submit an electronic medical form to enrol participants. Participants then receive an email with a unique, secure link to complete baseline measurements online in English, French, or Spanish. Subsequent online assessments are conducted by SPIN at 3-month intervals (27,28). The study included SPIN participants who completed all study measures in English or French during a regular assessment between August 2022 and January 2023, when the CD-RISC-10 was included in the SPIN Cohort. We did not include Spanish-language participants in this study because there were not enough individuals to conduct all study analyses.

To examine test–retest reliability, we administered the CD-RISC-10 to a subsample of participants 1–2 weeks following routine cohort assessment. We invited English- and French-speaking SPIN Cohort participants who completed the CD-RISC-10 as part of their regular SPIN Cohort assessment by email 7 days later (30,31) to complete the scale a second time via the online survey website Qualtrics. Invited participants had access to the questionnaire for 7 days, and they completed the retest assessments between 7 and 14 days after the initial assessment. We sent a reminder email to nonresponders 4 days after the initial invitation. As an incentive, we randomly selected 10 questionnaire respondents to win an Amazon gift card worth $100 CAD or the equivalent in their local currency. We emailed invitations until we reached our targeted sample size for test–retest reliability.

The SPIN Cohort study was approved by the Research Ethics Committee of the Centre intégré universitaire de santé et de services sociaux du Centre-Ouest-de-l'Île-de-Montréal (approval no. MP-05-2013-150) and by the ethics committees of all recruiting sites. The present study was approved as an amendment.

Measures. At baseline, SPIN Cohort participants report sociodemographic variables, including race or ethnicity, country, language, education, and marital status. Physician-reported data from the baseline data assessment included age, sex, height, weight, date of initial onset of non–Raynaud's phenomenon symptoms, SSc subtype, presence of GI involvement, digital ulcers anywhere on the fingers, current tendon friction rubs, presence of joint contractures, history of renal crisis, presence of pulmonary arterial hypertension, presence of interstitial lung disease, presence of primary biliary cirrhosis, and presence of overlap syndromes (rheumatoid arthritis, Sjögren's syndrome, systemic lupus erythematosus, idiopathic inflammatory myositis, autoimmune thyroid disease).

CD-RISC-10. CD-RISC-10 scores reflect multiple aspects of resilience, including flexibility, self-efficacy, regulation of emotion, optimism, and the ability to maintain focus under stress. Items assess the ability to tolerate and cope with experiences such as change, personal problems, illness, pressure, failure, and painful feelings (18). Item response options range from 0 (not true at all) to 4 (true nearly all the time). Participants respond to each statement in reference to the previous month. Evaluators score the scale by totalling item scores, resulting in possible scores of 0–40, with higher scores reflecting greater resilience. The correlation of the CD-RISC-10 with the 25-item CD-RISC was 0.92 in a sample of >500 undergraduate students (18). Researchers have validated a French version of the scale (21).

RS14. The 25-item Resilience Scale (RS25) was initially developed by researchers in a sample of older women who had recently experienced but successfully coped with a loss (e.g., loss of a spouse) (32). The scale received the second-highest score level in the review of resilience measures (16) and the highest possible rating for content and construct validity. The shortened form of the RS25, the RS14 (23), is based on a 1-factor structure and focuses on aspects of resilience such as self-reliance, purpose, equanimity, perseverance, and authenticity. Items are rated using a 7-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree). Evaluators sum item scores to a total (possible range 14–98), and higher scores reflect greater resilience. Researchers have validated the RS14 in numerous populations. It exhibits similar measurement properties compared to the original Resilience Scale, including evidence of high reliability and good validity in clinical and nonclinical settings (23). The correlation of the RS14 with the original 25-item Resilience Scale was 0.97 in a sample of 776 middle-aged and older adults (23). A French version of the scale has been validated by researchers (33).

Rosenberg Self-Esteem Scale. The RSES (24) assesses self-esteem, which reflects confidence in one's abilities or worth. It measures both positive and negative feelings about oneself. Researchers originally developed the scale in a sample of high school juniors and seniors (24). Since then, the scale has been applied in studies across a wide range of samples and has

demonstrated high reliability and good validity (34). The scale contains 10 items rated on a 4-point Likert scale, with response options from 0 (strongly disagree) to 3 (strongly agree). Evaluators calculate scoring the scale by first reverse scoring the negatively worded items (items 2, 5, 6, 8, and 9) and then totalling item scores, resulting in a possible range of 0 to 30, with higher scores reflecting greater self-esteem. Researchers previously validated a French version of the scale (35).

PROMIS Depression 4a version 2.0 and PROMIS Anxiety 4a version 2.0. The PROMIS Depression 4a version 2.0 and PROMIS Anxiety 4a version 2.0 scales (25) measure patient-reported depression and anxiety symptoms over the previous 7 days. Participants rate 4 statements for each domain on a 5-point scale ranging from 1 ("never") to 5 ("always"). The sum of item scores for each domain yields a score ranging from 4 to 20, which is converted by evaluators into a T score adjusted to the US general population (mean ± SD 50 ±10). Higher scores indicate greater severity of depression or anxiety symptoms. The SPIN research team previously validated the English and French versions of PROMIS Depression 4a version 2.0 and PROMIS Anxiety 4a version 2.0 in SSc (36).

Statistical analysis. We calculated descriptive sample statistics as the mean ± SD for continuous variables and frequencies and percentages for categorical variables for the total sample and separately for the English- and French-speaking samples.

CFA. We conducted a CFA to evaluate the single-factor structure of the CD-RISC-10 (18). Item responses for the CD-RISC-10 are ordinal Likert data. We modelled the responses using a weighted least squares estimator, a diagonal weight matrix, and robust standard errors. We used the Tucker–Lewis Index (TLI), comparative fit index (CFI), and root mean square error of approximation (RMSEA) to assess model fit. Well-fitting models are indicated by a TLI and CFI of ≥0.95 and RMSEA of ≤0.06 (37), although a CFI of ≥0.90 and an RMSEA of ≤0.08 (38) are often regarded as indicators of acceptable model fit. We used modification indices to identify pairs of items for which model fit would improve if error estimates were freed to covary and for which there were theoretically justifiable shared method effects (e.g., similar wording) if the original model did not achieve adequate model fit.

DIF analysis. We performed a DIF analysis using the Multiple Indicators Multiple Causes (MIMIC) model to identify possible differences in measurement properties between English and French versions of the CD-RISC-10. DIF analysis compares patterns of item responses in subgroups and tests whether individuals with similar levels of a latent construct respond to each item similarly, regardless of group affiliation. For DIF assessment, MIMIC models are based on structural equation models, in which the group variable (English versus French) is added to the basic CFA model as an observed variable. Thus, the base MIMIC model consists of the CFA factor model with the additional regression of the latent factor on group to control for group differences at the latent factor level. We then identified DIF by first separately regressing items, one at a time, on group. If there was DIF for ≥1 item in this first step, the item with the largest magnitude of statistically significant DIF was considered to have DIF, and the link between the language group variable and that item was included in the model. In a second step, we again separately regressed remaining items on language group one at a time and included the item with the largest DIF in the model. This procedure was repeated until none of the remaining items showed significant DIF. Once all items with significant DIF had been identified, the potential magnitude of DIF items collectively was evaluated by comparing the difference of the latent factor between language groups in the baseline CFA model and after controlling for DIF. Because we did not encounter DIF of a meaningful magnitude, item analyses and reliability and convergent validity were done with the whole sample and not separated by language.

Item analyses. We reported the mean ± SD, item intercorrelations, and item–rest correlations for each item of the CD-RISC-10. The item–rest correlation is the correlation of an item score with the total score after removing the item from the total score. In addition, we examined floor and ceiling effects, defined as ≥15% of the participants having the lowest or highest possible score (39).

Reliability and convergent validity. We computed Cronbach's alpha to determine internal consistency (40) and the intraclass correlation coefficients (ICC) to measure test–retest reliability (41). We chose the ICC as the measure of test–retest reliability because it reflects both the degree of correlation and agreement between measurements (42). We calculated ICC estimates and 95% confidence intervals (95% Cis) based on absolute agreement and a 2-way mixed-effects model.

To examine the convergent validity of the CD-RISC-10, we formulated hypotheses regarding the direction and magnitude of Pearson's correlations with other outcome measures a priori based on existing evidence from convergent validity comparisons for the CD-RISC-10 (20). The magnitude of correlations was interpreted as small ($|r| \leq 0.3$), moderate ($0.3 < |r| < 0.5$), or large ($|r| \geq 0.5$) (43). We hypothesized that all correlations between measures would be moderate to large and that the CD-RISC-10 would be more strongly related to another resilience measure, the RS14, than with other measures. We conducted CFA and DIF using Mplus version 7 (44). All other statistical analyses were performed using SPSS version 29 (45).

Sample size calculation. *Confirmatory factor analysis.* Recommendations for CFA sample size vary. In the present study, we performed a single-factor CFA with 10 indicators using a sample that we expected would include ∼1,000 participants. This number substantially exceeds the minimum number recommended by all established recommendations and standards (46–48) for a sample size necessary to achieve excellent agreement between true model characteristics and estimates.

Convergent validity. Stable estimates of correlations are typically achieved with a sample size of ≥250, although smaller

correlations require larger samples. To assess a Pearson's correlation with a 95% CI with a width of 0.10, a sample size of ≥403 is required for a correlation of 0.30, and a size of ≥275 is required for a correlation of 0.50 (40).

Test–retest reliability. Although an ICC value of 0.70 is considered acceptable for test–retest reliability, a coefficient close to or exceeding 0.80 is preferable (49). A test–retest sample size of 200 individuals would be required for a precision level of 95% CI with a width of 0.10 for an estimated ICC of 0.80 (31). Therefore, we aimed for a retest sample size of 200 participants.

RESULTS

Sample characteristics. In total, 962 participants completed all items of the CD-RISC-10, RS14, RSES, and PROMIS Depression 4a version 2.0 and PROMIS Anxiety 4a version 2.0. Sociodemographic and disease characteristics were similar across English- and French-language samples, as shown in Table 1. The total sample consisted of 848 female participants (88%) with a mean ± SD age of 61.1 ± 11.6 years. Mean ± SD time since onset of first non–Raynaud's phenomenon symptoms was 15.7 ± 9.6 years, and 345 individuals had diffuse SSc (36%). Participants were from France (37%), Canada (26%), the US (25%), the UK (9%), and Australia (2%). Just over half (549 [57%]) completed assessments in English.

CD-RISC measurement properties. *Confirmatory factor analysis.* The results of the CFA are shown in Table 2. In the initial CFA, the model fit for the hypothesized single-factor model was somewhat suboptimal (TLI 0.97, CFI 0.98, RMSEA 0.11). Our examination of modification indices showed that freeing the error terms of items 1 and 2 to covary would improve model fit. Items

Table 1. Sample sociodemographic and disease characteristics for the full sample and by assessment language*

Characteristics	Full sample (n = 962)		English (n = 549)		French (n = 413)	
	No.	Mean ± SD or no. (%)	No.	Mean ± SD or no. (%)	No.	Mean ± SD or no. (%)
Sociodemographic variables						
Age, years	962	61.1 ± 11.6	549	62.4 ± 10.7	413	59.4 ± 12.5
Female sex	962	848 (88)	549	488 (89)	413	360 (87)
White race or ethnicity	955	816 (85)	546	471 (86)	409	345 (84)
Nationality	962		549		413	
Canada		254 (26)		197 (36)	57	57 (14)
US		245 (25)		245 (45)	–	–
UK		85 (9)		85 (16)	–	–
France		358 (37)		2 (<1)	356	356 (86)
Australia		20 (2)		20 (4)	–	–
Language, English language speaking	962	549 (57)				
Education, years	960	15.1 ± 3.6	549	15.6 ± 3.0	411	14.4 ± 4.1
Marital status single	960	106 (11)	549	54 (10)	411	52 (13)
BMI, kg/m^2	962	25.1 ± 5.2	549	25.6 ± 5.4	413	24.4 ± 5.0
Disease characteristics						
Time since first non–Raynaud's symptom	892	15.7 ± 9.6	505	17.6 ± 9.9	387	13.3 ± 8.8
Diffuse subtype	955	345 (36)	543	221 (41)	412	124 (30)
Gastrointestinal involvement	962	828 (86)	549	480 (88)	413	348 (84)
Digital ulcers	914	124 (14)	513	72 (14)	401	52 (13)
Current tendon friction rubs	846	86 (10)	468	46 (10)	378	40 (11)
Large joint contractures (moderate or severe)	891	98 (11)	499	41 (8)	392	57 (15)
Small joint contractures (moderate or severe)	906	224 (25)	504	107 (21)	402	117 (29)
History of SSc renal crisis	945	40 (4)	539	25 (5)	406	15 (4)
Interstitial lung disease	941	296 (32)	534	159 (30)	407	137 (34)
Pulmonary arterial hypertension	931	70 (8)	525	41 (8)	406	29 (7)
Primary biliary cirrhosis	926	18 (2)	527	10 (2)	399	8 (2)
Any overlap syndrome†	962	195 (20)	549	113 (21)	413	82 (20)
Psychological assessments						
CD-RISC-10	962	27.8 ± 7.3	549	28.6 ± 7.2	413	26.8 ± 7.18
RS14	962	78.6 ± 15.1	549	80.2 ± 14.3	413	76.6 ± 15.9
Rosenberg Scale	962	20.8 ± 5.5	549	21.6 ± 5.7	413	19.9 ± 5.2
PROMIS Depression	962	51.5 ± 9.2	549	50.6 ± 9.0	413	52.8 ± 9.4
PROMIS Anxiety	962	53.6 ± 9.8	549	52.8 ± 9.6	413	54.6 ± 10.0

* BMI = body mass index; CD-RISC-10 = 10-item Connor-Davidson Resilience Scale; PROMIS = Patient Reported Outcomes Measurement Information System; RS14 = 14-item Resilience-Scale; SSc = systemic sclerosis.
† Participant had ≥1 of the following disease: rheumatoid arthritis, Sjögren's syndrome, systemic lupus erythematosus, or idiopathic inflammatory myositis.

Table 2. Factor loadings on the CD-RISC-10*

Item†	CFA factor loading‡	95% CIs
1. I am able to adapt when changes occur	0.76	0.73, 0.80
2. I can deal with whatever comes my way	0.87	0.85, 0.89
3. I try to see the humorous side of things when I am faced with problems	0.74	0.71, 0.77
4. Having to cope with stress can make me stronger	0.76	0.74, 0.80
5. I tend to bounce back after illness, injury, or other hardships	0.84	0.82, 0.86
6. I believe I can achieve my goals, even if there are obstacles	0.85	0.83, 0.87
7. Under pressure, I stay focused and think clearly	0.83	0.80, 0.85
8. I am not easily discouraged by failure	0.70	0.67, 0.73
9. I think of myself as a strong person when dealing with life's challenges and difficulties	0.87	0.85, 0.89
10. I am able to handle unpleasant or painful feelings like sadness, fear, and anger	0.83	0.81, 0.86

* 95% CI = 95% confidence interval; CD-RISC-10 = 10-item Connor-Davidson Resilience Scale; CFA = confirmatory factor analysis.
† On a 5-point scale, where 0 = not true at all and 4 = true nearly all the time.
‡ Error terms of items 1 and 2 were freed to covary.

1 and 2 evaluate how well individuals can adapt to changes or deal with things coming their way, which are closely related experiences. Therefore, we refitted the model to allow the error terms of these items to covary, resulting in good fit (TLI 0.99, CFI 0.99, RMSEA 0.08).

DIF analysis. The 1-factor model, which included regression of the latent resilience factor on language, demonstrated good fit (TLI 0.99, CFI 0.99, RMSEA 0.07). Baseline CFA model parameters before correcting for DIF are shown in Table 3. We identified 6 items with statistically significant language-based DIF. Compared to English-language participants, French-language participants had higher scores than would be expected on item 3 (β = 0.14 [95% CI 0.04, 0.23]) and item 9 (β = 0.13 [95% CI 0.04, 0.21]) and lower scores on item 1 (β = –0.17 [95% CI –0.27, –0.08]), item 4 (β = –0.12 [95% CI –0.23, –0.03]), item 5 (β = –0.22 [95% CI –0.32, –0.14]), and item 6 (β = –0.17 [95% CI –0.26, –0.08]). The difference between the 2 language groups (English and French) on the mean latent factor level was not meaningfully different between the model with DIF adjustment (standardized mean differences [SMD] 0.31 [95% CI 0.17, 0.43]) and without adjustment (SMD 0.26 [95% CI 0.13, 0.37]) (see Table 3).

Item analysis. The mean item and total CD-RISC-10 scores in the full sample are shown in Table 4. Mean item scores ranged from 2.5 for item 4 ("Having to cope with stress can

Table 3. Factor loading for the CD-RISC-10 in combined English and French samples and DIF evaluation*

Item	Base model†		DIF-corrected model‡	
	CFA factor loading	95% CIs	CFA factor loading	95% CIs
Items				
1. I am able to adapt when changes occur	0.77	0.74, 0.79	0.77	0.74, 0.79
2. I can deal with whatever comes my way	0.87	0.85, 0.88	0.87	0.85, 0.88
3. I try to see the humorous side of things when I am faced with problems	0.74	0.70, 0.76	0.74	0.70, 0.76
4. Having to cope with stress can make me stronger	0.76	0.74, 0.79	0.76	0.74, 0.79
5. I tend to bounce back after illness, injury, or other hardships	0.84	0.82, 0.86	0.84	0.82, 0.86
6. I believe I can achieve my goals, even if there are obstacles	0.85	0.83, 0.87	0.85	0.83, 0.87
7. Under pressure, I stay focused and think clearly	0.82	0.80, 0.84	0.82	0.80, 0.84
8. I am not easily discouraged by failure	0.70	0.66, 0.72	0.70	0.66, 0.72
9. I think of myself as a strong person when dealing with life's challenges and difficulties	0.87	0.85, 0.89	0.87	0.85, 0.89
10. I am able to handle unpleasant or painful feelings like sadness, fear, and anger	0.83	0.81, 0.85	0.83	0.81, 0.85
Direct effects on items attributable to the French language				
1. I am able to adapt when changes occur	–	–	−0.17	−0.27, −0.08
3. I try to see the humorous side of things when I am faced with problems	–	–	0.14	0.04, 0.23
4. Having to cope with stress can make me stronger	–	–	−0.12	−0.23, −0.03
5. I tend to bounce back after illness, injury, or other hardships	–	–	−0.22	−0.32, −0.14
6. I believe I can achieve my goals, even if there are obstacles	–	–	−0.17	−0.26, −0.08
9. I think of myself as a strong person when dealing with life's challenges and difficulties	–	–	0.13	0.04, 0.21
Standardized mean difference (English and French) on latent resilience factor	0.26	0.13, 0.37	0.31	0.17, 0.43

* 95% CI = 95% confidence interval; CD-RISC-10 = 10-item Connor-Davidson Resilience Scale; CFA = confirmatory factor analysis.
† Unstandardized model with fixed variance and regression of the latent resilience factor on language, not corrected for differential item functioning (DIF).
‡ Unstandardized model with fixed variance and regression of the latent resilience factor on language, corrected for DIF on items 1, 3, 4, 5, 6, and 9.

Table 4. Characteristics of the CD-RISC-10*

Item	Mean ± SD score†	Item–rest correlation
Individual scores		
1. I am able to adapt when changes occur	3.1 ± 0.84	0.70
2. I can deal with whatever comes my way	2.9 ± 0.86	0.80
3. I try to see the humorous side of things when I am faced with problems	2.7 ± 0.97	0.67
4. Having to cope with stress can make me stronger	2.5 ± 1.00	0.69
5. I tend to bounce back after illness, injury, or other hardships	3.0 ± 0.88	0.75
6. I believe I can achieve my goals, even if there are obstacles	2.8 ± 0.88	0.76
7. Under pressure, I stay focused and think clearly	2.6 ± 0.97	0.75
8. I am not easily discouraged by failure	2.6 ± 0.98	0.62
9. I think of myself as a strong person when dealing with life's challenges and difficulties	3.0 ± 0.92	0.78
10. I am able to handle unpleasant or painful feelings like sadness, fear, and anger	2.7 ± 0.97	0.75
Total score	27.8 ± 7.3	–

* CD-RISC-10 = 10-item Connor-Davidson Resilience Scale.
† On a 5-point scale, where 0 = not true at all and 4 = true nearly all the time.

make me stronger") to 3.1 for item 1 ("I am able to adapt when changes occur"). Correlations between items ranged from r = 0.44 ($P < 0.001$ for items 3 and 8) to r = 0.73 ($P < 0.001$ for items 1 and 2). Item–rest correlations ranged from r = 0.62 (item 8) to r = 0.80 (item 2). There were 2 participants (0.2%) with the lowest possible score (score of 0) on the scale and 48 participants (5.0%) with the highest possible score (score of 40). Item response frequencies are shown in Supplementary Table 1 (available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25139/abstract).

Reliability. Cronbach's alpha was 0.93 (95% CI 0.92, 0.94). We assessed test–retest reliability in a subsample of 230 participants, whose characteristics were similar compared to the full sample (for subsample sociodemographic and medical data, see Supplementary Table 2, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25139/abstract), resulting in an ICC of 0.80 (95% CI 0.75, 0.85), indicating good 1–2-week test–retest reliability.

Convergent validity. As shown in Table 5, there were moderate-to-large correlations between the CD-RISC-10 and measures of resilience (RS14), self-esteem (RSES), depression (PROMIS depression 4a version 2.0), and anxiety (PROMIS anxiety 4a version 2.0). All correlations were consistent with convergent validity hypotheses.

DISCUSSION

We tested the unidimensional structure of the CD-RISC-10, examined whether there were meaningful differences in measurement properties between English- and French-language versions of the scale, and evaluated internal consistency, test–retest reliability, and convergent validity. We found that the hypothesized single-factor structure of the scale fit well, supporting the use of a single total score for the CD-RISC-10 scale. There was statistically significant DIF for 6 items between English- and French-language participants. However, the cumulative effect of DIF was minimal and did not meaningfully influence estimates of differences in resilience between English- and French-language respondents in unadjusted models (SMD 0.26 [95% CI 0.17, 0.43]) versus DIF-adjusted models (SMD 0.31 [95% CI 0.17, 0.43]), allowing us to conclude that CD-RISC-10 scores of English- and French-language participants can be compared and aggregated without concerns of language-based bias.

Internal consistency reliability (α = 0.93 [95% CI 0.92, 0.94]) and test–retest reliability (ICC 0.80 [95% CI 0.75, 0.85]) were good, and there were no floor or ceiling effects. In addition, indices of convergent validity were consistent with study hypotheses; CD-RISC-10 correlated moderately to highly with all measurements (RSES r = 0.69; PROMIS depression r = –0.60; PROMIS

Table 5. Correlation of measures using the CD-RISC-10 to assess convergent validity*

Convergent validity†	Pearson correlation	95% CIs
Large positive correlation		
Resilience (RS14)	0.78	0.76, 0.81
Moderate-to-large positive correlation		
Self-esteem (Rosenberg Self-esteem Scale)	0.69	0.65, 0.72
Moderate-to-large negative correlation		
Depression (PROMIS Depression)	−0.60	−0.64, −0.56
Anxiety (PROMIS Anxiety)	−0.57	−0.61, −0.52

All hypotheses were confirmed. 95% CI = 95% confidence interval; CD-RISC-10 = 10-item Connor-Davidson Resilience Scale; PROMIS = Patient Reported Outcomes Measurement Information System; RS14 = 14-item Resilience-Scale.
† Magnitude of correlations was defined as small ($|r| \leq 0.3$), moderate ($0.3 < |r| < 0.5$), or large ($= |r| \geq 0.5$).

anxiety r = –0.57) and the magnitude of correlation with the RS14, another measure of resilience, was the largest (r = 0.78).

Researchers initially validated the CD-RISC-10 in a sample of 1,743 undergraduate students from the US (18). The present study is the first to validate the scale among individuals with SSc and, to our knowledge, the first comparison of measurement properties between English- and French-language versions. The overall outcomes of our study were consistent with results from previous studies that examined measurement properties of the CD-RISC scale in other samples, including among individuals with chronic diseases (18,21,22). We believe that this is the first study to examine language-based DIF in the CD-RISC-10.

Our findings have important implications for research. We found that the CD-RISC-10 provides a valid and reliable method for evaluating resilience in individuals with SSc. A previous study (Wojack et al, unpublished observations) used latent profile analysis and found that some individuals with SSc report positive mental health, despite experiencing severe disease manifestations and high levels of pain, fatigue, and sleep disturbance, which could be associated with resilience (11,12). Resilience, using the CD-RISC-10, should be compared between classes of individuals with SSc who differ in mental health despite having similar disease burdens to further elucidate the possible role of resilience in the mental health of individuals with SSc. We plan to conduct these analyses in a second study, using a sample from the SPIN Cohort. In addition, researchers could conduct similar analyses in other chronic illness populations.

The results of our DIF analysis demonstrate the comparability and combinability of CD-RISC-10 scores across English and French languages in SSc, presenting opportunities for broader utilization in international patient cohorts, including the SPIN Cohort (27,28). Among individuals with chronic medical conditions, intervention strategies that improve resilience and adaptive coping have been found to be effective in improving psychological adaptation and reducing symptom burden (15). The CD-RISC-10 presents a valid outcome measure for testing similar interventions in SSc.

Our study has several notable strengths, including its international cohort with participants from 47 clinical sites, its large sample size, its assessment of test–retest reliability, and the comparison of measurement properties in English- and French-language participants with SSc. There are also limitations to consider. First, the SPIN Cohort is a convenience sample of individuals with SSc receiving treatment at SPIN recruiting centers who can complete online measures, since SPIN collects data digitally only. However, a comparison with the European Scleroderma Trials and Research Cohort and the Canadian Scleroderma Research Group Cohort indicated broad comparability of participant characteristics, which supports generalizability in SSc (27). Second, the examination of DIF was limited to English- and French-language versions of the CD-RISC-10 and adults with SSc, and the generalizability of the results to other populations is not known. Third, the MIMIC approach for DIF evaluates uniform, but not nonuniform, DIF.

Overall, the results of this study indicate that the CD-RISC is a valid and reliable measure of resilience in English and French languages in SSc, supporting its use as an outcome measure to assess resilience in this population. In addition, we found DIF to be negligible, suggesting that CD-RISC-10 scores are comparable across English- and French-language versions.

AUTHOR CONTRIBUTIONS

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

Study conception and design. Neyer, Henry, Thombs.

Acquisition of data. Neyer, Henry, Carrier, Kwakkenbos, Wojeck, Gietzen, Gottesman, Guillot, Lawrie-Jones, Mayes, Mouthon, Nielson, Richard, Worron-Sauvé, Harel, Malcarne, Bartlett, Thombs.

Analysis and interpretation of data. Neyer, Henry, Carrier, Kwakkenbos, Wojeck, Gietzen, Gottesman, Guillot, Lawrie-Jones, Mayes, Mouthon, Nielson, Richard, Worron-Sauvé, Harel, Malcarne, Bartlett, Thombs.

REFERENCES

1. Varga J, Denton CP, Wigley FM, et al. Scleroderma: from pathogenesis to comprehensive management. New York (NY): Springer; 2017.
2. Denton CP, Khanna D. Systemic sclerosis [review]. Lancet 2017;390:1685–99.
3. Bassel M, Hudson M, Taillefer SS, et al. Frequency and impact of symptoms experienced by patients with systemic sclerosis: results from a Canadian national survey. Rheumatology (Oxford) 2011;50:762–7.
4. Rubio-Rivas M, Royo C, Simeón CP, et al. Mortality and survival in systemic sclerosis: systematic review and meta-analysis [review]. Semin Arthritis Rheum 2014;44:208–19.
5. Park EH, Strand V, Oh YJ, et al. Health-related quality of life in systemic sclerosis compared with other rheumatic diseases: a cross-sectional study. Arthritis Res Ther 2019;21:61.
6. Hudson M, Thombs BD, Steele R, et al. Health-related quality of life in systemic sclerosis: a systematic review [review]. Arthritis Rheum 2009;61:1112–20.
7. Lee YC, Fox RS, Kwakkenbos L, et al. Pain levels and associated factors in the Scleroderma Patient-centered Intervention Network (SPIN) Cohort: a multicentre cross-sectional study. Lancet Rheumatol 2021;3:e844–54.
8. Kwakkenbos L, Delisle VC, Fox RS, et al. Psychosocial aspects of scleroderma [review]. Rheum Dis Clin North Am 2015;41:519–28.
9. Jewett LR, Kwakkenbos L, Delisle VC, et al. Psychosocial issues and care for patients with systemic sclerosis. In: Varga J, Denton CP, Wigley FM, et al, editors. Scleroderma: from pathogenesis to comprehensive management. New York (NY): Springer; 2017. p. 615–21.
10. Nassar EL, Abdulkareem DA, Thombs BD. Results from a living systematic review of the prevalence of mood and anxiety disorders and factors associated with symptoms in systemic sclerosis. Sci Rep 2023;13:5181.
11. Tamura S, Suzuki K, Ito Y, et al. Factors related to the resilience and mental health of adult cancer patients: a systematic review. Support Care Cancer 2021;29:3471–86.

12. Stewart DE, Yuen T. A systematic review of resilience in the physically ill [review]. Psychosomatics 2011;52:199–209.
13. Luthar SS, Cicchetti D. The construct of resilience: implications for interventions and social policies [review]. Dev Psychopathol 2000; 12:857–85.
14. Herrman H, Stewart DE, Diaz-Granados N. What is resilience? [review] Can J Psychiatry 2011;56:258–65.
15. Liu JJ, Ein N, Gervasio J, et al. Comprehensive meta-analysis of resilience interventions [review]. Clin Psychol Rev 2020;82:101919.
16. Windle G, Bennett KM, Noyes J. A methodological review of resilience measurement scales [review]. Health Qual Life Outcomes 2011;9:8.
17. Connor KM, Davidson JR. Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). Depress Anxiety 2003; 18:76–82.
18. Campbell-Sills L, Stein MB. Psychometric analysis and refinement of the Connor-Davidson Resilience Scale (CD-RISC): validation of a 10-item measure of resilience. J Traum Stress 2007;20:1019–28.
19. Cosco TD, Kaushal A, Richards M, et al. Resilience measurement in later life: a systematic review and psychometric analysis [review]. Health Qual Life Outcomes 2016;14:16.
20. Madewell AN, Ponce-Garcia E. Assessing resilience in emerging adulthood: the Resilience Scale (RS), Connor-Davidson Resilience Scale (CD-RISC), and Scale of Protective Factors (SPF). Pers Individ Dif 2016;97:249–55.
21. Hébert M, Parent N, Simard C, et al. Validation of the French Canadian version of the brief Connor-Davidson Resilience Scale (CD-RISC 10). Can J Behav Sci 2018;50:9–16.
22. Notario-Pacheco B, Solera-Martínez M, Serrano-Parra MD, et al. Reliability and validity of the Spanish version of the 10-item Connor-Davidson Resilience Scale (10-item CD-RISC) in young adults. Health Qual Life Outcomes 2011;9:63.
23. Wagnild G. The Resilience Scale user's guide for the US English version of the Resilience Scale and the 14-Item Resilience Scale (RS14). Worden (MT): The Resilience Center; 2009.
24. Rosenberg M. Society and the adolescent self-image. Princeton (NJ): Princeton University Press; 2016.
25. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. J Clin Epidemiol 2010;63:1179–94.
26. Gagnier JJ, Lai J, Mokkink LB, et al. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. Qual Life Res 2021;30:2197–2218.
27. Dougherty DH, Kwakkenbos L, Carrier M-E, et al. The Scleroderma Patient-Centered Intervention Network Cohort: baseline clinical features and comparison with other large scleroderma cohorts. Rheumatology (Oxford) 2018;57:1623–31.
28. Kwakkenbos L, Jewett LR, Baron M, et al. The Scleroderma Patient-centered Intervention Network (SPIN) Cohort: protocol for a cohort multiple randomised controlled trial (cmRCT) design to support trials of psychosocial and rehabilitation interventions in a rare disease context. BMJ Open 2013;3:e003563.
29. Van den Hoogen F, Khanna D, Fransen J, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/-European League Against Rheumatism collaborative initiative. Arthritis Rheum 2013;65:2737–47.
30. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. Control Clin Trials 1991;12 Suppl:142S–58S.
31. Polit DF. Getting serious about test–retest reliability: a critique of retest research and some recommendations. Qual Life Res 2014;23: 1713–20.
32. Wagnild GM, Young HM. Development and psychometric evaluation of the Resilience Scale. J Nurs Meas 1993;1:165–78.
33. Cénat JM, Hébert M, Karray A, et al. Psychometric properties of the Resilience Scale – 14 in a sample of college students from France. Encephale 2018;44:517–22.
34. Donnellan MB, Trzesniewski KH, Robins RW. Measures of self-esteem. In: Boyle GJ, Saklofske DH, Matthews G, editors. Measures of personality and social psychological constructs. Amsterdam: Elsevier; 2015. p. 131–57.
35. Vallieres EF, Vallerand RJ. Traduction et validation Canadienne-Française de l'Échelle de l'Estime de Soi de Rosenberg. Int J Psychol 1990;25:305–16.
36. Kwakkenbos L, Thombs BD, Khanna D, et al. Performance of the Patient-Reported Outcomes Measurement Information System-29 in scleroderma: a Scleroderma Patient-centered Intervention Network Cohort study. Rheumatology (Oxford) 2017;56:1302–11.
37. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 1999;6:1–55.
38. Browne MW, Cudeck R. Alternative ways of assessing model fit. Sociol Methods Res 1992;21:230–58.
39. Terwee CB, Bot SD, Boer MR de, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34–42.
40. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297–334.
41. Fisher RA. Statistical methods for research workers. In: Kotz S, Johnson NL, editors. Breakthroughs in statistics. New York (NY): Springer New York; 1992. p. 66–70.
42. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15: 155–63.
43. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale (NJ): L. Erlbaum Associates; 1988.
44. Muthén LK, Muthén BO. Mplus user's guide. 8th ed. Los Angeles (CA): Muthén & Muthén; 2017.
45. IBM SPSS Statistics for Macintosh, version 29.0. Armonk (NY): IBM Corp.
46. Wolf EJ, Harrington KM, Clark SL, et al. Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. Educ Psychol Meas 2013;76:913–34.
47. Mundfrom DJ, Shaw DG, Ke TL. Minimum sample size recommendations for conducting factor analyses. Int J Test 2005;5:159–68.
48. Kline RB. Beyond significance testing: statistics reform in the behavioral sciences. 2nd ed. Washington (DC): American Psychological Association; 2013.
49. De Vet HC, Terwee CT, Mokkink LB, et al. Measurement in medicine: a practical guide. New York (NY): Cambridge University Press; 2011.

APPENDIX A: SPIN INVESTIGATORS

SPIN Investigators, in addition to the authors are as follows: Catherine Fortuné: Ottawa Scleroderma Support Group, Ottawa, Ontario, Canada; Marie Hudson, Andrea Benedetti: McGill University, Montreal, Quebec, Canada; Laura K. Hummers: Johns Hopkins University School of Medicine, Baltimore, Maryland; Claire Elizabeth Adams, Mara Cañedo Ayala, Vanessa Cook, Sophie Hu, Bianca Matthews, Elsa-Lynn Nassar, Julia Nordlund, Sabrina Provencher: Jewish General Hospital, Montreal, Quebec, Canada; Shervin Assassi: University of Texas McGovern School of Medicine, Houston, Texas; Ghassan El-Baalbaki: Université du Québec à Montréal, Montreal, Quebec, Canada; Kim Fligelstone: Scleroderma & Raynaud's UK, London, UK; Tracy Frech: University of Utah, Salt Lake

City; Monique Hinchcliff: Yale School of Medicine, New Haven, Connecticut; Sindhu R. Johnson: Toronto Scleroderma Program, Mount Sinai Hospital, Toronto Western Hospital, and University of Toronto, Toronto, Ontario, Canada; Maggie Larche, Nader Khalidi: McMaster University, Hamilton, Ontario, Canada; Catarina Leite: University of Minho, Braga, Portugal; Christelle Nguyen, François Rannou: Université Paris Descartes, Université de Paris, Paris, France, and Assistance Publique - Hôpitaux de Paris, Paris, France; Karen Nielsen: Scleroderma Society of Ontario, Hamilton, Ontario, Canada; Janet Pope: University of Western Ontario, London, Ontario, Canada; Tatiana Sofia Rodriguez-Reyna: Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico; Anne A. Schouffoer: Leiden University Medical Center, Leiden, The Netherlands; Maria E. Suarez-Almazor: University of Texas MD Anderson Cancer Center, Houston; Christian Agard: Centre Hospitalier Universitaire - Hôtel-Dieu de Nantes, Nantes, France; Marc André, Louis Olagne: Centre Hospitalier Universitaire Gabriel-Montpied, Clermont-Ferrand, France; Elana J. Bernstein: Columbia University, New York, New York; Sabine Berthier: Centre Hospitalier Universitaire Dijon Bourgogne, Dijon, France; Lyne Bissonnette, Alessandra Bruns, Ariel Masetto, Sophie Roux: Université de Sherbrooke, Sherbrooke, Quebec, Canada; Carlotta Cacciatore, Benjamin Crichi, Dominique Farge-Bancel: Assistance Publique - Hôpitaux de Paris, Hôpital St-Louis, Paris, France; Patricia Carreira, Regina Fare, Maria Martin Lopez, Sheila Melchor Díaz, Esther Rodríguez Almazar: Servicio de Reumatologia del Hospital 12 de Octubre, Madrid, Spain; Marion Casadevall, Benjamin Chaigne, Bertrand Dunogue, Alexis Régent: Assistance Publique - Hôpitaux de Paris, Hôpital Cochin, Paris, France; Lorinda Chung: Stanford University, Stanford, California; Robyn Domsic: University of Pittsburgh, Pittsburgh, Pennsylvania; James V. Dunne, Pearce Wilcox: St. Paul's Hospital and University of British Columbia, Vancouver, British Columbia, Canada; Paul R. Fortin, Alena Ikic: CHU de Québec - Université Laval, Quebec, Quebec; Jessica Gordon, Kimberly Lakin, Robert Spiera: Hospital for Special Surgery, New York City, New York; Brigitte Granel-Rey: Université, and Assistance Publique - Hôpitaux de Marseille, Hôpital Nord, Marseille, France; Aurélien Guffroy, Thierry Martin, Vincent Poindron: Les Hôpitaux Universitaires de Strasbourg, Nouvel Hôpital Civil, Strasbourg, France; Genevieve Gyger: Jewish General Hospital and McGill University, Montreal, Quebec, Canada; Eric Hachulla: Centre Hospitalier Régional Universitaire de Lille, Hôpital Claude Huriez, Lille, France; Sabrina Hoa: Centre hospitalier de l'Université de Montréal – CHUM, Montreal, Quebec, Canada; Niall Jones: University of Alberta, Edmonton, Alberta, Canada; Marc Lambert, David Launay, Hélène Maillard, Vincent Sobanski: Centre Hospitalier Régional Universitaire de Lille, Hôpital Claude Huriez, Lille, France; Yvonne C. Lee: Northwestern University, Chicago, Illinois; Nancy Maltez: University of Ottawa, Ottawa, Ontario, Canada; Joanne Manning: Salford Royal NHS Foundation Trust, Salford, UK; Isabelle Marie: CHU Rouen, Hôpital de Bois-Guillaume, Rouen, France; François Maurier: Uneos - Groupe hospitalier associatif, Metz, France; Arsene Mekinian, Sébastien Rivière: Assistance Publique - Hôpitaux de Paris, Hôpital St-Antoine, Paris, France; Mandana Nikpour: St. Vincent's Hospital and University of Melbourne, Melbourne, Victoria, Australia; Susanna Proudman: Royal Adelaide Hospital and University of Adelaide, Adelaide, South Australia, Australia; David Robinson: University of Manitoba, Winnipeg, Manitoba, Canada; Perrine Smets: Centre Hospitalier Universitaire Gabriel-Montpied, Clermont-Ferrand, France; Virginia Steen: Georgetown University, Washington, DC; Evelyn Sutton: Dalhousie University, Halifax, Nova Scotia, Canada; Carter Thorne: Southlake Regional Health Centre, Newmarket, Ontario, Canada.

Arthritis Care & Research
Vol. 75, No. 11, November 2023, pp 2379–2388
DOI 10.1002/acr.25137

American College of Rheumatology
Empowering Rheumatology Professionals

Clinical Features Associated With Rate of Fractures in Patients With Systemic Sclerosis: A US Cohort Study

Bliss Rogers,[1] Sina Famenini,[1] Jamie Perin,[1] Maria I. Danila,[2] Kristin Wipfler,[3] Kaleb Michaud,[4] and Zsuzsanna H. McMahan[1]

Objective. Systemic sclerosis (SSc) is associated with several specific risk factors for fracture due to the complications of the disease and related medications. The present study was undertaken to examine the relationship between SSc-associated clinical features and fracture rate in a large US cohort.

Methods. Participants with SSc in FORWARD, The National Databank for Rheumatic Diseases, were included (1998–2019). Age- and sex-matched individuals with osteoarthritis (OA) from the same database were included as comparators. The primary end point was self-reported major osteoporotic fracture. Cox proportional hazards models were used to study the associations between risk factors and fractures.

Results. The study included 922 individuals (SSc patients, n = 154; OA patients, n = 768). Eighty-seven percent were female, with a mean age of 57.8 years. Fifty-one patients developed at least 1 fracture during a median of 4.2 years (0.5–22.0 years) of follow-up. Patients with SSc had more frequent fractures compared to OA comparators (hazard ratio [HR] 2.38 [95% confidence interval (95% CI) 1.47–3.83]). Among patients with SSc, a higher Rheumatic Disease Comorbidity Index score (HR 1.45 [95% CI 1.20–1.75]) and a higher Health Assessment Questionnaire disability index score (HR 3.83 [95% CI 2.12–6.93]) were associated with more fractures. Diabetes mellitus (HR 5.89 [95% CI 2.51–13.82]) and renal disease (HR 2.43 [95% CI 1.10–5.37]) were independently associated with fracture among SSc patients relative to SSc patients without these comorbidities.

Conclusion. Our findings highlight factors associated with fracture among patients with SSc. Disability as measured by the HAQ DI is a particularly strong indicator of fracture rate in SSc. Improving SSc patients' functional status, where possible, may lead to better long-term outcomes.

INTRODUCTION

Systemic sclerosis (SSc) is an autoimmune, connective tissue disease characterized by cutaneous fibrosis, progressive microvascular disease, and internal organ dysfunction. Prevention and management of SSc complications are important in minimizing progression to functional loss and disability, both of which negatively impact quality of life and survival (1,2). Notably, many SSc complications (e.g., chronic systemic inflammation, gastrointestinal [GI] malabsorption or malnutrition, low vitamin D levels and renal disease, and restricted physical activity due to contractures and/or weakness) are potential risk factors for osteoporosis and fracture (3,4). In addition, several of the medications used to treat manifestations of SSc, such as proton-pump inhibitors (PPIs), glucocorticoids, selective serotonin reuptake inhibitors (SSRIs), opioids, and nonsteroidal antiinflammatory drugs may also contribute to risk and/or rate of fracture (5–9). As a result, patients with SSc may have more fractures secondary to both disease-specific complications and exposure to medications used to manage multisystem dysfunction (10).

While a study of patients from France found that SSc is associated with an increased risk of osteoporosis and fracture compared to the general population (4), the risk of fracture in patients with SSc in the US and the relative impact that specific SSc-

Supported by the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases grant K23-AR-071473), the Scleroderma Research Foundation, the Rheumatology Research Foundation (grant K23 supplement), and the Jerome L. Greene Foundation (grant to Dr. McMahan).

[1]Bliss Rogers, DO, Sina Famenini, MD, MPH, Jamie Perin, PhD, MS, Zsuzsanna H. McMahan, MD, MHS: Johns Hopkins University, Baltimore, Maryland; [2]Maria I. Danila, MD, MSc, MSPH: University of Alabama at Birmingham and Birmingham VA Medical Center, Birmingham; [3]Kristin Wipfler, PhD: FORWARD, The National Databank for Rheumatic Diseases, Wichita, Kansas; [4]Kaleb Michaud, PhD: FORWARD, The National Databank for Rheumatic Diseases, Wichita, Kansas, and University of Nebraska Medical Center, Omaha.

Drs. Rogers and Famenini contributed equally to this work.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25137.

Address correspondence via email to Zsuzsanna H. McMahan, MD, MHS, at zmcmaha1@jhmi.edu.

Submitted for publication August 16, 2022; accepted in revised form April 18, 2023.

SIGNIFICANCE & INNOVATIONS

- Patients living with systemic sclerosis in the US are at a high risk of fractures given their unique risk factors.
- Functional disability is a strong and potentially modifiable factor associated with rate of fracture in patients with systemic sclerosis.
- A high comorbidity burden is present among patients with fracture(s) and systemic sclerosis.
- Diabetes mellitus and renal disease are specific comorbidities that increase fracture rate in patients with systemic sclerosis.

related comorbidities and other clinical features may have in these patients remain unclear. Furthermore, the identification of modifiable risk factors for fracture could improve patient quality of life while secondarily addressing the economic burden associated with osteoporotic fractures (11). Therefore, the aim of this study was to examine the relationship between clinical risk factors (e.g., medication use and clinical characteristics of SSc) and fractures in SSc patients in a large US cohort (Figure 1) and to determine how these factors compare to a group of patients with OA from the same cohort of similar age and sex but without SSc.

PATIENTS AND METHODS

Patients. Patients were part of an observational study as participants in FORWARD, The National Databank for Rheumatic Diseases (1998–2019). FORWARD is a longitudinal observational patient-driven database founded as a nonprofit research organization in 1998. A primary questionnaire is distributed to patients twice yearly. Over 50,000 patients with >100 rheumatic diseases followed by >1,500 rheumatologists have completed at least one 6-month questionnaire (12–14). We hypothesized that patients with SSc would have a higher rate of fracture than OA comparators given their unique clinical characteristics. Patients with SSc and age- and gender-matched patients with physician-diagnosed osteoarthritis (OA) from the same registry were included as comparators. Age matching was done in decades. Though SSc and OA were matched 1:5 when possible (from a pool of ~1:25), the 20s age group did not have sufficient numbers for 1:5 matching, and additional participants with OA in their 30s were matched to compensate. As a result, there were no unmatched patients with SSc. Matching was done without replacement. All participants lived in the US and completed ≥2 semiannual questionnaires. There were 162 (18%) participants (of 922) with only 2 follow-up visits, and 305 (33%) with >10 visits. The vast majority of participants (863 of 922) responded to the survey at least yearly during the period when they were active. The average time between responses across all patients was 0.56 years, with a median of 0.51 years. Patients with OA were chosen as the control group since they were recruited through the same methods and do not generally share the risk factors for fractures that are enriched among patients with SSc (e.g., severe GI disease, renal disease, glucocorticoid therapy, chronic systemic inflammation, chronic high-dose PPI therapy). Patients diagnosed with both SSc and OA, or patients with either condition and concomitant rheumatic diseases (e.g., rheumatoid arthritis), were excluded. This study was approved by Ascension Via Christi Hospitals Wichita Institutional Review Board.

Covariates. Our analysis considered a variety of covariates, detailed below, in the multivariable models. Demographic data including age, disease duration, sex, postmenopausal status, race, education, body mass index (BMI), and smoking status were obtained from enrollment and semiannual questionnaires. With the exception of sex and race, these factors were reported by patients at each questionnaire and so in general were varying over time for each patient, including BMI and PPI use.

Age (years) was examined as a categorical variable (20–50 years [referent], 51–64 years, ≥65 years). The age at first

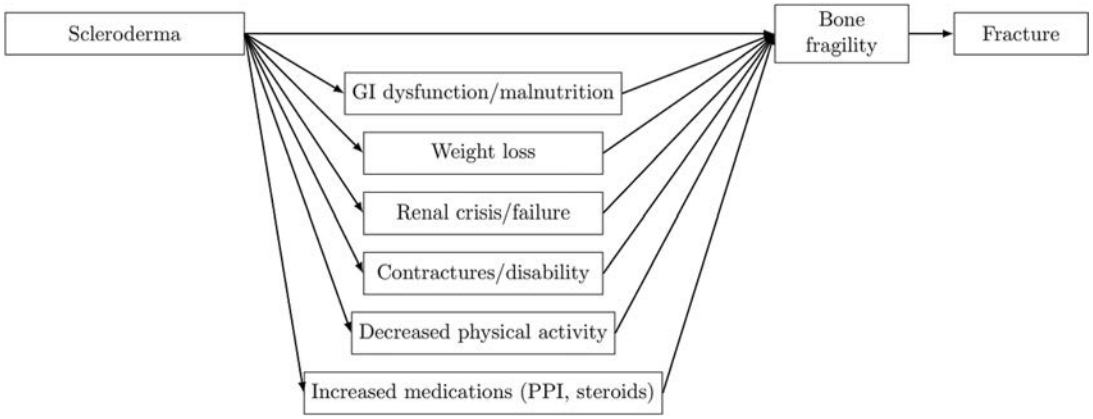

Figure 1. Causal diagram illustrating the relationships between scleroderma and its clinical features, bone fragility, and fracture. GI = gastrointestinal; PPI = proton-pump inhibitor.

symptoms (years) was used to calculate disease duration (years) from the date of first symptom to the date of the questionnaire. Sex was defined as a dichotomous variable (male/female). Race and ethnicity (i.e., White versus Black versus Asian or Pacific Islander; non-Hispanic versus Hispanic) were defined as dichotomous variables (non-Hispanic White versus other). Postmenopausal status was self-reported by the patient at the time of questionnaire completion. BMI was calculated and reported as both a numerical and categorical variable as described by the World Health Organization (kg/m^2; underweight <18.5 kg/m^2; normal weight between 18.5 and 25 kg/m^2; overweight between 25 and 30 kg/m^2; and obese >30 kg/m^2). Smoking history was studied as a dichotomous variable (past or current smoker versus never smoker). Other demographic data included education (years) and health insurance (yes/no).

Patients were asked about medical histories. If specific items in the medical history were unanswered by patients, they were assumed missing. Clinical data including Rheumatic Disease Comorbidity Index (RDCI; range 0–9) (15), Health Assessment Questionnaire disability index (HAQ DI; range 0–3) (16), diabetes mellitus, renal disease, osteoporosis, fracture risk assessment score estimating the probability of major osteoporotic fracture (MOF) or hip fracture within the next 10 years (17), GI disorder, GI scale (range 0–100), and history of GI symptoms (Table 1) were obtained using self-reported questionnaires. Diabetes mellitus was defined as ever having received a diagnosis of type 1 diabetes mellitus or type 2 diabetes mellitus (yes/no). Renal disease was defined as any history of or treatment for renal or kidney failure, reduced kidney function or elevated creatinine, and/or clinician diagnosed hematuria or proteinuria (yes/no). Osteoporosis was defined as having ever received a diagnosis of osteoporosis from a provider (yes/no). Gastrointestinal disorder was defined as "any history of GI disorder" (yes/no) and included liver disease, gallbladder disease, ulcers, and other stomach problems. Specific GI symptoms included irritable bowel syndrome, indigestion, vomiting, constipation, loss of appetite, or peptic ulcer disease. The GI scale (range 0–100), as defined in a previous study, utilized a visual analog scale to assess patient-reported GI symptom severity. A score of 0 indicated that a patient perceived no stomach problems, while a score of 100 indicated severe stomach problems in the past week. SSc was defined as having ever received a diagnosis of SSc (yes/no).

Medication exposure. Medication exposure was recorded at study enrollment and at 6-month intervals using questionnaires. Exposure to PPIs was examined in time-varying doses (omeprazole equivalents and using the following duration-combined categories: not-using [reference], low-dose PPI [≤20-mg omeprazole equivalents per day], and high-dose PPI [>20 mg omeprazole-equivalents per day]). Cumulative exposure to PPIs was also assessed and was defined as the number of previous surveys in which the participants indicated any use of PPI. It was measured using total omeprazole equivalents in grams of PPI taken during follow-up and was categorized as either low dose (≤20 mg) or high dose (>20 mg).

All medications were time varying, as participants were queried about their medication use at each follow-up. Glucocorticoid exposure was examined in terms of prednisone equivalents using the following categories: no use (reference); low dose (≤7.5 mg/day); and medium-high dose (>7.5 mg/day). Other medication exposures, including patient use of any osteoporosis medication(s), SSRIs, estrogen, opioid analgesia (weak or strong), nonopioid analgesia, and/or anticonvulsant(s), were defined as dichotomous variables (yes/no).

Measures of disease severity and quality of life. *Fracture risk assessment.* Fracture risk assessment was determined by a modified fracture risk assessment score as previously described in the Fracture Risk Assessment (FRAX) tool (version 4.1) (http://www.shef.ac.uk/FRAX) (17). The FRAX tool utilizes clinical risk factors for fracture to estimate the 10-year probability of major osteoporotic fracture or hip fracture incidence. Our FRAX score calculation was modified for the lack of data on available bone mineral density and unknown familial history of hip fracture as detailed in the referenced prior study (18).

HAQ DI. We utilized a validated self-administered questionnaire (16), which had patients quantify difficulty in performing activities of daily living over the previous week. Physical disability was assessed across 8 components: dressing and grooming; arising; eating; walking; hygiene; reach; grip; and activity. Questionnaires included 2–3 questions per component for a total of 20 questions specifically addressing activities of daily living. Patient response to each question was scored from 0 to 3 (score 0 = without any difficulty; score 1 = with some difficulty; score 2 = with much difficulty; score 3 = unable to do). The question with the highest score determined the component score (range 0–3). The disability index was calculated by dividing the total of the component scores by the number of components assessed. HAQ DI scores range from 0 to 3, with a higher score representing more severe disability (score 0 = no impairment in function/disability; 3 = maximal impairment in function/disability).

RDCI. We used patient-reported data to evaluate the burden of comorbidity across 11 comorbid conditions outlined in previous studies (19,20). Conditions included ulcer or stomach problem, hypertension, myocardial infarction, other cardiovascular disease (e.g., heart failure), lung disease, diabetes mellitus, depression, cancer, stroke, and fracture (i.e., spine, hip, or leg). The index was calculated based on the presence and severity of coexisting comorbid conditions, with impact assessed using 6 outcomes that were weighted as described in previous studies (15,19). The RDCI ranges from 0 to 9 and assesses quality of life and the anticipated effects of comorbidities on functional disability and mortality. A higher RDCI is predictive of increased physical disability and mortality (15).

Outcomes. The primary outcome was major osteoporotic fracture (i.e., fracture of the hip, which can include pelvis and/or femur, the humerus, clinical spine, and wrist) (21) among patients with SSc and OA. Fractures of the skull, hands, feet, fingers, and toes (21) were excluded. First fracture during the follow-up period and all subsequent fractures were included in the analysis such that some participants experienced multiple fractures during follow-up, all of which were incorporated in the analysis. Follow-up time commenced at cohort entry and continued until censoring at death, loss to follow-up, or until August 2019.

Statistical analysis. Baseline characteristics were compared between patients with SSc and OA comparators. We also used responses from the baseline survey to estimate associations

Table 1. Baseline demographic and clinical characteristics of patients with systemic sclerosis (SSc) and age- and sex-matched osteoarthritis (OA) (n = 922)*

Clinical feature	Patients with SSc (n = 154)		Patients with OA (n = 768)		
	No.	Value	No.	Value	P†
Female	154	87	768	87	0.991
Age, mean ± SD years	154	57.4 ± 12.6	768	57.9 ± 11.8	0.663
Age distribution	154		768		0.586
20–50		25.3		23.8	
51–64		46.1		45.3	
≥65		28.6		30.9	
Age at diagnosis, mean ± SD years	144	44.8 ± 15.3	679	45.8 ± 13.3	0.358
Non-Hispanic White	145	81.8	730	84	0.507
Postmenopause status	134	82.8	668	79.8	0.412
Ever smoker	154	47.4	767	42.9	0.303
Disease duration, mean ± SD years	144	12.4 ± 10.7	679	11.7 ± 10.2	0.518
BMI, mean ± SD kg/m^2	143	27.8 ± 6.2	693	30.8 ± 7.6	<0.001
BMI categories	143		693		<0.001
Underweight, <18.5		3.25		0.91	
Normal weight, 18.5–25		33.8		21.1	
Overweight, 25–30		29.9		28.5	
Obese, >30		33.1		49.6	
Health insurance	142	96.5	678	96.2	0.796
Education, mean ± SD years	154	14 ± 2.6	768	14.2 ± 2.4	0.618
RDCI score, mean ± SD (range 0–9)	153	2.4 ± 1.8	766	1.8 ± 1.5	<0.001
HAQ DI score, mean ± SD (range 0–3)	146	0.9 ± 0.6	709	0.8 ± 0.3	0.140
Diabetes mellitus	153	5.2	766	12.9	0.007
Renal disease	153	16.2	766	8.0	0.004
Osteoporosis	154	0.6	768	0.4	0.656
FRAX MOF score, mean ± SD (range 0–100)	151	8.3 ± 6.8	752	7.0 ± 5.2	0.046
FRAX hip score, mean ± SD (range 0–100)	151	2.3 ± 4.4	752	1.5 ± 2.6	0.027
PPI use	144	66	386	27.7	<0.001
Glucocorticoid use, %‡	59		291		<0.001
None		77.0		93.4	
Low, ≤7.5 mg/day		18.2		3.9	
Medium and high, >7.5 mg/day		4.7		2.7	
SSRI use, %	148	16.9	636	11.5	0.365
Osteoporosis medication use, %	154	18.2	768	27.2	0.020
GI disorder, %	153	54.6	766	41.4	0.007
GI scale, mean ± SD (range 0–100)	90	34.1 ± 31.1	496	19.1 ± 24.1	<0.001
History of diarrhea	146	28.8	722	15.8	<0.001
History of IBS	139	16.5	388	17.0	1.000
History of indigestion	146	43.8	722	30.3	0.002
History of vomiting	146	14.4	720	4.2	<0.001
History of constipation	146	30.8	722	24.0	0.094
History of loss of appetite	147	19.7	720	10.3	0.003
History of PUD	146	33.6	723	21.0	0.003

* Values are the percentage unless indicated otherwise. BMI = body mass index; FRAX = Fracture Risk Assessment Tool; GI = gastrointestinal; HAQ DI = Health Assessment Questionnaire disability index; IBS = irritable bowel syndrome; MOF = major osteoporotic fracture; OP = osteoporosis; PPIs = proton-pump inhibitors; PUD = peptic ulcer disease; RDCI = Rheumatic Disease Comorbidity Index; SSRIs = selective serotonin reuptake inhibitors.
† Determined by Wilcoxon's rank sum for continuous factors, Kruskal-Wallis for categorical factors, and chi-square or Fisher's exact test for binary factors, where appropriate.
‡ As prednisone does equivalents.

with rate of fracture. We included all fractures over the study period. Nonparametric Wilcoxon's rank sum tests and Pearson's chi-square tests or Fisher's exact tests were used to compare continuous and discrete variables, respectively. Bivariate/unadjusted and multivariable Cox proportional hazards models were used to investigate the risk factors for fractures or osteoporosis in the SSc and control groups. We examined raw unadjusted associations using Cox proportional hazards for the rate of fractures separately among those with SSc and those with OA, as well as in SSc and OA combined. For characteristics that were related to the rate of fracture in either SSc or OA or the combined group at a significance of 0.05, we also examined the association with Cox proportional hazards adjusted for age and BMI. In sensitivity analysis, we adjusted for smoking in addition to age and sex. We replaced the missing baseline values of BMI (missing for n = 86, 9%) and glucocorticoid use (missing for n = 572, 62%) with multiple imputations by chained equations to create multiple imputed data sets for analysis (22). We did the same for diabetes mellitus (yes/no), PPI use (any versus none), and renal disease. All tests were 2-sided and considered statistically significant when P values were less than 0.05. All statistical analyses were performed using Stata, version 16.0.

RESULTS

Clinical characteristics of SSc and OA patients. The study included 922 patients, 154 with SSc and 768 OA comparators. Fifty-one patients developed at least 1 fracture during a median of 4.2 years of follow-up, with a total of 82 fractures occurring. Of these 82, a total of 55 fractures occurred in the OA patients, while 27 fractures occurred among those with scleroderma. Clinical features of the SSc and OA participants including age, disease duration, sex, postmenopausal status, BMI, RDCI, HAQ DI scores, FRAX scores and history of diabetes mellitus, renal disease, osteoporosis, and GI disease are compared in Table 1. Patients with SSc had a significantly lower BMI compared to OA comparators (27.8 versus 30.8 kg/m^2; $P < 0.001$). As expected, patients with SSc also had more comorbidities when compared to OA comparators (Table 1). SSc patients had higher RDCI scores (2.4 versus 1.8; $P < 0.001$) and slightly higher FRAX MOF scores (8.3% versus7.0%; $P = 0.046$) and FRAX hip scores (2.3 versus 1.5; $P = 0.027$) when compared to OA comparators (FRAX MOF: moderate risk 10 to <20%; high risk ≥20; low risk <10; FRAX hip: high risk ≥3%; moderate risk ≥2 to <3%). However, there was no significant difference in baseline HAQ DI scores (0.9 versus 0.8; $P = 0.140$) between SSc and OA comparators. Both groups also had similar prevalence of preexisting osteoporosis (0.6% versus 0.4%; $P = 0.656$).

Compared to OA comparators, patients with SSc had a lower prevalence of diabetes mellitus (5.2% versus 12.9%; $P = 0.007$) but were more likely to have a history of renal disease (16.2% versus 8.0%; $P = 0.004$) and GI disorders (54.6% versus 41.4%; $P = 0.007$). Within GI disease, patients with SSc reported having more GI symptoms (Table 1), including diarrhea (28.8 versus 15.8; $P < 0.001$), indigestion (43.8 versus 30.3; $P = 0.002$), vomiting (14.4 versus 4.2; $P < 0.001$), loss of appetite (19.7 versus 10.3; $P = 0.003$), and peptic ulcer disease (33.6 versus 21.0; $P = 0.003$).

Univariate regression analysis. We sought to measure the strength of the association between distinct clinical factors and fractures in patients with SSc and OA comparators. We examined associations among SSc and OA comparators combined, as well as only among SSc patients and only among OA comparators. In the univariate analysis (Table 2), among the cohort of SSc and OA comparators, we found that patients with SSc were 2.27 times more likely to have fractures relative to OA comparators (hazard ratio [HR] 2.27 [95% confidence interval (95% CI) 1.43–3.61]) (Figure 2). No significant differences in association of demographic characteristics with fracture occurrence between patients with SSc and OA comparators were identified (Table 2).

Patients with more comorbidities tended to have more fractures. Among patients with SSc, a higher RDCI score, a higher HAQ DI score (Figure 3), the presence of diabetes mellitus, and the presence of renal disease were all associated with more fractures (Table 2). Among patients with SSc, a 1-point increase in RDCI score was associated with a 38% increased rate of fracture (HR 1.38 [95% CI 1.15–1.66]), while a 1-point increase in HAQ DI score was associated with a substantially increased risk of fractures (HR 3.16 [95% CI 1.84–5.41]). An even greater rate for fracture was seen in patients with both SSc and diabetes mellitus. These patients had significantly more fractures when compared to SSc patients without diabetes mellitus (HR 5.10 [95% CI 2.29–11.37]). Interestingly, among OA comparators, there was no significant difference in fractures between those with and without diabetes mellitus. A similar finding was observed when evaluating renal disease as a possible risk factor for fracture in patients with SSc and OA comparators. While no significant difference in the rate of fracture was seen in OA comparators when comparing those with and without renal disease, patients with SSc and renal disease had more fractures compared to SSc patients without renal disease (HR 2.32 [95% CI 1.07–5.00]).

Among SSc patients, those taking PPIs had a 55% lower rate of fracture (HR 0.45 [95% CI 0.21–0.99]) when compared to SSc patients not taking PPIs. Interestingly, this contrasts with the increase in fractures observed among OA comparators taking PPIs compared to those not taking PPIs (HR 2.00 [95% CI 1.16–3.46]). However, there was no significant association with rate of fracture when considering PPI cumulative dose in both SSc patients and OA comparators (Table 2).

The use of glucocorticoids, osteoporosis medications, SSRIs, estrogen, anticonvulsants, opioid analgesia, or nonopioid analgesia was not significantly associated with the rate of fracture

Table 2. Univariate risk factor analysis of fractures in patients with systemic sclerosis and osteoarthritis*

Variable	Whole cohort (n = 922)		Scleroderma patients only (n = 154)		Osteoarthritis patients only (n = 768)	
	Unadjusted HR (95% CI)	*P*	Unadjusted HR (95% CI)	*P*	Unadjusted HR (95% CI)	*P*
Demographic characteristics						
Scleroderma	2.27 (1.43–3.61)	0.001	–	–	–	–
Female	2.39 (0.97–5.92)	0.059	†	†	1.48 (0.59–3.71)	0.404
Age categories						
20–50 years	Ref.		Ref.		Ref.	
51–64 years	1.25 (0.51–3.06)	0.626	†	†	0.56 (0.21–1.52)	0.256
≥65 years	1.81 (0.76–4.30)	0.177	†	†	1.17 (0.48–2.86)	0.735
Non-Hispanic White	1.98 (0.72–5.43)	0.184	†	†	1.25 (0.45–3.47)	0.672
Postmenopausal status (vs. women only)	3.63 (0.89–14.87)	0.073	†	†	2.36 (0.57–9.78)	0.237
Ever smoker	1.37 (0.88–2.12)	0.165	1.98 (0.92–4.28)	0.083	1.00 (0.58–1.74)	0.999
Disease duration, years	1.01 (0.99–1.03)	0.379	1.00 (0.96–1.05)	0.862	1.01 (0.99–1.04)	0.297
BMI categories, kg/m^2						
Underweight, <18.5	1.33 (0.39–4.55)	0.646	1.21 (0.26–5.57)	0.808	0.97 (0.12–7.74)	0.976
Normal weight,18.5–25.0	Ref.		Ref.		Ref.	
Overweight, 25–30	1.04 (0.59–1.83)	0.884	0.52 (0.18–1.50)	0.227	1.55 (0.75–3.20)	0.237
Obese, >30	0.78 (0.45–1.36)	0.386	0.77 (0.32–1.87)	0.565	0.95 (0.46–1.98)	0.899
Comorbidities						
RDCI	1.32 (1.17–1.48)	<0.001	1.38 (1.15–1.66)	0.001	1.21 (1.03–1.43)	0.018
HAQ DI	2.56 (1.84–3.55)	<0.001	3.16 (1.84–5.41)	<0.001	2.07 (1.38–3.12)	<0.001
Diabetes mellitus	1.73 (1.08–2.78)	0.024	5.10 (2.29–11.37)	<0.001	1.29 (0.71–2.34)	0.406
Renal disease	1.46 (0.90–2.37)	0.121	2.32 (1.07–5.00)	0.032	1.02 (0.53–1.95)	0.952
GI disorder	0.98 (0.61–1.58)	0.937	1.21 (0.48–3.03)	0.682	0.83 (0.47–1.47)	0.531
GI scale (for 10-point increase)	0.93 (0.79–1.09)	0.349	0.79 (0.56–1.14)	0.209	0.92 (0.76–1.12)	0.413
Medications						
PPIs	1.56 (1.00–2.43)	0.050	0.45 (0.21–0.99)	0.046	2.00 (1.16–3.46)	0.013
PPI cumulative dose‡	1.03 (0.99–1.07)	0.106	0.95 (0.88–1.02)	0.166	1.04 (0.99–1.10)	0.086
Low dose (≤20 mg)	Ref.		Ref.		Ref.	
High dose (>20 mg)	0.28 (0.03–3.22)	0.310	†	†	0.15 (0.01–2.44)	0.183
Glucocorticoids						
None	Ref.		Ref.		Ref.	
Low, ≤7.5 mg/day	1.37 (0.59–3.16)	0.465	1.34 (0.53–3.41)	0.539	†	†
Medium and high, >7.5 mg/day	4.50 (1.80–11.25)	0.001	2.55 (0.58–11.19)	0.215	5.03 (1.55–16.36)	0.007
Osteoporosis medications	0.60 (0.08–4.29)	0.607	0.78 (0.10–5.79)	0.804	†	†
SSRIs	1.23 (0.69–2.20)	0.482	0.77 (0.18–3.37)	0.728	1.54 (0.81–2.94)	0.190
Estrogen	0.64 (0.33–1.24)	0.181	0.27 (0.04–2.02)	0.204	0.89 (0.43–1.82)	0.741
Opioid analgesia	1.46 (0.92–2.32)	0.106	0.52 (0.18–1.52)	0.232	2.11 (1.23–3.61)	0.006
Nonopioid analgesia§	0.60 (0.39–0.92)	0.020	0.76 (0.34–1.70)	0.509	0.65 (0.38–1.11)	0.113
Anticonvulsants	0.96 (0.52–1.79)	0.908	0.92 (0.32–2.67)	0.875	0.92 (0.43–1.95)	0.823

* 95% CI = 95% confidence interval; BMI = body mass index; GI = gastrointestinal; HAQ DI = Health Assessment Questionnaire disability index; HR = hazard ratio; PPIs = proton-pump inhibitors; RDCI = Rheumatic Disease Comorbidity Index; Ref. = reference; SSRIs = selective serotonin reuptake inhibitors.
† Not estimable due to low number of fractures within ≥1 categories.
‡ As omeprazole dose equivalents, in grams; cumulative PPI use was defined as the number of previous surveys where the participants indicated any use of PPI.
§ Nonsteroidal antiinflammatory drugs and cyclooxygenase 2 inhibitors.

among patients with SSc (Table 2). However, among OA comparators, those taking opioid analgesia had more fractures than those not taking opioid analgesia (HR 2.11 [95% CI 1.23–3.61]), and OA comparators taking medium- and high-dose glucocorticoids had more fractures than OA comparators not taking those medications (HR 5.03 [95% CI 1.55–16.36]).

Multivariable regression analysis. We sought to determine whether the associations between clinical features and fracture from the univariable analyses remained after adjusting for clinically relevant covariates and potential confounders. In the multivariable model, we adjusted for age and BMI and then reexamined the associations between fracture and each of the following variables: sex; scleroderma; GI disease; diabetes mellitus; renal disease; medication use; RDCI score; and HAQ DI score (Table 3). We determined that patients with SSc were more likely to have fracture relative to OA comparators (HR 2.38 [95% CI 1.47–3.83]). Furthermore, among patients with SSc, the associations between a significantly higher rate of fracture in patients with SSc and a higher RDCI score, higher HAQ DI score, and/or the presence of diabetes mellitus and renal disease from the unadjusted analysis (Table 2) remained significant in the multivariable

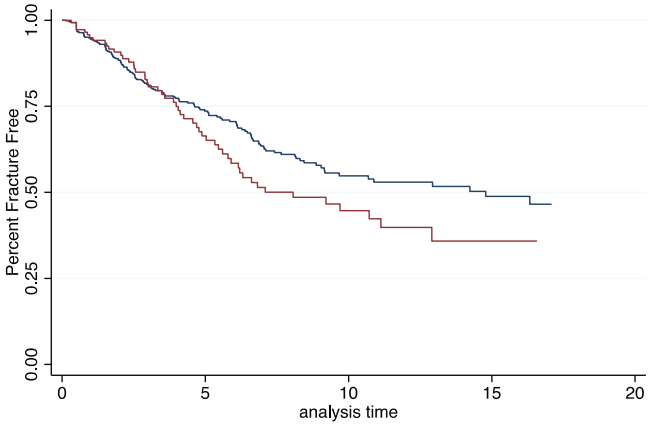

Figure 2. Kaplan-Meier estimates for the percent fracture free over the duration of the study (years), by Health Assessment Questionnaire disability index for systemic sclerosis (red line) and osteoarthritis (blue line) patients.

analyses (Table 3). In the patients with SSc, a 1-point increase in their RDCI score was associated with a 45% increased rate for fracture (HR 1.45 [95% CI 1.20–1.75]), and a 1-point increase in HAQ DI score was associated with an almost 3-fold increased rate for fracture (HR 3.83 [95% CI 2.12–6.93]). In patients with SSc, those with diabetes mellitus were 5.89 times as likely to have a fracture than those without diabetes mellitus (HR 5.89 [95% CI 2.51–13.82]), and those with renal disease were 2.43 times as likely to have fracture compared to SSc patients without renal disease (HR 2.43 [95% CI 1.10–5.37]). In a sensitivity analysis, we adjusted for smoking in addition to age and sex. We did not see any changes in the results for this sensitivity analysis (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25137).

DISCUSSION

Our study is the first to examine the rate of fracture among patients with SSc in the US by examining a large well-characterized sample of patients with rheumatic diseases. We observed a higher fracture rate in patients with SSc relative to OA comparators. We found that comorbidity burden and higher physical disability were strongly associated with a high fracture rate in both scleroderma and OA. Diabetes mellitus and renal disease were also determined to increase fracture rate in patients with SSc even after adjusting for age and BMI. After adjustment, significant disability and comorbidity burden were most strongly associated with fracture in patients with SSc.

Our finding that US patients with SSc are subject to an elevated rate of osteoporotic fracture compared to OA comparators is consistent with previous studies globally (23–25). For example, a French study by Avouac et al found that osteoporotic fractures are more prevalent among patients with SSc compared to a healthy control population (4), and Lai et al identified an almost 2 times higher rate for osteoporotic fracture in patients with SSc compared to the general population (24).

Another important result of our study is the finding that physical disability, a partially modifiable risk factor (26–29), is associated with a high fracture rate in patients with SSc. Musculoskeletal involvement such as joint contracture, myopathy, and arthritis are common features of SSc (30), which can lead to disability and impaired quality of life (1,23,31,32). Various local and global therapeutic rehabilitation programs (i.e., home exercises [29], manual lymphatic drainage [28], paraffin wax baths [33], connective tissue massage [34], joint manipulation [34], physical and occupational therapy [35,36], and aerobic exercises [37]) were noted to improve the function of patients with SSc. Specifically, improvements in mobility, flexibility, skin elasticity, aerobic capacity, and muscle endurance and decreased edema and stiffness were reported. Interventions targeting hand and upper extremity function may also reduce disability in SSc, and self-stretching programs may improve grip strength (38). Furthermore, home-based physical exercise regimens were also shown to decrease disability and improve biceps and quadriceps strength (29). As our study is the first to determine that high HAQ DI scores are strongly associated with high rate for fracture in patients with SSc, the timing and application of these interventions in high-risk patient subsets should be an important focus of future studies.

Our study also determined that a high comorbidity burden is an important risk factor for fracture in patients with SSc. Patients with high RDCI scores were more likely to have fractures when compared to patients with fewer comorbidities. These data provide additional insight to support the early recognition of SSc patients who have a high risk for fracture. Furthermore, it suggests that the early diagnosis and appropriate management of comorbidities in patients with SSc should be a priority, and that interventions, such as multidisciplinary rehabilitation (i.e., aimed

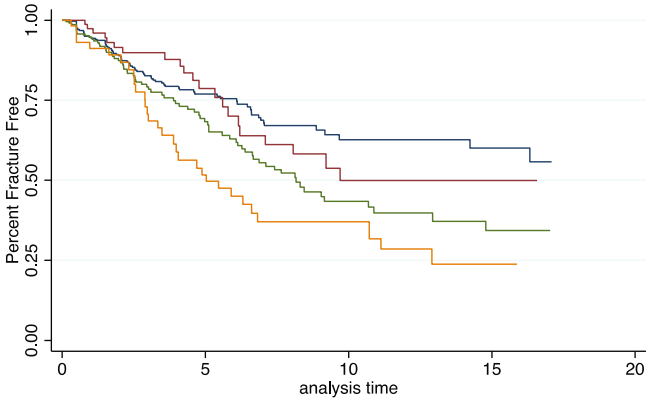

Figure 3. Kaplan-Meier estimates for the percent fracture free over the duration of the study (years), by Health Assessment Questionnaire disability index (HAQ DI) for systemic sclerosis patients with a HAQ DI score of <1 (red line) and a HAQ DI score of ≥1 (orange line), and for osteoarthritis patients with a HAQ DI score of <1 (blue line) and a HAQ DI score of ≥1 (green line).

Table 3. Multivariable analysis evaluating risk of fracture in patients with systemic sclerosis and osteoarthritis*

Variable	Whole cohort (n = 922)		Scleroderma patients only (n = 154)		Osteoarthritis patients only (n = 768)	
	Adjusted HR (95% CI)†	*P*	Adjusted HR (95% CI)†	*P*	Adjusted HR (95% CI)†	*P*
Demographic characteristics						
Scleroderma	2.38 (1.47–3.83)	<0.001	–	–	–	–
Female	2.57 (1.04–6.38)	0.042	‡	‡	1.69 (0.67–4.27)	0.268
Non-Hispanic White	1.99 (0.72–5.49)	0.184	‡	‡	1.32 (0.47–3.73)	0.594
Ever smoked	1.32 (0.85–2.05)	0.222	1.73 (0.79–3.77)	0.169	0.97 (0.56–1.69)	0.921
Comorbidities						
RDCI	1.33 (1.18–1.50)	<0.001	1.45 (1.20–1.75)	<0.001	1.23 (1.04–1.45)	0.015
HAQ DI	2.98 (2.12–4.19)	<0.001	3.83 (2.12–6.93)	<0.001	2.44 (1.58–3.76)	<0.001
Gastrointestinal disorder	1.02 (0.63–1.65)	0.930	1.38 (0.53–3.58)	0.508	0.87 (0.49–1.54)	0.638
Diabetes mellitus	1.92 (1.17–3.16)	0.010	5.89 (2.51–13.82)	<0.001	‡	‡
Renal disease	1.43 (0.88–2.32)	0.146	2.43 (1.10–5.37)	0.029	1.00 (0.52–1.92)	0.996
Medications						
PPIs	0.89 (0.33–2.36)	0.795	0.60 (0.21–1.72)	0.342	1.01 (0.26–3.92)	0.983
PPI cumulative dose§	1.03 (0.99–1.08)	0.097	0.93 (0.86–1.01)	0.100	1.04 (0.99–1.10)	0.087
Glucocorticoids						
None	Ref.		Ref.		Ref.	
Low, ≤7.5 mg/day	1.39 (0.60–3.22)	0.444	1.53 (0.57–4.10)	0.393	‡	‡
Medium and high, >7.5 mg/day	4.71 (1.87–11.86)	0.001	2.73 (0.60–12.41)	0.194	5.07 (1.55–16.63)	0.007
Estrogen	0.61 (0.31–1.20)	0.152	0.25 (0.03–1.91)	0.183	0.95 (0.46–1.98)	0.897
Opioid analgesia	1.50 (0.94–2.39)	0.089	0.53 (0.17–1.61)	0.263	2.15 (1.24–3.73)	0.006
Nonopioid analgesia	0.60 (0.39–0.93)	0.023	0.77 (0.34–1.76)	0.539	0.67 (0.39–1.14)	0.140

* All associations were adjusted for age and body mass index (BMI). 95% CI = 95% confidence interval; HAQ DI = Health Assessment Questionnaire Disability Index; HR = hazard ratio; PPIs = proton-pump inhibitors; RDCI = Rheumatic Disease Comorbidity Index.
† Adjusted for age and BMI in 3 categories.
‡ Not estimable due to low number of fractures within ≥1 categories.
§ Cumulative PPI use was defined as the number of previous surveys where the participants indicated any use of PPI.

to improve mobility and overall function) (39) may reduce long-term fracture risk.

To our knowledge, this is the first study to determine that diabetes mellitus is an independent risk factor for fracture in patients with SSc. Prior studies in the general population reported a 2–5 times higher risk for fracture in patients with diabetes mellitus when compared to those without diabetes mellitus (40) and an association between worse fracture outcomes in patients with diabetes mellitus when compared to normoglycemic individuals (41). Given the morbidity, mortality, and economic burden associated with fracture, our finding that diabetes mellitus is a risk factor for fracture in patients with SSc warrants further study. Our ability to provide more specific recommendations for risk stratifying SSc patients with diabetes mellitus is limited due to lack of data on the severity of diabetes mellitus and glycemic control within these participants.

Renal disease is associated with an increased rate of fracture when compared to patients without renal disease in the general population (range HR 1.16 [95% CI 1.01–1.33] to HR 5.04 [95% CI 1.38–18.45]) (42–44). However, previous studies examining risk factors for fracture in patients with SSc have failed to consider or excluded patients with comorbid renal disease (45). Our study found that patients with SSc and renal disease were significantly more likely to have fracture compared to SSc patients without renal disease (HR 2.43 [95% CI 1.10–5.37]). This finding is consistent with increased risk for fracture observed among patients with renal disease in the general population (42). As a result, particular attention should be directed toward monitoring and preventing bone loss in SSc-associated renal disease, as an increased risk for fracture exists among these patients.

Interestingly, we found that patients with SSc who take PPIs had a lower rate of fracture when compared to SSc patients not taking PPIs. This finding contrasts the association between PPI use and increased fracture risk reported in the general population (8,46). However, previous studies have not yet established a causative link between PPI use and fracture. Additional large prospective studies are needed to clarify the true risk of PPI use on fracture in the general population and in patients with other conditions such as SSc. We would also recommend interpreting this association with caution, as methodological imputations were utilized for missing values on PPI use in one-third of our patients.

Our study has several strengths. This is the first study to examine the rate of fracture among patients with SSc in the US. The FORWARD study includes well-characterized patients with SSc and a built-in comparator group of age- and sex-matched OA patients. Importantly, the OA comparators were not significantly distinct from patients with SSc in terms of demographic characteristics and lacked many of the risk factors for fractures that are enriched among patients with SSc. Additionally, we included and adjusted for age and BMI in patients with SSc to

limit possible confounding or unmeasured risk factors not addressed in previous studies. With regard to limitations, fractures were self-reported; therefore, a possibility exists for selection bias and underreporting events. We did not adjust for the history of fracture before study enrollment, as these data were not available to us. Although we had a large sample of well-defined patients, the number of SSc patients and patients with fracture(s) was relatively small, likely limiting our power to identify other important associations. Additionally, the patient questionnaire did not include information regarding the mechanism of fracture (i.e., low- versus high-energy trauma). Furthermore, given the small number of fractures, the effect size may have also been overestimated. Given the matching in decades and the adjustment in very large categories, residual confounding by age is a possibility, and there is always potential for other confounding due to unmeasured factors. There is also the potential that matching skewed the OA population to slightly younger people overall. As disease severity or duration of diabetes mellitus or renal disease were not quantified among patients, it is difficult to make any specific recommendations for risk stratification within these subpopulations of affected patients with SSc. We also recognize that our sample was primarily composed of non-Hispanic White female patients. More studies are needed to identify whether our findings are generalizable to other ethnic groups. Finally, our conclusions are limited in part by incomplete responses from participants (see Supplementary Table 1, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25137).

In conclusion, patients with SSc are at a high risk for fracture given their unique risk factors. Among patients with SSc, high comorbidity, high disability, diabetes mellitus, and renal disease appear to be independent risk factors for MOF. To our knowledge, this is the first study to identify disability as a direct, partially modifiable risk factor for fracture in patients with SSc. Disability in SSc, as measured by the HAQ DI, is a particularly strong indicator of future rate of fracture. Routine screening for disability associated with SSc, and integrating interventions aimed at improving patients' functional status, may together lead to improved clinical outcomes in SSc. We therefore suggest that clinicians consider utilizing tools such as the RDCI and the HAQ DI in clinical assessments for the early identification of patients with SSc who are at increased risk for major osteoporotic fracture. Multidimensional therapeutic rehabilitation (36) should be targeted at improving mobility in these patients to limit risk of fracture and fracture-related morbidity and mortality. Future prospective studies are needed to further elucidate the appropriate timing and modalities that are most appropriate for specific patient subgroups.

ACKNOWLEDGMENTS

We thank Drs. Helena Johansson, Mattias Lorentzon, Nicholas C Harvey, Eugene V. McCloskey, and John A. Kanis for their contributions and expertise related to the FRAX score analysis.

AUTHOR CONTRIBUTIONS

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

Study conception and design. Famenini, Perin, Danila, Wipfler, Michaud, McMahan.

Acquisition of data. Wipfler, Michaud.

Analysis and interpretation of data. Rogers, Perin, Danila.

REFERENCES

1. Lóránd V, Czirják L, Minier T. Musculoskeletal involvement in systemic sclerosis [review]. Presse Med 2014;43:e315–28.
2. Medsger TA, Rodnan GP, Moossy J, et al. Skeletal muscle involvement in progressive systemic sclerosis (scleroderma). Arthritis Rheum 1968;11:554–68.
3. Loucks J, Pope JE. Osteoporosis in scleroderma [review]. Semin Arthritis Rheum 2005;34:678–82.
4. Avouac J, Koumakis E, Toth E, et al. Increased risk of osteoporosis and fracture in women with systemic sclerosis: a comparative study with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2012;64:1871–8.
5. Bolton JM, Morin SN, Majumdar SR, et al. Association of mental disorders and related medication use with risk for major osteoporotic fractures. JAMA Psychiatry 2017;74:641–8.
6. Ping F, Wang Y, Wang J, et al. Opioids increase hip fracture risk: a meta-analysis. J Bone Miner Metab 2017;35:289–97.
7. Souverein PC, Webb DJ, Weil JG, et al. Use of antiepileptic drugs and risk of fractures: case-control study among patients with epilepsy. Neurology 2006;66:1318–24.
8. Yang YX, Lewis JD, Epstein S, et al. Long-term proton pump inhibitor therapy and risk of hip fracture. JAMA 2006;296:2947–53.
9. Vestergaard P, Hermann P, Jensen JE, et al. Effects of paracetamol, non-steroidal anti-inflammatory drugs, acetylsalicylic acid, and opioids on bone mineral density and risk of fracture: results of the Danish Osteoporosis Prevention Study (DOPS). Osteoporos Int 2012;23:1255–65.
10. Gale SL, Trinh H, Mathew N, et al. Characterizing disease manifestations and treatment patterns among adults with systemic sclerosis: a retrospective analysis of a US healthcare claims population. Rheumatol Ther 2020;7:89–99.
11. Williams SA, Chastek B, Sundquist K, et al. Economic burden of osteoporotic fractures in US managed care enrollees. Am J Manag Care 2020;26:e142–9.
12. Michaud K. The National Data Bank for Rheumatic Diseases (NDB). Clin Exp Rheumatol 2016;34 Suppl 101:S100–1.
13. Wolfe F, Michaud K. A brief introduction to the National Data Bank for Rheumatic Diseases. Clin Exp Rheumatol 2005;23 Suppl 39:S168–71.
14. Wolfe F, Michaud K. The National Data Bank for rheumatic diseases: a multi-registry rheumatic disease data bank. Rheumatology (Oxford) 2011;50:16–24.
15. England BR, Sayles H, Mikuls TR, et al. Validation of the rheumatic disease comorbidity index. Arthritis Care Res (Hoboken) 2015;67:865–72.
16. Poole JL, Steen VD. The use of the Health Assessment Questionnaire (HAQ) to determine physical disability in systemic sclerosis. Arthritis Care Res 1991;4:27–31.
17. FRAX. WHO fracture risk assessment tool. URL: http://www.shef.ac.uk/FRAX.

18. Ozen G, Pedro S, Wolfe F, et al. Medications associated with fracture risk in patients with rheumatoid arthritis. Ann Rheum Dis 2019;78: 1041–7.
19. Huang YJ, Chen JS, Luo SF, et al. Comparison of indexes to measure comorbidity burden and predict all-cause mortality in rheumatoid arthritis. J Clin Med 2021;10:5460.
20. Wolfe F, Michaud K, Li T, et al. Chronic conditions and health problems in rheumatic diseases: comparisons with rheumatoid arthritis, noninflammatory rheumatic disorders, systemic lupus erythematosus, and fibromyalgia. J Rheumatol 2010;37:305–15.
21. Danila MI, Outman RC, Rahn EJ, et al. Evaluation of a multimodal, direct-to-patient educational intervention targeting barriers to osteoporosis care: a randomized clinical trial. J Bone Miner Res 2018;33: 763–72.
22. Rubin DB. Multiple imputation for nonresponse in surveys. In: multiple imputation for nonresponse in surveys. New York (NY): John Wiley & Sons; 1987: p. i–xxix.
23. Avouac J, Walker U, Tyndall A, et al. Characteristics of joint involvement and relationships with systemic inflammation in systemic sclerosis: results from the EULAR Scleroderma Trial and Research Group (EUSTAR) database. J Rheumatol 2010;37:1488–501.
24. Lai CC, Wang SH, Chen WS, et al. Increased risk of osteoporotic fractures in patients with systemic sclerosis: a nationwide population-based study. Ann Rheum Dis 2015;74:1347–52.
25. Omair MA, Pagnoux C, McDonald-Blumer H, et al. Low bone density in systemic sclerosis: a systematic review [review]. J Rheumatol 2013; 40:1881–90.
26. Mugii N, Hamaguchi Y, Maddali-Bongi S. Clinical significance and usefulness of rehabilitation for systemic sclerosis. J Scleroderma Relat Disord 2018;3:71–80.
27. Schouffoer AA, Ninaber MK, Beaart-van de Voorde LJ, et al. Randomized comparison of a multidisciplinary team care program with usual care in patients with systemic sclerosis. Arthritis Care Res (Hoboken) 2011;63:909–17.
28. Bongi SM, del Rosso A, Passalacqua M, et al. Manual lymph drainage improving upper extremity edema and hand function in patients with systemic sclerosis in edematous phase. Arthritis Care Res (Hoboken) 2011;63:1134–41.
29. Filippetti M, Cazzoletti L, Zamboni F, et al. Effect of a tailored home-based exercise program in patients with systemic sclerosis: a randomized controlled trial. Scand J Med Sci Sports 2020;30:1675–84.
30. Sebastian MR, Eliana OC, Gerardo QL. Osteoarticular manifestations of systemic sclerosis: a systematic review of the literature. Revista Colombiana de Reumatología (English Edition) 2020;27:85–110.
31. Ouimet JM, Pope JE, Gutmanis I, et al. Work disability in scleroderma is greater than in rheumatoid arthritis and is predicted by high HAQ scores. Open Rheumatol J 2008;2:44–52.
32. Peytrignet S, Denton CP, Lunt M, et al. Disability, fatigue, pain and their associates in early diffuse cutaneous systemic sclerosis: the European Scleroderma Observational Study. Rheumatology (Oxford) 2018;57:370–81.
33. Sandqvist G, Akesson A, Eklund M. Evaluation of paraffin bath treatment in patients with systemic sclerosis. Disabil Rehabil 2004;26: 981–7.
34. Bongi SM, del Rosso A, Galluccio F, et al. Efficacy of connective tissue massage and Mc Mennell joint manipulation in the rehabilitative treatment of the hands in systemic sclerosis. Clin Rheumatol 2009;28: 1167–73.
35. Murphy SL, Barber M, Huang S, et al. Intensive and app-delivered occupational therapy to improve upper extremity function in early diffuse cutaneous systemic sclerosis: a pilot two-arm trial. Rheumatology (Oxford) 2021;60:5002–11.
36. Stefanantoni K, Sciarra I, Iannace N, et al. Occupational therapy integrated with a self-administered stretching program on systemic sclerosis patients with hand involvement. Clin Exp Rheumatol 2016;34 Suppl 100:157–61.
37. Alexanderson H, Bergegård J, Björnådal L, et al. Intensive aerobic and muscle endurance exercise in patients with systemic sclerosis: a pilot study. BMC Res Notes 2014;7:86.
38. Mugii N, Hasegawa M, Matsushita T, et al. The efficacy of self-administered stretching for finger joint motion in Japanese patients with systemic sclerosis. J Rheumatol 2006;33:1586–92.
39. Mouthon L, Poole J. Physical and occupational therapy. In: Varga J, Denton C, Wigley F, et al, eds. Scleroderma. New York (NY): Springer; 2016: p. 603–13.
40. Weber DR, Haynes K, Leonard MB, et al. Type 1 diabetes is associated with an increased risk of fracture across the life span: a population-based cohort study using The Health Improvement Network (THIN). Diabetes Care 2015;38:1913–20.
41. Sellmeyer DE, Civitelli R, Hofbauer LC, et al. Skeletal metabolism, fracture risk, and fracture outcomes in type 1 and type 2 diabetes [review]. Diabetes 2016;65:1757–66.
42. Nickolas TL, Leonard MB, Shane E. Chronic kidney disease and bone fracture: a growing concern [review]. Kidney Int 2008;74:721–31.
43. Ensrud KE, Lui LY, Taylor BC, et al. Renal function and risk of hip and vertebral fractures in older women. Arch Intern Med 2007;167:133–9.
44. Jamal SA, Bauer DC, Ensrud KE, et al. Alendronate treatment in women with normal to severely impaired renal function: an analysis of the fracture intervention trial. J Bone Miner Res 2007;22:503–8.
45. Chuealee W, Foocharoen C, Mahakkanukrauh A, et al. Prevalence and predictive factors of osteoporosis in Thai systemic sclerosis. Sci Rep 2021;11:9424.
46. Poly TN, Islam MM, Yang HC, et al. Proton pump inhibitors and risk of hip fracture: a meta-analysis of observational studies. Osteoporos Int 2019;30:103–4.

Arthritis Care & Research
Vol. 75, No. 11, November 2023, pp 2389–2398
DOI 10.1002/acr.25133

American College of Rheumatology
Empowering Rheumatology Professionals

REVIEW

Lung Transplantation: A Viable Option for Connective Tissue Disease?

Na Zhang,[1] Shizhou Liu,[1] Zhaoliang Zhang,[2] Ying Liu,[1] Liangyu Mi,[2] and Ke Xu[1]

Interstitial lung disease (ILD) and pulmonary hypertension (PH) caused by connective tissue disease (CTD) are one of the main causes of morbidity and death in patients. Although the International Society for Heart & Lung Transplant suggested that ILD and PH related to CTD are rare indications for lung transplantation in 2006, many lung transplantation centers are concerned that the multisystem involvement of CTD will affect survival outcomes after lung transplantation, and CTD is regarded as a relative contraindication for lung transplantation. However, long-term and short-term survival after lung transplantation in CTD patients is similar compared with survival in common indications for lung transplantation such as idiopathic pulmonary fibrosis (IPF), and no higher incidence of complications after transplantation in many lung transplant centers. This suggests that lung transplantation may be beneficial in CTD patients with disease that progresses to end-stage lung disease, and CTD should not be considered a contraindication for lung transplantation. In the future, more prospective studies are needed to analyze the risk factors of lung transplantation in CTD patients to improve survival rates and reduce the risk of complications. This narrative review summarizes the selection and evaluation of candidates for CTD before lung transplantation and describes the clinical outcomes in CTD after lung transplantation in large-capacity lung transplantation center. The purpose of this review is to help rheumatologists decide when to refer patients with CTD-related lung involvement to a lung transplantation center and the conditions to consider before transplantation and to provide confidence to lung transplant experts.

Introduction

Connective tissue diseases (CTDs) are a group of autoimmune diseases involving multiple systems and organs. Interstitial lung disease (ILD) and pulmonary hypertension (PH) are one of the main causes of morbidity and death in patients with CTDs (1). Nonspecific interstitial pneumonia (NSIP) is the most common pattern of most CTDs, except rheumatoid arthritis (RA), which is mainly related to usual interstitial pneumonia (UIP) (2). Treatment of CTD–ILD is still a challenge. At present, there is no global treatment guideline for CTD–ILD. Patients with CTD–ILD with disease that is difficult to treat with traditional medical treatments including glucocorticoids and immunosuppressants or with disease that is still progressing may need lung transplantation evaluation. Recently, 3 trials have shown a response to antifibrotic therapy in patients with other forms of progressive fibrotic ILD, including chronic hypersensitivity pneumonitis, autoimmune ILD, idiopathic nonspecific interstitial pneumonitis, unclassifiable idiopathic interstitial pneumonitis, and a group of other rarer fibrotic ILDs (3–5).

With the widespread use of antifibrotic drugs, it is more challenging to determine the timing of listing for lung transplantation, but these studies cannot prove the impact on death. Therefore, lung transplantation is the final treatment in CTD that has progressed to end-stage lung disease. The International Society for Heart & Lung Transplant (ISHLT) consensus on lung transplantation in patients with CTD has not changed the recommendation for referral and listing of patients who received treatment with new antifibrotic drugs (6). More data are needed to clarify the effect of antifibrotic drugs on the best time of transplantation. Since the successful introduction of lung transplantation in the early 1980's, there has been a great improvement in terms of candidate selection, management, and outcome. ISHLT published

Supported by the National Natural Science Foundation of China (grant 81871292) and the Key Research and Development Projects of Shanxi Province (grant 201803D31136).

[1]Na Zhang, MS, Shizhou Liu, MS, Ying Liu, MS, Ke Xu, MD: Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Tongji Shanxi Hospital, Third Hospital of Shanxi Medical University, Taiyuan, China and Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. [2]Zhaoliang Zhang, MS, Liangyu Mi, MD: Third Hospital of Shanxi Medical University, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences Tongji Shanxi Hospital,Taiyuan, China.
Na Zhang, Shizhou Liu, and Zhaoliang Zhang contributed equally to this work.

Author disclosures are available online at https://onlinelibrary.wiley.com/doi/10.1002/acr.25133.

Address correspondence via email to Ke Xu, MD, at zhaoxuke@hotmail.com.

Submitted for publication February 10, 2023; accepted in revised form April 11, 2023.

several consensus statements/guidelines on lung transplantation in 1998, 2006, 2015 and 2021 (6–9). The indications for lung transplantation mainly include IPF, chronic obstructive pulmonary disease (COPD), cystic fibrosis (CF), and idiopathic pulmonary arterial hypertension (IPAH). Until 2006, the guidelines of the ISHLT proposed that ILD and PAH related to CTD were rare indications for lung transplantation (0.5%) (8). Over the last 30 years, although there has been a significant increase in the number of lung transplants performed, CTD patients only account for 0.9% (10). Indications for adult lung transplantation are shown in Figure 1 (10).

Due to the multisystem involvement of CTD, CTD is a relative contraindication in many lung transplantation centers, and referral is often delayed because of concerns about the influence of preexisting conditions on the posttransplantation results. The potential factors of adverse outcomes include gastroesophageal reflux (GER) (thought to cause bronchiolitis obliteration syndrome [BOS]), kidney diseases (complicating the management of commonly used immunosuppressive agents and antibacterial drugs after transplantation) and extrapulmonary diseases such as myositis (complicating posttransplantation immunosuppression and rehabilitation management) (11). Because of these concerns, whether lung transplantation is a reasonable treatment in patients with CTD remains controversial. Compared with patients with other diseases, these patients receive less lung transplantation treatment, and different lung transplantation centers have differences in the selection and evaluation of CTD candidates (12). This narrative review summarizes the selection and evaluation of candidates for CTD lung transplantation, and describes the clinical outcome of CTD lung transplantation in large-capacity lung transplantation center, aiming to provide references for CTD patients to choose lung transplantation.

Methods

The study did not require approval from the medical ethics committee of Shanxi Bethune Hospital.

When to consider lung transplant. Lung transplantation should be considered in adults with chronic, end-stage lung disease who meet all the following general criteria: 1) high (>50%) risk of death from lung disease within 2 years if lung transplantation is not performed; 2) high (>80%) likelihood of surviving at least 90 days after lung transplantation; 3) high (>80%) likelihood of 5-year posttransplant survival from a general medical perspective provided that there is adequate graft function (7).

Referral and listing for CTD patients. For CTD patients, there is no independent guideline for referral and listing of lung transplantation. Table 1 shows the referral and listing timing formulated by ISHLT for ILD and pulmonary vascular diseases, which is applicable to CTD patients according to the major lung involvement (i.e., ILD or PH) (7,8).

Pretransplant considerations. CTD mainly includes RA, Sjögren's syndrome, systemic lupus erythematosus (SLE), systemic sclerosis (SSc), inflammatory myopathies (polymyositis [PM], dermatomyositis [DM], and antisynthetase syndrome), and antineutrophil cytoplasmic antibody (ANCA)–associated vasculitis (AAV). Extrapulmonary manifestations and specific transplant evaluation that should be considered before transplantation in each disease are shown in Supplementary Tables 1–6 (available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25133/abstract); the absolute contraindications (agreement strength of at least 80%) are shown in Supplementary Table 7 (13) (available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25133/abstract).

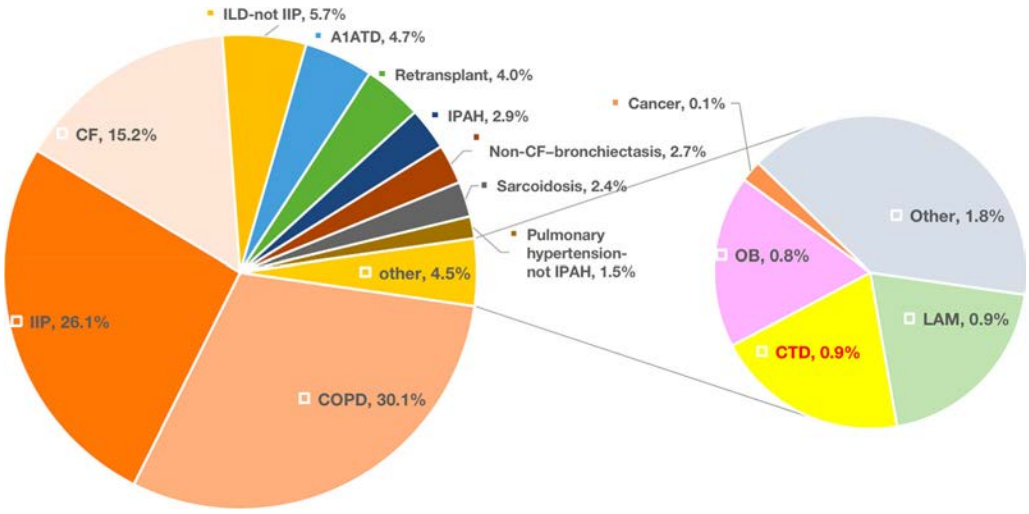

Figure 1. Primary indications in adult lung transplantation between January 1995 and June 2018. A1ATD = α_1 antitrypsin deficiency; CF = cystic fibrosis; COPD = chronic obstructive pulmonary disease; CTD = connective tissue disease; IIP = idiopathic interstitial pneumonia; ILD = interstitial lung disease; IPAH = idiopathic pulmonary hypertension; LAM = lymphangioleiomyomatosis; OB = obliterative bronchiolitis.

Table 1. Timing of referral and listing for CTD*

Timing	ILD	PH
Referral	Evidence of UIP or NSIP FVC <40% predicted Dyspnea or functional limitation Any O_2 requirement For inflammatory ILD, failure in improvement of dyspnea, O_2 requirement, or PFTs after medical therapy	NYHA class III–IV symptoms during escalating therapy Rapidly progressive disease Use of parenteral targeted PAH therapy Known or suspected PVOD or pulmonary capillary hemangiomatosis
Listing	≥10% decline in FVC, or ≥15% decline in DLco at 6-month follow-up O_2 saturation <88%, 6-minute walk test distance <250 meters 50-meter decline in 6-minute walk test distance at 6-month follow-up PH Hospitalization due to respiratory decline, pneumothorax, or acute exacerbation	NYHA class II–IV despite ≥3-month combination therapy including prostanoids CI <2 liters/minutes/m^2 mRAP >15 mm Hg 6-minute walk test distance <350 meters Significant hemoptysis, pericardial effusion, or progressive RHF

* CTD = connective tissue disease; CI = cardiac index; DLco = diffusing capacity for carbon monoxide; FVC = forced vital capacity; ILD = interstitial lung disease; mRAP = mean right atrial pressure; NSIP = nonspecific interstitial pneumonia; NYHA = New York Heart Association; PAH = pulmonary arterial hypertension; PFTs = pulmonary function tests; PH = pulmonary hypertension; PVOD = pulmonary venooclusive disease; RHF = right heart failure; UIP = usual interstitial pneumonia.

Outcome in CTD patients after lung transplantation

CTD–ILD and lung transplantation. Because the proportion of CTD patients receiving lung transplantation is smaller compared with that of other lung diseases (e.g., IPF, CF, COPD), there are few studies about the prognosis of lung transplantation in CTD patients. Recently, some studies have shown that the short-term and long-term survival rates of CTD–ILD patients posttransplant are comparable to those of IPF or non-CTD patients (14–20) (see Table 2). These studies confirmed the effectiveness of lung transplantation in patients with CTD–ILD. Takagishi et al found that the cumulative survival time in CTD–ILD patients was lower compared with the cumulative survival time in COPD patients, but there was no statistical difference in the increased risk of death (14). The author believes that this may be related to CTD-related comorbidity, which increases the risk of graft failure or death. Specific problems related to CTD that may increase the risk of graft dysfunction include esophageal dyskinesia, GER disease (GERD), neuromuscular weakness, and thromboembolism (12). Courtwright et al found that the mortality rate in non-scleroderma connective tissue–related ILD patients with walking distance <183 meters increased after transplantation (16), which is consistent with the previously reported relationship between 6-minute walking distance and survival rate after lung transplantation (21). Therefore, for patients below this threshold, the preoperative status should be optimized by improving the preoperative walking distance.

Prieto-Peña et al confirmed the existence of ILD using histologic features obtained from lung biopsy before lung transplantation (18). When evaluating the histologic characteristics of the recipient lung, all patients diagnosed with RA and receiving lung transplantation had UIP, while NSIP was the most common histologic subtype in the remaining CTD–ILD patients. In the study by Park et al (15), compared with IPF patients, there was no statistical difference in the duration of CTD–ILD patients in the intensive care unit (ICU) after surgery. However, Yang et al showed that PM/DM–ILD was related to longer stay in the ICU ($P < 0.001$) (20). In addition, a cohort study by Ratwani et al showed that the transplant-free survival rate of patients hospitalized for any reason was significantly lower compared with that of patients who had never been hospitalized (3-year survival rate 60% versus 94%; $P = 0.0001$) (22). The lower transplant-free survival rate was associated with age, male sex, RA, all-cause hospitalization, and cardiopulmonary hospitalization, which may be helpful in lung transplantation evaluation in case selection.

CTD–PH and lung transplantation. CTD-related PH was more common in SSc, mixed connective tissue disease, and SLE, with prevalence rates of 4.9–38%, 23–29%, and 2–14%, respectively (23). Compared with IPAH, CTD–PH has worse responsiveness to treatment and worse prognosis (24,25). Few studies have analyzed the prognosis of lung transplantation in PH patients. In a retrospective study at the Toronto General Hospital in Canada from January 1997 to September 2010 (26), including 123 IPAH patients, 77 with PAH associated with congenital heart disease, 102 with CTD–PH, and 14 with chronic thromboembolic disease, of the 16 patients with CTD–PH who received bilateral lung transplantation, 13 had scleroderma. The results of this study showed that CTD–PH had poor responsiveness to targeted drugs, made rapid progress, and had a worse prognosis compared with IPAH. The case fatality rate in the waiting list was 34%, which was much higher than CHD–PH and IPAH. However, it is very interesting that the survival of CTD–PH lung transplantation is better than that of IPAH group, especially the long-term survival rate of CTD–PAH patients at 5 and 10 years after lung transplantation (69% at 10 years). The results of this study encourage reexamination of lung transplantation in

Table 2. Outcomes in patients with CTDs after lung transplantation*

Author (time period)	Data sources	Study population	Years of survival, %			Complication, %			
			1	2	5	PGD	ACR	BOS	Infection
Takagishi et al, 2012 (1991–2009) (14)	OPTN and LUMC databases, US, multicenter	CTD–ILD (n = 284): SSc 174 (61.2%), RA 36 (12.7%), DM/PM 34 (12.0%), MCTD 22 (7.7%), SLE 11 (4%), SJS 7 (2.5%)	72.7†‡	66.3‡	46.1‡	NR	33‡§	NR	NR
		IPF (n = 4,190)	77.7	68.2	49.8	NR	39.7	NR	NR
		COPD (n = 6,720)	83.2	73.4	46.6	NR	46.2	NR	NR
Courtwright et al, 2017 (2005–2016) (16)	SRTR, US, multicenter	NS-CTLD (n = 275): MCTD 82 (29.8%), RA 68 (24.7%), PM 51 (18.5%), SJS 26 (9.5%); SLE 24 (8.7%), DM 7 (2.5%), other CTDs 5 (4.4%)	NS	NR	NR	NR	NS	NS	NR
		IPF (n = 6,346)	NR	NR	NR	NR	NR	NR	NR
Park et al, 2018 (2012–2016) (15)	Severance Hospital of Yonsei University, Korea, single center	CTD–ILD (n = 15); DM/PM 5 (33.3%), RA 4 (26.7%), SSc 3 (20%), SLE 1 (6.7%), SJS 1 (6.7%), UCTD 1 (6.7%)	80.0	NR	NR	26.7¶	NR	NR	NR
		IPF (n = 47)	59.6 (NS)	NR	NR	8.5 (NS)	NR	NR	NR
Prieto-Peña et al, 2020 (1998–2018) (18)	Northern Spain, single center	CTD–ILD (n = 26): RA 9 (34.6 %), SSc 6 (23.1%), SJS 4 (15.4%), ANCA-associated vasculitis 3 (11.5%), ASyS 2 (7.7 %), DM 1 (3.7%), SLE 1 (3.7%)	NR	NR	42.4	NR	32†	20	NR
		IPF (n = 26)	NR	NR	65.8 (NS)	NR	62.5	8.3 (NS)	NR
Csucska et al, 2021 (2012–2017) (19)	St. Joseph's Hospital and Medical Center, single center	CTD (n = 33): SSc 14 (43%), RA 13 (39%), DM/PM 3 (9%), SLE 2 (6%), MCTD 1 (3%)	84.8	NR	NR	NR	NR	NR	NR
		non-CTD (n = 461)	91.8 (NS)	NR	NR	NR	NR	NR	NR
Yang X et al, 2021 (2015–2019) (20)	Wuxi People's Hospital, China, single center	NM-CTLD (n=28): SS 8 (22.2%), RA 10 (27.8%), SSc 2 (5.6%), SLE 1 (2.8%), MCTD 6 (16.7%), UCTD 1 (2.8%)	NR	–	NR	–	NR	NR	66.7
		PM/DM–ILD (n = 8)	NR	19†	NR	75†	NR	NR	66.7
		IPF (n = 180)	NR	60	NR	21.7	NR	NR	65.3
Ju C et al, 2021 (2015–2020) (17)	First Affiliated Hospital of Guangzhou Medical University, China, single center	CTD–ILD (n = 31)	73.2	NR	71.4	90.3†	NR	NR	NR
		IPF (n = 98)	69.1 (NS)	NR	39.5 (NS)	70.4	NR	NR	NR

* ACR = acute cellular rejection; ASyS = myositis–antisynthetase syndrome; ANCA = antineutrophil cytoplasmic antibody; DM = dermatomyositis; LUMC = Leiden University Medical Center; MCTD = mixed connective tissue disease; NM-CTLD = non-myositis connective tissue-related ILD; NR = not reported; NS = not significant; OPTN = OPTN: Organ Procurement and Transplantation Network; PM = polymositis; RA = rheumatoid arthritis; SjS = sjögren's syndrome; SLE = systemic lupus erythematosus; SS = Sjögren's syndrome; SSc = systemic sclerosis; SRTR = Scientific Registry of Transplant Recipients; UCTD = undifferentiated connective tissue disease. See Table 1 for other definitions.
† $P < 0.05$ versus idiopathic pulmonary fibrosis (IPF).
‡ $P < 0.05$ versus chronic obstructive pulmonary disease (COPD).
§ Cumulative incidence of rejection.
¶ Grade 3 primary graft dysfunction (PGD).

recipients with CTD–PH. It indicates that CTD–PH patients have good early- and long-term prognosis after lung transplantation and should be treated more actively on the waiting list.

Complication. The survival rate in patients after lung transplantation may be affected by many complications, mainly graft dysfunction (primary graft dysfunction [PGD]), acute and chronic graft rejection, infection, and immunosuppressive drug–related side effects. Only a few studies have specifically reported the incidence of complications such as PGD, acute and chronic rejection, infection, etc. (see Table 2).

PGD. PGD is an acute lung injury that occurs early after lung transplantation, and it is the main cause of early death in lung transplant recipients (27). The current understanding of the pathogenesis of PGD emphasizes the multiple pathways leading to damage of pulmonary endothelium and alveolar epithelium, involving inflammation, innate immunity, platelet and coagulation dysfunction, fibrinolysis, and other pathways (26).The total incidence rate of PGD is estimated to be ∼30%, and the 30-day mortality rate of grade 3 PGD is ∼36%. This impact on survival has been confirmed 10 years after lung transplantation, especially for grade 3 PGD (28).

Compared with no or lower PGD, the highest PGD3 was associated with significantly longer mechanical ventilation time and hospital stay after transplantation (29–32). At present, there is no treatment to reverse PGD, which is the main cause of early death after lung transplantation (33). Additionally, it increases the risk of long-term death and chronic rejection among survivors (30,34).

The incidence of PGD in CTD patients after lung transplantation is shown in Table 2. Park et al reported that there was no significant difference in the incidence of PGD between CTD–ILD and IPF patients ($P = 0.154$) (15), and in 2021, findings from a multicenter retrospective study by Natalini et al demonstrated that the incidence of grade 3 PGD in patients with CTD–ILD is similar to that in patients with IPF (35), which was different from the study by Ju et al (17) that showed that PM/DM–ILD was associated with higher incidence of PGD (grade 3) ($P = 0.006$) and longer stay in the ICU ($P < 0.001$) (20). Cox proportional risk regression analysis after adjustment for age and sex showed that the occurrence of PGD after surgery and the length of stay in the ICU were independent risk factors for patient survival. Ju et al demonstrated that the incidence of PGD in CTD–ILD patients was significantly higher (90.3% versus 70.4%; $P = 0.03$), but there was no significant difference in mortality related to PGD between the 2 groups (6.5% versus 6.1%; $P = 0.95$) (17) (see Table 2).Therefore, there is no higher PGD-related mortality in CTD patients after lung transplantation, except in PM/DM–ILD patients. Careful consideration should be given to PM/DM–ILD patients before surgery to reduce the high mortality rate in these patients after lung transplantation.

Acute rejection and chronic rejection. Acute cellular rejection is a common complication after lung transplantation. Although clinically, acute cellular rejection is usually asymptomatic and rarely fatal, it has been recognized as a risk factor for chronic rejection (36). There are few treatment options for chronic rejection (37), which is the main cause of death in recipients who survived >1 year after transplantation, BOS is the most common form of chronic rejection in lung transplantation, ≤50% of the subjects have BOS within 5 years after transplantation (38). BOS is defined as a decrease in forced expiratory volume in 1 second (FEV1) ≥20% from the previous baseline, with evidence of airflow limitation measured by a ratio of FEV1/forced vital capacity of <0.7, and an absence of opacities on chest imaging (39,40). Compared with other types of solid organ transplantation, the incidence of acute and chronic rejection after lung transplantation increased (41).

These patients may also have a higher risk of acute and chronic rejection due to potential immune disorders and antibody-mediated allograft injury. Takagishi et al evaluated the incidence of graft rejection within 1 year after transplantation (14). Compared with COPD patients (46.2%), CTD patients had a lower cumulative rejection rate (33%; $P = 0.0004$), but there was no statistical difference compared with IPF patients (39.7%). Prieto-Peña et al found that the incidence of acute graft rejection in CTD–ILD patients was lower than that in IPF patients ($P = 0.032$), and the frequency of chronic graft rejection was not statistically significant ($P = 0.417$) (12). At the same time, Courtwright et al showed that there was no significant difference in acute and chronic rejection between patients with non-scleroderma connective tissue–related ILD and those with IPF (16). However, non-scleroderma connective tissue–related ILD is more likely to develop into BOS ≥2 ($P = 0.002$). Therefore, lung transplantation in CTD patients does not have a higher incidence of rejection.

Infection. Allograft infection after lung transplantation has a significant impact on the outcome. The increased susceptibility of allograft to infection is due to its direct contact with inhaled environmental microorganisms, immunosuppression, and impaired clearance mechanism after denervation of the transplanted lung. The possible microbial spectrum of allograft infection after lung transplantation is very broad, usually including *Pseudomonas aeruginosa*, cytomegalovirus, community-acquired respiratory virus, and *Aspergillus*. The prophylactic antibacterial treatment program after surgery can reduce the incidence of infection. However, the prevention strategies used by different transplant centers to reduce infection complications are still heterogeneous (42). Findings from one study suggested that lung infection was the main cause of death after lung transplantation. The incidence of IPF was 65.3% (47 of 72), non-myositis connective tissue–related ILD was 66.7% (8 of 12), and PM/DM–ILD was 66.7% (4 of 6) (20), but there was no statistical difference.

Malignancy. Immunosuppression can induce tolerance, increase the survival time of grafts, and prevent allograft rejection, but at the same time, it can reduce the natural antitumor immune response and increase the risk of malignant tumors (43). Cancer is still the third most common cause of death in lung transplant

recipients who have survived for >1 year. According to the data from the registry of the International Society of Heart and Lung Transplantation, the incidence of lung tumors in lung transplant patients at 1, 5 and 10 years is 5.1%, 18.2% and 28.7%, respectively. Its carcinogenic risk is considered to be directly related to the cumulative dose of immunosuppression (44).

The study by Ameye et al included 5 patients with idiopathic inflammatory myopathy (IIM)–associated ILD, 48 with IPF, and 37 with non-IPF– non-IIM–associated ILD, the incidence of malignant tumors after lung transplantation was 40%, 12.5% and 18.9%, respectively (45). Because fewer IIM patients were included in this study, the differences between groups were not analyzed.

Pulmonary and extrapulmonary recurrence of CTD after lung transplantation. Although CTD patients may benefit from lung transplantation, there are also concerns that recurrence of lung and extrapulmonary lesions may increase the posttransplant mortality rate in these patients. Takagishi et al did not find recurrence of lung lesions, but 4 cases showed acute or subacute arthritis attacks (3 had RA and 1 had SLE) after lung transplantation (14). Etanercept was stopped before transplantation and used again at onset, and arthritis subsided within 2 weeks. In the study by Park et al, during the follow-up period, a PM patient had weakness of the upper and lower limbs and carbon dioxide retention (15). The results of nerve conduction velocity showed sensorimotor polyneuropathy, and electromyography showed systemic myopathy; symptoms improved after 2 weeks of increasing glucocorticoid dosage. One SLE patient presented with dyspnea, fever, leukopenia, thrombocytopenia, and hypocomplementemia, and the symptoms improved after 2 weeks of steroid administration (1 mg/kg). This is not a serious complication. Courtwright et al suggested that there was no significant difference in extrapulmonary organ dysfunction between patients with non-scleroderma connective tissue–related ILD and those with IPF (16). Yang et al found no extrapulmonary recurrence (20). Generally speaking, no recurrence of lung lesions was found in these studies, and extrapulmonary condition improved after drug treatment.

The CTD–ILD subgroup and lung transplantation

SSc and lung transplantation. SSc accounts for the majority of patients with CTD–ILD, so there are many studies that regard SSc patients as a subgroup of CTD–ILD. The study showed that there was no significant difference in the cumulative survival time between SSc and IPF patients who received treatment with lung transplantation (46–55).The systematic review by Khan et al in 2012 summarized 7 observational studies from 1986 to 2012 and reported regarding 186 SSc patients (56). After transplantation, the survival rate in SSc was 69–91% at 30 days, 69–85% at 6 months, 5–93% at 1 year, and 49–80% at 2 years. Until now, the largest single-center study that had been published (57), included 72 SSc patients and 311 patients with pulmonary fibrosis caused by various causes except scleroderma. There is no significant difference in the survival rate of the 2 groups after transplantation in the first year (81% of scleroderma patients versus 79% of patients with pulmonary fibrosis; $P = 0.743$), and there was no significant difference in the 5-year survival rate (66% versus 58%; $P = 0.249$); a body mass index (BMI) of ≥35 kg/m^2 is a significant predictor of 1-year survival rate, which suggests that it is necessary for patients with a BMI of ≥35 kg/m^2 to lose weight before transplantation. Additionally, acute rejection and chronic rejection were similar between the 2 groups.

Some experts worry that esophageal dysfunction and gastroparesis may increase the risk of aspiration, resulting in a higher rejection rate and lower survival rate (58,59). Surprisingly, although moderate-to-severe esophageal dyskinesia is almost common in scleroderma patients, the detection rate of BOS in scleroderma patients did not increase in Crespo et al's study (57). Similarly, Sottile et al evaluated the posttransplant outcomes of SSc–ILD patients with esophageal involvement and those with non-CTD–ILD, survival was comparable in patients with SSc–ILD (n = 23) and those with non-CTD–ILD (n = 46) who underwent lung transplantation (55). In the SSc–ILD group, 1- and 5-year survival was compared with individuals in the non-CTD–ILD group (83% and 76%, 91% and 64%, respectively), there were no differences in terms of rates of BOS and acute cellular rejection; esophageal dysfunction was not associated with worse outcomes ($P > 0.55$). Therefore, severe GERD is not the contraindication of lung transplantation in SSc (55).

However, the above studies did not report the relationship between esophageal dyskinesia and survival of SSc patients. Csucska et al reported that in the CTD cohort, the 1-year and 3-year survival rates in the preserved esophageal motility group (100% and 87.5%, respectively) and ineffective esophageal motility group (100% and 85.7%, respectively) were significantly higher compared with those in the absent esophageal motility group (AEM) (50% and 20%, respectively; $P < 0.001$) (19). The 1-year and 3-year survival rates in the AEM group (50% and 20%, respectively) were significantly lower compared with those in the non-CTD patient cohort matched with the lung allocation score (92.5% and 65%, respectively; $P = 0.001$ and $P = 0.012$). Therefore, it may be important to improve esophageal motility before lung transplantation. Fisichella et al showed that esophageal pH monitoring can predict survival status the severity of reflux in patients with scleroderma awaiting lung transplantation; therefore, early consideration of esophageal pH monitoring may help to identify those in whom laparoscopic anti-reflux surgery should be performed faster to prevent GERD and harmful effects in patients waiting for lung transplantation (60).

On the contrary, Bernstein et al conducted a retrospective cohort study of 229 adult SSc patients, 201 PAH patients, and 3,333 ILD patients who received lung transplantation in the US (61). The data were provided by the American Organ Sharing Network. Compared with non–SSc-related ILD patients, the 1-year

mortality rate in adults with SSc who received lung transplantation increased by 48% after multivariate adjustment; however, the postoperative (i.e., 30 days) or intermediate (i.e., 3 years) mortality rate did not increase. There was no significant difference in 1-year mortality risk between SSc patients and patients with non–SSc-related PAH. The largest single-center retrospective cohort study in Europe included 15 patients with SSc (8 with ILD, 4 with ILD and PAH, and 3 with PAH), 198 patients with non–SSc-related ILD, and 18 patients with non–SSc-related PH (54). The 1-year and 3-year cumulative survival rates of SSc group were 80% and 65% respectively, and those in the non-SSc group were 78% and 63% respectively, regardless of the degree of esophageal involvement. Compared with non–SSc-related ILD or PH patients who underwent lung transplantation, SSc patients who underwent lung transplantation had no difference in 3-year mortality rate and cumulative survival rate, and the common complications of acute cell rejection and infection was not different between the 2 groups. PH had no recurrence after lung transplantation. Moreover, the main cause of death in all groups was sepsis. Therefore, compared with non-SSc–ILD and non-SSC–PH patients, lung transplantation in scleroderma patients has the same short-term and medium-term survival rates, and there was no higher incidence of complications.

RA and lung transplantation. Yazdani et al compared the posttransplant survival rates in 10 patients with RA–ILD, 53 with IPF, and 17 with SSc–ILD. The 1-year cumulative survival rates in the RA–ILD, IPF, and SSc–ILD groups were 67%, 69%, and 82% respectively, and there was no significant difference in the 1-, 2-, and 5-year survival rates after transplantation. Health-related quality of life scores in RA–ILD and IPF patients were compared before and after transplantation (62). It was found that the quality of life in both groups significantly improved after lung transplantation. Estimated using the Short Form 36 health survey scores and St. George's Respiratory Questionnaire (SGRQ) data, compared between the first time after transplantation and the last time before transplantation, SGRQ scores in 7 patients with RA–ILD improved significantly, with the total SGRQ score increasing from a mean ± SD of 70.4 ± 16.1 to 36.0 ± 18.5.

Inflammatory myopathies and lung transplantation. It is well known that DM is often associated with malignant tumors. For these reasons, lung transplantation is rarely reported. Ameye et al analyzed 90 ILD patients who received transplantation at the University Hospitals of Leuven from January 2004 to August 2013 (5 with IIM-related ILD, 48 with IPF-related ILD, 37 with non-IIM– non-IPF–related ILD) (45). The 1-year, 2-year and 5-year survival rates of patients with IIM-related ILD receiving lung transplantation treatment were 100%, 75%, and 75% respectively (mean ± SD follow-up period of 32.6 ± 4.4 months), the 1-year, 2-year and 5-year survival rates in IPF patients were 86%, 67%, and 58% respectively (mean ± SD follow-up period of 35.2 ± 3.9 months), and the 1-year, 2-year and 5-year survival rates in patients with non-IPF– and non-IIM–related ILD lung transplantation were 86%, 63%, and 57% respectively (mean ± SD follow-up period of 40.6 ± 20.5 months); the 1-year, 2-year, and 5-year cumulative survival rates in IIM patients are comparable to those in patients with non-IIM– non-IPF–related ILD or those with IPF receiving lung transplantation treatment.

In terms of complications, 1 of 5 DM/PM patients had acute rejection, and no chronic rejection was found. Two patients had grade 1 PGD, 1 patient had grade 2 PGD, and the other patient had grade 3 PGD. These results suggest that lung transplantation may be an effective choice in carefully selected patients with IIM-associated ILD. In terms of malignant tumors, 2 patients developed skin basal cell carcinoma after successful lung transplantation, and 1 patient developed bladder cancer 6 months after transplantation and received cystoprostatectomy. The patient developed esophageal cancer and right lung cancer again 16 months later and died 18 months later.

In 2022, Rivière et al suggested the 1-year, 3-year, and 5-year survival rates of IIM patients posttransplantation were comparable to other lung transplantation indications (63). Compared with patients with amyotrophic IIM, the survival rate of IIM with muscle involvement at 1, 3, and 5 years after lung transplantation is significantly worse. Dialysis during ICU hospitalization, grade 3 PGD at 72 hours, IIM with muscle involvement, skin involvement, and pre–lung transplantation immunosuppression lines were associated with worse survival. The main predictor was a history of muscle involvement. Five patients (8%) experienced a recurrence of IIM, 3 of whom had slight peripheral muscle involvement, and 2 of whom had specific skin damage. No patients experienced ILD recurrence. The most common causes of death in patients with IIM with muscle involvement are pneumonia and septic shock.

Shoji et al reported the first living-donor lobar lung transplantation in a patient with rapidly progressive diffuse interstitial pneumonia associated with clinically amyopathic DM (CADM) followed up for 7 years, no malignant disease was found (64). Besides, the anti–melanoma differentiation–associated gene 5 (anti–MDA-5) antibody positivity present in 10–35% of patients is frequently associated with CADM and a high risk of rapidly progressive ILD (RPILD) (64). Retrospective studies report a mortality rate of 75–84% despite maximal treatment (65,66). Recently, 2 case reports demonstrated the successful experience of lung transplantation in therapy-resistant RPILD associated with anti–MDA-5 antibody positive DM. Leclair et al reported that the anti–MDA-5 antibody disappeared after lung transplantation (67). After 12 years of follow-up, the patient was still in remission and had no signs of rejection. Some pathogenesis hypotheses in PM/DM indicate that after exposure to environmental factors such as viral infection or smoking, the lung will have initial immune response, leading to interruption of self-tolerance and autoimmune response (68). Under this assumption, lung transplantation will remove the source of autoantigens, which explains that anti–MDA-5 antibodies disappear in reported patients, and ILD or DM symptoms do not recur (67). In the case report by Marchiset

et al, no disease recurred after the 1-year follow-up (69), which indicates that lung transplantation can be used as a lifesaving means for RPILD related to anti–MDA-5 antibody positive DM.

SLE and lung transplantation. There is limited experience of lung transplantation in SLE-related lung diseases. A single-center retrospective cohort study conducted by Bush et al included 6 patients with SLE-related lung disease who underwent lung transplantation from 1994 to 2014, 4 patients with SLE–ILD, and 2 patients with SLE–PH (70). There were no active extrapulmonary manifestations of SLE before transplantation, 1 patient died posttransplant, 4 patients developed acute rejection, 1 patient developed BOS, and no SLE complications occurred after transplantation, with a 3-year survival rate of 83% and median follow-up time of 4 years. Therefore, in carefully selected patients, lung transplantation treatment for SLE-related lung disease can also achieve successful outcomes.

AAV and lung transplantation. There are few reports of lung transplantation in severe AAV. Weinkauf et al reported a case of rapidly progressive AAV with lung and kidney involvement undergoing emergency lung transplantation (71). The patient received an intravenous drip of methylprednisolone (1,000 mg/day for a total of 3 days), combined with cyclophosphamide treatment. Pulmonary hemoptysis worsened and alveolar hemorrhage developed. Subsequently, he underwent tracheal intubation, mechanical ventilation, and was admitted to the ICU and received continuous hemofiltration and 5 consecutive plasmapheresis; the patient received 1 rituximab (RTX) treatment (375 mg/m2), but pulmonary hemorrhage and respiratory failure continued to worsen. The patient began receiving extracorporeal membrane oxygenation (ECMO) support and was listed as an emergency lung transplant patient. Finally, the patient received bilateral lung transplantation. The ECMO support was removed during surgery. After lung transplantation, the patient received 3 RTX treatments once a week. Renal function returned to normal after hemodialysis, and cytoplasmic ANCA gradually became negative; he remained well within 1,450 days posttransplantation, with no recurrence of AAV. Therefore, lung transplantation may be an option in severe acute AAV that results in respiratory failure after conservative treatment.

Conclusion

Lung transplantation can be used as a feasible treatment in carefully selected patients with CTD. CTD should not be regarded as a contraindication of lung transplantation. CTD patients who meet the criteria for lung transplantation referral and listing should be referred to a lung transplantation center as early as possible, and the risk factors related to survival should be improved and optimized before surgery. At present, almost all studies are retrospective studies, and fewer patients were included. With the increase of lung transplantation in the world, lung transplantation in CTD patients is expected to increase in the future. Because of the heterogeneity of CTD, it is necessary to conduct more data collection, comparison, and analysis in future prospective and multicenter studies in each disease in CTD subgroups in order to determine the prognosis and risk factors related to survival rate in subgroups, which will help to select more suitable CTD candidates and improve outcomes after lung transplantation.

AUTHOR CONTRIBUTIONS

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

Study conception and design. N. Zhang, Xu.
Acquisition of data. Y. Liu, Mi.
Analysis and interpretation of data. S. Liu, Z. Zhang.

REFERENCES

1. Jablonski R, Dematte J, Bhorade S. Lung transplantation in scleroderma: recent advances and lessons. Curr Opin Rheumatol 2018; 30:562–9.
2. Kim EJ, Collard HR, King TE Jr. Rheumatoid arthritis-associated interstitial lung disease: the relevance of histopathologic and radiographic pattern. Chest 2009;136:1397–405.
3. Behr J, Prasse A, Kreuter M, et al. Pirfenidone in patients with progressive fibrotic interstitial lung diseases other than idiopathic pulmonary fibrosis (RELIEF): a double-blind, randomised, placebo-controlled, phase 2b trial. Lancet Respir Med 2021;9:476–86.
4. Assassi S, Distler O, Allanore Y, et al. Effect of nintedanib on progression of systemic sclerosis-associated interstitial lung disease over 100 weeks: data from a randomized controlled trial. ACR Open Rheumatol 2022;4:837–44.
5. Flaherty KR, Wells AU, Cottin V, et al. Nintedanib in progressive fibrosing interstitial lung diseases. N Engl J Med 2019;381:1718–27.
6. Leard LE, Holm AM, Valapour M, et al. Consensus document for the selection of lung transplant candidates: an update from the International Society for Heart and Lung Transplantation. J Heart Lung Transplant 2021;40:1349–79.
7. Weill D, Benden C, Corris PA, et al. A consensus document for the selection of lung transplant candidates: 2014–an update from the Pulmonary Transplantation Council of the International Society for Heart and Lung Transplantation. J Heart Lung Transplant 2015;34:1–15.
8. Orens JB, Estenne M, Arcasoy S, et al. International guidelines for the selection of lung transplant candidates: 2006 update–a consensus report from the Pulmonary Scientific Council of the International Society for Heart and Lung Transplantation. J Heart Lung Transplant 2006;25:745–55.
9. International guidelines for the selection of lung transplant candidates. The American Society for Transplant Physicians (ASTP)/American Thoracic Society(ATS)/European Respiratory Society(ERS)/International Society for Heart and Lung Transplantation (ISHLT). Am J Respir Crit Care Med 1998;158:335–9.
10. Chambers DC, Cherikh WS, Harhay MO, et al. The International Thoracic Organ Transplant Registry of the International Society for Heart and Lung Transplantation: thirty-sixth adult lung and heart-lung transplantation Report-2019; focus theme: donor and recipient size match. J Heart Lung Transplant 2019;38:1042–55.
11. Mathai SC, Danoff SK. Management of interstitial lung disease associated with connective tissue disease [review]. BMJ 2016;352:h6819.

12. Lee JC, Ahya VN. Lung transplantation in autoimmune diseases [review]. Clin Chest Med 2010;31:589–603.
13. Crespo MM, Lease ED, Sole A, et al. ISHLT consensus document on lung transplantation in patients with connective tissue disease: part I: epidemiology, assessment of extrapulmonary conditions, candidate evaluation, selection criteria, and pathology statements. J Heart Lung Transplant 2021;40:1251–66.
14. Takagishi T, Ostrowski R, Alex C, et al. Survival and extrapulmonary course of connective tissue disease after lung transplantation. J Clin Rheumatol 2012;18:283–9.
15. Park JE, Kim SY, Song JH, et al. Comparison of short-term outcomes for connective tissue disease-related interstitial lung disease and idiopathic pulmonary fibrosis after lung transplantation. J Thorac Dis 2018;10:1538–47.
16. Courtwright AM, El-Chemaly S, Dellaripa PF, et al. Survival and outcomes after lung transplantation for non-scleroderma connective tissue-related interstitial lung disease. J Heart Lung Transplant 2017; 36:763–9.
17. Ju C, Lian Q, Chen A, et al. Outcomes after lung transplantation among Chinese patients with connective tissue disease-associated interstitial lung disease and pulmonary hypertension: a retrospective cohort study. Clin Exp Rheumatol 2022;40:1666–73.
18. Prieto-Peña D, Martínez-Meñaca A, Calderón-Goercke M, et al. Long-term survival of lung transplantation for interstitial lung disease associated with connective tissue diseases: a study of 26 cases from a referral centre. Clin Exp Rheumatol 2020;38:615–20.
19. Csucska M, Razia D, Masuda T, et al. Bilateral lung transplant for a connective tissue disorder: esophageal motility and 3-year survival. Semin Thorac Cardiovasc Surg 2022;34:1065–73.
20. Yang X, Wei D, Liu M, et al. Survival and outcomes after lung transplantation for connective tissue disease-associated interstitial lung diseases. Clin Rheumatol 2021;40:3789–95.
21. Martinu T, Babyak MA, O'Connell CF, et al. Baseline 6-min walk distance predicts survival in lung transplant candidates. Am J Transplant 2008;8:1498–505.
22. Ratwani AP, Ahmad KI, Barnett SD, et al. Connective tissue disease-associated interstitial lung disease and outcomes after hospitalization: a cohort study. Respir Med 2019;154:1–5.
23. Ahmed S, Palevsky HI. Pulmonary arterial hypertension related to connective tissue disease: a review [review]. Rheum Dis Clin North Am 2014;40:103–24.
24. Zhang R, Dai LZ, Xie WP, et al. Survival of Chinese patients with pulmonary arterial hypertension in the modern treatment era. Chest 2011;140:301–9.
25. Mathai SC, Hassoun PM. Pulmonary arterial hypertension associated with systemic sclerosis [review]. Expert Rev Respir Med 2011;5: 267–79.
26. De Perrot M, Granton JT, McRae K, et al. Outcome of patients with pulmonary arterial hypertension referred for lung transplantation: a 14-year single-center experience. J Thorac Cardiovasc Surg 2012; 143:910–8.
27. Suzuki Y, Cantu E, Christie JD. Primary graft dysfunction [review]. Semin Respir Crit Care Med 2013;34:305–19.
28. Wilkey BJ, Abrams BA. Mitigation of primary graft dysfunction in lung transplantation: current understanding and hopes for the future [review]. Semin Cardiothorac Vasc Anesth 2020;24:54–66.
29. Christie JD, Bellamy S, Ware LB, et al. Construct validity of the definition of primary graft dysfunction after lung transplantation. J Heart Lung Transplant 2010;29:1231–9.
30. Christie JD, Sager JS, Kimmel SE, et al. Impact of primary graft failure on outcomes following lung transplantation. Chest 2005;127:161–5.
31. Diamond JM, Lee JC, Kawut SM, et al. Clinical risk factors for primary graft dysfunction after lung transplantation. Am J Respir Crit Care Med 2013;187:527–34.
32. Porteous MK, Lee JC. Primary graft dysfunction after lung transplantation [review]. Clin Chest Med 2017;38:641–54.
33. Shah RJ, Diamond JM. Primary graft dysfunction (PGD) following lung transplantation [review]. Semin Respir Crit Care Med 2018;39: 148–54.
34. Daud SA, Yusen RD, Meyers BF, et al. Impact of immediate primary lung allograft dysfunction on bronchiolitis obliterans syndrome. Am J Respir Crit Care Med 2007;175:507–13.
35. Natalini JG, Diamond JM, Porteous MK, et al. Risk of primary graft dysfunction following lung transplantation in selected adults with connective tissue disease-associated interstitial lung disease. J Heart Lung Transplant 2021;40:351–8.
36. Hachem RR. Acute rejection and antibody-mediated rejection in lung transplantation [review]. Clin Chest Med 2017;38:667–75.
37. Belperio JA, Weigt SS, Fishbein MC, et al. Chronic lung allograft rejection: mechanisms and therapy [review]. Proc Am Thorac Soc 2009;6: 108–21.
38. Arjuna A, Olson MT, Walia R, et al. An update on current treatment strategies for managing bronchiolitis obliterans syndrome after lung transplantation [review]. Expert Rev Respir Med 2021;15:339–50.
39. Estenne M, Maurer JR, Boehler A, et al. Bronchiolitis obliterans syndrome 2001: an update of the diagnostic criteria [review]. J Heart Lung Transplant 2002;21:297–310.
40. Meyer KC, Raghu G, Verleden GM, et al. An international ISHLT/ATS/ERS clinical practice guideline: diagnosis and management of bronchiolitis obliterans syndrome. Eur Respir J 2014;44:1479–503.
41. Knoop C, Estenne M. Acute and chronic rejection after lung transplantation [review]. Semin Respir Crit Care Med 2006;27:521–33.
42. Remund KF, Best M, Egan JJ. Infections relevant to lung transplantation [review]. Proc Am Thorac Soc 2009;6:94–100.
43. Ahmad M, Rees RC, Ali SA. Escape from immunotherapy: possible mechanisms that influence tumor regression/progression [review]. Cancer Immunol Immunother 2004;53:844–54.
44. Wang X, Dong M. Malignancy after lung transplantation: how to manage immunosuppression? [review] Transplant Proc 2020;52:315–20.
45. Ameye H, Ruttens D, Benveniste O, et al. Is lung transplantation a valuable therapeutic option for patients with pulmonary polymyositis? Experiences from the Leuven transplant cohort. Transplant Proc 2014;46:3147–53.
46. Massad MG, Powell CR, Kpodonu J, et al. Outcomes of lung transplantation in patients with scleroderma. World J Surg 2005;29: 1510–5.
47. Kubo M, Vensak J, Dauber J, et al. Lung transplantation in patients with scleroderma. J Heart Lung Transplant 2001;20:174–5.
48. Levine SM, Anzueto A, Peters JI, et al. Single lung transplantation in patients with systemic disease. Chest 1994;105:837–41.
49. Pigula FA, Griffith BP, Zenati MA, et al. Lung transplantation for respiratory failure resulting from systemic disease. Ann Thorac Surg 1997; 64:1630–4.
50. Rosas V, Conte JV, Yang SC, et al. Lung transplantation and systemic sclerosis [review]. Ann Transplant 2000;5:38–43.
51. Saggar R, Khanna D, Furst DE, et al. Systemic sclerosis and bilateral lung transplantation: a single centre experience. Eur Respir J 2010; 36:893–900.
52. Schachna L, Medsger TA Jr, Dauber JH, et al. Lung transplantation in scleroderma compared with idiopathic pulmonary fibrosis and idiopathic pulmonary arterial hypertension. Arthritis Rheum 2006;54: 3954–61.

53. Shitrit D, Amital A, Peled N, et al. Lung transplantation in patients with scleroderma: case series, review of the literature, and criteria for transplantation [review]. Clin Transplant 2009;23:178–83.
54. Fernández-Codina A, Berastegui C, Pinal-Fernández I, et al. Lung transplantation in systemic sclerosis: a single center cohort study. Joint Bone Spine 2018;85:79–84.
55. Sottile PD, Iturbe D, Katsumoto TR, et al. Outcomes in systemic sclerosis-related lung disease after lung transplantation. Transplantation 2013;95:975–80.
56. Khan IY, Singer LG, de Perrot M, et al. Survival after lung transplantation in systemic sclerosis. A systematic review [review]. Respir Med 2013; 107:2081–7.
57. Crespo MM, Bermudez CA, Dew MA, et al. Lung transplant in patients with scleroderma compared with pulmonary fibrosis. Short- and long-term outcomes. Ann Am Thorac Soc 2016;13:784–92.
58. Gasper WJ, Sweet MP, Golden JA, et al. Lung transplantation in patients with connective tissue disorders and esophageal dysmotility. Dis Esophagus 2008;21:650–5.
59. D'Ovidio F, Singer LG, Hadjiliadis D, et al. Prevalence of gastroesophageal reflux in end-stage lung disease candidates for lung transplant. Ann Thorac Surg 2005;80:1254–60.
60. Fisichella PM, Reder NP, Gagermeier J, et al. Usefulness of pH monitoring in predicting the survival status of patients with scleroderma awaiting lung transplantation. J Surg Res 2014;189:232–7.
61. Bernstein EJ, Peterson ER, Sell JL, et al. Survival of adults with systemic sclerosis following lung transplantation: a nationwide cohort study. Arthritis Rheumatol 2015;67:1314–22.
62. Yazdani A, Singer LG, Strand V, et al. Survival and quality of life in rheumatoid arthritis-associated interstitial lung disease after lung transplantation. J Heart Lung Transplant 2014;33:514–20.
63. Rivière A, Picard C, Berastegui C, et al. Lung transplantation for interstitial lung disease in idiopathic inflammatory myositis: a cohort study. Am J Transplant 2022;22:2990–3001.
64. Shoji T, Bando T, Fujinaga T, et al. Living-donor lobar lung transplantation for rapidly progressive interstitial pneumonia associated with clinically amyopathic dermatomyositis: report of a case. Gen Thorac Cardiovasc Surg 2013;61:32–4.
65. Vuillard C, Pineton de Chambrun M, de Prost N, et al. Clinical features and outcome of patients with acute respiratory failure revealing antisynthetase or anti-MDA-5 dermato-pulmonary syndrome: a French multicenter retrospective study. Ann Intensive Care 2018;8:87.
66. Chino H, Sekine A, Baba T, et al. Interstitial lung disease with antimelanoma differentiation-associated protein 5 antibody: rapidly progressive perilobular opacity. Intern Med 2019;58:2605–13.
67. Leclair V, Labirua-Iturburu A, Lundberg IE. Successful lung transplantation in a case of rapidly progressive interstitial lung disease associated with antimelanoma differentiation-associated gene 5 antibodies. J Rheumatol 2018;45:581–3.
68. Danoff SK, Casciola-Rosen L. The lung as a possible target for the immune reaction in myositis [review]. Arthritis Res Ther 2011;13:230.
69. Marchiset A, Neuville M, Voiriot G, et al. High-emergency lung transplantation for interstitial lung disease associated with anti-MDA5 dermatomyositis: a case report. Transplant Proc 2021;53: 2613–5.
70. Bush EL, Faust H, Lee J, et al. Is systemic lupus erythematosus (SLE) related lung disease a contraindication to lung transplantation? J Heart Lung Transplant 2015;34:S256.
71. Weinkauf J, Puttagunta L, Stewart K, et al. Lung transplantation for severe antineutrophilic cytoplasmic antibody-associated vasculitis. Transplant Proc 2010;42:2707–10.

Arthritis Care & Research
Vol. 75, No. 11, November 2023, pp 2399–2402
DOI 10.1002/acr.25190

American College of Rheumatology
Empowering Rheumatology Professionals

Erratum

In the article by Harkey et al in the March 2022 issue of *Arthritis Care & Research* (Prevalence of Early Knee Osteoarthritis Illness Among Various Patient-Reported Classification Criteria After Anterior Cruciate Ligament Reconstruction [pages 377–85]) all of the below corrected values are applicable to each uncorrected instance in the text. The authors state that this slight change in the prevalence of early OA symptoms does not change the overall discussion or conclusions.

Several values in the Results section of the Abstract were incorrectly shown as follows: A greater prevalence of participants with ACLR met the Luyten original criteria (n = 165 [54%]) compared to those who met the Englund original criteria (n = 128 [42%]; $\chi^2 = 19.3$, $P < 0.001$). When using the KOOS subscale PASS as thresholds, a significantly greater prevalence of participants with ACLR met the Luyten PASS criteria (n = 133 [43%]) compared to those who met the Englund PASS criteria (n = 85 [28%]; $\chi^2 = 48.0$, $P < 0.001$). When combining the Luyten and Englund KOOS criteria and using the original/PASS subscale thresholds, respectively, 40%/57% of participants met neither, 24%/15% met only 1, and 36%/28% met both KOOS criteria.

The correct Results section should read as follows: A greater prevalence of participants with ACLR met the Luyten original criteria (n = 107, 36%) compared to those who met the Englund original criteria (n = 72, 24%; $\chi^2 = 24.0$, $P < 0.001$). When using the KOOS subscale PASS as thresholds, a significantly greater prevalence of participants with ACLR met the Luyten PASS criteria (n = 126, 42%) compared to those who met the Englund PASS criteria (n = 73, 24%; $\chi^2 = 53.0$, $P < 0.001$). When combining the Luyten and Englund KOOS criteria and using the original/PASS subscale thresholds, respectively, 62%/58% of participants met neither, 17%/18% met only one, and 21%/24% met both KOOS criteria.

The following sentence in the Conclusion of the Abstract was incorrect: Regardless of the classification criteria used to define early OA illness, it is concerning that 28–54% of patients report considerable symptoms ∼6 months post-ACLR.

The correct sentence should read as follows: Regardless of the classification criteria used to define early OA illness, it is concerning that 24–42% of patients report considerable symptoms at ∼6 months post-ACLR.

The following sentences in the Significance & Innovations sections were incorrect:

At 5–7 months post–anterior cruciate ligament reconstruction (ACLR), 54% and 42% of people report a level of self-reported disability that meets the Luyten original and Englund original Knee Injury and Osteoarthritis Outcomes Score (KOOS) criteria for early knee osteoarthritis (OA) illness, respectively. After refining the Luyten and Englund KOOS criteria by using post-ACLR specific patient acceptable symptom state (PASS) as KOOS subscale thresholds,43% and 28% of people, respectively, present with self-reported early OA illness. By combining the Luyten original and Englund original KOOS criteria into a single composite early knee OA illness variable and using the original/PASS subscale thresholds, respectively, we identified that 36%/28% of participants meet both, 24%/15% meet one, and 40%/57% meet neither of the original/PASS KOOS criteria.

The correct sentences should read as follows: At 5–7 months post–anterior cruciate ligament reconstruction (ACLR), 36% and 24% of people report a level of self-reported disability that meets the Luyten original and Englund original Knee Injury and Osteoarthritis Outcomes Score (KOOS) criteria for early knee osteoarthritis (OA) illness, respectively. After refining the Luyten and Englund KOOS criteria by using post-ACLR specific patient acceptable symptom state (PASS) as KOOS subscale thresholds,42% and 24% of people, respectively, present with self-reported early OA illness. By combining the Luyten original and Englund original KOOS criteria into a single composite early knee OA illness variable and using the original/PASS subscale thresholds,

respectively, we identified that 21/24% of meet both, 17/18% meet one, and 62/58% meet neither of the Original/PASS KOOS criteria.

The following sentence in the Methods section was incorrect: This was a secondary analysis of patient-reported outcomes collected for ongoing research at Michigan State University (n = 123), the University of Virginia (n = 55), and Creighton University (n = 128).

The correct sentence should read as follows: This was a secondary analysis of patient-reported outcomes collected for ongoing research at Michigan State University (n = 117), the University of Virginia (n = 55), and Creighton University (n = 128).

The following sentence in the Methods section was incorrect: Because time since ACLR may influence patient-reported outcomes, we performed a post hoc stratified analysis that repeated the analyses in participants who were at the 5-month time point post-ACLR (n = 109), as well as in participants who were at the 6- or 7-month time point post-ACLR (n = 197).

The correct sentence should read as follows: Because time since ACLR may influence patient-reported outcomes, we performed a post hoc stratified analysis that repeated the analyses in participants who were at the 5-month time point post-ACLR (n = 106), as well as in participants who were at the 6- or 7-month time point post-ACLR (n = 194).

The following sentences in the Discussion section were incorrect: Interestingly, the percentage of participants from this study who fit into the probable early knee OA illness category using the original KOOS criteria (36%) is similar to the prevalence of magnetic resonance imaging evidence of early knee OA at 1 year post-ACLR (31%) (27), radiographic knee OA at 10 years post-ACLR (36%) (5), and unacceptable symptoms at 1, 2, and 6 years post-ACLR (33–43%) (7,22). An important next step will be to determine if the presence of probable early knee OA illness is associated with a greater risk of developing incident OA–related structural pathology.

The correct sentences should read: Interestingly, the percentage of participants from this study who fit into the early knee OA illness category using the Luyten original KOOS criteria (36%) is similar to the prevalence of magnetic resonance imaging evidence of early knee OA at 1 year post-ACLR (31%) (27), radiographic knee OA at 10 years post-ACLR (36%) (5), and unacceptable symptoms at 1, 2, and 6 years post-ACLR (33–43%) (7,22). An important next step will be to determine if the presence of early knee OA illness is associated with a greater risk of developing incident OA–related structural pathology.

The following sentences in the Discussion section were incorrect: A lower prevalence of definite and probable early knee OA illness when using the PASS KOOS criteria (28% and 15%) compared to the prevalence using the original KOOS criteria (36% and 24%; Table 2), respectively. Additionally, 35% of people were classified differently between the original KOOS criteria and our refined PASS KOOS criteria (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24809/abstract). Specifically, 22% of participants characterized as having possible/probable early knee OA illness with the original KOOS criteria were characterized as having no early knee OA illness when using the PASS KOOS criteria.

The correct sentences should read: A similar prevalence of probable and possible early knee OA illness when using the PASS KOOS criteria (24% and 18%) compared to the prevalence using the original KOOS criteria(21% and 17%; Table 2), respectively. Additionally, 29% of people were classified differently between the original KOOS criteria and our refined PASS KOOS criteria (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24809/abstract). Specifically, 13% of participants characterized as having possible/probable early knee OA illness with the original KOOS criteria were characterized as having no early knee OA illness when using the PASS KOOS criteria.

The values in Tables 2–5 were incorrect. Below are corrected Tables 2–5.

Table 2. Prevalence of the Luyten original and Englund original KOOS criteria for early knee OA illness in participants post-ACLR*

	Englund original KOOS criteria		
	No early OA illness	Early OA illness	Total
Luyten original KOOS criteria			
No early OA illness	185 (62)	8 (3)	193 (64)
Early OA illness	43 (14)	64 (21)	107 (36)
Total	228 (76)	72 (24)	300

* Values are the number (%). ACLR = anterior cruciate ligament reconstruction (see Table 1 for other definitions).

Table 3. Demographic characteristics across each combination of meeting the Luyten original and Englund original KOOS criteria for early OA illness*

Demographic characteristics	Overall cohort	Did not meet Luyten or Englund criteria	Met Luyten/did not meet Englund criteria; Did not meet Luyten/met Englund criteria	Met Luyten and Englund criteria
No. of participants	300	185	51	64
Female, no. (%)	160 (53)	100 (54)	26 (51)	34 (53)
Age, years	20.0 ± 4.9	19.0 ± 3.9	19.9 ± 4.9	23.2 ± 6.0
BMI (kg/m^2)	24.8 ± 4.5	24.6 ± 4.5	24.3 ± 3.6	25.9 ± 5.0
Time post-ACLR, months	6.2 ± 0.6	6.2 ± 0.6	6.2 ± 0.7	6.2 ± 0.6
Preinjury Tegner score (range 0–10), median (IQR)†	9 [7,9]	9 [8,10]	9 [7,10]	8 [7,9]
IKDC (range 0–100)	81.6 ± 11.7	88.6 ± 8.2	78.7 ± 7.7	69.3 ± 12.9
KOOS subscales (range 0–100)‡				
QOL	65.7 ± 19.1	72.5 ± 17.3	59.3 ± 15.6	51.1 ± 17.0
Pain	92.2 ± 8.3	96.2 ± 4.0	91.1 ± 6.3	81.5 ± 9.4
Symptoms	86.1 ± 12.1	92.9 ± 6.4	78.8 ± 10.9	72.2 ± 10.3
ADL	97.4 ± 6.6	99.2 ± 1.6	97.5 ± 3.2	92.0 ± 12.4
Sport	91.7 ± 9.2	93.4 ± 6.6	92.2 ± 7.7	86.6 ± 13.8

* Values are the mean ± SD unless indicated otherwise. ACLR = anterior cruciate ligament reconstruction; BMI = body mass index; IKDC = International Knee Documentation Committee Subjective Knee Form (lower score indicates worse function); IQR = interquartile range (see Table 1 for other definitions).
† Based on the Tegner Activity Scale, which includes a one-item score, grading based on level of work and sports activities.
‡ For the Knee Injury and Osteoarthritis Outcome Score (KOOS), lower score indicates worse outcome.

Table 4. Prevalence of the Luyten and Englund PASS KOOS criteria for early knee OA illness in participants post-ACLR using PASS as thresholds for each KOOS subscale*

	Englund PASS KOOS criteria		
	No early OA illness	Early OA illness	Total
Luyten PASS KOOS criteria			
No early OA illness	174 (58)	0 (0)	174 (58)
Early OA illness	53 (18)	73 (24)	126 (42)
Total	227 (76)	73 (24)	300

* Values are the number (%) of participants. ACLR = anterior cruciate ligament reconstruction; KOOS = Knee Injury and Osteoarthritis Outcome Score; OA = osteoarthritis; PASS = patient acceptable symptom state.

Table 5. Demographic characteristics across each combination of meeting the Luyten PASS and Englund PASS KOOS criteria for early OA illness*

Characteristics	PASS KOOS criteria		
	Did not Luyten or Englund criteria	Met Luyten, did not meet Englund criteria	Met Luyten and Englund criteria
No. of participants	174	53	73
Female sex, no. (%)	94 (54)	29 (55)	37 (51)
Age, years	19.2 ± 4.4	19.7 ± 4.3	22.2 ± 5.8
BMI (kg/m^2)	24.6 ± 4.5	25.0 ± 4.0	25.3 ± 4.9
Time post-ACLR, months	6.2 ± 0.7	6.2 ± 0.6	6.1 ± 0.6
Preinjury Tegner (range 0–10), median (IQR)†	9 [8,10]	9 [7,9]	9 [7,9]
IKDC (range 0–100)	87.4 ± 8.1	79.9 ± 6.4	68.9 ± 11.6
KOOS subscales (range 0–100)‡			
QOL	75.2 ± 16.1	59.4 ± 13.4	47.5 ± 13.6
Pain	96.7 ± 3.4	91.2 ± 6.1	82.1 ± 9.0
Symptoms	90.8 ± 9.0	84.0 ± 10.2	76.4 ± 13.6
ADL	99.5 ± 1.5	95.9 ± 12.4	93.4 ± 6.1
Sport	94.2 ± 4.8	92.5 ± 6.4	85.3 ± 14.4

* Values are the mean ± SD unless indicated otherwise. ACLR = anterior cruciate ligament reconstruction; ADL = activities of daily living; BMI = body mass index; IKDC = International Knee Documentation Committee Subjective Knee Form (lower score indicates worse function); IQR = interquartile range; QOL = quality of life.
† Based on the Tegner Activity Scale, which includes a one-item score, grading based on level of work and sports activities.
‡ For the Knee Injury and Osteoarthritis Outcome Score (KOOS), lower score indicates worse outcome.