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 74 No. 7 July 2022

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 74 • JULY 2022 • NO. 7

Cover image: The figure on the cover (from Weiss et al, page 1058) shows a coronal oblique STIR image of the sacroiliac joints of a 15-year-old female, demonstrating metaphyseal-equivalent hyperintense signal, a normal variant, that could be mistaken for subchondral inflammation by less experienced reviewers.

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1039–1040
DOI 10.1002/acr.24863

American College of Rheumatology
Empowering Rheumatology Professionals

EDITORIAL

Clinical Academic Rheumatology: Still Getting More Than You Pay For

Sterling G. West and V. Michael Holers

In this issue of *Arthritis Care & Research*, D'Anna et al found that rheumatology clinician-educators generate more than 12 times (more than 9 times in 2005 dollars) greater downstream revenue that benefits their academic medical center than they are paid for an outpatient office visit (1). This larger and more robust study confirms a 2005 study by Wickersham et al, which similarly showed that academic rheumatologists contribute significantly to the overall financial health of their health care system (2). Both studies suggest that these data can be used to argue for higher compensation and/or more protected time for clinical academic rheumatologists based on their financial value. They also argue for consideration of supplementation of academic rheumatologist salary income by the academic health care system, especially as the downstream revenue can compare quite favorably to specialists in primary care where this approach is not uncommon (3).

The study by D'Anna et al (1) hypothesized that changes in rheumatology clinical practice, such as musculoskeletal ultrasound and more infusible biologics, would generate significantly more downstream revenue than the amount found in the previous 2005 study. This hypothesis was not confirmed, as they reported $12.14 ($9.37 in 2005 dollars) in downstream revenue compared to $10.02 in the Wickersham et al study (2) for every $1.00 received for a rheumatology outpatient office visit charge. The larger data set and different study design may explain these results. However, it should be pointed out that the office visit charge (Current Procedural Terminology codes) that a clinician-educator billed (i.e., the denominator) was far less in the 2005 Wickersham study due to institutional undercoding for evaluation and management (E&M) visits, for fear of potential Medicare audits at the time. In addition, over ~15 years, there has been an increase in Medicare outpatient E&M visit payments for cognitive subspecialties, and Southern California has a higher local reimbursement rate than Colorado (the location of the study by Wickersham et al). Finally, the study by D'Anna and colleagues did not include revenue generated by any patients treated with intravenous gamma globulin for conditions such as inflammatory myositis, which accounted for a large percentage (60%) of infusible drug costs in the study by Wickersham et al. If D'Anna and colleagues would have had such patients, the downstream revenue generated (i.e., the numerator) would likely have been higher.

So, what should rheumatology division chiefs and practice directors do with this data? Following publication of the Wickersham et al study, an editorial was published entitled "Academic Rheumatology: Like It, Leave It, or Fight to Change It" (4). In the division of rheumatology at the University of Colorado School of Medicine, we decided to "fight to change it." Specifically, our changes have included the following measures:

1. Intensive education of institutional coders and clinicians to prevent E&M coding errors, yet assure that clinicians get the appropriate payment for outpatient visits.
2. Inclusion of rheumatology clinician-educator salaries that are pegged to the national median for academic nephrology clinician-educators (who are typically compensated at a higher rate) based on their academic rank and that are normalized through additional increases accounting for the number of years at rank.
3. Establishment of a Gender Equity Council by the university to assure female and male faculty, including clinician-educators at each academic rank, get equal pay, benefits, promotions, and leadership opportunities for equal time, work, and productivity.
4. Eligibility of rheumatology clinician-educators who meet their clinical billing benchmarks to share in a yearly clinical salary bonus based on patients seen and not on the amount of downstream income generated.
5. Creation of written guidelines by the university that are used by academic promotion committees for clinician-educators in which clinical output is only 1 factor. Ratings on teaching

Sterling G. West, MD, MACP, MACR, V. Michael Holers, MD, MACR: University of Colorado Anschutz Medical Campus, Aurora.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.24863&file=acr24863-sup-0001-Disclosureform.pdf.

Address correspondence to Sterling G. West, MD, MACP, MACR, University of Colorado Anschutz Medical Campus, Division of Rheumatology, Department of Medicine, 1775 Aurora Court, Mailstop B-115, Aurora, Colorado 80045. Email: sterling.west@cuanschutz.edu

Submitted for publication June 2, 2021; accepted in revised form January 13, 2022.

effectiveness by fellows and residents are important. Scholarly excellence can be demonstrated through multiple means (including publications, curriculum development, leadership on education committees), as well as other methods.

6. Incorporation of a non-proceduralist agreement that stipulates that the university hospital will supplement clinical income to the rheumatology division on a sliding scale based on prespecified work relative value units generated utilizing 8 half days per week of outpatient clinical time as 1.0 full-time equivalent workload.
7. Receipt of credit for clinician-educators for academic production (such as lectures and published articles) and a monetary bonus as a reward that can only be used for nonsalary purposes, including costs of medical licenses and educational expenses (such as meeting costs and travel reimbursement).
8. Functioning of the hospital infusion center as a pharmacy benefit manager; the hospital administration is aware and regularly reminded of the income generated by infusion of biologics ordered by clinical rheumatologists, which lessens the pressure to continually see more patients.
9. Screening of all outpatient rheumatology consults and scheduling of only patients with an inflammatory rheumatic disease in the university outpatient rheumatology clinic, assuring that patients most in need of rheumatologic care are seen in a timely manner and that more complicated cases are available for training fellows and residents. The university hospital administration recognizes the advantage of scheduling these patients who generate higher E&M codes and significantly more (44 times as much) downstream revenue for the hospital compared to patients with noninflammatory musculoskeletal problems (5).

All of these changes required significant discussions with the chief of medicine and the university hospital administration, and these modifications are constantly at risk of being rescinded. The studies by D'Anna et al and Wickersham et al (1,2) provide hard financial data that can aid in such discussions at other universities so that the benefit of rheumatology clinician-educators to the financial health of the academic medical center is recognized and rewarded.

There are many rewards and challenges to practicing rheumatology as a clinician-educator at an academic medical center (6). It is a high (likely too high) expectation for a rheumatology clinician-educator to utilize 8 half-day outpatient clinics and as many patients as a rheumatologist in private practice and still be academically productive in order to advance in academic rank. Most clinician-educators get some help in meeting their clinical requirements by being an attending in the fellows' and/or rotating residents' clinics and on inpatient ward rounds where the fellows and residents do much of the time-consuming work, such as recording the evaluation in the electronic medical record after the patient is evaluated by the faculty attending. In addition, administrative and educational roles can be compensated by hospitals and other sources to decrease the time spent in clinical activities. The fellows and residents also rate the clinician-educator's educational effectiveness during these encounters, which are used for faculty advancement. Moreover, fellows (and residents) are frequently included in the faculty's research, which can help the clinician-educator meet their scholarly goals for advancement while helping the fellowship meet their Accreditation Council for Graduate Medical Education requirements for accreditation.

Unfortunately, there will be situations where talented clinician-educators will fail to meet these high expectations and this will lead to stress, burnout, and/or loss of the clinician-educator to private practice where they can receive a higher salary. It is important, however, to point out that private practice rheumatologists face their own unique problems in their clinical practices without having a larger institution to provide financial backstops in case of unforeseen challenges, such as a global pandemic (7,8). The continued support and advocacy by the American College of Rheumatology will be important to meet the future challenges of both academic and private practice rheumatologists.

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. D'Anna KM, Lynch CS, Cabling M, Torralba KD, Downey C. Clinical academic rheumatology: a boon for health systems. Arthritis Care Res (Hoboken) 2022;74:1041–8.
2. Wickersham P, Golz D, West SG. Clinical academic rheumatology: getting more than you pay for. Arthritis Rheum 2005;53:149–54.
3. Fahey P, Cruz-Huffmaster D, Blincoe T, Welter C, Welker MJ. Analysis of downstream revenue to an academic medical center from a primary care network. Acad Med 2006;81:702–7.
4. Wortmann RL. Academic rheumatology: like it, leave it, or fight to change it. Arthritis Rheum 2005;53:643–5.
5. West SG, Pearson DW, Striebich CC, Goecker R, Kolfenbach JR. The effect of pre-appointment consultation triage on patient selection and revenue generation in a university rheumatology practice. Arthritis Care Res (Hoboken) 2019;71:689–93.
6. Hassan S, Smith MM, Block JA, Jolly M. Challenges to practicing rheumatology in an academic center. Rheum Dis Clin N Am 2019;45:27–37.
7. Eisenberg GM. The focused musculoskeletal factory. Rheum Dis Clin North Am 2019;45:53–66.
8. Hamburger M, Concoff AL, Tardio DK. Opening remarks. On survival and resilience: managing a rheumatology practice through the COVID-19 pandemic. Presented at the United Rheumatology webinar; 2020 April 17.

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1041–1048
DOI 10.1002/acr.24864

American College of Rheumatology
Empowering Rheumatology Professionals

Clinical Academic Rheumatology: A Boon for Health Systems

Kathleena M. D'Anna, Carlos Silva Lynch, Marven Cabling, Karina D. Torralba, and Christina Downey

Objective. Finding a balance between clinical and scholarly productivity is a challenge for many academic clinician-educator rheumatologists. An examination of workload and downstream revenue determines if the financial value generated by services rendered by rheumatologists are proportionate to the financial value created for a health system. A 2005 study found that academic rheumatologists generate $10.02 for every $1.00 they receive for an office visit.

Methods. A retrospective analysis of ordering and billing practices of 5 full-time clinician-educator rheumatologists from August 2017 to February 2019 was conducted. Individual workload is defined as averaged full-time equivalent workload based on time spent on clinical and academic duties. Academic productivity was reviewed. Revenue-generating activities that benefited the division directly and downstream revenue were collected. Revenue was extrapolated based on volumes of referrals, publicly available drug costs, and estimated Medicare reimbursement values (average sales price) of representative drugs.

Results. The total revenue by physician that benefited the division directly was $597,203, with evaluation and management codes accounting for $174,456. Downstream revenue by physician totaled $2,119,437. The largest contributor was from referrals to the hospital-based infusion center, at $1,287,496. The downstream revenue generated by rheumatologist per dollar of evaluation and management services was found to be $12.14 ($9.37 in 2005 dollars).

Conclusion. For every $1 generated though office visits by 5 practicing academic rheumatologists at our institution, $12.14 was generated through downstream revenue, which, when adjusted for inflation, shows stability in the value generated by academic rheumatologists ($10.02 versus $9.37).

INTRODUCTION

A major challenge faced by clinical rheumatologists practicing in an academic medicine setting is finding a balance between clinical productivity and academic pursuits (1). As rheumatology continues to be one of the least compensated subspecialties in medicine (2), an examination of the workload of academic faculty and the downstream revenue generated in the faculty practice serves to determine if the value generated by rheumatology services are proportionate to salaries. An analysis performed by Wickersham et al in 2005 found that clinical rheumatologists generate >$10.00 for every $1.00 they receive for an office visit (3). When adjusted for inflation, this becomes $13.00 for every $1.00.

Emphasis on early diagnosis and treatment, and the expansion of the aging population has led to an increase in demand for rheumatologists; however, the supply of such physicians has not been able to keep pace with demand. According to the 2007 American College of Rheumatology (ACR) Workforce Study (4), there is currently a shortage of practicing rheumatologists in the US, and this will become more dire as time goes on (4). It is likely this shortage will also affect academic rheumatology practices. The ACR Workforce Study estimates a 138% increase in the demand for full-time practicing rheumatologists by the year 2030 (5). This problem is further exacerbated by the distance one is from a major metropolitan center or coast, where the majority of large academic medical centers are located and where most physicians choose to practice (6). Outside of regional practice areas, a rheumatologist transitioning from fellowship to independent practice has a general choice to make when determining what career path they will pursue; private practice or academic practice. One major factor influencing whether a physician

Kathleena M. D'Anna, DO, Carlos Silva Lynch, Marven Cabling, MD, Karina D. Torralba, MD, Christina Downey, MD: Loma Linda University School of Medicine, Loma Linda, California.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.24864&file=acr24864-sup-0001-Disclosureform.pdf.

Address correspondence to Christina Downey, MD, Address: 11234 Anderson Street, MC-1516, Loma Linda, California 92354. Email: cdowney@llu.edu.

Submitted for publication December 2, 2021; accepted in revised form January 13, 2022.

SIGNIFICANCE & INNOVATIONS

- The present study explores the relationship between direct clinical services rendered by a rheumatologist through patient care and the downstream revenue a hospital system garners from these encounters.
- We show the dollar value generated by rheumatologists' ordering practices as compared to their direct evaluation, and management services have remained fairly stable from a similar study published ~15 years ago, using a more robust data set.
- This study is valuable to those in positions of leadership in academic health systems as it may be used to argue for higher compensation or more protected time for clinical academic rheumatologists based on the overall value brought to the health system.

chooses private practice or academia is compensation. Traditionally, physicians practicing in teaching environments are paid lower salaries than those who work in a private practice setting, on the order of $100,000 per year less (2,7). This financial reality may be driving some rheumatologists to work in a higher income–generating setting. Aside from family obligation or other personal circumstances, the cost of pursuing a career in medicine is rising, as is the debt incurred to cover those expenses. According to the American Association of Medical Colleges, 45% of graduating medical students owe $200,000 or more in student loans (8), which continues to generate high interest monthly upon graduation.

Incentives to enter academic practice may therefore be lacking for many graduating fellows; however, in 2016 it was noted that <50% of physicians in general owned their own practices and have opted to join large medical groups, including those who are part of major academic centers (9). As health systems have evolved to include more community-based practice centers for academic medical centers, with stable compensation that is often higher than those at academic centers, as opposed to traditional academic faculty who rely on competing for research grants for their salary or on compensation from Medicare and Medicaid, there might be greater opportunity for rheumatologists in academically affiliated settings to be more involved in teaching and training the next generation of rheumatologists.

While the line from services rendered to direct financial benefit is clearly visible in specialties that generate large reimbursements through services such as cardiac catheterizations or colonoscopies and that largely provide hospital-based services (10), the value generated by rheumatologists is less concrete, as most of the services provided by rheumatologists are clinic based and reimbursements are much smaller. There is a lack of available national benchmarks for rheumatology value to health care systems, and therefore it is up to each individual rheumatology division to assert the worth of the division to hospital leadership. Those rheumatologists who are drawn to academic practices may not have the knowledge base or business sense to argue their financial value to administration.

The objective of this study is to determine the financial gains a health system can expect from clinical academic rheumatology practices and to address additional aspects of productivity. Our hypothesis is that with changes in rheumatology clinic practices, including the increase in infusion drug availability and cost, the amount generated per office visit has surpassed that found by Wickersham et al more than 15 years ago (3).

MATERIALS AND METHODS

A retrospective review of 5 full-time academic rheumatologists was conducted, analyzing the ordering and billing patterns over an 18-month period (August 2017 to February 2019) in a non hospital-based faculty clinic practice. The workload of each individual was averaged across the group and described as full-time equivalents (FTEs). FTEs are based on time spent completing inpatient consultation work, outpatient faculty practice income-generating activities, as well as academic and administrative duties. Salary information was disclosed by administrators and averaged across all faculty. No part of physician salary is supported by grants.

Revenue-generating activities were classified as those that directly benefited the hospital (downstream revenue) and the rheumatology division directly through the faculty practice. All laboratory tests, radiologic studies, treatments, and new and established patient encounters were performed within the Loma Linda University Health System and recorded in the electronic medical record. The values of Current Procedural Terminology (CPT) codes were assigned based on facility price Medicare allowable charges for 2018. When a CPT code was not available in the Medicare allowable charges list, a value was assigned based on the billings used for our individual institution (Loma Linda University School of Medicine).

The total revenue generated from downstream referrals (referred procedures to other specialties, referrals for consultations from other specialists, laboratory testing, radiographic testing, and the hospital infusion center profits) were summed. Within-division revenue was calculated by adding the in-office procedure, in-office infusion, and evaluation and management (E&M) code revenue. We then calculated the ratio of downstream revenue to E&M code revenue to determine if the value found by Wickersham and colleagues has changed over ~15 years (3). When making comparisons to previous data, values were adjusted for inflation using the American Institute for Economic Research calculator.

Academic productivity. Academic productivity was noted based on scholarly work, publications, and professorial

status. Scholarly work was defined as posters presented at national professional society meetings. Publications were counted if they were associated with a PubMed ID number. In-process work was identified by the investigator portals at each teaching site.

Faculty practice description. During the study period, the practice comprised 5 full-time faculty at a nonhospital-based faculty practice. As of 2017, there has been a designated faculty practice director and a faculty practice infusion director. There are 4 chairs en-suite for infusion. Four of 5 faculty perform ultrasounds at the point of care with 2 ultrasound machines. It is important to note that the location of this faculty practice serves patients from the counties of San Bernardino and Riverside, where many Indigenous patients reside. The group also provides inpatient and outpatient services to the Riverside County Health System and a federally qualified health care center in San Bernardino, but data from those services were excluded. This study exclusively examined the university health system data.

Office-based revenue. Physicians' charges for E&M coded office visits and office-based procedures, including ultrasound, inpatient consultation billings, and in-office infusions, during the same 18-month period were recorded. Reimbursements for in-office infusions were calculated based on infusion procedure codes, including non-oncologic chemotherapy drug J codes and injection (CPT codes 96372–96379) and infusion (CPT codes 96372–96379).

Downstream revenue. Generated downstream revenue was calculated based on orders placed, which included laboratory tests, radiologic studies, office-based procedures, referred procedures to other specialties, and consultations from referrals to other specialists (including physical therapy and hospital-based ambulatory center infusions). For every dollar generated, the downstream hospital revenue was estimated. CPT codes for all charges were assigned based on facility price Medicare allowable charges for 2018. Referrals to other specialists were accounted for by assigning reimbursement for a level 4, the new patient E&M code.

The practice feeds referrals into 2 infusion centers, 1 of which is in-office while the other is hospital based. Attempts were made to track revenue by the rheumatology-ordered infusions from the hospital-based infusion center; however, this information was not disclosed by hospital administration. Instead, we tracked the number of infusions given at the hospital-based infusion center over an 18-month period, using rituximab, abatacept, infliximab, tocilizumab, and denosumab as reference drugs. Published wholesale acquisition prices for the reference drugs were estimated based on the typical dose and frequency for an adult weighing 75 kg and were extrapolated to an 18-month period.

The rituximab dosage used was 2 infusions of 1,000 mg every 14 days, with 2 infusions to be repeated every 6 months over 18 months, totaling 6,000 mg. The abatacept dosage was estimated to be 14 infusions of 750 mg per infusion per year, equaling 15,750 mg per 18-month period. The infliximab dosage was estimated to be 5.5 mg/kg at 75 kg every 6 weeks, totaling 7 infusions per year and an 18-month total dosage of 4,537.5 mg. The tocilizumab total dosage was calculated to be 6 mg/kg at 75 kg every 4 weeks, or 14 doses per year, totaling 9,450 mg. The denosumab dosage was 60 mg every 6 months, totaling 180 mg during the study period.

Our hospital participates in the 340B Drug Pricing Program; however, discounted prices are proprietary and therefore were not shared with us. We estimated 340B cost to be 60% of wholesale acquisition prices based on consultation with private practice colleagues in the area. These were chosen as representative infusion medications due to the availability of wholesale acquisition prices. We did not take in to account overhead costs, as these were not disclosed to us. Further, the Medicare reimbursement scheme takes overhead into account by reimbursing the average sales price (ASP) and an additional 6%. This 6% is intended to cover the costs of administering the medications, including overhead. To simplify our assumptions, we did not include the additional 6% when calculating the reimbursement rate to the infusion center. Profit was calculated as the ASP minus the estimated 340B price multiplied by the number of patients found on retrospective chart review to have received those medications at the hospital-based infusion center and who were referred by rheumatology.

Table 1. Academic productivity by 5 academic rheumatologists over an 18-month period*

Academic marker	Outcome
Promotion/rank	1 full professor (achieved within 5 years of start of employment), 1 associate professor (within 4 years), 3 assistant professors
Teaching time	2 half-days of trainee supervision (on average, per week), 1 monthly lecture for trainees
Peer-reviewed publications	12 publications in high-quality rheumatology journals (in total)
ACR abstracts	12 accepted and presented (on average, generated yearly)
Book chapters	1 book chapter each by 2 faculty
Ongoing research projects including clinical trial	1 NIH-funded research 1 investigator-initiated pharmaceutical company–sponsored research 4 industry-sponsored trials 7 nonfunded investigator-initiated research
Centers development	7 centers approved and under development (lupus, inflammatory arthritis, bone health and osteoporosis, vasculitis, scleroderma, myositis, Sjögren's disease)

* ACR = American College of Rheumatology; NIH = National Institutes of Health.

Table 2. Calculation of E&M billings*

CPT code	No. of encounters	Facility price†	Total billings
99203	88	$79.06	$6,957.28
99204	896	$133.74	$119,831.04
99205	430	$174.63	$75,090.90
99213	444	$53.21	$23,625.24
99214	3,593	$81.53	$292,937.29
99215	3,071	$115.22	$353,840.62
Total	–	–	$872,280.00
Total per physician	–	–	$174,456.00

* CPT = Current Procedural Terminology; E&M = evaluation and management code.
† Taken from physician fee schedule for 2018 Medicare administrative contractor, Riverside-San Bernardino-Ontario.

RESULTS

Individual workloads, averaged as the FTE workload based on time spent on inpatient consultation, outpatient faculty practice income-generating activities, and academic and administrative duties, were noted (Table 1). Time for industry-sponsored clinical trials was considered academic time. Average FTE allocation is 0.2 FTEs for academic time and 0.8 FTEs for clinical duties, with 0.4 FTEs allocated for clinic care with trainees and 0.4 FTEs for faculty practice clinic care. Faculty practice care was reduced for on-call hospital consultations (0.5 FTEs). Additional FTE allowances were given for clinical directorships (0.1 FTEs), division chief (0.1 FTEs), and fellowship program directorship (0.3 FTEs). Academic productivity was also noted by the number of publications generated by each faculty member.

Academic rank of the 5 faculty members included 1 full professor (ranking achieved within 5 years), 1 associate professor (ranking achieved within 4 years), and 3 assistant professors. In total, the faculty published 12 original research papers and 2 book chapters. At the time of this study, the faculty members were supporting 1 National Institutes of Health–funded research project, 1 investigator-initiated, pharmaceutical company–sponsored trial, 4 industry-sponsored trials, and 7 nonfunded investigator-initiated research projects (Table 1).

The faculty office practice generated an average of $174,456 per physician for office visits (E&M codes) (Table 2). On average, faculty saw 9 patients per half-day clinic. In-office procedures, including ultrasound, generated $25,180 per physician. Inpatient billings totaled an average of $349,195 per physician. The in-office infusion center generated $48,372 per physician. The sum of revenue which benefits the rheumatology practice directly was $597,203 per physician (Table 3).

Downstream revenue from referred procedures to other specialties totaled $60,310, and referrals for consultations from other specialties totaled $300,189, each reported as per physician totals. Laboratory testing and imaging studies were found to total $347,542 and $123,900 per physician, respectively. The largest piece of downstream revenue by far was the profit to the hospital-based infusion center, which totaled $6,437,480 in whole or $1,287,496 per physician.

The most widely used medication of those tracked in our practice was rituximab, with 220 patients receiving treatment at an estimated profit of $4,972,414. There were 21 patients who received abatacept, with an estimated profit of $745,794.

Table 3. Revenues generated per clinical activity by rheumatologists over an 18-month period*

Activity	Average revenue generation per physician, hospital/downstream	Average revenue generation per physician, within division
In-office procedures (arthrocentesis, point-of-care ultrasound)	–	$25,180
In-office infusions†	–	$48,372
E&M codes	–	$174,456
Inpatient billing	–	$349,195
Referred procedures to other specialties (echocardiograms, pulmonary function tests, etc.)	$60,310	–
Referrals for consultations to other specialties	$300,189	–
Laboratory tests	$347,542	–
Radiology (MRI, CT, x-rays, DXA)	$123,900	–
Hospital infusion center‡	$1,287,496	–
Total	$2,119,437	$597,203
Ratio of downstream generated hospital revenue to E&M revenue	$12.14	–

* CT = computed tomography; DXA = dual x-ray absorptiometry; E&M = evaluation and management codes; MRI = magnetic resonance imaging.
† Non hospital-based infusion clinic, based only on administration codes as the facility does not buy and bill, and is not able to utilize 340B pricing.
‡ Based on a representative number of medications administered during the study period estimated by subtracting the published average sales price from the estimated 340B price to find the profit or loss per dose of medication.

Table 4. Calculation of hospital-based infusion center revenue

Drug	No.	WAC for 18-month course*	Predicted 340B pricing (cost)†	ASP‡	Estimated profit per patient (ASP-cost)	Total profit during study period
Rituximab	220	$50,113.50	$30,068.10	$52,669.98	$22,601.88	$4,972,413.60
Abatacept	21	$62,182.50	$37,309.50	$72,823.50	$35,514.00	$745,794.00
Infliximab	99	$64,230.10	$36,786.60	$38,936.74	$2,150.14	$212,863.98
Tocilizumab	26	$47,815.50	$28,689.30	$42,808.50	$14,119.20	$367,099.2
Denosumab	156	$3,836.37	$2,301.82	$3,194.82	$893.00	$139,307.69
Total	–	–	–	–	–	$6,437,478.47
Per physician	–	–	–	–	–	$1,287,496

* Wholesale Acquisition Cost (WAC) of representative drugs obtained from Schmier et al (ref. 25). Doses of drugs calculated based on a 75 kg patient with stable rheumatoid arthritis or osteoporosis.
† Predicted 340B Drug Pricing Program price calculated as WAC (from Schmier et al) with 40% discount.
‡ From the January 2018 Centers for Medicare and Medicaid Services 2018 Average Sales Price (ASP) Drug Pricing Files (ref. 26).

Infliximab was estimated to yield $212,864 for the 99 patients who were treated with this drug for rheumatologic indications. A total of 26 patients received tocilizumab at a profit of $367,099, and 156 patients were referred to the hospital-based infusion center for denosumab by rheumatology for a total profit of $139,308 (Table 4).

For every dollar our practice generated in E&M codes, we generated $12.14 in downstream revenue, ~$2 more than the initial study examining this ratio. However, when adjusted for inflation, this difference ($0.65) is negligible from that found ~15 years ago.

Figure 1 shows succinctly the utilization of health care dollars by the academic practice. The largest proportion was on infused medications, followed by laboratory tests and E&M codes.

DISCUSSION

In 2005, it was determined that academic rheumatologists generate $10.02 for every $1.00 they bill for an office visit (3). According to our findings, this number has largely held stable when adjusted for inflation. We did find, however, an overall increase in the dollar amount generated from laboratory studies, radiology, and consults to other specialists when compared to the earlier study. However, we also found an increase in revenue generated from patient E&M services. Both the numerator and denominator have increased. In 2005, Wickersham et al found that physician office visit billing generated $36,297 total from 730 encounters over an 18-month period. This equates to $49.72 per encounter ($65.09 in 2019) (3). Our study found $229,042 total from 1,704 encounters over an 18-month period. This equates to $134.41 per encounter, which is more per encounter than the rate found in the study by Wickersham and colleagues in 2005, even when adjusted for inflation.

The 2005 study was conducted in Colorado, where reimbursement rates based on Medicare data are lower than in our study, which was conducted in Southern California, meaning regional differences could account for some of the difference. One could be tempted to equate some of this difference to the change in reimbursement for CPT codes 99214 and 99215; however, this does not have bearing. In 2005, the local reimbursement for CPT code 99214 was $90.31 and $130.64 for CPT code 99215. In 2019 these rates grew to $120.98 and $161.34, respectively. However, when adjusted for inflation, the reimbursement rate has remained stable in the case of CPT code 99214 ($118.22 versus $120.98) and has actually decreased for CPT code 99215 ($170.01 versus $161.34).

While in essence we attempted to have the same goals as the study by Wickersham et al (3), it is important to note that there are distinct differences between the 2 studies. The study by Wickersham and colleagues included purely clinical revenue without accounting for academic productivity, did not specify FTE allocation, and included only 127 consecutive patients seen over

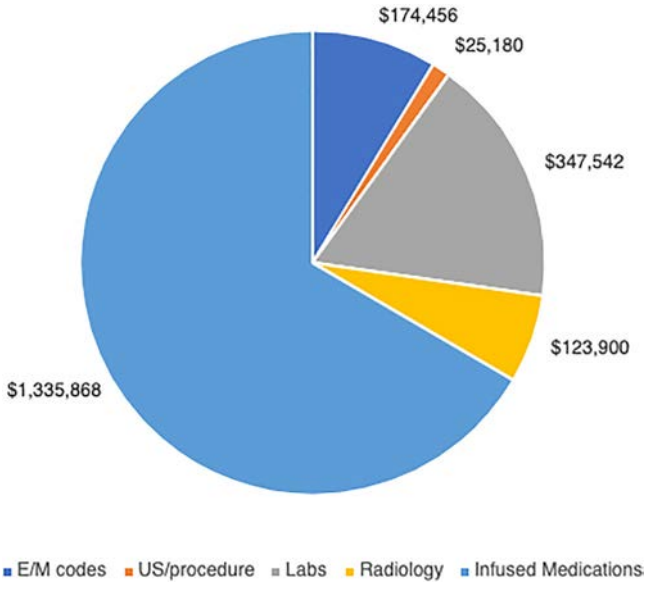


Figure 1. Proportion of health care dollars utilized by an academic rheumatology practice. * E/M = evaluation and management; US = ultrasound. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24864/abstract.

20 half-day clinics by a single rheumatologist in a single center, with the same patients over an 18-month period resulting in 730 visits. In contrast, we included data from 8,522 patient encounters total, or ~1,704 per physician. This increase in volume and variety of data may paint a more robust picture than that which was painted 15 years ago, thus adding more weight to the conclusion of the original study. The increase in health care spending overall in the interim 15 years between the studies is concerning; however, based on our review of our practice patterns, the largest area in which health care dollars are being spent is medications (50%), not E&M codes (6.4%) (Figure 1).

We believe reducing health care costs should come from focusing on reducing the costs of medications administered to our patients and not devaluing E&M codes. If E&M coding valuation is increased, we predict rheumatologists would then be more likely to stay in academia to train the next generations of rheumatologists. Because the E&M code reimbursement scheme has held steady or decreased, reimbursement has not been adequate to recoup the indirect overhead costs of doing business (11,12). A 2016 study by Lee et al found "indirect overhead costs, such as information technology, administrative staff, hospital operations, and maintenance are generally estimated to represent almost half of total hospital costs, increasing more rapidly than medical inflation during the past decade" (13). Because the E&M code reimbursement scheme hasn't kept up with the cost of running a practice, clinician educators are driven by administration to see more and more patients to keep up with rising indirect costs, taking precious time away from research, teaching, and mentorship.

Practicing academic rheumatologists treat chronic and complex medical issues, offering substantial benefits not only to their patients but to the community as a whole, through research and education. In the past decade the field of rheumatology has grown considerably, with an increased global prevalence of rheumatologic and musculoskeletal diseases by ~2 million individuals (14). The demand has accelerated in part due to the aging US patient population (15), along with the rise of new diseases and spread of existing diseases.

The growing disparity of trained rheumatologists among the ever-increasing patient pool leaves little time for academic activities outside direct clinical care. A lack of appropriate compensation and proper allocation of protected time is another driving factor for this disparity. Currently, rheumatology remains one of the lowest compensated specialties, according to the 2019 Medscape Physician Compensation Report (2). Cognitive-based specialties like rheumatology face a huge disparity when compared to high-earning, procedural-based specialties, particularly for the practicing academic rheumatologist. Based on data collected for the 2019 Medscape Physician Compensation report, academic rheumatologists see an average salary of $154–$164K annually versus $255–260K for their private-practice–based colleagues (2).

The field of rheumatology has seen many advancements and changes in the last several decades (16), from an increase in the number of new biologic agents available to the further incorporation of diagnostic tools such as ultrasound (17) and targeted blood markers (18,19). The improved availability of clinical blood markers in rheumatology could be one of the reasons our study found laboratory testing to be one of the highest portions of downstream revenue generated. As new and improved tests become available, and as research progresses on previously unexplored biologic and immunomodulation therapies, the treatment of rheumatologic conditions continues to be revolutionized, but that progress is stunted.

Academic rheumatologists face great challenges balancing a demanding outpatient clinical practice with the pressures and expectations of scholarly activity. Most clinicians who pursue academic medicine are driven by a desire to complete research, to teach, and by the intellectual stimulation offered by the academic setting (20); however, financial pressures and other obligations distract and skew the appeal. Clinician educators face compounding difficulty contributing to scholarly activity given increasing patient load, degree of complex cases at academic centers, and duties with fellow and resident education (1). These stressors with the added burden of inadequate financial rewards, have led to a flux of practitioners leaving academic medicine for private practice or early retirement (15, 20), adding further to the disproportion of academic rheumatologists.

According to the Becker's Hospital Review 2015 Physician Compensation Report, private practice rheumatologists generate on average 4,821 work relative value units (RVUs) (21). In our practice, academic rheumatologists averaged an annual production of 4,755 work RVUs, comparable to their private practice counterparts. In addition to working RVU production, faculty still were able to generate an average annual manuscript publication rate of 8 publications per year, not including book chapter writing, mentorship for learner projects, and teaching time. Most faculty are also active in volunteer positions with the American College of Rheumatology. Faculty perform clinically at a level of a private practice rheumatologist, generate a high level of scholarly activity, participate nationally in medical societies, and contribute to medical education.

The present study has several limitations. First, this was a retrospective study. Given the nature of the study, a prospective study could have altered practice patterns of physicians and would not be a true snapshot of the natural ordering practices. Second, only revenue generated by available Medicare allowable charges was able to be accurately included in the final total and several CPT codes were found to be missing from Medicare allowable charges data. Further, our data were collected at a single site, which does not take into consideration regional variations of Medicare allowable charges. Also, clinical trials revenues, which can be a major source of income, were not included due to a major restructuring of the research administration during the

study period. Fourth, hospital-based infusion center revenue, which accounts for the largest source of downstream revenue in both the 2005 Wickersham study (3) and in this current study, was not readily available. Information from the hospital on 340B pricing is considered proprietary. The 340B pricing system is in place at our institution, since a significant number of underinsured patients are seen in our health system; this pricing system was put into place to allow the monies offset by the system to be reinvested in supporting other services that would help the same population. In contrast, 340B pricing for medicines given at the non-hospital–based rheumatology clinic is currently not in place.

Other more specific information directly derived from the hospital about the volumes from our clinics, such as the number of patients who received joint replacement or other major surgeries, medication ordered by each rheumatologist, revenues related to services provided to patients with rheumatic diseases at the rehabilitation hospital, and hospital-based home health and home-based infusion services are not wholly retrievable. We extrapolated data based on infusion drug costs and reimbursement values of some of the medications based on publicly published information, as detailed previously in the present study. Further studies would be needed to adequately address this aspect of revenue, especially since it represents such a large portion of a rheumatologist's billed expenses. Our study likely underestimates infusion revenue, which was the highest source of downstream revenue in both studies.

The lack of transparency across cost centers in a hospital system makes research such as this difficult to conduct. It is unknown how much revenue the hospital system truly generates from its rheumatologists, especially when taking into consideration infusion medications ordered by rheumatologists, but even based on the 2005 study, we predict that our sums are grossly underestimated. Lastly, we only included data related to services provided at the university health system, which can also underestimate our value overall. Part of the lack of transparency may be in effort to reduce the potential for conflicts of interest, as suggested by Seaman's letter addressing the 2005 article by Wickersham et al (3,20). However, as voiced by the original authors of the study, we assert that knowing the value of downstream infusion revenue would allow academic rheumatologists to better communicate the financial value they bring to health care systems, thus allowing more leverage when negotiating hospital support for rheumatology divisions. It is well known that interventional cardiologists are compensated handsomely in part due to the revenue brought to hospital systems from their procedures; why not increase compensation for rheumatologists based on the revenue they generate for a health system? We do, however, agree that conflicts of interest and diagnostic/procedural bias due to earning potential should be investigated and minimized.

Currently, fee-for-service encourages physicians to see large volumes of patients or perform many procedures rather than compensate based on the complexity of disease observed in patients (23). The proposed 2021 Physician Fee Schedule aims to increase compensation for cognitive specialists, acknowledging the value of caring for patients with complex medical profiles rather than skewing reimbursements toward procedure-based specialties (24). We are hopeful that if finalized, the increase in compensation for cognitive specialties will translate to increased salary for academic rheumatologists or an increase in protected time for clinician educators to pursue scholarly activity.

In conclusion, we argue that downstream revenue generated in addition to direct billings should be strongly considered in providing value to the work rheumatologists provide to an entire academic health system. We also suggest developing a compensation package that considers teaching and other scholarly activities, since time demand for these activities may not correlate to the amount of corresponding compensation obtained. We encourage new consideration for the downstream revenue generated by infusion products, laboratory testing, procedures, and radiographic imaging in overall salary earned by academic rheumatologists.

REFERENCES

1. Hassan S, Smith MM, Block JA, Jolly M. Challenges to practicing rheumatology in an academic center. Rheum Dis Clin North Am 2019;45:27–37.
2. Kane L. Medscape Physician Compensation Report 2019. URL: https://www.medscape.com/slideshow/2019-compensation-overview-6011286.
3. Wickersham P, Golz D, West SG. Clinical academic rheumatology: getting more than you pay for. Arthritis Rheum 2005;53:149–54.
4. Deal CL, Hooker R, Harrington T, Birnbaum, Hogan P, Bouchery E, et al. The United States rheumatology workforce: supply and demand, 2005–2025. Arthritis Rheum 2007;56:722–29.
5. Battafarano DF, Ditmyer M, Bolster MB, Fitzgerald JD, Deal C, Bass AR, 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.
6. American College of Rheumatology Committee on Rheumatology Training and Workforce Issues, FitzGerald JD, Battistone M, Brown CR Jr, Cannella AC, Chakravarty E, et al. Regional distribution of adult rheumatologists. Arthritis Rheum 2013;65:3017–25.
7. Differences between academic and community medical centers. George Washington University. 2021. URL: https://healthcaremba.gwu.edu/blog/the-differences-between-community-and-academic-medical-centers/.
8. Minder CM. Student debt in American medicine: I am not a loan! J Am Coll Cardiol 2016;67:885–8.
9. Schmier J, Ogden K, Nickman N, Halpern MT, Cifaldi M, Ganguli A, et al. Costs of providing infusion therapy for rheumatoid arthritis in a hospital-based infusion center setting. Clin Ther 2017;39:1600–17.
10. 2019 Physician Inpatient/Outpatient Revenue Survey. Merritt Hawkins. URL: https://www.merritthawkins.com/uploadedFiles/MerrittHawkins_RevenueSurvey_2019.pdf.
11. American Association of Medical Colleges. AAMC Faculty Salary Report. URL: https://www.aamc.org/data-reports/workforce/report/aamc-faculty-salary-report.
12. Coffron M, Zlatos C. Medicare physician payment on the decline: it's not your imagination. 2019. URL: https://bulletin.facs.org/

2019/09/medicare-physician-payment-on-the-decline-its-not-your-imagination/.

13. Lee VS, Kawamoto K, Hess R, Park C, Young J, Hunter C, et al. Implementation of a value-driven outcomes program to identify high variability in clinical costs and outcomes and association with reduced cost and improved quality. JAMA 2016;316:1061–72.
14. Maini MA, Adelowo F, Al-Saleh J, Al Weshahi Y, Burmester GR, Cutolo M, et al. The global challenges and opportunities in the practice of rheumatology: white paper by the World Forum on Rheumatic and Musculoskeletal Diseases. Clin Rheumatol 2015;34:819–29.
15. Lawrence-Wolff K, Hildebrand B, Monrad S, Ditmyer M, Fitzgerald J, Erickson A, et al. 2015 ACR/ARHP Workforce Study in the United States: A Maldistribution of Adult Rheumatologists [abstract]. Arthritis Rheumatol 2016;68 Suppl 10. URL: https://acrabstracts.org/abstract/2015-acrarhp-workforce-study-in-the-united-states-a-maldistribution-of-adult-rheumatologists/.
16. Romão VC, Fonseca JE. Major challenges in rheumatology: will we ever treat smarter, instead of just harder? Front Med 2019;6:144.
17. Amorese-O'Connell L, Gutierrez M, Reginato AM. General applications of ultrasound in rheumatology practice. Fed Pract 2015;32 Suppl:8–20S.
18. Cuppen BV, Welsing PM, Sprengers JJ, Bijlsma JW, Marijnissen AC, van Laar JM, et al. Personalized biological treatment for rheumatoid arthritis: a systematic review with a focus on clinical applicability. Rheumatology (Oxford) 2016;55:826–39.
19. Smolen JS, et al. Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. Ann Rheum Dis 2016;75:3–15.
20. Straus SE, Straus C, Tzanetos K. Career choice in academic medicine: systematic review. J Gen Intern Med 2006;21:1222–9.
21. Rappleye E. 2015 Physician Compensation, Work RVU by specialty. Becker's Healthcare. URL: https://www.beckershospitalreview.com/compensation-issues/2015-physician-compensation-work-rvu-by-specialty.html.
22. Seaman WE. Clinical academic rheumatology: comment on the article by Wickersham et al. Arthritis Rheum 2005;53:800.
23. Anandarajah A, Ritchlin CT. Problems with fee for service payments for academic rheumatology practices: a need for payment reform [abstract]. Arthritis Rheumatol 2014. URL: https://acrabstracts.org/abstract/problems-with-fee-for-service-payments-for-academic-rheumatology-practices-a-need-for-payment-reform/.
24. US Centers for Medicare and Medicaid Services. Fact sheet proposed policy payment, and quality provisions changes to the Medicare physician fee schedule for calendar year 2021. 2020. URL: https://www.cms.gov/newsroom/fact-sheets/proposed-policy-payment-and-quality-provisions-changes-medicare-physician-fee-schedule-calendar-year-4.
25. Schmier J, Ogden K, Nickman N, Halpern MT, Cifaldi M, Ganguli A, et al. Costs of providing infusion therapy for rheumatoid arthritis in a hospital-based infusion center setting. Clin Ther 2017;39:1600–17.
26. Centers for Medicare and Medicaid Services. 2018 ASP Drug Pricing Files. URL: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Part-B-Drugs/McrPartBDrugAvgSalesPrice/2018ASPFiles.

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1049–1057
DOI 10.1002/acr.24860

American College of Rheumatology
Empowering Rheumatology Professionals

Patient Perceptions and Preferences Regarding Telemedicine for Autoimmune Rheumatic Diseases Care During the COVID-19 Pandemic

Maria I. Danila,[1] Kelly Gavigan,[2] Esteban Rivera,[2] W. Benjamin Nowell,[2] Michael D. George,[3] Jeffrey R. Curtis,[1] Lisa Christopher-Stein,[4] Shubhasree Banerjee,[3] Peter A. Merkel,[3] Kalen Young,[5] Dianne G. Shaw,[5] Jennifer Gordon,[5] and Shilpa Venkatachalam[2]

Objective. To assess the perceptions and preferences of telemedicine among patients with autoimmune rheumatic diseases during the COVID-19 pandemic.

Methods. We conducted an online survey among patients with autoimmune rheumatic diseases. Attitudes about telemedicine (i.e., telemedicine acceptability), evaluated using the validated Telemedicine Perception Questionnaire (TMPQ), and visit satisfaction were assessed for different telemedicine experiences and types of autoimmune rheumatic disease.

Results. Of 3,369 invitations, 819 responses were received. Participants had a mean ± SD age of 58.6 ± 11.6 years and were mostly White (n = 759, or 92.7%) and female (n = 702, or 85.7%). Of the 618 participants who said that telemedicine was available to them, 449 (72.7%) reported having a telemedicine visit, with 303 (67.5%) reporting attending a telemedicine video visit. On a 0 to 10 scale, the mean ± SD visit satisfaction score was 7.3 ± 1.8, with 25.8% of respondents being very satisfied (scores of 9 or 10). Video visits and higher TMPQ scores were associated with higher satisfaction. Compared to those who did not experience a telemedicine visit, patients who did were more likely to prefer telemedicine (video or phone) for routine visits (73.7% versus 44.3%; $P < 0.001$), reviewing test results (64.8% versus 53.8%; $P < 0.001$), when considering changing medications (40.5% versus 26.8%; $P < 0.001$), and when starting a new injectable medication (18.9% versus 12.7%; $P = 0.02$).

Conclusion. During the COVID-19 pandemic, patients with autoimmune rheumatic diseases frequently had telemedicine visits, with the majority held via video, and were satisfied with these visits. These results suggest that because patients prefer telemedicine for certain visit reasons, maximizing effective use of telemedicine will require personalized patient scheduling.

INTRODUCTION

Outpatient health care delivery has been significantly transformed due to the COVID-19 pandemic (1,2). The pandemic disrupted nonessential in-person outpatient visits (3) and led to a dramatic uptake in remotely delivered diagnostic and treatment services (e.g., telemedicine) for patients with chronic conditions, including autoimmune rheumatic diseases. These patients are distinctly at risk for worse COVID-19 outcomes due to multimorbidity (4) and the use of immunosuppressive drugs, such as glucocorticoids and biologics (5,6) that predispose them to infections (7–9) and require close monitoring for side effects (10,11).

Although communication technologies can facilitate timely assessment, treatment, and health education for people living with

The Arthritis and Rheumatic Disease COVID-19 Project was supported by the Patient-Centered Outcomes Institute (PCORI), Eli Lilly and Company, and Janssen Pharmaceuticals. The Vasculitis Patient-Powered Research Network was partially supported by a PCORI Award (PPRN-1306-04758) and by GlaxoSmithKline. Dr. George's work was supported by the NIH, National Institute of Arthritis and Musculoskeletal and Skin Diseases (grant K23-AR-073931-01).

[1]Maria I. Danila, MD, MSc, MSPH, Jeffrey R. Curtis, MD, MPH, MS: University of Alabama at Birmingham; [2]Kelly Gavigan, MPH, Esteban Rivera, MS, W. Benjamin Nowell, PhD, Shilpa Venkatachalam, PhD, MPH: Global Healthy Living Foundation, Upper Nyack, New York; [3]Michael D. George, MD, MSCE, Shubhasree Banerjee, MD, Peter A. Merkel, MD, MPH: University of Pennsylvania, Philadelphia; [4]Lisa Christopher-Stein, MD, MPH: Johns Hopkins University School of Medicine, Baltimore, Maryland; [5]Kalen Young, MA, Dianne G. Shaw, MA, Jennifer Gordon, PhD: Vasculitis Patient-Powered Research Network, Kansas City, Missouri.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.24860&file=acr24860-sup-0001-Disclosureform.pdf.

Address correspondence to Maria I. Danila, MD, MSc, MSPH, Associate Professor of Medicine, University of Alabama at Birmingham, 2nd Avenue S, Birmingham, AL 35226. Email: mdanila@uabmc.edu.

Submitted for publication March 9, 2021; accepted in revised form January 11, 2022.

SIGNIFICANCE & INNOVATIONS

- During the COVID-19 pandemic, members of patient communities who had autoimmune rheumatic diseases frequently had telemedicine visits, with the majority held via video, and were satisfied with these visits.
- Compared to patients who did not have a telemedicine visit, those who had experienced telemedicine care were more likely to prefer telemedicine for routine visits, reviewing test results, and when considering changing medications, including new injectable medication.
- Because patients prefer telemedicine for certain visit reasons, maximizing effective use of telemedicine will require personalized patient scheduling.

chronic conditions, much remains to be learned about the impact of the telemedicine expansion on the access and quality of care that patients with autoimmune rheumatic diseases have received in the COVID-19 era. For example, while some patients were able to successfully engage in telemedicine visits for rheumatology care during the rapid transition to telemedicine (12), socially vulnerable populations (as defined by race, income, education, rural residence, computer literacy, and internet access) may experience unintended consequences from these factors that shape access to and the effectiveness of telemedicine (13–15).

To support policy-level changes and promote patient- and clinician-informed decisions about optimal rheumatology care via telemedicine during and beyond the COVID-19 pandemic, it is critical to understand patients' experiences with telemedicine, patients' access to different types of telemedicine visits (e.g., video, phone), and how availability of telemedicine may affect patient preferences for receiving care in-office or virtually. Thus, in June 2020, as part of the Autoimmune COVID-19 Project of the Autoimmune Research Collaborative, we launched an online survey focused on telemedicine for members of patient communities who have autoimmune or inflammatory conditions, including autoimmune rheumatic diseases. The goal of this study was to gain insight on uptake and utilization of telemedicine by video or phone among this medically vulnerable patient population and to better understand patients' perceptions and attitudes about telemedicine visits and factors that influence these perceptions.

PATIENTS AND METHODS

Study setting and population. Adults ages ≥19 years with an autoimmune/rheumatic condition participating in the Autoimmune COVID-19 Project (www.rheumcovid.com) conducted by the Autoimmune Research Collaborative were invited to participate in the present study (16). The Autoimmune Research Collaborative is an alliance of patient-powered research networks (PPRNs) including the Inflammatory Bowel Disease Partners, Multiple Sclerosis PPRN, ArthritisPower PPRN, and Vasculitis PPRN (17). Participants in the Autoimmune COVID-19 Project also include members of the following patient organizations: Myositis Support and Understanding, Lupus Allied Disease Association, American Bone Health, and the International Foundation for Autoimmune and Autoinflammatory Arthritis. Launched on March 28, 2020, the goal of the Autoimmune COVID-19 project is to understand the COVID-19–related concerns and behaviors of patients in the US and Canada who have autoimmune and rheumatic conditions and to collect information from patients about their experiences with medical care during the COVID-19 pandemic. We included participants who were ages ≥19 years because the ArthritisPower Registry has institutional review board (IRB) approval to recruit US participants who are ≥19 years of age. The protocol was approved by the Advarra IRB (protocol no. Pro00042873).

The cross-sectional survey specifically about telemedicine (e.g., access, satisfaction, perceptions about telemedicine, and preference for next visit type) was conducted between June 18, 2020, and August 10, 2020. We sent survey invitations to all participants in the Autoimmune COVID-19 Project (n = 3,369) (18), and the following results are of the participants with autoimmune rheumatic diseases who completed this telehealth survey.

Data collection. As part of the Autoimmune COVID-19 Project, participants completed questions about their age, race/ethnicity, sex, state and 5-digit zip code of their residence, smoking habits, comorbidities, Patient-Reported Outcomes Measurement Information System (PROMIS) anxiety score (18), type of autoimmune or rheumatic condition, and use of immunosuppressive/immunomodulatory therapies, glucocorticoids, and nonsteroidal antiinflammatory drugs. For participants indicating multiple autoimmune rheumatic conditions, a hierarchical approach was used to categorize their autoimmune rheumatic condition considering the relative specificity of various diagnoses (antineutrophil cytoplasmic antibody [ANCA]–associated vasculitis > other vasculitis or relapsing polychondritis > myositis > lupus > psoriatic arthritis [PsA] > ankylosing spondylitis [AS] > rheumatoid arthritis [RA]), similar to previous studies (19). For example, participants reporting diagnoses of PsA and RA were categorized as having PsA, given the expectation of greater specificity for that diagnosis. Participants' residence in a rural versus urban county was defined using the Centers for Disease Control 2013 National Center for Health Statistics classification (20).

In the cross-sectional telemedicine survey, the participants were asked, "Is your rheumatologist/specialist that manages your rheumatic/autoimmune condition offering telephone or telehealth visits?" with possible response options being "Yes," "No," or "I don't know." Participants self-reported whether they had a telemedicine visit and its type (e.g., phone only or video), reported satisfaction with their telemedicine visit using the 1-item overall visit satisfaction (0–10 scale, with 0 representing "worst possible

visit" and 10 representing "best possible visit") from the validated Agency for Healthcare Research and Quality Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey (21,22). Patients also reported preference for type of visit at the next appointment (i.e., "If you had a choice, what type of visit would you prefer?" with survey choices for types of visits including in-office, videoconference, phone, or videoconference or phone visit, with the last choice listed indicating no preference for video or phone visit [23]) as well as attitudes about telemedicine using (with permission) a modified version of the validated Telemedicine Perception Questionnaire (TMPQ) (24). The TMPQ score is a validated measure to assess patient acceptability of health care delivered via telemedicine that takes into account perceptions of benefits and limitations of in-home telemedicine monitoring (24). A total TMPQ score (range 17–85) was calculated for each of the respondents, with higher scores showing higher acceptability.

All participants were also asked to indicate their preference for a future telemedicine visit compared to an in-office visit with their rheumatologist or autoimmune disease specialist (choices included preferences for an in-office visit, preference for a telemedicine visit, no preference, and not sure) for specific clinical scenarios (i.e., reasons for visits) including routine visit when feeling well, during a disease flare, for reviewing test results, for having medication side effects, for a new problem, when considering changing therapy, and when starting a new injectable medication. Because the telemedicine survey was deployed during the COVID-19 pandemic, respondents answered questions regarding future telemedicine visits in relation to the pandemic being ongoing.

Statistical analysis. To summarize the data, we used the mean ± SD for continuous variables, and frequencies and proportions for categorical variables. Because visit satisfaction ratings were positively skewed, we standardized visit satisfaction ratings (mean = 0 and variance = 1). We used *t*-tests and multivariable linear regression analysis to compare the satisfaction and TMPQ scores in respondents who reported participating in video and phone-only visits and by disease type, grouping together RA, AS, and PsA as "inflammatory arthritis" versus other conditions. Chi-square tests were used to compare preferences for telemedicine visits for specific clinical scenarios between those who had experience receiving care with telemedicine versus those who did not have experience with telemedicine (i.e., our comparator group comprised those who did not have a telemedicine visit irrespective of whether they were aware or not of the fact that they had access to such visits). We built multivariable logistic regression models that included age, sex, place of residence, and diagnosis to determine patient factors associated with preference for telemedicine visits versus in-person visits as a future visit type among those who experienced a telemedicine visit. We categorized disease type as follows: other autoimmune condition (group 1), RA, PsA, and AS (group 2), myositis and systemic lupus erythematosus (group 3), and ANCA-associated vasculitis and other vasculitis (group 4).

We built a multinomial logistic regression model evaluating preference for telemedicine (phone or video) versus in-office visit for multiple different clinical scenarios (i.e., routine visit, disease flare, reviewing test results, discussing medication side effects, discussing a new problem, changing medications, and starting a new injectable medication). The clinical scenario was included in the model as a factor variable where each scenario served as a category and where the "discuss new problem" scenario was the referent. This model included age, residence (rural versus urban), and visit type for the patient's previous telemedicine visit (video versus phone). Because we thought that the type of telemedicine visits a patient may have already experienced might also influence their preference for telemedicine versus in-office visit across different clinical scenarios, we focused our analysis on the group of participants who previously experienced telemedicine visits for each clinical situation. To control for correlations within patients, we estimated the model with clustered SEs. We validated the estimates from the multinomial logistic regression analysis using bootstrap resampling (25). The referent group for the dependent variable in this model was preference for in-office visit. Odds ratios (ORs) shown are for comparisons made between telemedicine visits and in-office visits, with an OR of >1 indicating that a patient had a higher preference for a telemedicine visit and an OR of <1 indicating a lower preference for a telemedicine visit compared to an in-office visit. ORs for "no preference" versus in-office visit are not shown. All analyses were conducted in SAS (version 9.3; Enterprise Guide version 4.3) and R (version 4.0.2).

RESULTS

Of the 3,369 invitations sent for our online telemedicine survey, 1,852 individuals (55.0%) opened the invitation email, and 819 people with self-reported autoimmune rheumatic diseases completed the survey. Compared to nonrespondents, respondents were older, more likely to have RA, and more likely to be receiving a biologic disease-modifying antirheumatic drug, a JAK inhibitor, methotrexate, and/or hydroxychloroquine (Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24860). A total of 618 respondents (75.5%) said that they were aware that telemedicine was available to them if they needed it. Among respondents who already had a telemedicine visit (n = 449), the most common reported autoimmune rheumatic conditions were RA (41.6%), ANCA-associated vasculitis (16.3%), PsA (11.8%), and AS (7.8%). Those who self-reported experiencing a telemedicine visit had a mean ± SD age of 57.7 ± 12.1 years and were mostly White (92.2%) and female (86.0%); a minority of respondents resided in a rural area (11.0%). The mean ± SD T score on the PROMIS Anxiety questionnaire among people who reported

Table 1. Demographic and clinical characteristics of respondents, stratified by type of telemedicine visit*

Characteristic	All participants (n = 819)	Video telemedicine visit (n = 303)	Phone telemedicine visit (n = 146)	No telemedicine visit (n = 370)	*P*
Age, mean ± SD years	58.6 ± 11.6	56.7 ± 12.6	59.8 ± 10.6	59.7 ± 10.8	0.01†
Female sex	702 (85.7)	264 (87.1)	122 (83.6)	316 (85.4)	0.82
White	759 (92.7)	278 (91.8)	136 (93.2)	345 (93.2)	0.57
Hispanic	37 (4.5)	11 (3.6)	4 (2.7)	22 (6.0)	0.07
Rural residence	99 (13.2)	28 (9.8)	16 (13.8)	55 (15.9)	0.05†
Autoimmune condition					
Rheumatoid arthritis	353 (43.1)	128 (42.2)	59 (40.4)	166 (44.9)	0.35
ANCA-associated vasculitis	115 (14.0)	47 (15.5)	26 (17.8)	42 (11.4)	0.04†
Psoriatic arthritis	108 (13.2)	32 (10.6)	21 (14.4)	55 (14.9)	0.20
Ankylosing spondylitis	66 (8.1)	26 (8.6)	9 (6.2)	31 (8.4)	0.76
Other autoimmune rheumatic disease‡	54 (6.6)	16 (5.3)	10 (6.9)	28 (7.6)	0.31
Other vasculitis or relapsing polychondritis	54 (6.6)	24 (7.9)	10 (6.9)	20 (5.4)	0.21
Lupus	38 (4.6)	16 (5.3)	6 (4.1)	16 (4.3)	0.70
Myositis	31 (3.8)	14 (4.6)	5 (3.4)	12 (3.2)	0.46
Medications					
Biologic DMARD	376 (45.9)	147 (48.5)	75 (51.4)	154 (41.6)	0.03†
JAK inhibitor	70 (8.6)	24 (7.9)	11 (7.5)	35 (9.5)	0.40
Methotrexate	250 (30.5)	101 (33.3)	52 (35.6)	97 (26.2)	0.02†
Hydroxychloroquine	195 (23.8)	77 (25.4)	36 (24.7)	82 (22.2)	0.32
Glucocorticoids	241 (29.4)	101 (33.3)	47 (32.2)	93 (25.1)	0.01†
NSAIDs	285 (34.8)	103 (34.0)	52 (35.6)	130 (35.1)	0.85
Comorbidities					
Hypertension	354 (43.2)	136 (44.9)	69 (47.3)	149 (40.3)	0.12
Lung disease§	299 (36.5)	111 (36.6)	52 (35.6)	136 (36.8)	0.90
Diabetes mellitus	101 (12.3)	41 (13.5)	14 (9.6)	46 (12.4)	0.95
Kidney disease	81 (9.9)	28 (9.2)	20 (13.7)	33 (8.9)	0.40
Heart disease	72 (8.8)	21 (6.9)	12 (8.2)	39 (10.5)	0.11
Current smoking	60 (7.3)	21 (6.9)	6 (4.1)	33 (8.9)	0.11
Malignancy	17 (2.1)	5 (1.7)	3 (2.1)	9 (2.4)	0.52
PROMIS anxiety, mean ± SD T score¶	58.2 ± 8.8	58.9 ± 8.2	58.1 ± 9.0	57.6 ± 9.1	0.06

* Except where indicated otherwise, values are the number (%) of respondents. Rural residence status is shown for participants who had available zip codes. *P* values were calculated based on differences between the characteristics of respondents who had a telemedicine visit versus those who did not. ANCA = antineutrophil cytoplasmic antibody; DMARD = disease-modifying antirheumatic drug; NSAIDs = nonsteroidal anti-inflammatory drugs; PROMIS = Patient-Reported Outcomes Measurement Information System.
† Statistically significant at $P = 0.05$.
‡ Other rheumatic diseases included antiphospholipid antibody syndrome, anti–glomerular basement membrane antibody disease, juvenile idiopathic arthritis, mixed connective tissue disease, psoriasis, sarcoidosis, scleroderma, and Sjögren's syndrome.
§ Lung disease included asthma, emphysema, chronic obstructive pulmonary disease, pulmonary hypertension, and other chronic lung disease.
¶ Anxiety was measured using the PROMIS anxiety short form (score range 1–100). For reference, the mean ± SD PROMIS anxiety T score in the US adult population is 50 ± 0.

having a telemedicine visit was 58.7 ± 8.5, and among those who did not report experiencing telemedicine, it was 57.6 ± 9.1, which is 8.7 SDs and 7.6 SDs higher, respectively, than normative values for people living in the US ($P = 0.06$) (18). We observed statistically significant differences between those who experienced a telemedicine visit (n = 449) and those who did not have a telemedicine visit (n = 370) in age, place of residence, ANCA-associated vasculitis diagnosis, and types of medication used (e.g., methotrexate and glucocorticoids) (Table 1).

Clinical scenarios favored for telemedicine visits. Among the survey respondents, a majority preferred a telemedicine visit by video or phone for routine visits (n = 495, or 60.4%) or for review of test results (n = 490, or 59.8%) (Figure 1). However, a minority of the respondents also preferred telemedicine visits for evaluation of a new problem (n = 176, or 21.5%), during a disease flare (n = 150, or 18.3%), or when starting a new injectable medication (n = 132, or 16.1%) (Figure 1). The proportion of survey respondents preferring telemedicine versus in-office visits among different clinical scenarios were similar between those who had video visits versus phone visits (data not shown).

More participants who had experienced a video or phone telemedicine visit reported preferring a telemedicine visit for a routine check-in when feeling well (73.7%) than participants who had not had a telemedicine visit (44.3%; $P < 0.001$). The results were similar for a visit to review blood work or other tests (64.8% versus 53.8%; $P < 0.001$), when considering changing medications

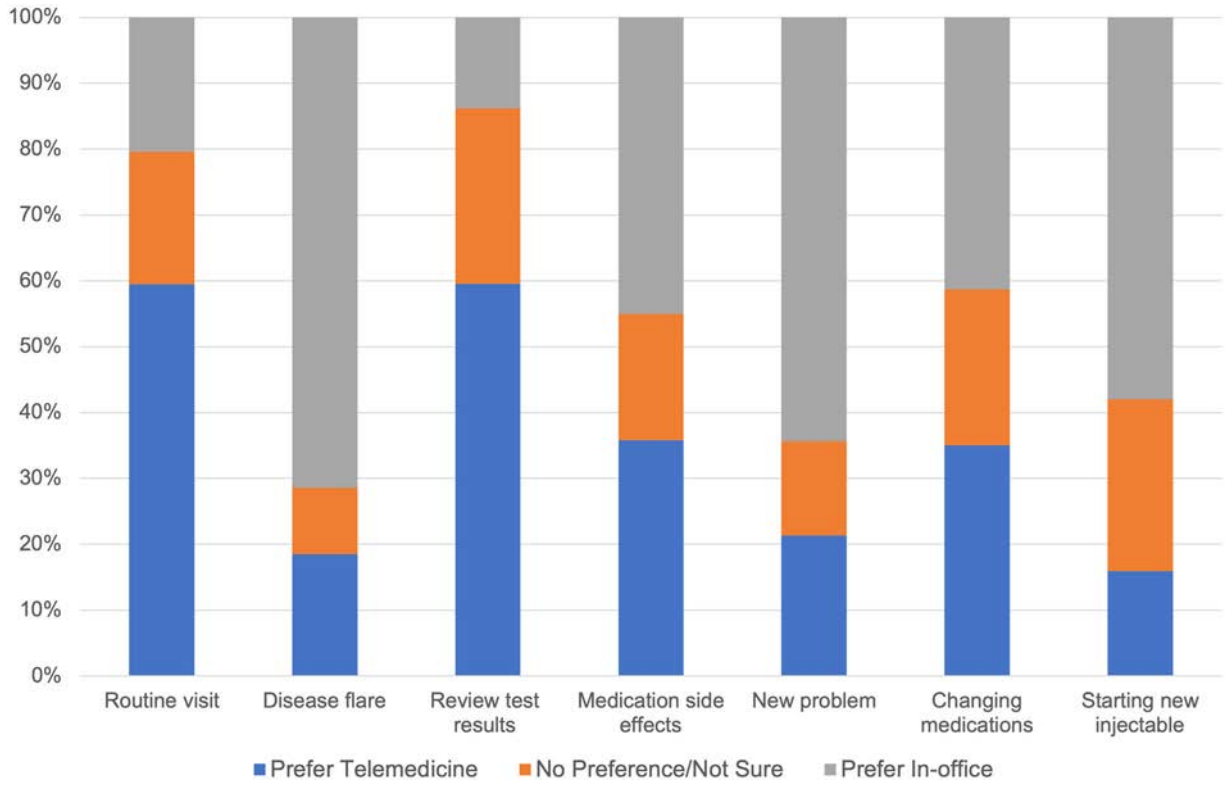


Figure 1. Preference for telemedicine visits versus in-office visits among all survey respondents (n = 819), stratified by reason for visit.

(40.5 versus 26.8; $P < 0.001$), and when starting a new injection medication (18.9% versus 12.7%; $P = 0.02$) (Table 2).

Visit satisfaction and telemedicine acceptability among respondents who experienced a telemedicine visit. The mean ± SD satisfaction rating was 7.3 ± 1.8, and 25.8% of the respondents (n = 116) reported high levels of satisfaction with the telemedicine visit (score of 9 or 10) (Table 3). The mean ± SD telemedicine acceptability from the validated TMPQ score was 62.8 ± 10.7, supporting a favorable attitude toward telemedicine among patients who participated in a telemedicine visit. Among survey respondents who experienced a telemedicine visit, satisfaction and telemedicine perception scores were similar between respondents with autoimmune inflammatory arthritis (e.g., RA, AS, and PsA) and those with other autoimmune conditions (7.2 ± 1.9 versus 7.4 ± 1.7 for satisfaction scores and 62.5 ± 10.5 versus 63.4 ± 10.9 for telemedicine perception scores). Levels of agreement/disagreement with specific statements within the TMPQ are presented in Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24860. More participants who had a video visit agreed with the statement that telemedicine was a convenient form of health care delivery for them (78.2%) versus participants who had a phone visit (69.9%; $P = 0.05$). Similarly, more participants who had video visits agreed that telemedicine saves time (90.4% versus 82.9%; $P = 0.02$) (Supplementary Table 2).

Table 2. Participant preference for telemedicine visit for different visit reasons based on prior experience with telemedicine*

Reason for clinic visit	All participants (n = 819)	Had a telemedicine visit (n = 449)	Did not have a telemedicine visit (n = 370)	*P*
Routine care	495 (60.4)	331 (73.7)	164 (44.3)	<0.001†
Disease flare	150 (18.3)	79 (17.6)	71 (19.2)	0.56
Review of test results	490 (59.8)	291 (64.8)	199 (53.8)	<0.001†
Medication side effects	293 (35.8)	169 (37.6)	124 (33.5)	0.22
New problem	176 (21.5)	101 (22.5)	75 (20.3)	0.44
Medication change	281 (34.3)	182 (40.5)	99 (26.8)	<0.001†
Starting a new injectable medication	132 (16.1)	85 (18.9)	47 (12.7)	0.02†

* Values are the number (%) of patients.
† Statistically significant at $P = 0.05$.

Table 3. Respondents' perceptions about telemedicine by type of telemedicine visit experienced*

Characteristic	All telemedicine visits (n = 449)	Video visit (n = 303)	Phone visit (n = 146)
Mean ± SD satisfaction rating on a 0–10 scale	7.3 ± 1.8	7.5 ± 1.7	7.0 ± 2.0
<6	83 ± 18.5	49 ± 16.2	34 ± 23.3
6–8	250 ± 55.7	173 ± 57.1	77 ± 52.7
9–10	116 ± 25.8	81 ± 26.7	35 ± 24.0
Mean ± SD score on the TMPQ†	62.8 ± 10.7	63.3 ± 10.4	61.8 ± 11.1
Office visit is better	258 (57.5)	167 (55.1)	91 (62.3)
Telemedicine visit is better	41 (9.1)	31 (10.2)	10 (6.9)
No difference/unsure	150 (33.4)	105 (34.7)	45 (30.8)
Preference for next visit type‡			
In-office	255 (56.8)	166 (54.8)	89 (61.0)
Video or phone	57 (12.7)	32 (10.6)	25 (17.1)
Phone	23 (5.1)	6 (2.0)	17 (11.6)
Video	114 (25.4)	99 (32.7)	15 (10.3)

* Except where indicated otherwise, values are the number (%) of respondents.
† Telemedicine perception questionnaire (TMPQ) scores range from 17 to 85, with higher values indicating a more favorable perception of telemedicine.
‡ Next visit type preference was assessed using answers to the following question, "If you had a choice, what type of visit would you prefer."

Patient factors associated with preference for telemedicine versus in-office as the next visit type among respondents who had a telemedicine visit. Among respondents who experienced telemedicine visits (n = 449), 255 (56.8%) stated that they preferred an in-office visit as their next visit type, and 194 (43.2%) preferred a telemedicine visit (Table 3). More than half of respondents (57.5%) said that an office visit is better than a telemedicine visit. In multivariable logistic regression models, we found that patient diagnosis, place of residence, age, or sex were not associated with preference for telemedicine as the next type of visit (data not shown).

Furthermore, in an adjusted multinominal logistic regression model for preference of the next visit type (with in-office visit as the referent), compared to having a visit to discuss a new problem, ORs (95% confidence intervals [95% CIs]) were higher when the patient would indicate preferring a telemedicine visit versus an in-office visit when the reason for the visit was to review test results (OR 18.45 [95% CI 12.25–25.75]), for routine care (OR 17.76 [95% CI 12.25–25.75], to discuss medication change (OR 3.30 [95% CI 2.41–4.52]), and to discuss medication side effects (OR 2.35 [95% CI 1.73–3.20]). In contrast, compared to having a visit to discuss a new problem, ORs (95% CIs) were lower when the patient would prefer a telemedicine visit versus an in-office visit for evaluation of a disease flare (OR 0.66 [95% CI 0.47, 0.92]) (Table 4).

Table 4. Multinomial regression model evaluating preference for telemedicine visit versus in-office visit for specific clinical scenarios among respondents who already had a telemedicine visit*

Parameter	OR (95% CI)	*P*
Video visit	1.27 (1.04–1.55)	0.017†
Rural residence	1.03 (0.692–1.520)	0.899
Age	0.994 (0.987–1.002)	0.139
Review test results	18.45 (12.25–25.75)	<0.0001†
Medication change	3.30 (2.41–4.52)	<0.0001†
Start a new injectable medication	0.98 (0.70–1.37)	0.902
Disease flare	0.66 (0.47–0.92)	0.026†
Routine care	17.76 (12.25–25.75)	<0.0001†
Medication side effects	2.35 (1.73–3.20)	<0.0001†

* 95% CI = 95% confidence interval; OR = odds ratio. Clinical scenarios included in the analysis were modeled using a "discuss new problem" scenario as the referent.
† Statistically significant at $P = 0.05$.

Patient factors associated with telemedicine visit satisfaction and telemedicine perception score. Among participants who reported having a telemedicine visit (n = 449), most of these visits occurred by videoconferencing (n = 303, or 67.5%). Compared to those who had phone-only telemedicine visits, respondents who experienced video visits were slightly younger, resided in urban areas, and reported higher satisfaction with the telemedicine visit; a higher proportion of these respondents also expressed a preference for a video telemedicine visit as a future visit (Table 2). However, there were no differences in the TMPQ score between those who had video telemedicine visits and those who had phone telemedicine visits.

In multivariable linear regression models after controlling for age and place of residence (rural versus urban), we found that compared to those who had phone-only visits, those who had video visits expressed higher satisfaction (an average of 0.145 units on the standardized scale). Similarly, a positive relationship existed between TMPQ score and telemedicine visit satisfaction rating. A 0.068-unit increase in the TMPQ score led to a 1-unit increase in satisfaction with visit rating (Table 5). There was no association between type of telemedicine visit (phone versus video) and TMPQ score (Table 5).

Table 5. Factors associated with telemedicine visit satisfaction and telemedicine perception score (n = 449)*

Variable	Outcome: patient satisfaction			Outcome: telemedicine perception score		
	Estimate	SE	*P*	Estimate	SE	*P*
Rural residence	0.029	0.111	0.796	−1.321	1.571	0.401
Video visit	0.145	0.073	0.047†	−0.147	0.785	0.852
Age	0.002	0.003	0.371	−0.018	0.03	0.546
TMPQ score	0.068	0.003	<0.0001†	NA	NA	NA
Satisfaction score	NA	NA	NA	7.613	0.356	<0.0001†

* NA = not applicable; TMPQ = telemedicine perception questionnaire score.
† Statistically significant at $P = 0.05$.

DISCUSSION

Early in the COVID-19 pandemic, we found that approximately three-fourths of respondents in this population of patients with autoimmune rheumatic diseases had access to telemedicine visits, that the majority of respondents had already had at least 1 telemedicine visit, and that they reported overall good levels of satisfaction with both video and phone-only home-based telemedicine visits. We found that those who had ANCA-associated vasculitis or those who were receiving methotrexate or glucocorticoids were more likely to have experienced telemedicine, possibly indicating they may have had to use telemedicine sooner than other groups. A plurality of respondents thought that telemedicine was as good as or better than in-office visits. The respondents in our present study were much more likely to prefer telemedicine for certain types of clinical scenarios (e.g., routine visits, review of test results, among others), although for all scenarios some patients preferred telemedicine. These results highlight that patients with autoimmune rheumatic diseases have rapidly embraced the expansion of telemedicine for care of chronic diseases, as necessitated by the COVID-19 pandemic, including the use of video and phone-only telemedicine visits.

In our study, we found that the mean ± SD home telemedicine visit satisfaction score was 7.3 ± 1.8, with one-fourth of participants being very satisfied (score 9 or 10 on the 0–10 patient satisfaction scale). The level of satisfaction we observed in the present study was slightly lower than in that in a study of veterans with inflammatory arthritis who received in-facility telemedicine visits (26). While telemedicine may somewhat mitigate current and likely future rheumatology workforce issues, including geographic maldistribution of rheumatologists (27)—and although home-based telemedicine expanded dramatically due to the COVID-19 pandemic due to major health policy changes (28,29)—patients' perceptions, attitudes, and perspectives about telemedicine for rheumatology care are understudied. Home-based telemedicine visits conducted via phone or video-conferencing enable rheumatology care for socially and medically vulnerable groups (30–32) and allow patients to avoid travel that increases COVID-19 risk (33). We collected data on patient experience with telemedicine visits because patient satisfaction is a key quality of care outcome, and satisfaction has been tied to Medicare reimbursement for clinical services (34). We found that the respondents were satisfied with telemedicine visits irrespective of whether the visits were phone-only or conducted via video-conferencing, a finding that is in line with previous studies on patient satisfaction with telemedicine in rheumatology (35) and supports continued access to telemedicine after the COVID-19 pandemic.

Phone-only telemedicine visits have expanded access to care to patients who may have experienced barriers due to health policy (e.g., insurance coverage) and factors such as age, rural residence, lack of broadband internet, or limited digital literacy (13–15). Given the differences in the type and extent of physical examination that can be performed during phone-only visits compared to video telemedicine visits, we explored whether patients' acceptability of telemedicine differed among those who reported participating in these 2 types of home-based telemedicine visits. We found that the acceptability of home-based telemedicine was good and similar for both video visits and phone-only visits, which supports the perceived value of both types of telemedicine visits for patients with autoimmune rheumatic diseases. In addition, visit satisfaction and telemedicine acceptability were correlated with one another. Compared to a phone-only visit, having a video visit was associated with higher visit satisfaction rating after adjustment for TMPQ score, which is a measure that accounts for the benefits and limitations of different types of home-based telemedicine. This observation is not surprising given that interpersonal communication through phone-only visits is limited to verbal cues and thus lacks the additional visual information that video visits bring into conversations (e.g., non-verbal cues, elements of physical examination). However, this result does not negate the utility of phone-only telemedicine visits for both patients and their rheumatologists as means to preserve access to limited chronic disease care when videoconferencing capability is lacking and in areas that may have limited broadband access, particularly if other data (e.g., electronically collected patient-reported outcomes or passive data from health tracker devices) might be available to supplement the information available to the medical team (36). Our findings can be used by patients, clinicians, and policy makers as they make decisions about participating in and supporting access to telemedicine, both video and phone visits, in the future.

Previous studies have reported rheumatologists' views on the appropriateness of a clinical situation for telemedicine, but to

our knowledge, our study is the first to evaluate the patient perspective on the suitability of particular scenarios for a telemedicine visit. Importantly, the patients surveyed in our study favored telemedicine versus in-office visits in some specific clinical contexts, such as for reviewing results of blood work and other testing and for routine visits when feeling well, perhaps because these visits may not require a full hands-on physical examination. These clinical scenarios are commonly encountered in clinical practice and conducting such visits via telemedicine could reduce the burden on patients associated with travel for in-person office visits, alleviate illness-related work productivity loss, and mitigate other social impacts such as the need to arrange for care of children or other family members. As telemedicine for rheumatology care grows, future research needs to address best practices for delivering care remotely. For example, for video visits, expanding access to high-speed internet, defining the appropriate audiovisual equipment needed (computer versus smartphone), visit setting (in-home versus a facility close to a patient's home), environmental characteristics (e.g., a quiet and well-lit space), and training of patients and medical teams on how to participate in and guide with physical examination are key for enabling best quality of care.

Conversely, patients who had used telemedicine at least once favored in-office visits to a telemedicine visit for the evaluation of a new problem, during a disease flare or when starting a new injectable medication, although a sizeable minority preferred telemedicine even in these situations. These patients' views are remarkably concordant with the perspectives of a group of academic rheumatologists who participated in center-based telemedicine visits and deemed that telemedicine visits were not optimal because of unclear diagnosis (e.g., disease flare in the context of another rheumatic condition), complexity of the disease process (e.g., requiring physical examination that could not be performed remotely), or previous poor engagement in care (e.g., lack of recent in-person evaluation) (35). However, for those patients who have a good relationship with their medical team and who understand their disease well, telemedicine for a disease flare might be appropriate.

Our study had several strengths, including a large sample size and use of validated instruments to measure patients' perceptions about different types of telemedicine visits, satisfaction, and acceptability of telemedicine in rheumatology care. Despite its strengths, the study also had some limitations. We surveyed members of an online community of patients with rheumatic diseases, and hence, they may be more comfortable with using technology compared to patients who are not active online.

Although participants in this survey live in different geographical areas, they are primarily White, and their perceptions and attitudes about different types of telemedicine visits may not reflect those of people from other racial/ethnic groups who live with rheumatic diseases in the US. In addition, most respondents reported residing in an urban area, and so the present findings might not be generalizable to people with autoimmune rheumatic diseases who live in rural areas. We did not collect data about patient satisfaction with in-office visits, and thus, we were not able to compare satisfaction with home-based telemedicine versus that with an in-office visit. Because we did not collect information on educational attainment, income level, or whether participants were able to choose the type of telemedicine visit they reported as part of this cross-sectional survey, our study did not examine the association of these factors with perceptions and attitudes about phone or video telemedicine visits. While our survey was conducted in the first 4–6 months of the COVID-19 pandemic, it is possible that the participants who had experienced telemedicine may have had multiple telemedicine visits, both by video and phone, and their responses may reflect these experiences overall, rather than one experience in particular. We chose as a comparator group those who did not have a telemedicine visit, irrespective of their knowledge of telemedicine availability, rather than those who had access to telemedicine but did not have a telemedicine visit because we could not ascertain the reasons why this group did not experience a telemedicine visit (e.g., did not need a visit, had an in-office visit). Thus, our results are not generalizable to those patients with autoimmune rheumatic diseases who did not experience a telemedicine visit. Furthermore, attitudes about both the limitations and benefits of telemedicine versus in-person visits may be different in the middle of the COVID-19 pandemic from what patients may feel once the pandemic has subsided. Although our findings are subject to recall bias, which may affect the estimates of satisfaction with the telemedicine visit, this is less likely to impact assessment of telemedicine benefits and limitations for each modality (video versus phone). Importantly, it is unclear how preferences for types of visits might change when the pandemic is better controlled, especially since phone visits were not considered telehealth/telemedicine and were not reimbursed in the same way as video visits in the past.

In conclusion, during the COVID-19 pandemic, patients with autoimmune rheumatic diseases that were members of an online patient community frequently had telemedicine visits, with the majority held via video, and were satisfied with these visits. Patient preference for telemedicine versus in-office visits depended on the reasons for a visit, past experiences with telemedicine, and attitudes about different types of telemedicine visits. These findings highlight the need to ensure equitable access to telemedicine and to integrate telemedicine into clinical practice in a way that maximizes effectiveness of and satisfaction with visits, with a focus on the reason for a patient's visit and patient preferences.

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. Danila had full access to all data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Danila, Gavigan, Nowell, George, Curtis, Merkel, Young, Shaw, Gordon, Venkatachalam.
Acquisition of data. Gavigan, Nowell, Merkel, Young, Venkatachalam.
Analysis and interpretation of data. Danila, Gavigan, Rivera, Nowell, George, Christopher-Stein, Banerjee, Venkatachalam.

REFERENCES

1. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020;382:1679–81.
2. Bachireddy C, Chen C, Dar M. Securing the safety net and protecting public health during a pandemic: Medicaid's response to COVID-19. JAMA 2020;323:2009–10.
3. Kuy S, Gupta R, Correa R, Tsai R, Vohra S. Best practices for a Covid-19 preparedness plan for health systems. NEJM Catalyst Innovations in Care Delivery. 2020. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0108.
4. Radner H, Yoshida K, Smolen JS, Solomon DH. Multimorbidity and rheumatic conditions—enhancing the concept of comorbidity. Nat Rev Rheumatol 2014;10:252–6.
5. Michaud K, Wipfler K, Shaw Y, Simon TA, Cornish A, England BR, et al. Experiences of patients with rheumatic diseases in the United States during early days of the COVID-19 Pandemic. ACR Open Rheumatol 2020;2:335–43.
6. Mikuls TR, Johnson SR, Fraenkel L, Arasaratnam RJ, Baden LR, Bermas BL, et al. American College of Rheumatology guidance for the management of rheumatic disease in adult patients during the COVID-19 pandemic: version 1. Arthritis Rheumatol 2020;72:1241–51.
7. Dixon WG, Suissa S, Hudson M. The association between systemic glucocorticoid therapy and the risk of infection in patients with rheumatoid arthritis: systematic review and meta-analyses. Arthritis Res Ther 2011;13:R139.
8. Sciascia S, Mompean E, Radin M, Roccatello D, Cuadrado MJ. Rate of adverse effects of medium- to high-dose glucocorticoid therapy in systemic lupus erythematosus: a systematic review of randomized control trials. Clin Drug Investig 2017;37:519–24.
9. Singh JA, Cameron C, Noorbaloochi S, Cullis T, Tucker M, Christensen R, et al. Risk of serious infection in biological treatment of patients with rheumatoid arthritis: a systematic review and meta-analysis. Lancet 2015;386:258–65.
10. Singh JA, Saag KG, Bridges SL Jr, Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. Arthritis Rheumatol 2016;68:1–26.
11. Saag KG, Teng GG, Patkar NM, Anuntiyo J, Finney C, Curtis JR, et al. American College of Rheumatology 2008 recommendations for the use of nonbiologic and biologic disease-modifying antirheumatic drugs in rheumatoid arthritis. Arthritis Rheum 2008;59:762–84.
12. Bashshur R, Doarn CR, Frenk JM, Kvedar JC, Woolliscroft JO. Telemedicine and the COVID-19 pandemic, lessons for the future. Telemed J E Health 2020;26:571–3.
13. Kruse CS, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: a systematic review. J Telemed Telecare 2018;24:4–12.
14. Weinstein RS, Lopez AM, Joseph BA, Erps KA, Holcomb M, Barker GP, et al. Telemedicine, telehealth, and mobile health applications that work: opportunities and barriers. Am J Med 2014;127:183–7.
15. Nouri S, Khoong EC, Lyles CR, Karliner L. Addressing equity in telemedicine for chronic disease management during the Covid-19 pandemic. NEJM Catalyst Innovations in Care Delivery. 2020. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0123.
16. Nowell WB, Merkel PA, McBurney RN, Young K, Venkatachalam S, Shaw DG, et al. Patient-powered research networks of the autoimmune research collaborative: rationale, capacity, and future directions. Patient 2021;14:699–710.
17. Daugherty SE, Wahba S, Fleurence R. Patient-powered research networks: building capacity for conducting patient-centered clinical outcomes research. J Am Med Inform Assoc 2014;21:583–6.
18. Liu H, Cella D, Gershon R, Shen J, Morales LS, Riley W, et al. Representativeness of the Patient-Reported Outcomes Measurement Information System Internet panel. J Clin Epidemiol 2010;63:1169–78.
19. George MD, Venkatachalam S, Banerjee S, Baker JF, Merkel PA, Gavigan K, et al. Concerns, healthcare use, and treatment interruptions in patients with common autoimmune rheumatic diseases during the COVID-19 pandemic. J Rheumatol 2020;48:603–7.
20. Centers for Disease Control and Prevention. NCHS urban-rural classification scheme for counties. 2013. URL: https://www.cdc.gov/nchs/data/series/sr_02/sr02_166.pdf.
21. Mukherjee S, Rodriguez HP, Elliott MN, Crane PK. Modern psychometric methods for estimating physician performance on the Clinician and Group CAHPS survey. Health Serv Outcomes Res Methodol 2013;13:109–23.
22. Agency for Healthcare Research and Quality. Consumer Assessment of Healthcare Providers and Systems. URL: https://www.ahrq.gov/cahps/about-cahps/index.html.
23. Walter SD, Turner RM, Macaskill P, McCaffery KJ, Irwig L. Estimation of treatment preference effects in clinical trials when some participants are indifferent to treatment choice. BMC Med Res Methodol 2017;17:29.
24. Demiris G, Speedie S, Finkelstein S. A questionnaire for the assessment of patients' impressions of the risks and benefits of home telecare. J Telemed Telecare 2000;6:278–84.
25. Suzuki R, Shimodaira H. Hierarchical clustering with P-values via multiscale bootstrap resampling. R package. 2013.
26. Wood PR, Caplan L. Outcomes, satisfaction, and costs of a rheumatology telemedicine Program: a longitudinal evaluation. J Clin Rheumatol 2019;25:41–4.
27. Ward IM, Schmidt TW, Lappan C, Battafarano DF. How critical is telemedicine to the rheumatology workforce? Arthritis Care Res (Hoboken) 2016;68:1387–9.
28. Centers for Medicare & Medicaid Services. Trump Administration issues second round of sweeping changes to support U.S. healthcare system during COVID-19 pandemic. 2020. URL: https://www.cms.gov/newsroom/press-releases/trump-administration-issues-second-round-sweeping-changes-support-us-healthcare-system-during-covid.
29. Centers for Medicare & Medicaid Services. Medicare Telemedicine Health Care Provider Fact Sheet. 2020. URL: https://www.cms.gov/newsroom/fact-sheets/medicare-telemedicine-health-care-provider-fact-sheet.
30. Hayes BL, Curtis JR, Laster A, Saag K, Tanner SB, Liu C, et al. Osteoporosis care in the United States after declines in reimbursements for DXA. J Clin Densitom 2010;13:352–60.
31. Heath B, Salerno R, Hopkins A, Hertzig J, Caputo M. Pediatric critical care telemedicine in rural underserved emergency departments. Pediatr Crit Care Med 2009;10:588–91.
32. Feldman CH, Ramsey-Goldman R. Widening disparities among patients with rheumatic diseases in the COVID-19 era: an urgent call to action. Arthritis Rheumatol 2020;72:1409–11.
33. Uscher-Pines L, Fischer S, Tong I, Mehrotra A, Malsberger R, Ray K. Virtual first responders: the role of direct-to-consumer telemedicine in caring for people impacted by natural disasters. J Gen Intern Med 2018;33:1242–4.
34. Centers of Medicare and Medicaid Services. Quality Measures Requirements. 2020. URL: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures.
35. Kulcsar Z, Albert D, Ercolano E, Mecchella JN. Telerheumatology: a technology appropriate for virtually all. Semin Arthritis Rheum 2016;46:380–5.
36. Nowell WB, Curtis D, Thai M, Wiedmeyer C, Gavigan K, Venkatachalam S, et al. Digital interventions to build a patient registry for rheumatology research. Rheum Dis Clin North Am 2019;45:173–86.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1058–1064
DOI 10.1002/acr.24529

REVIEW

Children With Enthesitis-Related Arthritis and Possible Benefits From Treatments for Adults With Spondyloarthritis

Pamela F. Weiss,[1] Robert C. Fuhlbrigge,[2] Emily von Scheven,[3] Daniel J. Lovell,[4] Robert A. Colbert,[5] and Hermine I. Brunner,[4] for the PRCSG Advisory Council and the CARRA Executive Committee

This review will summarize clinical, genetic, and pathophysiologic characteristics that are shared between children with enthesitis-related arthritis (ERA) with axial involvement and adults with nonradiographic (and in some cases radiographic) axial spondyloarthritis (SpA), as well as between children with ERA and primarily peripheral disease manifestations and adults with peripheral SpA. Due to the differences in classification criteria for children with ERA and adults with axial and peripheral SpA, the US Food and Drug Administration (FDA) granted automatic full waivers of studies in children for new medications for "axial spondyloarthropathies including ankylosing spondylitis" up until July 2020. Thus, although current juvenile idiopathic arthritis treatment guidelines recommend the use of biologic disease-modifying antirheumatic drugs as part of the early treatment for patients with ERA, none of the FDA-approved therapies for peripheral SpA or nonradiographic axial SpA (certolizumab pegol, ixekizumab, and secukinumab) have been studied or are labeled for use in children with ERA. Considering the similarities between adult SpA and ERA in terms of etiology, genetics, pathogenesis, and clinical manifestations summarized in this review, medications approved for axial SpA or peripheral SpA should also be studied in children with active ERA involving axial or peripheral joints, respectively, with the intent to achieve labeling for use in children. Considering the current lack of effective FDA-approved therapies for ERA, the FDA should also consider requiring pediatric studies for medications that have already been approved for the treatment of adults with SpA.

Introduction

Juvenile idiopathic arthritis (JIA) is a group of chronic pediatric rheumatic diseases of unknown etiology that present by the age of 16 years. JIA is classified into 6 mutually exclusive categories by the International League of Associations for Rheumatology (ILAR) criteria (1); a seventh category, "undifferentiated," is for children fulfilling criteria for more than 1 category. Patients categorized as having extended oligoarticular JIA or polyarticular JIA are accepted as the pediatric extensions of rheumatoid arthritis for US Food and Drug Administration (FDA) drug approval, and, likewise, those categorized as having juvenile psoriatic arthritis are the extensions of psoriatic arthritis in adults, respectively. Enthesitis-related arthritis (ERA) was the JIA category applied to children with spondyloarthritis (SpA), recognizing enthesitis as a

Partially supported by the NIH/National Institute of Arthritis and Musculoskeletal and Skin Diseases (grant P30-AR-076316-01; Dr. Brunner is principal investigator). Dr. Weiss' work was supported by the NIH/National Institute of Arthritis, Musculoskeletal, and Skin Diseases (grant R01-AR-074098). Dr. Colbert's work was supported by the National Institute of Arthritis, Musculoskeletal, and Skin Diseases Intramural Research Program (grant Z01-AR-041184).

[1]Pamela F. Weiss, MD, MSCE: Children's Hospital of Philadelphia, Perlman School of Medicine at the University of Pennsylvania School of Medicine, Philadelphia; [2]Robert C. Fuhlbrigge, MD, PhD: Children's Hospital Colorado, University of Colorado School of Medicine, Aurora; [3]Emily von Scheven, MD, MAS: University of California, San Francisco; [4]Daniel J. Lovell, MD, MPH, Hermine I. Brunner, MD, MSc, MBA: Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, Ohio; [5]Robert A. Colbert, MD, PhD: NIH/National Institute of Arthritis, Musculoskeletal, and Skin Diseases, Bethesda, Maryland.

Dr. Weiss has received consultant fees and honoraria from Pfizer and Lilly (less than $10,000 each). Dr. Lovell has received consultant fees and honoraria from AstraZeneca, Wyeth, Amgen, Abbott, Pfizer, F. Hoffmann-LaRoche, Novartis, UBC, Takeda, Janssen, GlaxoSmithKline, Boehringer Ingelheim, Celgene, Bristol Myers Squibb, AbbVie, and Forest Research (less than $10,000 each). Dr. Brunner has received consultant fees, speaking fees, and/or honoraria from Novartis, Roche, (more than 10,000 each), Ablynx, AbbVie, Astra Zeneca-Medimmune, Biogen, Boehringer, Bristol Myers Squibb, Celgene, Eli Lilly, EMD Serono, Genzyme, GlaxoSmithKline, F. Hoffmann-La Roche, Merck, Novartis, R-Pharm, and Sanofi (less than $10,000 each). Dr. Brunner is an employee of Cincinnati Children's Hospital, which has received contributions from the following companies in the past 3 years: Bristol Myers Squibb, Eli Lilly, GlaxoSmithKline, F. Hoffmann-La Roche, Janssen, Novartis, and Pfizer (more than $10,000 each). This funding has been reinvested for the research activities of the hospital in a fully independent manner, without any commitment to third parties.

Address correspondence to Pamela F. Weiss, MD, MSCE, Children's Hospital of Philadelphia, Perlman School of Medicine at the University of Pennsylvania School of Medicine, 2716 South Street, Room 11121, Philadelphia, PA 19146. Email: weisspa@email.chop.edu.

Submitted for publication September 21, 2020; accepted in revised form December 1, 2020.

defining characteristic. The prevalence of JIA is estimated at 20 to 45 per 100,000 children, of which 15–20% have ERA (2). The ILAR criteria for ERA are arthritis plus enthesitis, or arthritis or enthesitis plus at least 2 of the following: sacroiliac tenderness or inflammatory back pain, HLA–B27 positivity, first-degree relative with HLA–B27–associated disease, acute anterior uveitis, and arthritis in a male individual older than 6 years (1). The ERA criteria do not specifically account for inflammatory bowel disease arthropathy, ankylosing spondylitis, or reactive arthritis, which are clinical conditions included with adult SpA; children with these conditions may or may not meet the ERA criteria depending upon what disease features are present.

This review will summarize clinical, genetic, and pathophysiologic characteristics shared between children with ERA with axial involvement and adults with nonradiographic, and in some cases radiographic, axial SpA, as well as between children with ERA and peripheral disease manifestations and adults with peripheral SpA. Further, insights into validated outcome measures and therapy for ERA and adult SpA are provided.

Evidence that ERA and SpA are similar diseases based on biology

Much of our understanding of ERA pathogenesis is derived from studies of HLA–B27, a risk allele for adult and juvenile SpA. HLA–B27 is linked to activation of the interleukin 23 (IL-23/IL-17) axis through noncanonical mechanisms not involving antigen presentation to CD8+ T cells (3). A population of CD4/CD8-negative T cells in the entheses was shown to mediate IL-23–driven SpA (4,5). These cells were first identified in mice, and an equivalent type 3 innate-like lymphocyte has been described in human entheses (6). Juvenile SpA, like its adult counterpart, may also have an extra synovial basis of disease (7–9). The overlap in genetic susceptibility to ERA and SpA also includes endoplasmic reticulum aminopeptidase 1 (10), a peptidase specialized to produce peptides presented on class I major histocompatibility complex molecules, and a major risk gene for ankylosing spondylitis (11).

Subsets of adults with SpA and children with ERA have bowel inflammation (12). This has been studied more in adults (13), as access to intestinal tissue from children with subclinical inflammation is limited by ethical concerns. A number of different cell types have been implicated, and studies have emphasized the potential importance of bacterial dysbiosis, although cause and effect relationships remain unclear.

Similarity of clinical features

SpA develops on a continuum with a major peak of onset in young adulthood (14). Although sacroiliitis is well documented in ERA (15), the ILAR classification criteria focus on the importance of extra-axial manifestations, i.e., peripheral arthritis and enthesitis. Conversely, SpA classification in adults considers the presence of axial disease and peripheral disease (1). For reasons that remain unclear, common presenting features of juvenile-onset disease localize more to hips and peripheral joints (16), while adults experience predominantly inflammatory back pain (17).

Table 1 highlights the similarities and differences between the ERA classification criteria, the Assessment of SpondyloArthritis international Society (ASAS) criteria for nonradiographic axial SpA, and the ASAS criteria for peripheral SpA (18). The principal commonalities of children with ERA and axial arthritis, and adults with nonradiographic axial SpA, include enthesitis, arthritis, inflammatory back pain, anterior uveitis, HLA–B27 positivity, and family history of HLA–B27–associated disease. Magnetic resonance imaging (MRI) is increasingly used to confirm the presence of subchondral bone marrow edema around the sacroiliac joints; many patients have elevated C-reactive protein (CRP) levels, and the majority of patients experience some response to nonsteroidal antiinflammatory drugs (NSAIDs). One study reported 62% of ERA patients had axial disease at the time of diagnosis, and 63% of patients with only peripheral arthritis at the time of diagnosis developed axial involvement within 5 years (19). Figure 1 demonstrates that the inflammatory changes in the sacroiliac joints are indistinguishable between adults and children. In children, maturational changes may be mistaken for inflammatory changes by those with less experience evaluating the pediatric sacroiliac joint (Figure 2) (20). Unlike nonradiographic axial SpA, ERA is exclusive of psoriasis, while inflammatory bowel disease and reactive arthritis are largely ignored. Taken together, despite common clinical, laboratory, and radiographic features, differences in the classification between ERA and adult SpA can unduly complicate communication between providers, insurance carriers, and regulatory agencies (including the FDA), the transition from pediatric to adult care, and access to medications.

Similar outcomes in imaging outcome measures between pediatric and adult disease

For children and adults, evaluation for axial disease often includes MRI. Pediatric studies (15,21) utilize the ASAS MRI lesion definitions (22). Further, there are validated tools for assessment of axial joint inflammation and damage in adults and children. The Spondyloarthritis Research Consortium of Canada (SPARCC) sacroiliac joint inflammation score (SIS; range 0–72) considers site, extent, and severity of sacroiliac joint inflammation and has been validated for use in adults and children to capture response to therapy (23,24). A change in SIS score of 2.5 is considered clinically relevant in both populations (23). Damage in the sacroiliac joint can be quantified by the SPARCC sacroiliac structural score (SSS), which features 4 domains, including erosion (0–40), fat metaplasia (0–40), backfill (0–20), and ankylosis (0–20); there is no total score. The SPARCC SSS is validated in children and adults (24,25).

Table 1. Comparison of classification criteria used in children and adults*

	ERA	Nonradiographic axial SpA	Peripheral SpA
Criteria set	ILAR	ASAS	ASAS
Inclusion or entry criteria	Arthritis and enthesitis OR arthritis or enthesitis plus ≥2 supporting features	≥3 months of back pain starting before age 45 years AND sacroiliitis on imaging plus ≥1 SpA feature OR ≥2 SpA features	Arthritis OR enthesitis OR dactylitis† OR plus ≥1 group A feature OR ≥2 group B features
Supporting features			
Enthesitis	X	X	X (group B)
Arthritis	X	X	X (group B)
Dactylitis		X	X (group B)
Sacroiliac tenderness or IBP	X	X‡	X (group B)‡
Anterior uveitis	X	X	X (group A)
Psoriasis		X	X (group A)
IBD		X	X (group A)
Preceding infection§			X (group A)
Imaging			X¶
HLA–B27 positivity	X		
Family history	HLA–B27–associated disease in 1st-degree relative	1st- or 2nd-degree relative with SpA	Group B: 1st- or 2nd-degree relative with SpA
Markers of inflammation/ elevated C-reactive protein		X	
Therapeutic response to NSAIDs		X	

* ASAS = Assessment of SpondyloArthritis international Society; ERA = enthesitis-related arthritis; IBD = inflammatory bowel disease; IBP = inflammatory back pain; ILAR = International League Against Rheumatism; NSAIDs = nonsteroidal antiinflammatory drugs; SpA = spondyloarthritis.
† Arthritis, enthesitis, or dactylitis must be present at the time of evaluation.
‡ Inflammatory back pain only.
§ Urethritis/cervicitis or diarrhea within 1 month prior to onset of symptoms.
¶ Sacroiliitis on imaging (bilateral grade 2 to 4 or unilateral grade 3 to 4 on radiographs or active sacroiliitis on magnetic resonance imaging).

Similarity of response to therapy in children and adults

Algorithms to treat ERA with axial arthritis and nonradiographic axial SpA are similar, as is evidenced by published American College of Rheumatology (ACR) treatment recommendations for both conditions (26,27). The recommended initial treatment of both is NSAIDs, followed by tumor necrosis factor (TNF) inhibitors if NSAIDs are not tolerated or ineffective. Numerous trials in adults have shown that conventional disease-modifying antirheumatic drugs (cDMARDs) do not improve axial disease (28). Although similar trials have not been conducted in ERA, ACR pediatric treatment recommendations strongly advise against methotrexate monotherapy and moving directly to anti-TNF therapy, based on extrapolation from the adult studies and clinical experience (26).

Treatment algorithms for children with ERA and adults with SpA and peripheral disease depend upon the number of affected joints and risk factors present. For peripheral disease affecting fewer than 5 joints, intraarticular joint injections with or without NSAIDs are considered first-line therapy (26). For peripheral disease affecting 5 or more joints, cDMARDs including methotrexate are first-line therapy and may be used with TNF inhibitors, if joint damage is present or if there is involvement of high-risk joints (cervical spine, wrist, hip) (26). While there are no formal guidelines for treatment of adults with peripheral SpA, treatment algorithms are analogous to those used in children with ERA and inflammation of peripheral joints.

Response to therapies is also similar in ERA and adults with SpA. Randomized placebo-controlled clinical trials in adults (29–31) and data from children (32,33) show the efficacy of TNF inhibitors for peripheral arthritis, enthesitis, and axial arthritis. However, as many as half of adults with axial disease are unable to achieve remission with TNF inhibitors, with 15% of adults with axial SpA failing to show any improvement with TNF inhibitors (34). Similarly, 33% of children with ERA treated with TNF inhibitors and NSAIDs lack response to therapy (19). In 1 study, only 24% of children with ERA achieved inactive disease during the initial 12 months of treatment (35), and fewer than 20% achieved remission within 5 years (36). Additionally, physical function limitations and moderate chronic pain are more prevalent with ERA than with other JIA

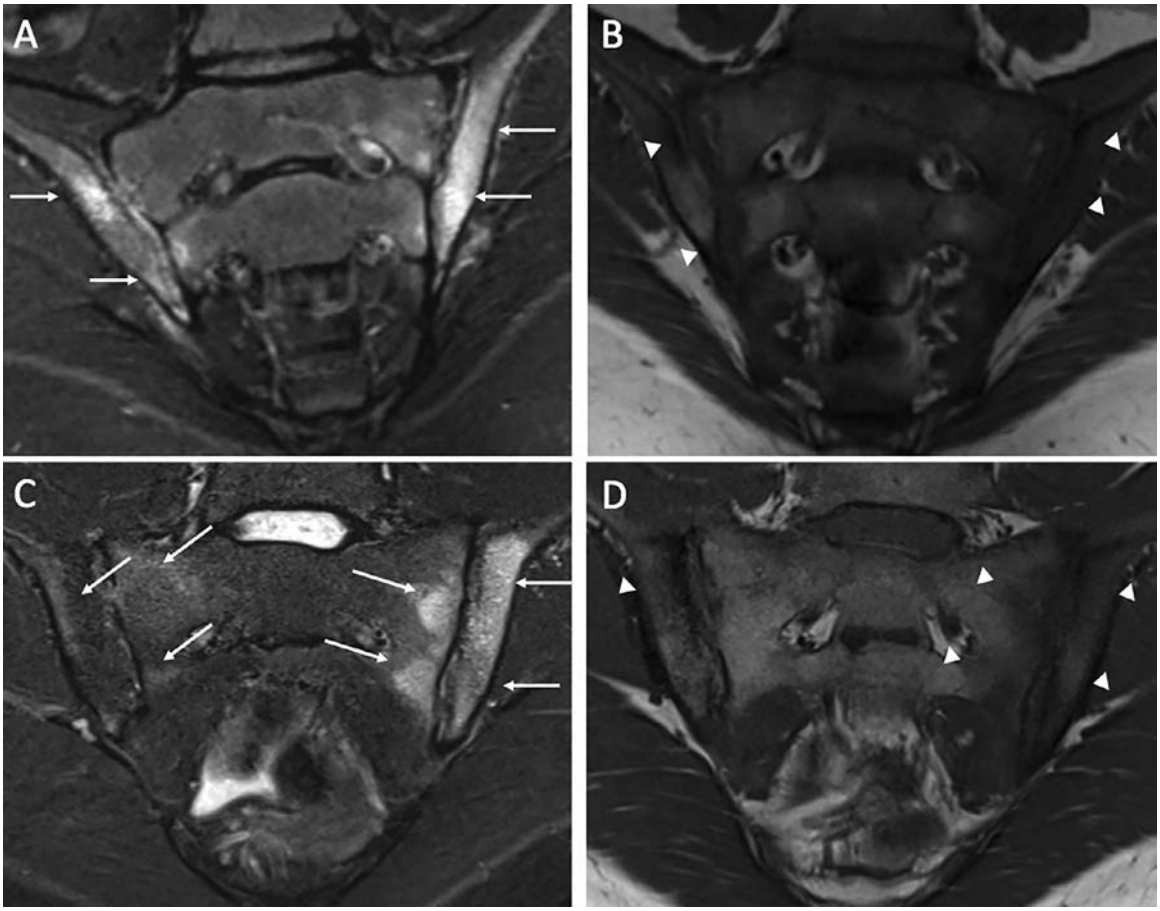


Figure 1. Coronal oblique STIR (**A** and **C**) and coronal oblique T1-weighted (**B** and **D**) images of the sacroiliac joints of a 7-year-old, HLA–B27–positive female patient (**A** and **B**) and a 20-year-old HLA–B27–positive male patient (**C** and **D**). There is active sacroiliitis with periarticular bone marrow edema within the iliac aspect of both joints as demonstrated by increased signal intensity on STIR imaging (**A**; arrows) and decreased signal intensity on T1-weighted imaging (**B**; arrowhead). There is active sacroiliitis with periarticular bone marrow edema within the sacral and iliac bones, much more intensely on the left than the right, as demonstrated by increased signal intensity on STIR imaging (**C**; arrows) and decreased signal intensity on T1-weighted imaging (**D**; arrowheads).

categories (37). Thus, achieving inactive disease status or clinical remission is difficult for children with ERA, and many continue to have disease activity despite off-label use of existing therapies.

Regulatory environment for medication approval

In the US, the FDA is the federal agency charged with overseeing drug manufacturing, labeling, advertisement, and safety of medications and biological products. The Best Pharmaceuticals for Children Act (BPCA) (38,39) and the Pediatric Research Equity Act (PREA) (40) govern medication approval for children in the US. While the BPCA encourages drug companies to test their products in children, the PREA necessitates the study of new drugs and biologic DMARDs (bDMARDs) in children if there is a pediatric disease similar to the non-orphan adult disease, and if it is likely that the new agent will be used in children (41).

The FDA gives automatic full waivers from conducting studies in children under the PREA if the pediatric equivalent of the adult disease "rarely or never occurs in pediatrics." This is because studies in children would be highly impractical. ERA is common, comprising 15–20% of JIA cases in the US. Indeed, ERA is at least as common as systemic JIA, for which clinical trials have been successfully completed (42). However, due to the differences in classification criteria outlined above, the FDA has granted automatic full waivers of studies in children for new medications for "axial spondyloarthropathies including ankylosing spondylitis" up until July 2020. Thus, although current JIA treatment guidelines recommend the use of bDMARDs as part of the early treatment for patients with ERA (43), none of the FDA-approved therapies for peripheral SpA or nonradiographic axial SpA (certolizumab pegol [2019], ixekizumab [2020], and secukinumab [2020]) have been studied or are labeled for use in children with ERA.

Recommendations to improve treatment options for children with ERA

Evidence of uncontrolled disease despite a trial of NSAIDs could identify children with ERA who require advanced therapies

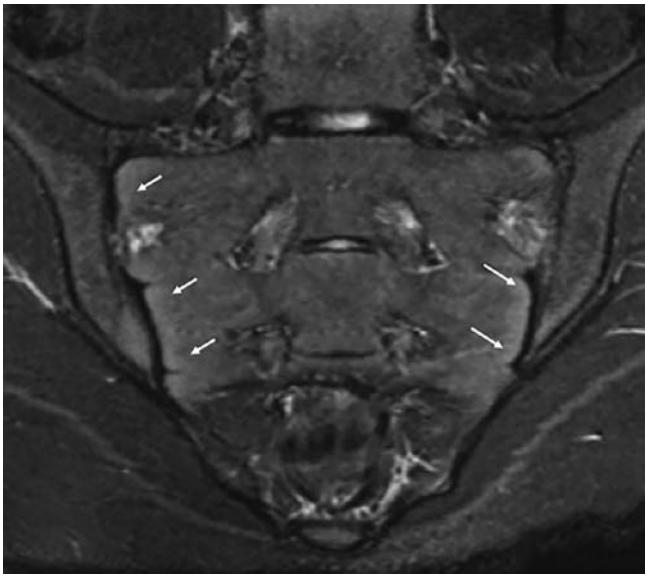


Figure 2. Coronal oblique STIR image of the sacroiliac joints of a 15-year-old female patient, demonstrating metaphyseal-equivalent hyperintense signal (arrows), a normal variant, that could be mistaken for subchondral inflammation by less experienced reviewers.

and may participate in clinical trials, irrespective of the presence of axial or peripheral involvement. Clinical trials in ERA should capture and evaluate response of axial and peripheral disease separately. This may be done via subanalysis of axial and peripheral disease response. Similar to trials of nonradiographic axial SpA (44), eligibility criteria for children with ERA and axial features could include the presence of some of the following disease features: active inflammatory sacroiliitis based on typical MRI changes according to ASA/Outcome Measures in Rheumatology criteria; elevated CRP level; and inadequate response or intolerance to NSAIDs. Because axial disease does not respond to treatment with cDMARDs and ~40% of children are HLA–B27 negative (45), absence of these features should not be exclusionary. Presence of acute uveitis should also not be exclusionary, as this is generally treatable with topical medications. The FDA grants partial waivers for study conduct in certain pediatric age groups. With respect to ERA, a partial waiver for studies of children younger than age 6 years seems sensible as disease onset prior to this age is unusual.

Similar to trials of adults with peripheral SpA (46), active disease in children with ERA and peripheral disease can be defined by a combination of the following: persistence of active arthritis in 1 or more joints, active enthesitis, and/or dactylitis despite NSAID exposure; evidence of systemic inflammation; physician global assessment of disease activity reflective of active disease; and patient global assessment of pain indicating ongoing ERA-related pain. Efficacy could be assessed using clinically meaningful change in validated composite disease activity scores or patient-reported outcomes. Given the challenges of entheseal assessment in children (47) and the lack of a validated pediatric enthesitis index, we caution against the use of enthesitis as a primary outcome.

The FDA encourages extrapolation of effectiveness from adult to pediatric populations when appropriate. With regard to ERA, extrapolation of effectiveness of a medication to control signs and symptoms should assume that an appropriate pediatric dose can be established either through achieving a similar exposure in children as the proven therapeutic exposure in adults, or by using an appropriate pharmacodynamic or clinical end point to achieve the targeted effect (48). Conversely, the ability to extrapolate safety from adults with SpA to children with ERA is limited, and special consideration should be made to utilize trial designs that allow for the assessment of unique pediatric toxicities, including the potential impact of the drug on growth and development (48).

To ensure the most appropriate dosing and confirm anticipated efficacy of a medication to be used in children with ERA, sufficient data need to be available. As is detailed in the Center for Drug Evaluation and Research document (49), the types of studies needed will depend on what is already known about pediatric dosing (pharmacokinetics) and whether there are differences between pediatric and adult pharmacodynamics, and therefore potential differences in efficacy. Study needs will have to be determined on a case-by-case basis. Depending on the available knowledge base, no additional studies may be required, or a randomized double-blinded study might be needed.

In summary, despite FDA-approved treatments for adult axial and peripheral SpA, there remains an unmet need for effective medications for children with spondyloarthropathies. Considering the similarities between adult SpA and ERA in terms of etiology, genetics, pathogenesis, and clinical manifestations (50), it is evident that medications approved for axial or peripheral SpA should be studied in children with ERA involving axial or peripheral joints, respectively, with the intent to achieve labeling for use in children. Considering the current lack of effective therapies for ERA, the FDA should consider requiring pediatric studies for medications that have already been approved for the treatment of adults with SpA. The design of trials in ERA will depend on the amount of prior knowledge about a given drug and could entail full and partial extrapolation strategies in support of achieving an indication for the treatment of ERA.

AUTHOR CONTRIBUTIONS

All authors drafted the article or revised it critically for important intellectual content, and approved the final version to be published.

REFERENCES

1. Petty RE, Southwood TR, Manners P, Baum J, Glass DN, Goldenberg J, et al. International League of Associations for Rheumatology classification of juvenile idiopathic arthritis: second revision, Edmonton, 2001. J Rheumatol 2004;31:390–2.

2. Juvenile idiopathic arthritis. In: Petty RE, Laxer RM, Lindsley CB, Wedderburn LR, editors. Textbook of pediatric rheumatology. 7th ed. Philadelphia: Elsevier; 2016. p. 188–204.
3. Taurog JD, Chhabra A, Colbert RA. Ankylosing spondylitis and axial spondyloarthritis. N Engl J Med 2016;374:2563–74.
4. Colbert RA, Navid F, Gill T. The role of HLA-B*27 in spondyloarthritis. Best Pract Res Clin Rheumatol 2017;31:797–815.
5. Sherlock JP, Joyce-Shaikh B, Turner SP, Chao CC, Sathe M, Grein J, et al. IL-23 induces spondyloarthropathy by acting on ROR-gammat+ CD3+CD4-CD8- entheseal resident T cells. Nat Med 2012;18: 1069–76.
6. Cuthbert RJ, Fragkakis EM, Dunsmuir R, Li Z, Coles M, Marzo-Ortega H, et al. Brief report: group 3 innate lymphoid cells in human enthesis. Arthritis Rheumatol 2017;69:1816–22.
7. Jacques P, McGonagle D. The role of mechanical stress in the pathogenesis of spondyloarthritis and how to combat it. Best Pract Res Clin Rheumatol 2014;28:703–10.
8. Lee EY, Sundel RP, Kim S, Zurakowski D, Kleinman PK. MRI findings of juvenile psoriatic arthritis. Skeletal Radiol 2008;37:987–96.
9. Tuttle KS, Vargas SO, Callahan MJ, Bae DS, Nigrovic PA. Enthesitis as a component of dactylitis in psoriatic juvenile idiopathic arthritis: histology of an established clinical entity. Pediatr Rheumatol Online J 2015;13:7.
10. Hinks A, Martin P, Flynn E, Eyre S, Packham J, Childhood Arthritis Prospective Study, et al. Subtype specific genetic associations for juvenile idiopathic arthritis: ERAP1 with the enthesitis related arthritis subtype and IL23R with juvenile psoriatic arthritis. Arthritis Res Ther 2011;13:R12.
11. Evans DM, Spencer CC, Pointon JJ, Su Z, Harvey D, Kochan G, et al. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. Nat Genet 2011;43:761–7.
12. Mielants H, Veys EM, Cuvelier C, De Vos M, Goemaere S, Maertens M, et al. Gut inflammation in children with late onset pauciarticular juvenile chronic arthritis and evolution to adult spondyloarthropathy: a prospective study. J Rheumatol 1993;20:1567–72.
13. Rizzo A, Ferrante A, Guggino G, Ciccia F. Gut inflammation in spondyloarthritis. Best Pract Res Clin Rheumatol 2017;31:863–76.
14. Colbert RA. Classification of juvenile spondyloarthritis: enthesitis-related arthritis and beyond. Nat Rev Rheumatol 2010;6:477–85.
15. Weiss PF, Xiao R, Biko DM, Chauvin NA. Assessment of sacroiliitis at diagnosis of juvenile spondyloarthritis by radiography, magnetic resonance imaging, and clinical examination. Arthritis Care Res (Hoboken) 2016;68:187–94.
16. Burgos-Vargas R, Vazquez-Mellado J. The early clinical recognition of juvenile-onset ankylosing spondylitis and its differentiation from juvenile rheumatoid arthritis. Arthritis Rheum 1995;38:835–44.
17. Riley MJ, Ansell BM, Bywaters EG. Radiological manifestations of ankylosing spondylitis according to age at onset. Ann Rheum Dis 1971;30:138–48.
18. Rudwaleit M, van der Heijde D, Landewe R, Listing J, Akkoc N, Brandt J, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. Ann Rheum Dis 2009;68:777–83.
19. Goirand M, Breton S, Chevallier F, Duong NP, Uettwiller F, Melki I, et al. Clinical features of children with enthesitis-related juvenile idiopathic arthritis/juvenile spondyloarthritis followed in a French tertiary care pediatric rheumatology centre. Pediatr Rheumatol Online J 2018;16:21.
20. Weiss PF, Brandon TG, Bohnsack J, Heshin-Bekenstein M, Francavilla ML, Jaremko JL, et al. Variability in magnetic resonance imaging interpretation of the pediatric sacroiliac joint. Arthritis Care Res (Hoboken) 2021;73:841–8.
21. Bray TJ, Lopes A, Fisher C, Ciurtin C, Sen D, Hall-Craggs MA. Sacroiliac joint ankylosis in young spondyloarthritis patients receiving biologic therapy: observation of serial magnetic resonance imaging scans. Arthritis Rheumatol 2019;71:594–8.
22. Rudwaleit M, Haibel H, Baraliakos X, Listing J, Marker-Hermann E, Zeidler H, et al. The early disease stage in axial spondylarthritis: results from the German Spondyloarthritis Inception Cohort. Arthritis Rheum 2009;60:717–27.
23. Maksymowych WP, Lambert RG, Brown LS, Pangan AL. Defining the minimally important change for the SpondyloArthritis Research Consortium of Canada spine and sacroiliac joint magnetic resonance imaging indices for ankylosing spondylitis. J Rheumatol 2012;39: 1666–74.
24. Weiss PF, Maksymowych WP, Xiao R, Biko DM, Francavilla ML, Lambert RG, et al. Spondyloarthritis Research Consortium of Canada sacroiliac joint inflammation and structural scores: change score reliability and recalibration utility in children. Arthritis Res Ther 2020;22:58.
25. Weiss PF, Maksymowych WP, Lambert RG, Jaremko JL, Biko DM, Paschke J, et al. Feasibility and reliability of the Spondyloarthritis Research Consortium of Canada sacroiliac joint structural score in children. J Rheumatol 2018;45:1411–7.
26. Ringold S, Angeles-Han ST, Beukelman T, Lovell D, Cuello CA, Becker ML, 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.
27. Ward MM, Deodhar A, Gensler LS, Dubreuil M, Yu D, Khan MA, et al. 2019 update of the American College of Rheumatology/Spondylitis Association of America/Spondyloarthritis Research and Treatment Network recommendations for the treatment of ankylosing spondylitis and nonradiographic axial spondyloarthritis. Arthritis Care Res (Hoboken) 2019;71:1285–99.
28. Haibel H, Brandt HC, Song IH, Brandt A, Listing J, Rudwaleit M, et al. No efficacy of subcutaneous methotrexate in active ankylosing spondylitis: a 16-week open-label trial. Ann Rheum Dis 2007;66:419–21.
29. Molto A, Sieper J. Peripheral spondyloarthritis: concept, diagnosis and treatment. Best Pract Res Clin Rheumatol 2018;32:357–68.
30. Paramarta JE, De Rycke L, Heijda TF, Ambarus CA, Vos K, Dinant HJ, et al. Efficacy and safety of adalimumab for the treatment of peripheral arthritis in spondyloarthritis patients without ankylosing spondylitis or psoriatic arthritis. Ann Rheum Dis 2013;72:1793–9.
31. Sieper J, van der Heijde D, Dougados M, Mease PJ, Maksymowych WP, Brown MA, et al. Efficacy and safety of adalimumab in patients with non-radiographic axial spondyloarthritis: results of a randomised placebo-controlled trial (ABILITY-1). Ann Rheum Dis 2013;72: 815–22.
32. Burgos-Vargas R, Tse SM, Horneff G, Pangan AL, Kalabic J, Goss S, et al. A randomized, double-blind, placebo-controlled multicenter study of adalimumab in pediatric patients with enthesitis-related arthritis. Arthritis Care Res (Hoboken) 2015;67:1503–12.
33. Horneff G, Fitter S, Foeldvari I, Minden K, Kuemmerle-Deschner J, Tzaribacev N, et al. Double-blind, placebo-controlled randomized trial with adalimumab for treatment of juvenile onset ankylosing spondylitis (JoAS): significant short-term improvement. Arthritis Res Ther 2012; 14:R230.
34. Deodhar A, Strand V, Conaghan PG, Sullivan E, Blackburn S, Tian H, et al. Unmet needs in ankylosing spondylitis patients receiving tumour necrosis factor inhibitor therapy; results from a large multinational real-world study. BMC Rheumatol 2020;4:19.
35. Donnithorne KJ, Cron RQ, Beukelman T. Attainment of inactive disease status following initiation of TNF-alpha inhibitor therapy for juvenile idiopathic arthritis: enthesitis-related arthritis predicts persistent active disease. J Rheumatol 2011;38:2675–81.

36. Flato B, Aasland A, Vinje O, Forre O. Outcome and predictive factors in juvenile rheumatoid arthritis and juvenile spondyloarthropathy. J Rheumatol 1998;25:366–75.
37. Taxter AJ, Wileyto EP, Behrens EM, Weiss PF. Patient-reported outcomes across categories of juvenile idiopathic arthritis. J Rheumatol 2015;42:1914–21.
38. Rumsey DG, Guzman J, Rosenberg AM, Huber AM, Scuccimarri R, Shiff NJ, et al. Characteristics and course of enthesitis in a juvenile idiopathic arthritis inception cohort. Arthritis Care Res (Hoboken) 2018; 70:303–8.
39. Best Pharmaceuticals for Children Act of 2007. Food and Drug Administration Amendments Act (FDAAA). Title V. 2018. URL: https://www.fda.gov/science-research/pediatrics/best-pharmaceuticals-children-act-and-pediatric-research-equity-act.
40. Pediatric Research Equity Act. PUBLIC LAW 108–155. 2003. URL: https://www.congress.gov/108/plaws/publ155/PLAW-108publ155.pdf.
41. Field MJ, Boat TF, Institute of Medicine (US), National Research Council (US). Safe and effective medicines for children: pediatric studies conducted under the Best Pharmaceuticals for Children Act and the Pediatric Research Equity Act. Washington: National Academies Press; 2012.
42. Ruperto N, Brunner HI, Quartier P, Constantin T, Wulffraat N, Horneff G, et al. Two randomized trials of canakinumab in systemic juvenile idiopathic arthritis. N Engl J Med 2012;367:2396–406.
43. Ringold S, Angeles-Han ST, Beukelman T, Lovell D, Cuello CA, Becker ML, 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 Rheumatol 2019;71:846–63.
44. Deodhar A, van der Heijde D, Gensler LS, Kim TH, Maksymowych WP, Ostergaard M, et al. Ixekizumab for patients with non-radiographic axial spondyloarthritis (COAST-X): a randomised, placebo-controlled trial. Lancet 2020;395:53–64.
45. Gmuca S, Xiao R, Brandon TG, Pagnini I, Wright TB, Beukelman T, et al. Multicenter inception cohort of enthesitis-related arthritis: variation in disease characteristics and treatment approaches. Arthritis Res Ther 2017;19:84.
46. Carron P, Varkas G, Cypers H, van Praet L, Elewaut D, van den Bosch F, et al. Anti-TNF-induced remission in very early peripheral spondyloarthritis: the CRESPA study. Ann Rheum Dis 2017;76: 1389–95.
47. Weiss PF, Chauvin NA, Klink AJ, Localio R, Feudtner C, Jaramillo D, et al. Detection of enthesitis in children with enthesitis-related arthritis: dolorimetry compared to ultrasonography. Arthritis Rheumatol 2014; 66:218–27.
48. US Food and Drug Administration. Pediatric study plans: content of and process for submitting initial pediatric study plans and amended initial pediatric study plans: guidance for industry. Draft Guidance Revision 2016. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/pediatric-study-plans-content-and-process-submitting-initial-pediatric-study-plans-and-amended.
49. Programs affecting safety and innovation in pediatric therapies: hearing before the Subcommittee on Health of the Committee on Energy and Commerce, House of Representatives, One Hundred Tenth Congress, First Session, May 22, 2007.
50. Brunner HI, Schanberg LE, Kimura Y, Dennos A, Co DO, Colbert RA, et al. New medications are needed for children with juvenile idiopathic arthritis. Arthritis Rheumatol 2020;72:1945–51.

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1065–1069
DOI 10.1002/acr.24535

American College of Rheumatology
Empowering Rheumatology Professionals

BRIEF REPORT

Association of p155/140 Autoantibody With Loss of Nailfold Capillaries but not Generalized Lipodystrophy: A Study of Ninety-Six Children With Juvenile Dermatomyositis

Amer Khojah,[1] Victoria Liu,[2] Sonia I. Savani,[3] Gabrielle Morgan,[2] Richard Shore,[4] Jackie Bellm,[4] and Lauren M. Pachman[1]

Objective. Myositis-specific antibodies (MSAs) facilitate grouping children with juvenile dermatomyositis (DM) into distinct phenotypes. The first aim of this study was to investigate the link between anti-p155/140 and lipodystrophy as determined by dual x-ray absorptiometry (DXA) assessment of fat distribution. The second aim was to examine the relationship between anti-p155/140 and damage to the nailfold capillary system.

Methods. Children with juvenile DM followed for a minimum of 5 years were included. The study population was divided into 3 groups (anti-p155/140, other MSA, and MSA negative). Lipodystrophy was assessed by physician assessment and DXA fat distribution (trunk-to-leg fat ratio). Documentation of nailfold capillary end row loops (ERLs) was obtained at diagnosis.

Results. A total of 96 subjects (44% anti-p155/140, 23% other MSA, 33% MSA negative) were included. There was no significant difference in age, disease activity scores, or lipodystrophy between the 3 groups. The trunk-to-leg fat ratios were similar among the 3 groups at different time points. However, the anti-p155/140 group had significantly decreased ERL counts (P = 0.006) at baseline as well as a prolonged duration of untreated disease at diagnosis (P = 0.027). Also, the anti-p155/140 group had fewer patients with a monophasic disease course than the other 2 groups (P = 0.008).

Conclusion. Generalized lipodystrophy frequency was equivalent in all 3 groups based on physician assessments and trunk-to-leg fat ratios. The anti-p155/140 group had a greater loss of ERLs, suggesting that this MSA may impact the vascular component of juvenile DM.

INTRODUCTION

Juvenile dermatomyositis (DM) is a multisystemic autoimmune vasculopathy characterized by skin and muscle inflammation (1). One manifestation of the small vessel destruction associated with juvenile DM is the loss of nailfold capillary end row loops (ERLs) (2), which can be evaluated by nailfold capillaroscopy (NFC), a noninvasive imaging technique. More severe nailfold capillary changes have been associated with a more aggressive disease course (3), and others have previously suggested that nailfold capillary density may be used as a marker of skin (4) and muscle disease activity in juvenile DM (2). The role of myositis-specific antibodies (MSAs) in DM is well established. More recently, there has been an effort to subcategorize juvenile DM into multiple distinct clinical phenotypes utilizing MSAs (1), which may be used as biomarkers to anticipate disease outcomes in juvenile DM (5). Of these MSAs, anti-p155/140 (or anti-TIF1γ antibody) is the most common in children and appears to be associated with more severe skin disease with cutaneous ulcerations (5). Anti-p155/140 was also reported to be associated with generalized lipodystrophy in case reports and in one small retrospective study with a limited sample size (5,6).

Lipodystrophy, a condition in which individuals lose subcutaneous fat and possibly gain visceral fat, is a serious

[1]Amer Khojah, MD (current address: Umm Al-Qura University, Mecca, Saudi Arabia), Lauren M. Pachman, MD: Ann & Robert H. Lurie Children's Hospital of Chicago and Cure JM Center of Excellence, Stanley Manne Research Center, Chicago, Illinois; [2]Victoria Liu, Gabrielle Morgan, MA: Cure JM Center of Excellence, Stanley Manne Research Center, Chicago, Illinois; [3]Sonia I. Savani, MD: Medical University of South Carolina, Charleston; [4]Richard Shore, MD, Jackie Bellm, NMT: Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois.

No potential conflicts of interest relevant to this article were reported.

Address correspondence to Amer Khojah, MD, Pediatric Department, College of Medicine, Umm Al-Qura University, 101 Road, Al-Abdeyah, Mecca 24381, Saudi Arabia. Email: khojah.a@gmail.com.

Submitted for publication July 9, 2020; accepted in revised form December 3, 2020.

SIGNIFICANCE & INNOVATIONS

- To our knowledge, this the first study to show an association of loss of nailfold capillary end row loops with anti-p155/140 myositis-specific autoantibodies (MSAs) in children with juvenile dermatomyositis.
- The frequency of generalized lipodystrophy as measured by physician evaluation and dual x-ray absorptiometry assessment of fat distribution was the same in patients with the anti-p155/140 compared to the other groups (other MSA and MSA negative).

complication of juvenile DM and is of clinical significance given its medical and psychological influence on children with juvenile DM. Lipodystrophy additionally predisposes patients to serious metabolic complications such as insulin resistance, diabetes mellitus, and hypertriglyceridemia, each of which can lead to an increased risk of cardiovascular disease, the leading cause of mortality in myositis patients (7). Furthermore, studies of HIV patients with lipodystrophy showed a significant impact of lipodystrophy on self-esteem and body image (8). It is estimated that between 10% and 40% of juvenile DM patients have lipodystrophy (6,7,9). This wide range reflects the difficulty in making the diagnosis by physical examination, which may lead to underestimation of the disease. Various imaging technologies, such as dual x-ray absorptiometry (DXA), computed tomography (CT), and magnetic resonance imaging (MRI), have been used to aid in the diagnosis by providing a far more objective assessment of fat distribution. However, the cost has been a limiting factor in the widespread use of these methods. DXA has been used widely to monitor bone density in patients with an increased risk of osteoporosis. Given the increased risk of pathologic fractures in juvenile DM patients from active inflammation, decrease mobility, and the chronic use of steroids, routine monitoring of bone mineral density by DXA is recommended (10). Measurement of percent body fat and regional fat distribution can also be obtained from the same whole-body DXA images with no additional radiation exposure.

The first aim of this study was to characterize the phenotype of anti-p155/140 positive juvenile DM by first investigating the relationship of anti-p155/140 and lipodystrophy as measured by subjective physician assessment and objective assessment of body fat distribution by DXA. The second aim was to examine vascular compromise by exploring the relationship between anti-p155/140 and nailfold changes.

PATIENTS AND METHODS

Patient population. This report is of a retrospective chart review study, conducted at the CureJM Center of Excellence in Juvenile Myositis Care and Research at the Ann & Robert H. Lurie Children's Hospital of Chicago between 2000 and 2017 (IRB number 2012-14858). We included all subjects with a definite or probable juvenile DM diagnosis based on the criteria of Bohan and Peter, who had access to at least 5 years of follow-up data, multiple DXA results, and studies of nailfold capillary ERLs during the study period.

DXA. DXA was performed using a Lunar iDXA bone densitometer (GE Healthcare), and data were analyzed by Encore software, version 16. To study the natural history of body fat changes, we divided the DXA data into 3 groups based on the duration of time between the date of first medication use and the time of assessment (visit 1 >1.5 years; visit 3 = 1.51–3.49 years; visit 5 = 3.5–5.0 years). Generalized lipodystrophy was evaluated using the clinical assessment of the provider during the visit and by measuring the trunk-to-leg fat ratio using the DXA results. Most patients with generalized lipodystrophy have atrophy of peripheral fat and hypertrophy of visceral fat, leading to the increased trunk-to-leg fat ratio. This ratio is widely used in the literature on HIV (11).

MSAs. MSAs were measured via immunoprecipitation and immunodiffusion at the Oklahoma Medical Research Foundation. We divided the study population into 3 groups based on their MSAs (anti-p155/140 antibody, other MSA, and MSA negative). Patients with juvenile DM with >1 MSA or overlap syndrome (such as subjects with positive anti-PM/Scl, anti–U1 RNP, or anti–U2 RNP) were excluded from the analysis.

Nailfold capillary ERLs. After obtaining age-appropriate written informed consent/assent (IRB number 2010-14117), images of each of the 8 digits, excluding thumbs, were captured and recorded for analysis later. Earlier photos were taken using freeze-frame video microscopy (12×) and printed in real-time on photo paper, where they were analyzed. Digital images were obtained via a Dermlite II ProHR dermatoscope (18×) with a Nikkon camera adapter, and analysis was performed utilizing Photoshop CS5 (Adobe). After standardizing for magnification, ERL/mm was quantified by counting the number of end row capillary loops per 3-mm section on each of the 8 fingers, dividing this by 3, and transforming this count into ERL/mm. A mean ERL/mm of the number of fingers analyzed was then obtained for each patient.

Disease activity and disease course assessment. We also collected disease activity markers on presentation including the Disease Activity Score (DAS) for juvenile DM skin subscale (DAS-S) scores, muscle subscale (DAS-M) scores, and total (DAS-T) scores (12), Childhood Myositis Assessment Scale (CMAS) scores, and neopterin levels. The disease course was defined as either monophasic (having achieved remission within any length of time without subsequent flares requiring reinitiation of treatment), polycyclic (remission achieved within any length of

time with at least 1 flare requiring treatment), and chronic (those with at least 36 months of data and who remained on treatment).

Statistical analysis. SPSS Statistics software was used to perform one-way analysis of variance, and chi-square testing was used to compare the baseline characteristics, disease activity markers, ERL counts, and fat distribution (trunk-to-leg fat ratio) among the 3 groups. The figures were generated using Prism 8 software (GraphPad).

RESULTS

A total of 96 children with juvenile DM (78% female, 70% White) were included. These subjects were divided into 3 groups based on their MSA: 42% anti-p155/140 antibody; 23% other MSA; and 35% MSA negative. There was no significant difference between the groups in terms of sex, race, or age at onset of symptoms (Table 1). The duration of untreated disease in the anti-p155/140 group was twice as long as in the other 2 groups (P = 0.027). On the initial assessment, the DAS-M score was slightly lower in the anti-p155/140 group, with a mean value of 4.3 versus 5.8 and 5.9 in the other MSAs and MSA negative groups, respectively (P = 0.049). On the other hand, neopterin level, CMAS scores, DAS-S scores, and DAS-T scores were not statistically different in the 3 groups. The anti-p155/140 group had a lower nailfold capillary ERL count (P = 0.006) at the initial assessment (Figure 1). Post hoc comparisons using Tukey's honest significant difference test indicate that the significant difference was between the anti-p155/140 group and MSA negative group. This difference

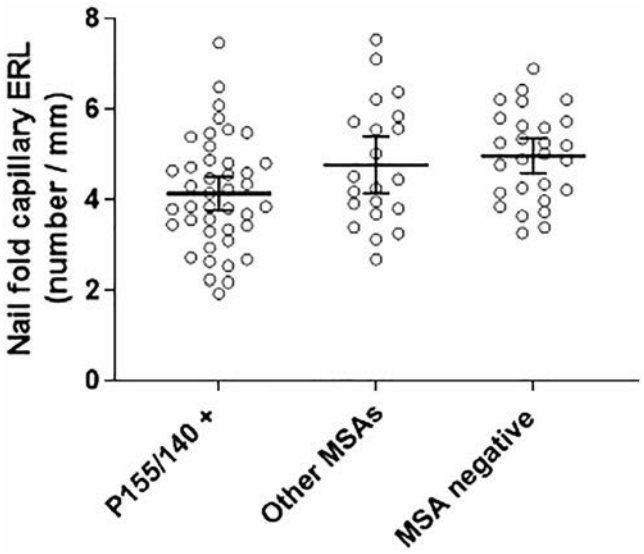


Figure 1. Nail fold capillary end row loops (ERLs) at the first visit. Each circle represents an ERL measurement from a patient. Horizontal lines indicate the mean. Error bars indicate the 95% confidence interval. MSA = myositis-specific autoantibody.

was significant (adjusted P = 0.01) even after controlling the duration of untreated disease. Of note, there was no significant correlation between the duration of untreated disease and ERL count. Also, the anti-p155/140 group had fewer patients with monophasic disease course than the other 2 groups (P = 0.008) (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24535). The percentage of total body fat (see Supplementary Table 2,

Table 1. Baseline demographic characteristics and disease activity markers of the study cohort*

Characteristic	Anti-p155/140 positive	Other MSAs (MJ, Mi-2, MDA5)	MSA negative	*P*
No. of subjects	43	21	32	
Sex, no.				0.482
Female	34	18	23	
Male	9	3	9	
Race, no.				0.298
White	34	13	20	
Hispanic	7	4	6	
African American	0	3	4	
Others	2	1	2	
Age at onset of symptoms, years	6.2 ± 3.5	6.8 ± 2.9	6.5 ± 3.5	0.815
Duration of untreated disease, months	9.5 ± 10.0	4.3 ± 5.0	5.5 ± 7.1	0.027†
Treatment status on conical presentation, no. of untreated/treated‡	20/23	6/15	11/21	0.322
DAS for juvenile DM total score‡	10.2 ± 3.6	11.3 ± 3.3	11.3 ± 4.6	0.382
DAS for juvenile DM skin score‡	5.9 ± 1.8	5.5 ± 1.3	5.4 ± 2.3	0.511
DAS for juvenile DM muscle weakness score‡	4.3 ± 2.8	5.8 ± 3.2	5.9 ± 3.2	0.049†
CMAS score‡	37.9 ± 13.5	32.3 ± 14.7	33.5 ± 14.9	0.343
ERLs‡	4.1 ± 1.2	4.6 ± 1.4	5.0 ± 1.1	0.006†
Neopterin level‡	17.9 ± 12.6	13.0 ± 5.9	17.5 ± 8.7	0.217

* Values are the mean ± SD unless indicated otherwise. CMAS = Childhood Myositis Assessment Scale; DAS = Disease Activity Score; DM = dermatomyositis; ERLs = end row loops; MSA = myositis-specific autoantibody.
† Significant.
‡ On initial clinical presentation.

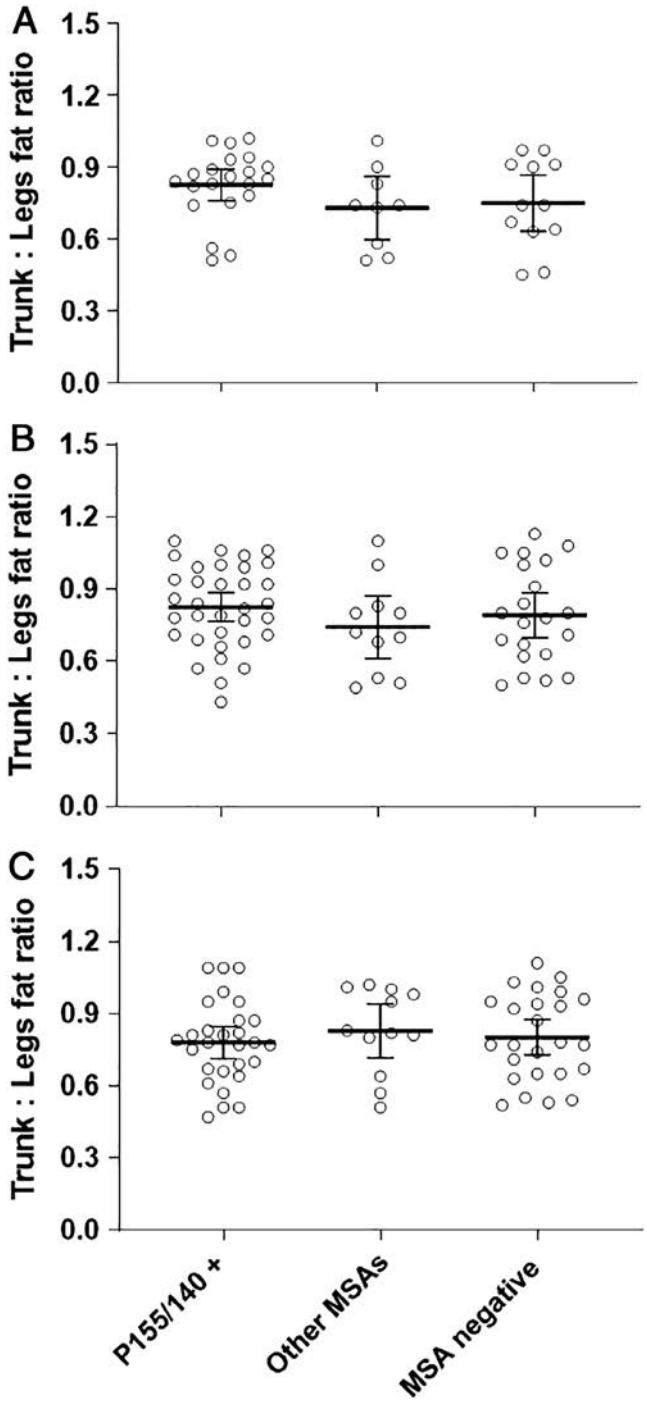


Figure 2. Fat distribution 0–18 months from the first visit (**A**), 19–41 months from the first visit (**B**), and 42–60 months from the first visit (**C**). Each circle represents the trunk:legs fat ratio from a single patient. Horizontal lines indicate the mean. Error bars indicate the 95% confidence interval. MSA = myositis-specific autoantibody.

available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24535/abstract) and trunk-to-leg fat ratio were similar among the different MSA groups at all 3 time points (visit 1 >1.5 years; visit 3 = 1.51–3.49 years; visit 5 = 3.5–5.0 years) (Figure 2 and Supplementary Table 2, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24535). Also, the frequency of lipodystrophy by physician evaluation was similar among the 3 groups (33% of the anti-p155/140 antibody group, 24% of the other MSA group, and 34% of the MSA negative group; $P = 0.697$). There was no significant difference between the 3 MSA groups in terms of total cholesterol, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, triglycerides, fasting insulin, fasting glucose, and homeostasis model assessment–estimated insulin resistance (see Supplementary Table 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24535).

DISCUSSION

To our knowledge, this is the largest study of fat distribution utilizing DXA as a screening tool for lipodystrophy in juvenile DM. Because lipodystrophy has been reported as a late complication of juvenile DM, we only included patients with at least 5 years of data. The previous study of predictors of lipodystrophy showed a significant association between anti-p155/140 positive autoantibodies and generalized lipodystrophy (6). However, the sample size was limited, as there were only 7 subjects with generalized lipodystrophy, and 6 of them were anti-p155/140 positive. In contrast to that study, we did not find any evidence of increased generalized lipodystrophy in the anti-p155/140 group by clinical assessment or DXA assessment of fat distribution (Figure 2). The limitations of this study include the lack of measurement of subcutaneous and visceral fat by MRI or CT scan for the assessment of lipodystrophy and the possibility of missing cases of lipodystrophy that presented after 5 years of the onset of therapy. This study did not assess focal lipodystrophy, which is typically associated with panniculitis, because DXA does not detect focal fat changes.

Interestingly, the anti-p155/140 positive group demonstrated significantly worse nailfold capillary ERL counts when compared to the other 2 groups. Within the juvenile DM literature, reduced ERL count is correlated with the severity and chronicity of skin disease (4,13), suggesting a role for early and aggressive treatment for these patients (13). Although in our study, the anti-p155/140 group did not demonstrate changes in DAS-S score compared to the other 2 groups, a decrease in ERL count has been implicated in other studies as a potential marker for disease activity, including skin activity at diagnosis (4) and after 3 years of follow-up (13). Another longitudinal study demonstrated an association between low ERL count and disease activity of both skin and muscle (2). These studies suggest that low ERL count may be considered a marker of disease activity in juvenile DM and may be helpful in guiding treatment decisions early in the disease course (2,4,13). The present study also found that the p155/140 group was less likely to have a monophasic disease course. The duration of untreated disease in the anti-p155/140 group was twice as long as in the other 2 groups, suggesting that a delay in diagnosis is

common in this group. This corroborates the findings that the presence of end row capillary loop dropout is an important feature of untreated disease (4). In the adult population with idiopathic inflammatory myopathies, patients with anti–MDA-5 and anti-p155/140 autoantibodies have a greater prevalence of nailfold abnormalities when compared to anti–aminoacyl–transfer RNA synthetase autoantibodies (14). However, there have not been any similar studies correlating NFC parameters with MSAs in the pediatric population.

Demonstrating that the anti-p155/140 MSA group has a lower ERL count when compared to the other 2 groups, as documented in this study, is a novel finding. It has previously been demonstrated by our study center that juvenile DM patients with decreased ERL counts have impaired absorption of oral steroids when compared to the equivalent dose of steroids administered by the intravenous route (15). Thus, our findings suggest that the anti-p155/140 positive MSA group is more likely to experience decreased absorption of oral steroids, which in turn can contribute to suboptimal control of disease if oral glucocorticoids are used for therapy. This may explain the discrepancy between our finding that the anti-p155/140 antibody did not increase general lipodystrophy and the findings of Bingham and colleagues (6), as our study center aggressively uses intravenous steroids, which facilitates drug absorption, early in the treatment for juvenile DM in comparison to other centers. This is an intriguing concept and provides further support for consideration of early aggressive treatment in anti-p155/140 positive patients with juvenile DM.

In conclusion, our findings suggest that while the prevalence of generalized lipodystrophy as measured by physician assessment and DXA assessment of fat distribution was the same in patients with anti-p155/140 antibody compared to the other subgroups (other MSA and MSA negative), this autoantibody was associated with loss of nailfold ERL capillaries and a longer duration of untreated disease at the time of diagnosis. These findings suggest that the anti-p155/140 MSA subgroup may warrant the initiation of more aggressive therapy. Further studies are needed to identify nailfold capillary changes relative to MSA subtype to determine if additional patterns exist, which may provide much needed aid to clinicians in making treatment decisions in this population of patients.

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. Khojah 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. Khojah, Pachman.

Acquisition of data. Liu, Morgan.

Analysis and interpretation of data. Khojah, Liu, Savani, Shore, Bellm, Pachman.

REFERENCES

1. Pachman LM, Khojah AM. Advances in juvenile dermatomyositis: myositis specific antibodies aid in understanding disease heterogeneity. J Pediatr 2018;195:16–27.
2. Schmeling H, Stephens S, Goia C, Manlhiot C, Schneider R, Luthra S, et al. Nailfold capillary density is importantly associated over time with muscle and skin disease activity in juvenile dermatomyositis. Rheumatology (Oxford) 2011;50:885–93.
3. Silver RM, Maricq HR. Childhood dermatomyositis: serial microvascular studies. Pediatrics 1989;83:278–83.
4. Smith RL, Sundberg J, Shamiyah E, Dyer A, Pachman LM. Skin involvement in juvenile dermatomyositis is associated with loss of end row nailfold capillary loops. J Rheumatol 2004;31:1644–9.
5. Tansley SL, Simou S, Shaddick G, Betteridge ZE, Almeida B, Gunawardena H, et al. Autoantibodies in juvenile-onset myositis: their diagnostic value and associated clinical phenotype in a large UK cohort. J Autoimmun 2017;84:55–64.
6. Bingham A, Mamyrova G, Rother KI, Oral E, Cochran E, Premkumar A, et al. Predictors of acquired lipodystrophy in juvenile-onset dermatomyositis and a gradient of severity. Medicine (Baltimore) 2008;87:70–86.
7. Huemer C, Kitson H, Malleson PN, Sanderson S, Huemer M, Cabral DA, et al. Lipodystrophy in patients with juvenile dermatomyositis: evaluation of clinical and metabolic abnormalities. J Rheumatol 2001;28:610–5.
8. Blanch J, Rousaud A, Martinez E, de Lazzari E, Milinkovic A, Peri JM, et al. Factors associated with severe impact of lipodystrophy on the quality of life of patients infected with HIV-1. Clin Infect Dis 2004;38:1464–70.
9. McCann LJ, Juggins AD, Maillard SM, Wedderburn LR, Davidson JE, Murray KJ, et al. The Juvenile Dermatomyositis National Registry and Repository (UK and Ireland): clinical characteristics of children recruited within the first 5 yr. Rheumatology (Oxford) 2006;45:1255–60.
10. Castro TC, Terreri MT, Szejnfeld VL, Len C, Fonseca AS, Hilario MO. Bone mineral density of Brazilian girls with juvenile dermatomyositis. Braz J Med Biol Res 2005;38:309–13.
11. Freitas P, Carvalho D, Santos AC, Mesquita J, Matos MJ, Madureira AJ, et al. Lipodystrophy defined by fat mass ratio in HIV-infected patients is associated with a high prevalence of glucose disturbances and insulin resistance. BMC Infect Dis 2012;12:180.
12. Bode RK, Klein-Gitelman MS, Miller ML, Lechman TS, Pachman LM. Disease activity score for children with juvenile dermatomyositis: reliability and validity evidence. Arthritis Rheum 2003;49:7–15.
13. Christen-Zaech S, Seshadri R, Sundberg J, Paller AS, Pachman LM. Persistent association of nailfold capillaroscopy changes and skin involvement over thirty-six months with duration of untreated disease in patients with juvenile dermatomyositis. Arthritis Rheum 2008;58:571–6.
14. Kubo S, Todoroki Y, Nakayamada S, Nakano K, Satoh M, Nawata A, et al. Significance of nailfold videocapillaroscopy in patients with idiopathic inflammatory myopathies. Rheumatology (Oxford) 2019;58:120–30.
15. Rouster-Stevens KA, Gursahaney A, Ngai KL, Daru JA, Pachman LM. Pharmacokinetic study of oral prednisolone compared with intravenous methylprednisolone in patients with juvenile dermatomyositis. Arthritis Rheum 2008;59:222–6.

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1070–1078
DOI 10.1002/acr.24548

American College of Rheumatology
Empowering Rheumatology Professionals

Predictors of Unsuccessful Hydroxychloroquine Tapering and Discontinuation: Can We Personalize Decision-Making in Systemic Lupus Erythematosus Treatment?

Celline C. Almeida-Brasil,[1] Christian A. Pineau,[2] Evelyne Vinet,[2] John G. Hanly,[3] Christine A. Peschken,[4] Ann E. Clarke,[5] Paul R. Fortin,[6] Michal Abrahamowicz,[7] and Sasha Bernatsky[1]

Objective. Hydroxychloroquine (HCQ) is a key systemic lupus erythematosus (SLE) drug, making concerns of drug shortages grave. Our objective was to evaluate factors associated with poor outcomes after HCQ taper or discontinuation in SLE.

Methods. We studied 5 Canadian SLE cohorts between 1999 and 2019, following patients from the date of HCQ tapering (cohort 1) or discontinuation (cohort 2). A composite outcome was defined as any of the following: a need for therapy augmentation, an increase (of at least 4 points) in the Systemic Lupus Erythematosus Disease Activity Index 2000 score, or hospitalization for SLE. In each cohort, multivariable Cox regression was used to identify demographic and clinical factors associated with time to the earliest of these events. A third cohort continuing to receive HCQ was also studied, to assess whether the same factors influenced the outcome even when the HCQ dose was unchanged.

Results. The poor outcome rate, per 100 person-years, was 35.7 (95% confidence interval [95% CI] 31.6–40.3) in the HCQ taper cohort (n = 398), 29.0 (95% CI 25.5–33.0) in the discontinuation cohort (n = 395), and 16.1 (95% CI 13.2–19.6) in the maintenance cohort (n = 395). In patients tapering HCQ, baseline prednisone use was independently associated with greater risk of poor outcomes. In the discontinuation cohort, the risk of poor outcomes was greater for Black patients and those diagnosed with SLE at age ≤25 years. Among those maintaining HCQ, baseline immunosuppressive use and First Nations ethnicity were associated with poor outcomes.

Conclusion. We identified demographic and clinical factors associated with poor outcomes after HCQ taper/discontinuation. This information is critical in the current setting of potential shortages, but over the long term, such information could inform personalized therapies.

INTRODUCTION

Hydroxychloroquine (HCQ) is a cornerstone medication for systemic lupus erythematosus (SLE) (1), and sustained HCQ use might greatly reduce disease flares (2–4). However, there is concern over retinal toxicity, an irreversible complication that may affect 20% of patients after long-term exposure (5). Uncertainties about the relative risks/benefits of long-term treatment are a primary concern voiced by patients with SLE (6), and almost one-third of patients with SLE discontinue HCQ treatment by 5–8 years (7).

Supported by a Canadian Institutes of Health Research (CIHR) project grant (PJT-156395). Dr. Almeida-Brasil's work was supported by postdoctoral salary awards from the Fonds de Recherche du Québec Santé, the Research Institute of the McGill University Health Centre, and the CIHR Drug Safety and Effectiveness Cross-Disciplinary Training Program. Dr. Fortin has a tier 1 Canada Research Chair at the Systemic Autoimmune Rheumatic Diseases Centers in Winnipeg and Quebec City, and he contributed data thanks to their participation in the ongoing cohort of the Canadian Network for Improved Outcomes in Systemic Lupus Erythematosus, also supported by unrestricted funds from Lupus Canada.

[1]Celline C. Almeida-Brasil, PhD, Sasha Bernatsky, MD, PhD: Research Institute of McGill University Health Centre, Montreal, Quebec, Canada; [2]Christian A. Pineau, MD, Evelyne Vinet, MD, PhD: McGill University Health Centre, Montreal, Quebec, Canada; [3]John G. Hanly, MD: Queen Elizabeth II Health Sciences Center, Halifax, Nova Scotia, Canada; [4]Christine A. Peschken, MD: University of Manitoba College of Medicine, Winnipeg, Manitoba, Canada; [5]Ann E. Clarke, MD: University of Calgary, Calgary, Alberta, Canada; [6]Paul R. Fortin, MD: CHU de Québec–Université Laval, Quebec, Quebec, Canada; [7]Michal Abrahamowicz, PhD: McGill University, Montreal, Quebec, Canada.

No potential conflicts of interest relevant to this article were reported.

Address correspondence to Sasha Bernatsky, MD, PhD, Department of Medicine, Division of Rheumatology and Division of Clinical Epidemiology and Centre for Outcomes Research and Evaluation, Research Institute of the McGill University Health Centre, 5252 Boulevard de Maisonneuve West, room 3F.51, Montreal, Quebec H4A 3S5, Canada. Email: sasha.bernatsky@mcgill.ca.

Submitted for publication May 12, 2020; accepted in revised form December 17, 2020.

SIGNIFICANCE & INNOVATIONS

- Although some patients may do well after reducing therapy, others will have potentially life-threatening complications related to systemic lupus erythematosus (SLE) flares, and there is no information available to guide individual decision-making.
- Three cohorts of patients with SLE who were tapering, stopping, or maintaining hydroxychloroquine (HCQ) were evaluated. The crude flare rate was significantly lower in patients maintaining HCQ therapy than in those tapering or discontinuing the drug.
- Non-White patients (especially Black, Asian, or First Nations patients), those age ≤25 years at SLE diagnosis, and those with active disease, including patients receiving prednisone or immunosuppressors, are at higher risk of having flares.
- Our results suggest caution in tapering or discontinuation of HCQ in some groups of patients with SLE. The identification of these predictors is an important approach to promote personalized medicine to avoid unnecessary toxicities, as well as to monitor for flares in situations such as the current setting of potential HCQ shortages due to interest in this drug as a therapy for COVID-19.

On an individual level, patients with SLE and clinicians struggle with many treatment decisions, since there is little information available on tapering or stopping HCQ to guide individual decision-making. Some patients may do well after HCQ withdrawal, but others will have potentially life-threatening complications (8). Recently, new concerns have arisen regarding HCQ shortages for patients with SLE due to potential COVID-19 treatment (9,10). Clearly, we need better predictors of flare risk after HCQ is lowered or discontinued. The aim of this study was to identify baseline factors associated with a poor outcome once HCQ is tapered or discontinued in SLE.

PATIENTS AND METHODS

Data sources. Our study combined data from 5 clinical SLE cohorts in Canada (McGill University Health Centre [MUHC] in Montreal, CHU de Québec–Université Laval in Quebec City, Dalhousie University in Halifax, University of Manitoba in Winnipeg, and the Southern Alberta Registry for Lupus Erythematosus at the University of Calgary). The cohorts enrolled unselected patients age ≥18 years who met American College of Rheumatology criteria for SLE (11) at the time that they presented to each center (including both incident and prevalent cases). The enrollment and follow-up period spanned January 1999 to January 2019.

Data on demographic, medication, and clinical variables were collected in a standardized manner at enrollment as well as annually and were submitted to the coordinating center at the MUHC for data harmonization. Ethics approval was obtained from the institutional review boards at all participating sites.

Study population. We studied adult patients with SLE exposed to HCQ during the study period. Starting from the first visit with HCQ exposure, we identified patients receiving a lower dose or discontinuing HCQ at a follow-up visit. We created 1 cohort to study patients from the time they lowered their HCQ dose: in that case, time zero was the date of the first reduction of HCQ dose. Patients were right censored if they discontinued HCQ completely (for any reason), as they then entered the cohort of patients who had discontinued HCQ (where the date of first HCQ discontinuation was defined as time zero). Patients who discontinued HCQ but started chloroquine right away were not included in the discontinuation cohort, as they were still taking an antimalarial. Patients were followed until the outcome of interest, end of the study period (February 2019), death, or loss to follow-up (see Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24548).

HCQ use was assumed to be constant for the entire 1-year period between 2 adjacent visits; if, for example, a participant was taking HCQ at the first and second visits but not at the third visit, then they were considered an HCQ user during the period between the first and third visits (approximately 2 years) and a nonuser from the third visit on (unless HCQ was reintroduced, which qualified as part of our outcome of interest). However, in sensitivity analyses, we repeated the primary analysis, reassigning the date that the patients tapered/stopped HCQ (from the first study visit where the reduction/stop was originally recorded) to a date 6 months prior (i.e., half-way between study visits).

Outcome. The primary composite outcome was time to the first of the following events indicating an SLE flare: 1) an increase of at least 4 points (above baseline score) in the SLE Disease Activity Index 2000 (SLEDAI-2K) score; 2) hospitalization for SLE; and/or 3) augmented SLE therapy, defined as an increase in HCQ (or restart if discontinued) or a new start or increase in any of the following: prednisone, immunosuppressive agents (azathioprine, methotrexate, or mycophenolate mofetil), biologics (rituximab or belimumab), cyclophosphamide, or start of chloroquine.

One center (Halifax) was unable to provide information on hospitalizations for 50% of their participants, thus the primary composite outcome for these patients was based on an increase in disease activity and therapy augmentation only. We also performed sensitivity analyses leaving out these patients.

Potential risk factors. Sociodemographic variables included sex, race/ethnicity (White versus Asian, Black, First

Nations, or other), education (high school education or less versus college or university education), and age at SLE diagnosis (dichotomized at ≤25 versus >25 years, to prevent collinearity with disease duration). Other baseline variables included body mass index (continuous), currently smoking (yes versus no), SLE duration (continuous), disease activity (≥4 points in SLEDAI-2K score, which is a validated definition of active SLE) (12,13), time taking HCQ since study enrollment (continuous), use of prednisone, immunosuppressive agents (azathioprine, methotrexate, or mycophenolate mofetil), and biologic agents (rituximab or belimumab), and the presence of renal damage based on the Systemic Lupus International Collaborating Clinics (SLICC) damage index (i.e., scores for heavy proteinuria and/or reduced glomerular filtration rate, or end-stage renal failure) (14).

Statistical analysis. Crude event rates and 95% confidence intervals (95% CIs) were calculated for the primary composite outcome (i.e., the earliest event indicating an SLE flare). We used multivariable Cox proportional hazards regression to estimate the adjusted hazard ratios (HR_{adj}) and 95% CIs for the associations between patient characteristics and the primary outcome. The proportional hazards assumption was verified using the cumulative sums of Martingale residuals and the Kolmogorov-type supremum test. Multicollinearity was assessed using collinearity indices, eigenvalues, and variable decomposition proportions.

We performed several secondary analyses. First, we repeated the primary analysis separately for each specific component of the composite outcome (i.e., increase in disease activity, SLE-related hospitalizations, and therapy augmentation), while censoring patients who had one of the other events prior to the component of interest. Second, as mentioned before, we conducted sensitivity analysis reassigning the date that the patients tapered/stopped HCQ from the first study visit where the reduction/stop was originally recorded to a date half-way between study visits. Third, we also conducted a sensitivity analysis removing patients without hospitalization data (Halifax) from the primary analysis.

Finally, we compared all results with a third cohort of patients with SLE: those who did not reduce or stop HCQ. The purpose here was to explore whether risk factors influenced the outcome even when HCQ dose was unchanged or whether risk factors were specific among those who decreased or interrupted HCQ treatment. Since the date of taper/discontinuation was used as time zero in our first 2 cohorts, time zero in the third cohort (HCQ maintenance) was defined

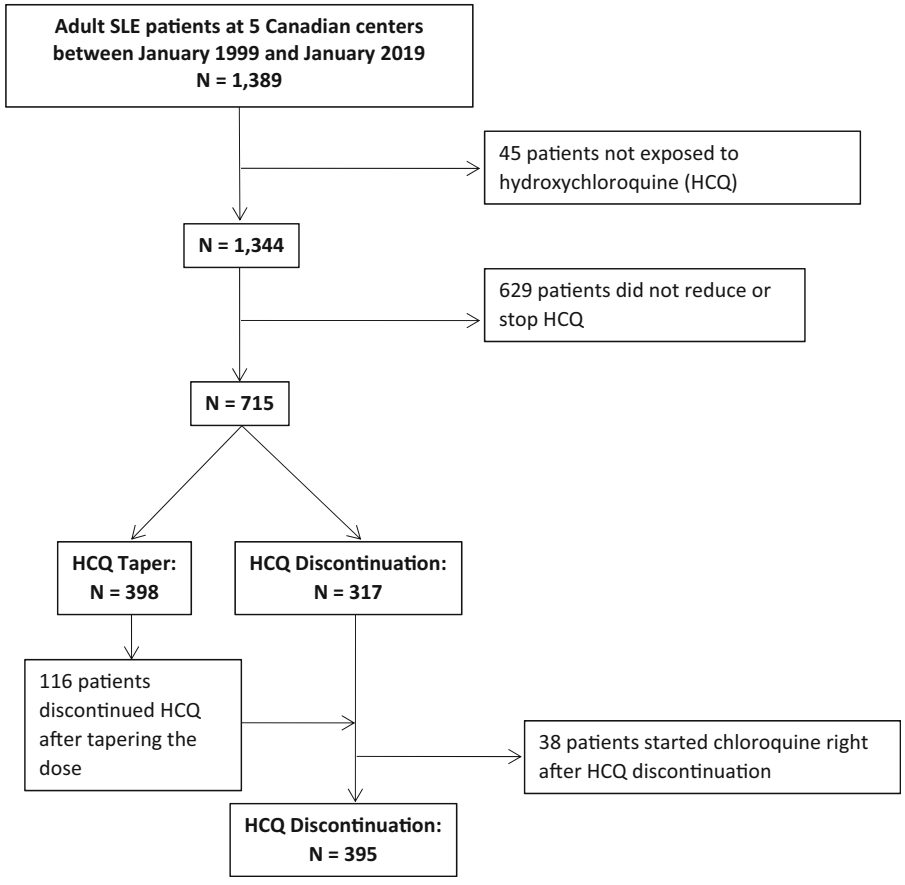


Figure 1. Cohort selection. SLE = systemic lupus erythematosus.

Table 1. Baseline characteristics of patients with SLE tapering, discontinuing, or maintaining HCQ*

Characteristic	Taper (n = 398)	Discontinuation (n = 395)	Maintenance (n = 395)
Female	368 (92.5)	361 (91.4)	347 (87.8)
Race/ethnicity			
White	297 (74.6)	307 (77.7)	295 (75.3)
Asian	55 (13.8)	32 (8.1)	33 (8.4)
Black	31 (7.8)	37 (9.4)	35 (8.9)
First Nations	9 (2.3)	16 (4.0)	20 (5.1)
Others†	6 (1.5)	3 (0.8)	9 (2.3)
Age at SLE diagnosis, years			
Median (IQR)	30.6 (23.2–41.8)	31.9 (24.1–42.0)	32.4 (22.1–46.4)
Age ≤25 years	129 (32.4)	114 (28.9)	127 (32.2)
Age at time zero, median (IQR) years	43.7 (33.5–55.4)	48.5 (37.4–59.0)	46.6 (34.1–57.8)
No college/university education	110 (28.1)	113 (30.1)	85 (22.6)
Center			
Montreal	224 (56.3)	202 (51.1)	109 (27.6)
Halifax	76 (19.1)	84 (21.3)	100 (25.3)
Calgary	45 (11.3)	17 (4.3)	49 (12.6)
Winnipeg	32 (8.0)	83 (21.0)	72 (18.2)
Quebec	21 (5.3)	9 (2.3)	65 (16.5)
SLE duration, median (IQR) years	7.9 (3.6–16.6)	12.9 (6.4–20.6)	6.4 (4.1–17.1)
Disease activity			
Median SLEDAI-2K score (IQR)	2.0 (0.0–6.0)	1.0 (0.0–4.0)	2.0 (0.0–4.0)
SLEDAI-2K score ≥4	186 (46.7)	125 (31.6)	159 (40.2)
Renal damage	30 (7.5)	36 (9.1)	29 (7.3)
Current smoker	150 (37.7)	145 (36.7)	98 (24.8)
Body mass index, median (IQR) kg/m^2	24.4 (21.7–28.3)	25.4 (22.5–30.1)	25.4 (22.1–30.3)
Current prednisone	79 (19.8)	42 (10.6)	103 (26.1)
Current immunosuppressors‡	144 (36.2)	76 (19.2)	176 (44.6)
Current biologic agents‡	17 (4.3)	6 (1.5)	9 (2.3)
Time on HCQ, median (IQR) years§	2.3 (1.2–4.5)	3.0 (1.2–6.1)	1.6 (1.0–3.1)

* Values are the number (%) unless indicated otherwise. HCQ = hydroxychloroquine; IQR = interquartile range; SLE = systemic lupus erythematosus; SLEDAI-2K = Systemic Lupus Erythematosus Disease Activity Index 2000.
† Arab peoples and those of mixed ethnicity.
‡ Immunosuppressors included mycophenolate, azathioprine, and methotrexate; biologics included belimumab and rituximab.
§ Since study entry.

as the visit date when both SLE duration and time taking HCQ individually matched the respective baseline values of patients tapering/discontinuing HCQ. Since patients in the HCQ taper and discontinuation cohorts could not enter that cohort unless they had had at least 2 visits (one visit with baseline HCQ and the second visit when HCQ was tapered/stopped), we also required at least 2 visits for patients in the HCQ maintenance cohort. Statistical analyses were performed using SAS software, version 9.4.

RESULTS

Among 1,389 individuals receiving care in the participating lupus clinics between January 1999 and January 2019, 1,344 (96.8%) were exposed to HCQ (Figure 1). We identified 398 patients (1,740 person-years) who reduced the HCQ dose, and 395 (2,120 person-years) who discontinued HCQ. Among those who maintained HCQ therapy (n = 629), 395 patients (792 person-years) were successfully matched to patients tapering or discontinuing HCQ therapy on previous disease duration and time on HCQ. Overall, a total of 240 patients were lost to follow-up, 62 withdrew consent, and 35 died during the follow-up. All these patients were censored at the corresponding times.

The baseline characteristics of each of the 3 cohorts of patients with SLE are shown in Table 1. As expected, approximately 90% of the participants were female, and most were White. The primary composite outcome occurred in 261 of the 398 patients who tapered HCQ (35.7 events per 100 person-years [95% CI 31.6–40.3]), in 226 of the 395 patients who discontinued HCQ (29.0 per 100 person-years [95% CI 25.5–33.0]), and in 97 of the 395 patients who remained on HCQ (16.1 events per 100 person-years [95% CI 13.2–19.6]).

The most common poor outcome was therapy augmentation (52.8% after tapering, 48.9% after stopping HCQ, and 17.2% in those maintaining HCQ), followed by SLEDAI-2K score increase of ≥4 points (19.4% after tapering, 20.2% after stopping HCQ, and 10.3% in those maintaining HCQ) and hospitalization for SLE (0.8% after tapering, 0.6% after stopping HCQ, and 0.3% in those maintaining HCQ).

Table 2. Cox regression analysis: baseline characteristics and the primary outcome (earliest poor outcome)*

Baseline characteristic	HCQ taper (n = 398)		HCQ discontinuation (n = 395)		HCQ maintenance (n = 395)	
	HR (95% CI)	HR_{adj} (95% CI)†	HR (95% CI)	HR_{adj} (95% CI)†	HR (95% CI)	HR_{adj} (95% CI)†
Male	1.22 (0.73–2.03)	1.40 (0.83–2.37)	0.60 (0.34–1.06)	0.65 (0.36–1.15)	0.65 (0.31–1.34)	0.52 (0.23–1.19)
Race/ethnicity (vs. White)						
Asian	1.26 (0.87–1.84)	1.14 (0.76–1.72)	1.12 (0.65–1.94)	1.04 (0.58–1.89)	2.05 (0.94–4.47)	1.33 (0.54–3.25)
Black	1.37 (0.90–2.09)	1.03 (0.65–1.65)	1.79 (1.20–2.67)‡	1.61 (1.03–2.51)‡	1.49 (0.69–3.24)	1.29 (0.56–2.97)
First Nations	0.76 (0.24–2.38)	0.75 (0.23–2.40)	1.19 (0.56–2.54)	0.90 (0.40–2.03)	2.81 (1.34–5.86)‡	2.87 (1.21–6.76)‡
Others	0.42 (0.10–1.69)	0.35 (0.09–1.44)	3.22 (0.79–13.0)	2.01 (0.47–8.54)	1.71 (0.24–12.4)	1.44 (0.16–12.9)
Age ≤25 years at SLE diagnosis	1.29 (0.99–1.67)	1.22 (0.92–1.62)	1.61 (1.22–2.13)‡	1.75 (1.29–2.38)‡	1.37 (0.91–2.07)	1.18 (0.73–1.91)
No college/university education	0.97 (0.74–1.26)	0.92 (0.69–1.22)	1.10 (0.83–1.45)	1.18 (0.87–1.60)	0.77 (0.47–1.25)	0.86 (0.49–1.49)
SLE duration, years	1.00 (0.99–1.01)	1.00 (0.99–1.02)	0.99 (0.98–1.01)	1.00 (0.98–1.01)	0.99 (0.96–1.01)	0.99 (0.96–1.02)
SLEDAI-2K score ≥4	1.23 (0.96–1.57)	1.16 (0.89–1.50)	1.08 (0.82–1.42)	1.01 (0.74–1.36)	1.33 (0.88–2.01)	1.10 (0.70–1.73)
Baseline renal damage	1.30 (0.86–1.99)	1.04 (0.65–1.67)	1.05 (0.69–1.59)	0.84 (0.52–1.35)	1.12 (0.45–2.75)	1.32 (0.51–3.44)
Body mass index, kg/m^2	1.01 (0.99–1.04)	1.02 (0.99–1.04)	1.02 (0.99–1.04)	1.01 (0.98–1.03)	0.99 (0.96–1.03)	0.98 (0.95–1.02)
Smoker at baseline	0.71 (0.56–0.92)‡	0.81 (0.61–1.07)	0.62 (0.47–0.82)‡	0.66 (0.49–0.89)‡	0.71 (0.46–1.10)	0.88 (0.52–1.50)
Baseline prednisone	1.84 (1.34–2.53)‡	1.74 (1.23–2.45)‡	1.30 (0.88–1.91)	1.16 (0.72–1.87)	1.40 (0.89–2.19)	1.00 (0.58–1.71)
Baseline immunosuppressors	1.19 (0.92–1.53)	1.06 (0.80–1.41)	1.04 (0.76–1.42)	0.93 (0.64–1.33)	1.90 (1.27–2.85)‡	1.72 (1.08–2.71)‡
Baseline biologics	1.06 (0.56–2.00)	0.89 (0.45–1.75)	2.46 (0.78–7.72)	2.20 (0.64–7.52)	2.36 (0.96–5.83)	1.54 (0.55–4.29)
Time receiving HCQ, years	0.99 (0.95–1.04)	1.00 (0.96–1.05)	0.97 (0.93–1.00)	0.98 (0.94–1.01)	0.86 (0.76–0.97)‡	0.89 (0.78–1.01)

* Baseline medication use represents whether or not patients were taking the drug at time zero. 95% CI = 95% confidence interval; HCQ = hydroxychloroquine; HR = hazard ratio; SLE = systemic lupus erythematosus; SLEDAI-2K = Systemic Lupus Erythematosus Disease Activity Index 2000.
† Variables adjusted concomitantly for all others.
‡ Statistically significant.

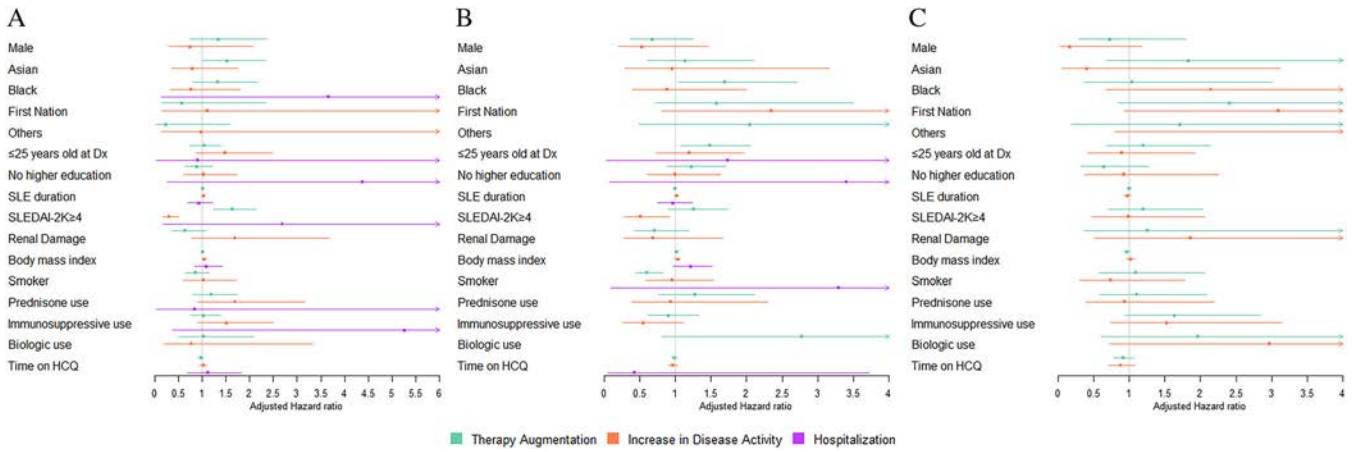


Figure 2. Forest plot showing associations between baseline characteristics and each secondary outcome among patients: **A**, hydroxychloroquine (HCQ) tapering; **B**, HCQ discontinuation; and **C**, HCQ maintenance. Baseline medication use represents whether patients were taking the drug at time zero. Variables are adjusted concomitantly for all others. Omitted lines indicate that no event occurred among exposed patients. Dx = diagnosis; SLE = systemic lupus erythematosus; SLEDAI-2K = Systemic Lupus Erythematosus Disease Activity Index 2000.

Table 2 shows associations between patients' baseline characteristics and the earliest poor outcome. In multivariable analyses, patients using prednisone at time zero had an increased risk of experiencing a poor outcome after HCQ was tapered. After discontinuing HCQ, Black patients, those diagnosed with SLE at a younger age (≤25 years), and nonsmokers had a higher risk of a poor outcome. None of these factors were clearly associated with the composite outcome among those maintaining HCQ, although First Nations ethnicity and baseline immunosuppressive use were.

Evaluations of each outcome separately are shown in Figure 2. Asian patients (HR_{adj} 1.52 [95% CI 0.99–2.32]) and those with active disease (HR_{adj} 1.62 [95% CI 1.22–2.14]) at the time of HCQ taper were more likely to need therapy augmentation (Figure 2A). In the HCQ discontinuation cohort, Black race/ethnicity (HR_{adj} 1.69 [95% CI 1.05–2.71]), younger age at SLE diagnosis (HR_{adj} 1.48 [95% CI 1.07–2.06]), and nonsmoking (HR_{adj} 0.59 [95% CI 0.43–0.82]) were predictors of therapy augmentation (Figure 2B). Patients with baseline SLEDAI-2K score ≥4 were less likely to have a subsequent increase in disease activity after HCQ was tapered or discontinued. We did not identify clear predictors of any separate outcome among patients who maintained HCQ (Figure 2C). No clear associations were observed between patients' characteristics and SLE-related hospitalization, possibly because statistical power was limited due to the relatively low number of hospitalizations.

The results of sensitivity analyses where we reassigned the date that the patients tapered/stopped HCQ from the first study visit where the reduction/stop was originally recorded to a date 6 months prior (i.e., half-way between study visits) were consistent with the primary analyses (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24548), as were the results leaving patients without hospitalization information out from the analysis (see Supplementary Table 2).

DISCUSSION

Although our study was originally motivated by the desire to better understand personalized therapy in SLE, our findings take on new importance in the current setting, where physicians and patients may face shortages of HCQ, due to interest in this drug as a potential therapy for COVID-19. Although current evidence shows that HCQ is not effective for COVID-19, patients in many countries, including the US, Brazil, and India, are still using HCQ to prevent or treat the infection. In addition, HCQ shortages may occur in other circumstances, such as the prolonged manufacturing shortage faced by the US in 2015 (15). Moreover, stopping HCQ by choice is not that rare; studies have shown that over 30% of patients with SLE discontinue HCQ by choice (7,16).

In this clinical cohort of patients with SLE exposed to HCQ, we observed that multiple demographic and baseline clinical factors are associated with poor outcomes, such as an increase in disease activity and a need for therapy augmentation, after HCQ taper/discontinuation. Tapering HCQ in Asian patients with SLE or when the patient is still on prednisone or with a SLEDAI-2K score of ≥4 may result in poorer outcomes. Discontinuing HCQ may be associated with poorer outcomes in Black patients with SLE, and patients age ≤25 years at SLE diagnosis. None of these factors were associated with the outcome among those maintaining HCQ, although First Nations patients and baseline immunosuppressive use were.

We did not aim to compare the 3 cohorts directly in terms of flare rate. However, the crude flare rate was significantly lower in the HCQ maintenance cohort (16.1 events per 100 person-years

[95% CI 13.2–19.6]) than in the taper cohort (35.7 [95% CI 31.6–40.3]) and in the discontinuation cohort (29.0 [95% CI 25.5–33.0]). Of course, these crude rates do not consider the fact that variables like disease activity and concomitant medications may differ considerably between the cohorts. If patients tapered or discontinued HCQ solely because of inactive disease, we would expect lower outcome rates in the taper than in the discontinuation group. However, reasons for lowering the dose or discontinuing a medication are multifactorial and may have included patient tolerance, adherence, or even changes in guidelines, including the 2016 American Academy of Ophthalmology (AAO) recommendations, which cautioned against cumulative use of HCQ and lowered dosing to 5 mg/kg per day (5). Thus, not all patients were in remission when HCQ was tapered or discontinued. In fact, we observed that the proportion of patients with active disease (SLEDAI-2K score ≥4) at baseline was higher in the HCQ taper cohort (46.7% [95% CI 41.8–51.6]) than in the discontinuation cohort (31.6% [95% CI 27.0–36.2]). Given this finding, the relatively high number of patients in the taper group with a poor outcome makes sense. The fact that the maintenance group had higher baseline disease activity (and more use of immunosuppressives, corticosteroids, and biologics) than the discontinuation group, but a significantly lower (not higher) flare rate, suggests that HCQ is beneficial is this group.

Prednisone is a marker of more severe and active SLE (17,18). In our sample, among patients using prednisone while tapering HCQ, 60.8% (95% CI 49.9–71.6) had active SLE, compared with 43.3% (95% CI 37.8–48.7) not taking prednisone. A baseline SLEDAI-2K score of ≥4 was also identified as a predictor of therapy augmentation among patients tapering HCQ. These findings confirm a clinical intuition that patients with active disease are more likely to have poor outcomes, especially a need for therapy augmentation.

Among patients who remained on HCQ, immunosuppressive use was associated with our composite outcome (i.e., the earliest poor outcome). Immunosuppressors are also a marker of more severe disease (19,20). However, we did not find significant interactions between prednisone or immunosuppressors and disease activity in relation to our outcomes. Although multicollinearity between baseline prednisone, immunosuppressors, and SLEDAI-2K score could theoretically be an issue, diagnostic tests showed no threat of multicollinearity in our multivariate models.

In both tapering and discontinuation cohorts, we observed a negative association between a baseline SLEDAI-2K score of ≥4 and the specific outcome related to "increase in SLEDAI-2K score." This finding may represent a ceiling effect of the SLEDAI-2K assessment tool (21,22), which prevented the detection of a worsening in disease activity in patients with a baseline SLEDAI-2K score of ≥4, and/or a regression toward the mean, where more active patients with SLE may get better over time, and less active patients may get worse.

Non-White patients were more likely than White patients to have poor outcomes in all 3 cohorts. This finding was especially true in Black patients discontinuing HCQ, Asian patients tapering HCQ, and First Nations patients remaining on HCQ. Non-White patients, especially Black and First Nations patients, not only may have more severe SLE due to innate disease characteristics, but may also face barriers to optimal health outcomes, including access to care issues (even in the context of Canada's comprehensive health care system, which does not cover the cost of out-of-hospital medications for all individuals) and poor medication adherence (23,24). In general, non-White patients with SLE may have poorer outcomes due to sociocultural and psychosocial issues (25,26), including a higher risk for flares (27). As mentioned before, patients possibly discontinued the drug against physician advice; those patients may also have been nonadherent with other medications and physician advice, which could explain the findings of higher flare risk with Black patients who discontinued HCQ.

The risk of a poor outcome after HCQ discontinuation was higher in patients with SLE diagnosed at age ≤25 years. Younger SLE onset is generally more driven by genetic factors, which may correspond to a more severe SLE phenotype (28). Previous studies also identified younger age at SLE diagnosis as a strong predictor of lupus flares (20,27,29), including pediatric-onset SLE (30). At the same time, treatment toxicity (primarily retinal) has the potential to accumulate over a long period (30), which creates difficulty balancing the risks and benefits of long-term HCQ use.

We observed that baseline current smoking was inversely associated with poor outcomes after HCQ discontinuation, a finding that was not apparent in adjusted analyses of HCQ tapering or maintenance. Antimalarials are known to have decreased efficacy among smokers (31), probably due to tobacco's effect on the cytochrome P-450 enzyme system (32). Indeed, we observed that smokers already had worse disease activity at baseline than nonsmokers (data not shown). Thus, discontinuation of HCQ in smokers may not have the same clinical impact as in nonsmokers. On the other hand, since we did not update smoking status over time, some of those patients who smoked at the time of HCQ discontinuation may have stopped smoking, whereas very few of the nonsmokers would have started smoking over time. This nondifferential misclassification of smoking exposure may have contributed to the unexpected inverse relationship between smoking and poor outcomes (particularly if stopping smoking was associated with other nonmeasured variables, such as adherence to other medications).

We acknowledge important potential limitations in the current study. First, due to its exploratory nature, we did not adjust our analyses for multiple comparisons (33). Therefore, subsequent research with preplanned hypotheses should be conducted to confirm the observed associations. Second, data on HCQ use before the beginning of the study were not available, and the variable "time on HCQ" was calculated using the study

entry date, which may underestimate the real time of exposure. We estimate that the real duration of HCQ may be approximately 2 years less than SLE duration, since patients usually start HCQ therapy 2 years after SLE diagnosis. Third, we do not know for sure the reasons for reducing the dose or discontinuing HCQ. Therefore, drawing conclusions about effectiveness of HCQ by comparing the 3 cohorts is not possible, although patients remaining on HCQ had a significantly lower outcome rate than those tapering or discontinuing HCQ. In addition, although the identified demographic and baseline factors are warning signs of patients who might not do well on a taper/discontinuation independently of the reason, the reason for tapering/stopping HCQ may influence later flare risk and/or the reason for subsequent therapy augmentation. As mentioned before, HCQ tapering may have occurred because the patient was doing well (stable disease), or because the physician was following the 2016 AAO recommendations. HCQ discontinuation, on the other hand, may be due to retinal toxicity or the patient's choice (nonadherence), besides the cases where patients were in prolonged disease remission.

To exclude the possibility that the reasons for tapering/stopping HCQ may be biasing our results, we evaluated the effects of the calendar year, considering the date that the AAO guideline was published, and retinal damage in the respective cohorts. Among patients tapering HCQ, 30% had their dose reduced after the AAO guideline was published, and the inclusion of the calendar year variable in the multivariable model for the taper cohort did not yield different estimations. Similarly, among those stopping HCQ, 8% had retinal damage (identified using the SLICC damage index) at baseline, and adjusting for it did not change the multivariable model. Although we did not evaluate adherence, by adjusting the analyses for sex, age, race/ethnicity, and multiple medications, we accounted for factors that are themselves strong predictors of adherence in SLE. At Canadian centers, measurements of HCQ levels are not part of usual care and are rarely obtained. Nevertheless, the literature indicates that approximately 30% of patients with SLE are nonadherent to HCQ treatment (7,16), so most patients with active disease at baseline who stopped HCQ probably did so on their own (possibly due to side effects or other concerns), since physicians do not commonly discontinue treatment in patients with active SLE.

Our multivariate analyses suggested that prednisone use and a SLEDAI-2K score of ≥4 at the time of HCQ tapering were associated with a greater risk of a poor outcome, as was Asian race/ethnicity. Among those discontinuing HCQ, the risk of a poor outcome was greater for Black patients and those diagnosed with SLE at age ≤25 years. In patients who maintained HCQ therapy, baseline immunosuppressive use and First Nations ethnicity were associated with poor outcomes. The identification of multiple demographic and clinical predictors of poor outcomes after HCQ taper/discontinuation may be useful in personalizing decisions for patients with SLE (and their physicians) around medication de-escalation or maintenance, as well as monitoring for flares when HCQ tapering or stopping is needed, such as in the current setting of potential HCQ shortages due to interest in this drug as a therapy for COVID-19.

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. Bernatsky 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. Pineau, Vinet, Hanly, Peschken, Clarke, Fortin, Abrahamowicz, Bernatsky.

Acquisition of data. Pineau, Vinet, Hanly, Peschken, Clarke, Fortin, Bernatsky.

Analysis and interpretation of data. Almeida-Brasil, Abrahamowicz, Bernatsky.

REFERENCES

1. Keeling SO, Bissonauth A, Bernatsky S, Vandermeer B, Fortin PR, Gladman DD, et al. Practice variations in the diagnosis, monitoring, and treatment of systemic lupus erythematosus in Canada. J Rheumatol 2018;45:1440–7.
2. Akhavan PS, Su J, Lou W, Gladman DD, Urowitz MB, Fortin PR. The early protective effect of hydroxychloroquine on the risk of cumulative damage in patients with systemic lupus erythematosus. J Rheumatol 2013;40:831–41.
3. Esdaile JM, Abrahamowicz M, MacKenzie T, Hayslett JP, Kashgarian M. The time-dependence of long-term prediction in lupus nephritis. Arthritis Rheum 1994;37:359–68.
4. Ruiz-Irastorza G, Ramos-Casals M, Brito-Zeron P, Khamashta MA. Clinical efficacy and side effects of antimalarials in systemic lupus erythematosus: a systematic review. Ann Rheum Dis 2010;69:20–8.
5. Marmor MF, Kellner U, Lai TY, Melles RB, Mieler WF, American Academy of Ophthalmology. Recommendations on screening for chloroquine and hydroxychloroquine retinopathy (2016 revision). Ophthalmology 2016;123:1386–94.
6. Haag H, Liang T, Avina-Zubieta JA, De Vera MA. How do patients with systemic autoimmune rheumatic disease perceive the use of their medications: a systematic review and thematic synthesis of qualitative research. BMC Rheumatol 2018;2:9.
7. Mehat P, Atiquzzaman M, Esdaile JM, Aviña-Zubieta A, De Vera MA. Medication nonadherence in systemic lupus erythematosus: a systematic review. Arthritis Care Res (Hoboken) 2017;69:1706–13.
8. Tsakonas E, Joseph L, Esdaile JM, Choquette D, Senecal JL, Cividino A, et al. A long-term study of hydroxychloroquine withdrawal on exacerbations in systemic lupus erythematosus. The Canadian Hydroxychloroquine Study Group. Lupus 1998;7:80–5.
9. Zoellner D. Coronavirus: lupus sufferers facing drug shortage after spike in prescriptions for potential Covid-19 treatments. The Independent March 20, 2020. URL: https://www.independent.co.uk/news/world/americas/coronavirus-treatment-drugs-covid-19-lupus-shortage-pharmacy-trump-a9415516.html.
10. Touret F, de Lamballerie X. Of chloroquine and COVID-19. Antiviral Res 2020;177:104762.
11. Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1982;25:1271–7.
12. Gladman DD, Urowitz MB, Kagal A, Hallett D. Accurately describing changes in disease activity in systemic lupus erythematosus. J Rheumatol 2000;27:377–9.

13. Yee CS, Farewell VT, Isenberg DA, Griffiths B, Teh LS, Bruce IN, et al. The use of Systemic Lupus Erythematosus Disease Activity Index-2000 to define active disease and minimal clinically meaningful change based on data from a large cohort of systemic lupus erythematosus patients. Rheumatology (Oxford) 2011;50:982–8.
14. Gladman DD, Ibanez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. J Rheumatol 2002;29:288–91.
15. Erkan D, Unlu O, Sciascia S, Belmont HM, Branch DW, Cuadrado MJ, et al. Hydroxychloroquine in the primary thrombosis prophylaxis of antiphospholipid antibody positive patients without systemic autoimmune disease. Lupus 2018;27:399–406.
16. Aouhab Z, Hong H, Felicelli C, Tarplin S, Ostrowski RA. Outcomes of systemic lupus erythematosus in patients who discontinue hydroxychloroquine. ACR Open Rheumatol 2019;1:593–9.
17. Ad Hoc Working Group on Steroid-Sparing Criteria in Lupus. Criteria for steroid-sparing ability of interventions in systemic lupus erythematosus: report of a consensus meeting. Arthritis Rheum 2004;50:3427–31.
18. Hoes JN, Jacobs JW, Boers M, Boumpas D, Buttgereit F, Caeyers N, et al. EULAR evidence-based recommendations on the management of systemic glucocorticoid therapy in rheumatic diseases. Ann Rheum Dis 2007;66:1560–7.
19. Hernandez-Cruz B, Tapia N, Villa-Romero AR, Reyes E, Cardiel MH. Risk factors associated with mortality in systemic lupus erythematosus: a case-control study in a tertiary care center in Mexico City. Clin Exp Rheumatol 2001;19:395–401.
20. Ines L, Duarte C, Silva RS, Teixeira AS, Fonseca FP, da Silva JA. Identification of clinical predictors of flare in systemic lupus erythematosus patients: a 24-month prospective cohort study. Rheumatology (Oxford) 2014;53:85–9.
21. Nantes SG, Strand V, Su J, Touma Z. Comparison of the sensitivity to change of the 36-item Short Form health survey and the Lupus Quality of Life Measure using various definitions of minimum clinically important differences in patients with active systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2018;70:125–33.
22. Rodriguez-Gonzalez MG, Valero-Gaona GA, Vargas-Aguirre T, Amezcua Guerra LM. Performance of the Systemic Lupus Erythematosus Disease Activity Score (SLE-DAS) in a Latin American population [letter]. Ann Rheum Dis 2020;79:e158.
23. Feldman CH, Yazdany J, Guan H, Solomon DH, Costenbader KH. Medication nonadherence is associated with increased subsequent acute care utilization among Medicaid beneficiaries with systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2015;67:1712–21.
24. Hurd K, Barnabe C. Mortality causes and outcomes in indigenous populations of Canada, the United States, and Australia with rheumatic disease: a systematic review. Semin Arthritis Rheum 2018;47:586–92.
25. Alarcón GS, McGwin G Jr, Bastian HM, Roseman J, Lisse J, Fessler BJ, et al. Systemic lupus erythematosus in three ethnic groups. VIII: predictors of early mortality in the LUMINA cohort. LUMINA Study Group. Arthritis Rheum 2001;45:191–202.
26. Peschken CA, Esdaile JM. Systemic lupus erythematosus in North American Indians: a population based study. J Rheumatol 2000;27:1884–91.
27. Peng L, Wang Z, Li M, Wang Y, Xu D, Wang Q, et al. Flares in Chinese systemic lupus erythematosus patients: a 6-year follow-up study. Clin Rheumatol 2017;36:2727–32.
28. Webb R, Kelly JA, Somers EC, Hughes T, Kaufman KM, Sanchez E, et al. Early disease onset is predicted by a higher genetic risk for lupus and is associated with a more severe phenotype in lupus patients. Ann Rheum Dis 2011;70:151–6.
29. Cervera R, Doria A, Amoura Z, Khamashta M, Schneider M, Guillemin F, et al. Patterns of systemic lupus erythematosus expression in Europe. Autoimmun Rev 2014;13:621–9.
30. Aggarwal A, Srivastava P. Childhood onset systemic lupus erythematosus: how is it different from adult SLE? Int J Rheum Dis 2015;18:182–91.
31. Chasset F, Frances C, Barete S, Amoura Z, Arnaud L. Influence of smoking on the efficacy of antimalarials in cutaneous lupus: a meta-analysis of the literature. J Am Acad Dermatol 2015;72:634–9.
32. Schein JR. Cigarette smoking and clinically significant drug interactions. Ann Pharmacother 1995;29:1139–48.
33. Althouse AD. Adjust for multiple comparisons? It's not that simple. Ann Thorac Surg 2016;101:1644–5.

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1079–1088
DOI 10.1002/acr.24546

American College of Rheumatology
Empowering Rheumatology Professionals

Systemic Lupus Erythematosus Symptom Clusters and Their Association With Patient-Reported Outcomes and Treatment: Analysis of Real-World Data

Zahi Touma,[1] Ben Hoskin,[2] Christian Atkinson,[2] David Bell,[2] Olivia Massey,[2] Jennifer H. Lofland,[3] Pamela Berry,[3] Chetan S. Karyekar,[3] and Karen H. Costenbader[4]

Objective. To identify discrete clusters of systemic lupus erythematosus (SLE) patients based on symptoms and investigate differences across clusters.

Methods. Data were collected in the US and 5 European countries via the Adelphi Real World Lupus Disease Specific Programme, a cross-sectional survey. Rheumatologists provided data for 5 consecutively consulting adult patients with SLE, who were invited to participate. Identified SLE symptoms were reduced to factors based on commonly concurrent symptoms, using principal-component factor analysis. Factors were used as covariates in a latent-class cluster analysis to identify discrete patient clusters. Patient-reported outcomes and physician-reported data were compared across clusters.

Results. Among 1,376 patients, 87% were female and 74% were White. We identified 4 patient clusters (very mild, mild, moderate, and severe) based on 39 signs/symptoms. Physician-reported symptom burden, organ involvement, disease activity, and the number of flares increased with increasing cluster severity ($P < 0.0001$). Patient-reported impact (health status, fatigue, work productivity impairment, anxiety/depression, and emotional impact) increased with increasing cluster severity ($P < 0.0001$). Glucocorticoid and immunosuppressant use increased, and antimalarial use decreased, with increasing cluster severity. In all clusters, <20% of patients received biologics; >15% of patients not receiving biologics were considered eligible for treatment by their physician. The proportion of physicians and patients satisfied with treatment decreased with increasing cluster severity ($P < 0.0001$).

Conclusion. Our large, international, real-world survey of SLE patients and physicians demonstrated strong associations between increased impairment, organ involvement, and humanistic burden in SLE, highlighting an unmet need for effective treatment options in patients with high disease activity.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a chronic autoimmune disease affecting multiple body systems, including the skin, musculoskeletal, cardiovascular, respiratory, renal, hematologic, and central nervous systems (1,2). The clinical presentation of SLE is heterogenous, and the pathogenesis of the disease is complex, with activation of autoreactive T and B cells. The activation of these T and B cells leads to the production of pathogenic autoantibodies, and excessive plasmacytoid dendritic cell activation and interferon production, amplifying the inflammatory response (3,4).

A number of studies have demonstrated clusters of symptoms and clinical and laboratory features of SLE, possibly

Supported by Janssen, Inc. Dr. Touma's work was supported by the Young Operating and the Young Investigator salary award of The Arthritis Society, the New Investigator research grant of the PhysiciaIncorporated Foundation, the CRA (CIORA)–Arthritis Society Clinician Investigator Award, the Rocca Family, and the Kaiser Family. Dr. Costenbader's work was supported by the National Institutes of Health and the Lupus Foundation of America.

[1]Zahi Touma, MD, PhD: University of Toronto, Toronto, Ontario, Canada; [2]Ben Hoskin, MSc, Christian Atkinson, MRes, David Bell, BSc, Olivia Massey, BSc: Adelphi Real World, Bollington, UK; [3]Jennifer H. Lofland, PharmD, MPH, PhD, Pamela Berry, MSc, Chetan S. Karyekar, MD, PhD: Janssen Global Services, Horsham, Pennsylvania; [4]Karen H. Costenbader, MD, MPH: Brigham and Women's Hospital, Boston, Massachusetts.

Mr. Hoskin, Mr. Atkinson, Mr. Bell, and Ms. Massey are employed by Adelphi Real World, and they have received payment from Janssen Inc. Ms. Lofland, Ms. Berry, and Mr. Karyekar are employees and stockholders of Janssen Inc. Dr. Costenbader has received consulting fees from AstraZeneca, GlaxoSmithKline, Neutrolis, and UpToDate (less than $10,000 each), and has received research funding from Merck, Eli Lilly and Company, AstraZeneca, GlaxoSmithKline, and Janssen. No other disclosures relevant to this article were reported.

Address correspondence to Christian Atkinson, MRes, Adelphi Real World, Adelphi Mill, Grimshaw Lane, Bollington, SK10 5JB, UK. Email: christian.atkinson@adelphigroup.com.

Submitted for publication June 5, 2020; accepted in revised form December 17, 2020.

SIGNIFICANCE & INNOVATIONS

- This was an analysis of real-world data collected from 1,376 patients with systemic lupus erythematosus and their physicians in the US and 5 European countries.
- Latent-class cluster analysis identified 4 discrete patient clusters based on patients' symptoms.
- The results suggested an unmet need for effective treatment options in each of the 4 defined clusters.
- Significant differences across clusters in disease characteristics, patient-reported outcomes, and treatments received indicated the highest level of unmet need in the severe cluster of patients with high disease activity and organ involvement.

reflecting varying patterns of disease expression (5–7). However, each of these was a single-center study with limited patient numbers, only 1 study examined differences in patient-reported outcomes, specifically depression and sleep quality, across clusters (7), and only 1 study explored differences between clusters in SLE treatment (6).

The objective of the current study was to identify symptom clusters from a large cohort of SLE patients from the US and 5 European countries using real-world data collected from patients with SLE and their physicians. By exploring differences in demographics characteristics, clinical characteristics, organ involvement, humanistic burden, and treatment across clusters, we aimed to improve the understanding of diverse manifestations of this disease and provide insight into the disease burden and unmet need that can inform decisions on management and treatment of patients with SLE.

PATIENTS AND METHODS

Study design. This was an analysis of data drawn from the Adelphi Real World Lupus Disease Specific Programme (DSP), collected in 2015 in the US and Europe (France, Germany, Spain, Italy, and the UK). The Lupus DSP is a real-world, noninterventional, cross-sectional survey of rheumatologists and their patients with SLE; the full DSP methodology has been published (8).

Rheumatologists from a broad geographic spread across the US and Europe were identified from publicly available lists and invited to participate in the DSP if they were actively managing ≥5 patients with SLE in a typical month. Participating rheumatologists completed a patient record form for the next 5 consecutively consulting patients age ≥18 years with a confirmed diagnosis of SLE. Patients for whom physicians had completed a patient record form were invited to complete a patient self-completion form.

Data collection. Physician-reported data included patient demographic and clinical characteristics, SLE management history, and physician satisfaction with current treatment; information on data reported by the physician is shown in Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24546. Physician-reported data included an indication of which symptoms from a list of 39 signs, symptoms, and laboratory findings the patient was currently experiencing or had previously experienced, together with an option of Other and a specification of what sign/symptom Other referred to. The list was initially developed with input from clinical experts and was refined based on Other signs/symptoms reported during repeated waves of data collection. Information was obtained by the rheumatologists through review of patients' medical records.

Patient-reported data focused on similar data to that reported by the physician, but included patient-reported outcome questionnaires assessing disease burden and a number of questions relating to the patients' feelings about SLE, the impact of the condition on patients' health-related quality of life (HRQoL), and their satisfaction with current treatment for SLE. Details of patient-reported data are shown in Supplementary Table 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24546. The patient-reported outcome questionnaires included were the EuroQoL 5-domain 3-level questionnaire (EQ-5D-3L), the Functional Assessment of Chronic Illness Therapy–Fatigue (FACIT-F) scale, and the Work Productivity and Activity Impairment questionnaire; information on these questionnaires is available in Supplementary Table 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24546.

Statistical analysis. To determine the underlying structure of SLE signs/symptoms (i.e., to identify a small number of sign/symptom factors), principal-component factor analysis was used to extract linear composites of the SLE signs/symptoms identified and reduce them into uncorrelated linear combinations of symptoms that accounted for the majority of variance (9). Factor solutions were derived for 6, 7, and 8 factors. Additionally, a final factor solution was obtained by retaining all factors with an eigenvalue ≥0. To facilitate the interpretation of the factor loadings, orthogonal varimax rotation was performed. A pragmatic interpretation of each of the 4 loading matrices (that show the correlations between factors and signs/symptoms) was made by identifying which signs/symptoms relate to which factors (those with a high correlation), and the model with the most intuitive interpretation was selected. Factor variables were then created based on the interpreted relationships from the loading matrix, i.e., each factor was taken to be the mean of the related signs/symptoms. These symptom factors were used as covariates in latent-class cluster analysis, to define discrete clusters of patients. Cluster solutions were created for varying numbers of clusters (1 to 8). The Bayes information criterion

Table 1. Demographic and clinical characteristics by cluster*

Characteristic	Overall (n = 1,376)	Very mild (n = 325)	Mild (n = 743)	Moderate (n = 143)	Severe (n = 165)	P†
Age, years, no.	1,375	324	743	143	165	0.4318
Mean ± SD	42.1 ± 13.6	41.2 ± 13.7	42.6 ± 13.7	41.6 ± 13.3	42.0 ± 13.0	–
Female, no./total no. (%)	1,196/1,374 (87)	276/324 (85)	644/742 (87)	125/143 (87)	151/165 (92)	0.2640
Ethnicity, no.	1,363	323	735	141	164	0.2758
White	1,003 (74)	230 (71)	560 (76)	94 (67)	119 (73)	–
African ancestry	183 (13)	43 (13)	95 (13)	21 (15)	24 (15)	–
Hispanic/Latino	85 (6)	22 (7)	41 (6)	11 (8)	11 (7)	–
Other‡	92 (7)	28 (9)	39 (5)	15 (11)	10 (6)	–
Employment status, no.	1,329	308	724	136	161	–
Working full-time	566 (42)	150 (49)	305 (42)	58 (42)	53 (33)	0.0121
Working part-time	170 (13)	34 (11)	93 (13)	14 (10)	29 (18)	0.1360
Student	99 (7)	27 (9)	55 (8)	7 (5)	10 (6)	0.5317
Homemaker	201 (15)	49 (16)	107 (15)	25 (18)	20 (12)	0.5211
Retired	142 (11)	33 (11)	83 (12)	13 (10)	13 (8)	0.6169
Unemployed	160 (12)	16 (5)	87 (12)	20 (15)	37 (23)	<0.0001
Time since diagnosis, years, no.	1,351	320	731	141	159	0.2021
Mean ± SD	5.5 ± 6.1	5.9 ± 6.6	5.4 ± 6.0	4.7 ± 5.8	6.0 ± 6.0	–
Current symptoms, no.	1,376	325	743	143	165	<0.0001
Mean ± SD	4.8 ± 4.1	0.6 ± 0.9	5.2 ± 2.7	5.1 ± 2.9	11.4 ± 4.3	–
Current organs affected, no.	1,376	325	743	143	165	<0.0001
Mean ± SD	2.3 ± 1.4	1.4 ± 1.2	2.4 ± 1.2	2.6 ± 1.3	3.6 ± 1.5	–
Disease activity index, no.§	1,327	312	718	138	159	–
Scored with disease activity index	443 (33.4)	105 (33.7)	241 (33.6)	47 (34.1)	50 (31.4)	0.9567
SLEDAI	373 (28.1)	87 (27.9)	202 (28.1)	41 (29.7)	43 (27.0)	0.9651
Current disease severity (patient perceived), no.¶	852	195	473	81	103	–
Mild	516 (61)	164 (84)	260 (55)	51 (63)	41 (40)	<0.0001
Moderate	268 (32)	29 (15)	171 (36)	20 (25)	48 (47)	<0.0001
Severe	68 (8)	2 (1)	42 (9)	10 (12)	14 (14)	<0.0001
Current disease severity (physician perceived), no.	852	195	473	81	103	–
Mild	594 (70)	183 (94)	305 (65)	55 (68)	51 (50)	<0.0001
Moderate	224 (26)	11 (6)	153 (32)	18 (22)	42 (41)	<0.0001
Severe	34 (4)	1 (1)	15 (3)	8 (10)	10 (10)	<0.0001
Concordance of patient/physician perceived disease severity, no.	852	195	473	81	103	–
Agree	678 (79.6)	170 (87.2)	365 (77.2)	70 (86.4)	73 (70.9)	0.008
Patient rates more severe	136 (16.0)	22 (11.3)	84 (17.8)	8 (9.9)	22 (21.4)	0.008
Physician rates more severe	38 (4.5)	3 (1.5)	24 (5.1)	3 (3.7)	8 (7.8)	0.008
Current disease progression (physician perceived), no.	1,375	325	742	143	165	–
Improving	372 (27.1)	115 (35.4)	188 (25.3)	42 (29.4)	27 (16.4)	<0.0001
Stable	788 (57.3)	199 (61.2)	435 (58.6)	75 (52.4)	9 (47.9)	<0.0001
Deteriorating slowly	160 (11.6)	8 (2.5)	96 (12.9)	14 (9.8)	42 (25.5)	<0.0001
Deteriorating rapidly	15 (1.1)	2 (0.6)	4 (0.5)	3 (2.1)	6 (3.6)	<0.0001
Unstable	40 (2.9)	1 (0.3)	19 (2.6)	9 (6.3)	11 (6.7)	<0.0001
Currently experiencing a flare, no./total no. (%)	88/1,371 (6.1)	7/324 (2.2)	43/739 (5.8)	12/143 (8.4)	26/165 (15.8)	<0.0001
≥1 flare in past 12 months, no./total no. (%)	430/804 (54)	45/150 (30)	242/442 (55)	51/82 (62)	92/130 (71)	<0.0001
No. of flares in past 12 months (all patients), no.	1,365	323	737	142	163	
Mean ± SD	0.7 ± 1.3	0.3 ± 0.9	0.7 ± 1.2	0.6 ± 0.9	1.5 ± 2.0	<0.0001
No. of flares in past 12 months (patients with ≥1 flare), no.	424	44	240	50	90	
Mean ± SD	2.2 ± 1.5	2.0 ± 1.7	2.1 ± 1.4	1.7 ± 0.8	2.7 ± 2.0	0.0013
Patients also diagnosed with lupus nephritis, no./total no. (%)	223/1,376 (16)	44/325 (14)	104/743 (14)	37/143 (26)	38/165 (23)	0.0002

* Values are the number (%) unless indicated otherwise. SLEDAI = Systemic Lupus Erythematosus Disease Activity Index.
† Comparing across clusters.
‡ Includes ethnicities comprising <3% of patients overall: Native American, Middle Eastern, Chinese, Asian–Indian, Asian–other, and mixed race.
§ Patients reported to have been scored using a valid disease activity index at some time; actual scores were not collected; indices include British Isles Lupus Assessment Group Assessment, SLEDAI, European Consensus Lupus Activity Measurement, Systemic Lupus Erythematosus Responder Index, Cutaneous Lupus Erythematosus Disease Area and Severity Index, and Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index.
¶ Patient-reported data, with matching physician-reported data.

was used to determine which cluster solutions had similar goodness-of-fit, and then the most appropriate solution was selected.

The number of patients in each cluster, based on rheumatologist-reported data on specific organ/body area (musculoskeletal, mucocutaneous, neuropsychiatric, cardiorespiratory,

gastrointestinal, ophthalmic, renal, constitutional, or hematologic) involvement, was calculated. The level of anxiety/depression experienced by patients was taken from their responses to the relevant item on the EQ-5D-3L. The impact of SLE on patients' emotional state was taken from their response to the question "During the past seven days, how much impact has lupus had on your emotional state?" (responses were on a numeric rating scale ranging 0–10, where 0 = no impact, 10 = highest impact). Treatments were analyzed on the basis of the treatment class to which they belonged: antimalarials, glucocorticoids, immunosuppressants, nonsteroidal antiinflammatory drugs (NSAIDs), and biologics. Chi-square and Kruskal-Wallis statistical tests were conducted to compare physician-reported data and patient-reported outcomes and treatment across these clusters. All analyses were conducted in Stata software, version 15.1 or later (10).

Ethical considerations. The DSP adheres to the European Pharmaceutical Market Research Association code of conduct. This code of conduct states that research meeting the definition relating to market or consumer behavior of the sort that pharmaceutical companies routinely commission, whether involving health care professionals, patients, carers, or members of the public, does not require Clinical Research Ethics Committee or Independent Review Board approval.

All data collection from the DSP was undertaken through third-party fieldwork agencies, ensuring that the identity of health care professionals and patients was not known to Adelphi Real World or any subscribers to the data. Furthermore, data were analyzed and provided to subscribers in an aggregated format. Patients who participated in the study provided consent for their self-completion data to be used by selecting a checkbox on the patient self-completion form and by returning the form for use. Physicians provided written consent for their data to be used via the online survey they completed. Physicians were paid a fair market rate for their time involved in completing the survey.

RESULTS

Participants. Rheumatologists in Europe and the US reported data for a total of 1,376 patients; demographic and clinical characteristics for the total study population are shown in Table 1. Patients' mean age was 42 years, most patients were female (87%) and White (74%), and over half of patients (55%) were in full- or part-time employment. The majority of patients currently had mild SLE as reported by both physicians (70%) and patients (61%), and their SLE was considered to be stable by their physician. Over half of patients (54%) had experienced a flare in the 12 months prior to data collection. Physician-reported data were available for 502 patients from the US and 874 patients from Europe (France, n = 175; Germany, n = 175; Spain, n = 174; Italy, n = 175; UK, n = 175). Data were provided by 859 patients, 311 patients from the US and 548 patients from Europe (France, n = 97; Germany, n = 141; Spain, n = 107; Italy, n = 126; UK, n = 77). Differences were observed in demographic and clinical characteristics across countries (see Supplementary Table 4, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24546).

Identification of symptom clusters. A total of 39 unique SLE symptoms were identified and first reduced to 8 factors with common, concurrent symptoms, using principal-component factor analysis (see Supplementary Table 5, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24546). Latent-class cluster analysis provided 4 discrete patient clusters (Figure 1). Of the 1,376 patients, 24%, 54%, 10%, and 12% were included in the very mild, mild, moderate, and severe clusters, respectively. The proportions of patients who comprised each cluster were comparable between the US and Europe. Small differences in proportions in different clusters were observed in individual countries compared to global or regional clusters: fewer patients in the very mild and severe clusters, and more patients in the mild cluster in France, Germany, and Italy; more patients in the very mild cluster and fewer patients in the mild cluster in Spain and the UK; and more patients in the severe cluster in the UK (see Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24546). Based on patients with time since the diagnosis recorded, the proportions of patients in each cluster were broadly similar in early and established disease, although patients diagnosed for less than a year were numerically more likely to be in the very mild cluster and less likely to be in the mild cluster than patients with more established disease (see Supplementary Figure 2, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24546).

The very mild cluster had the lowest symptom burden, characterized by skin involvement, while the mild cluster was characterized by joint and skin involvement (Figure 1). The moderate and severe clusters had heavier symptom burdens than the very mild and mild clusters (Figure 1), with high skin and joint manifestations, particularly in the severe cluster. There was high cardiovascular involvement in the moderate cluster, and renal and constitutional/mental factor involvement were high in the severe cluster, relative to other clusters (Figure 1).

An analysis of the organs or areas of the body affected by SLE for each patient showed good alignment with the factor analysis, with significant differences across clusters in the proportions of patients considered by their physician to have musculoskeletal, mucocutaneous, neuropsychiatric, cardiorespiratory, gastrointestinal, ophthalmic, renal, constitutional, and hematologic involvement ($P < 0.0001$, except gastrointestinal, $P < 0.01$).

Demographic and clinical characteristics by cluster. Demographic and clinical characteristics for each cluster are shown in Table 1. Patients were comparable across clusters for

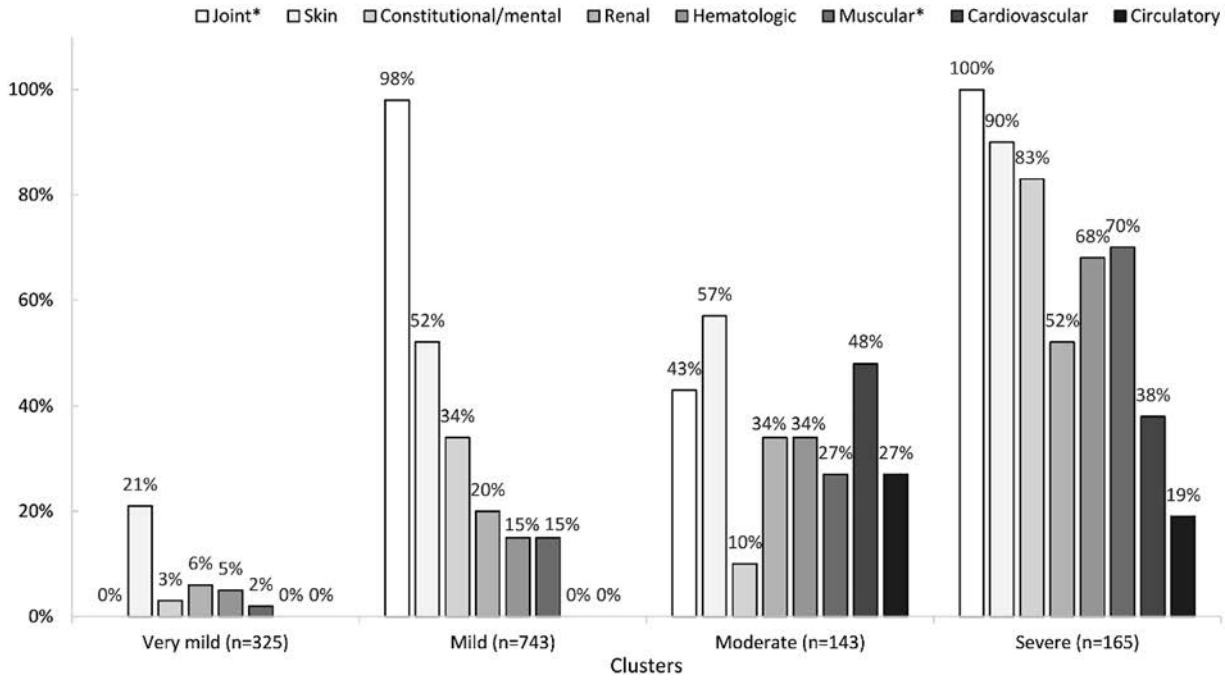


Figure 1. Identification of symptom clusters, with overall cluster composition. * = joint factor characterized by joint tenderness, joint stiffness, and joint swelling; muscular factor characterized by muscle inflammation (see Supplementary Table 5, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24546).

most demographic data; however, there was a significant difference in the proportion of unemployed individuals across clusters ($P < 0.0001$), with higher proportions of patients unemployed as cluster severity increased.

Time since diagnosis was similar across clusters. The mean number of symptoms experienced at the time of data collection differed significantly across clusters ($P < 0.0001$), with patients having increasing numbers of symptoms with increasing cluster severity. Higher proportions of patients with a diagnosis of lupus nephritis were observed in the moderate and severe clusters, where higher rates of renal involvement were found, compared with the very mild and mild clusters. Physicians reported that SLE disease activity had been captured for approximately one-third of their patients on a validated SLE disease activity index at some time (prior to the study or at the clinic visit when data were collected for the study), and that fewer than 30% had been scored using the Systemic Lupus Erythematosus Disease Activity Index; scores from disease activity indices were not collected. Physicians and patients perceived a significant difference across clusters in disease severity ($P < 0.0001$); disease severity increased with increasing cluster severity. Physicians and patients agreed on disease severity in ~80% of cases, with higher concordance in the very mild and moderate clusters than in the mild and severe clusters. There was a trend for patients to perceive their SLE as more severe than their physicians, with 40% of patients overall considering their disease as moderate or severe, compared with 30% of physicians; a higher proportion of patients compared with physicians reported moderate/severe SLE across all clusters. There was a significant difference across clusters in physician perception of disease progression ($P < 0.0001$), with a lower proportion of patients considered by their physicians to be improving, and a higher proportion of patients considered to be deteriorating or unstable, with increasing cluster severity. The proportion of patients experiencing a flare at the time of data collection, the proportion of patients who had experienced at least 1 flare in the past 12 months and the mean number of flares all differed significantly across clusters ($P < 0.0001$), with flares increasing with increasing cluster severity.

Patient-reported outcomes. Significant differences were seen across clusters for all patient-reported outcome questionnaire end points ($P < 0.0001$) (Figure 2). The EQ-5D-3L health utility index decreased, indicating poorer health status, across clusters from the very mild to the severe cluster (Figure 2A). Patients in the severe cluster recorded lower EQ-5D-3L visual analog scale (VAS) scores, indicating worse health status, than those in the very mild cluster; patients in the mild and moderate clusters had similar scores that were between those of the very mild and severe clusters (Figure 2B). The FACIT-F score was lower in the severe cluster, indicating greater fatigue, than in the very mild cluster; the mild and moderate clusters had similar scores that were between those of the very mild and severe clusters (Figure 2C). The level of overall work productivity impairment due to SLE increased with increasing cluster severity (Figure 2D). Similar trends in patient-reported outcome end points across clusters were seen when data were analyzed on a regional or country level (data not shown).

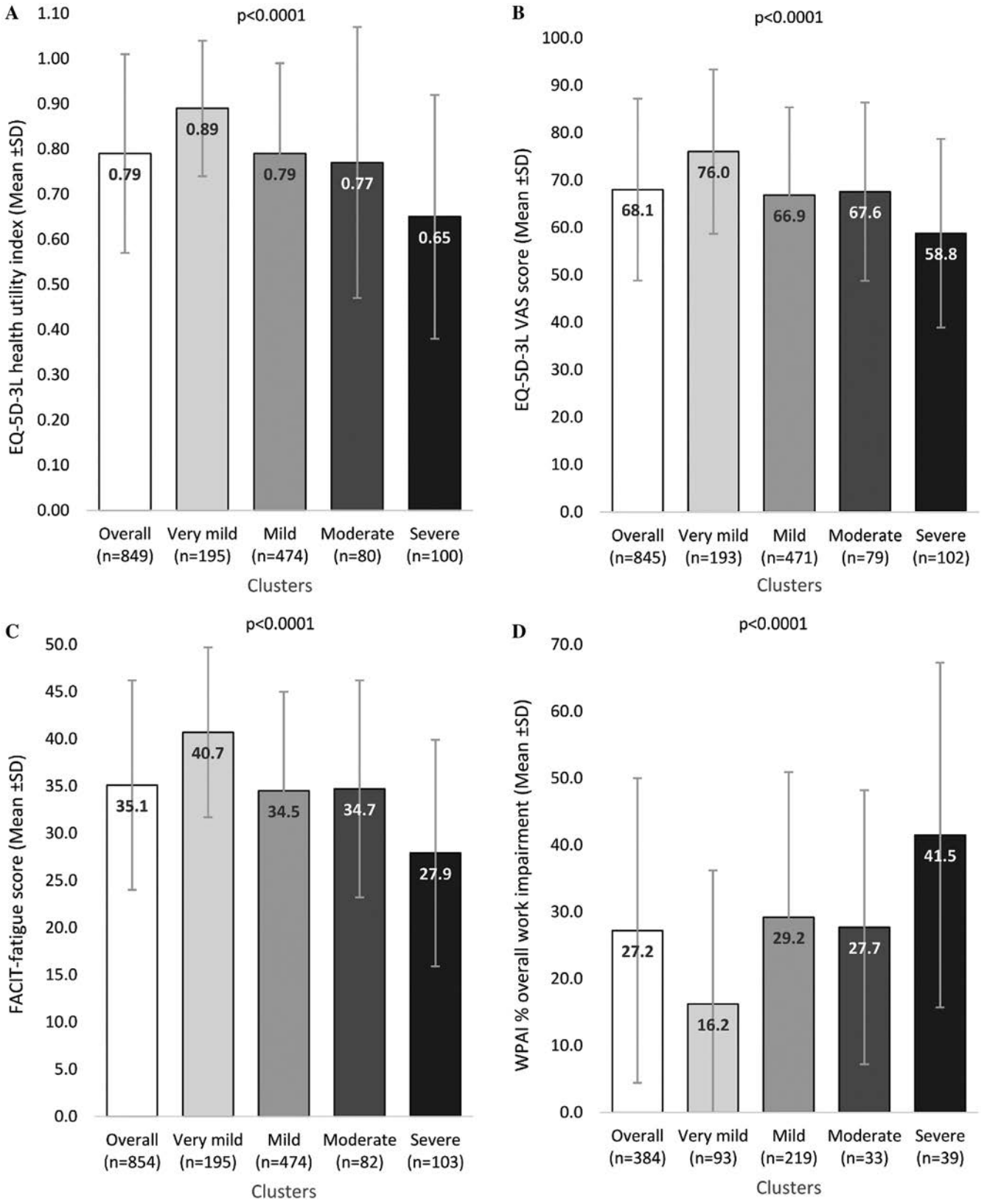


Figure 2. Patient-reported outcomes: **A**, EuroQoL 5-domain 3-level questionnaire (EQ-5D-3L) health utility index (maximum score of 1, higher score more favorable); **B**, EQ-5D-3L visual analog scale (VAS) score (maximum score of 100, higher score more favorable); **C**, Functional Assessment of Chronic Illness Therapy–Fatigue scale (FACIT-fatigue) score (maximum score of 52, higher score more favorable); **D**, Work Productivity and Activity Impairment questionnaire (WPAI) overall work impairment due to SLE (percentage, lower score more favorable). Error bars indicate SE.

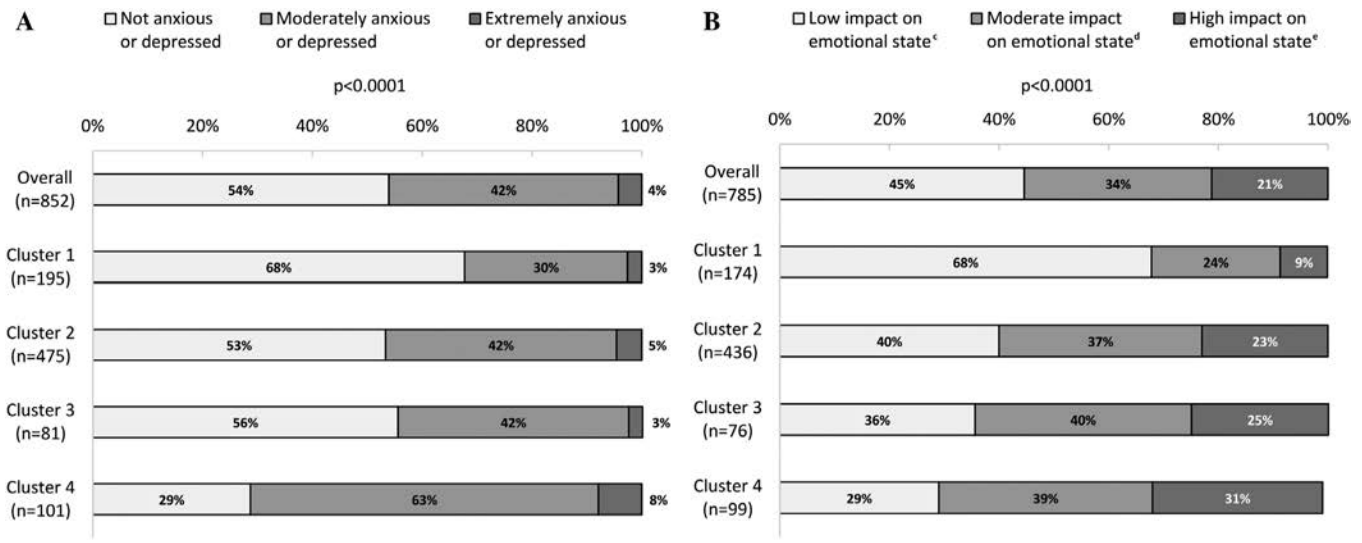


Figure 3. Psychological outcomes: **A**, Anxiety/depression, response to anxiety/depression item on EuroQoL 5-domain 3-level questionnaire; **B**, Impact on emotional state, response to question "During the past 7 days, how much impact has lupus had on your emotional state?" where 0 = no impact, 10 = high impact; [c]Response 0–3; [d]Response 4–6; [e]Response 7–10. *P* value reflects significant differences for distribution of all values, not aggregate levels.

The impact of SLE on patients' psychological well-being differed significantly across clusters ($P < 0.0001$) (Figure 3). Patients in the severe cluster reported higher levels of anxiety/depression than those in the very mild cluster; patients in the mild and moderate clusters reported similar levels that were between those reported in the very mild and severe clusters (Figure 3A). Patients reported an increasing impact of SLE on their emotional state with increasing cluster severity (Figure 3B).

Treatment. At the time of data collection, the number of individual treatment regimens prescribed across clusters differed statistically, with patients in the moderate and severe clusters prescribed a higher mean number of treatments than those in the very mild and mild clusters ($P < 0.05$) (Figure 4A). When considering treatments prescribed to ≥10% of patients overall, regardless of whether they were taken as monotherapy or in combination with other treatment, significant differences were observed across clusters for all treatments ($P < 0.01$) (Figure 4B). Use of glucocorticoids and immunosuppressants increased with increasing cluster severity ($P < 0.0001$ for both). Antimalarial use was the highest in the very mild cluster (71% of patients), with similar proportions of patients in all other clusters (56–61%) receiving this form of treatment. Biologics and NSAIDs were prescribed in <20% of patients in all clusters, with no clear pattern to usage across clusters (Figure 4B). Analysis of the types of treatment showed generally similar patterns across countries, although there were some variations. In Spain, no patient in the severe cluster was receiving a biologic, and in Italy only 1 patient in the mild cluster and 2 in the moderate cluster were receiving a biologic. Also, >90% of patients in all clusters in Spain were receiving steroids.

A high number of treatments were taken in combination with a treatment of a different class; treatment combinations differed significantly across clusters ($P < 0.0001$) (Figure 4C). The highest proportions of patients receiving both antimalarials and immunosuppressants as monotherapy were in the very mild cluster, while glucocorticoid monotherapy was the most common in the moderate cluster (Figure 4C). The highest proportions of patients receiving glucocorticoid plus immunosuppressant combination therapy, and antimalarial plus glucocorticoid plus immunosuppressant combination therapy, were in the severe cluster (Figure 4C).The levels of satisfaction with treatment reported by physicians and patients were broadly similar to each other, but differed significantly across clusters for both physicians and patients ($P < 0.0001$) (see Supplementary Figure 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24546). The proportion of physicians and patients who were satisfied (total of very satisfied, satisfied, or somewhat satisfied) decreased with increasing cluster severity, with 90% and 91% of physicians and patients, respectively, in the very mild cluster being satisfied, compared with 58% and 65%, respectively, in the severe cluster (see Supplementary Figure 3, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24546).

For patients not receiving a biologic, in response to the question "Does this patient's overall condition warrant a biologic?" physicians confirmed the patient's eligibility for biologic treatment in 15%, 25%, 31%, and 42% of patients in the very mild, mild, moderate, and severe clusters, respectively.

DISCUSSION

This study used international, real-world data to generate evidence supporting the concept of discrete clusters of SLE

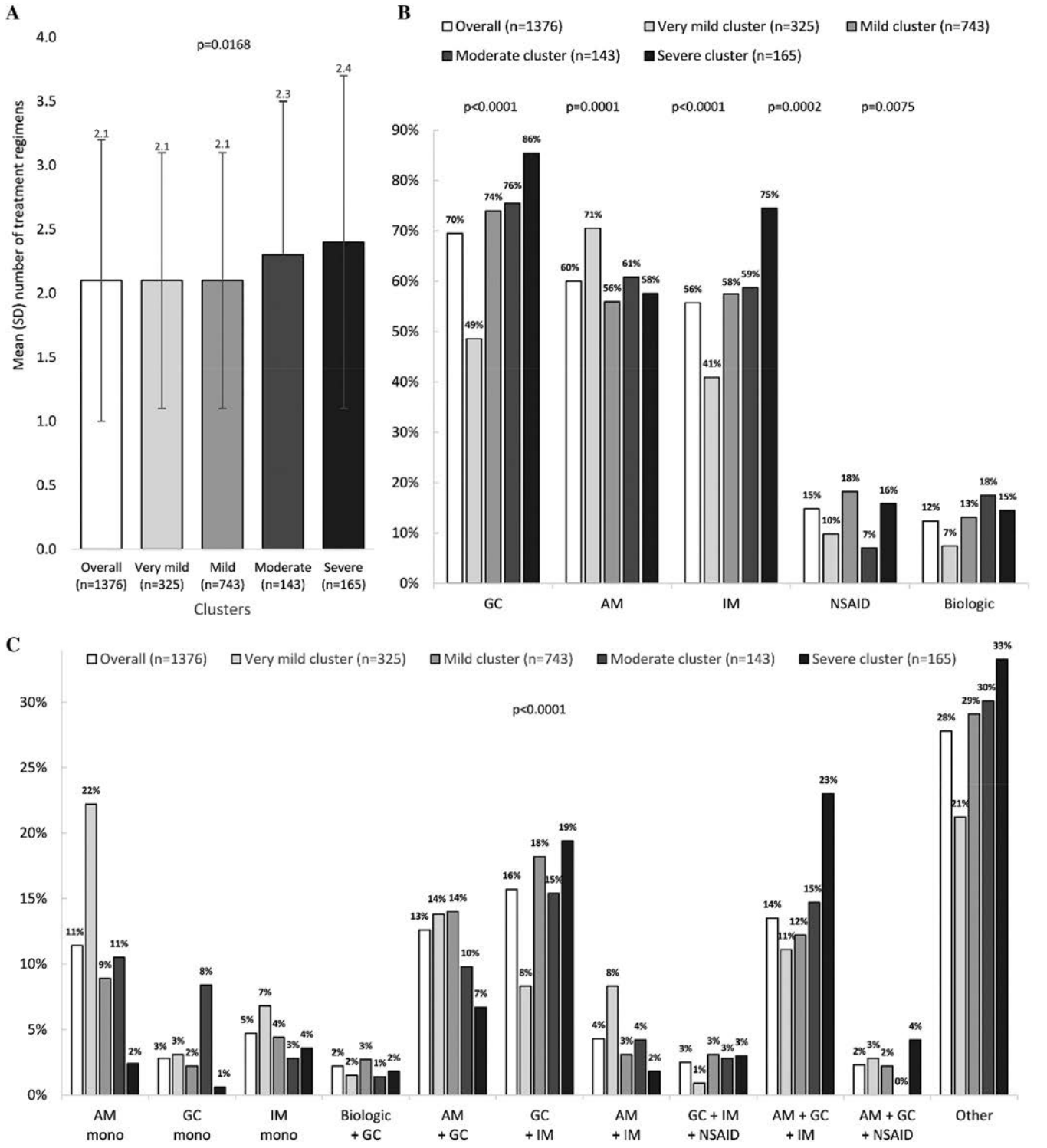


Figure 4. Treatment: **A**, Number of current treatment regimens; **B**, Individual treatments currently prescribed, any treatment class prescribed to ≥10% of patients overall; **C**, Treatment combinations currently prescribed. Error bars indicate SE. AM = antimalarial; GC = glucocorticoid; IM = immunosuppressant; mono = monotherapy; NSAID = nonsteroidal antiinflammatory drug.

symptoms, with varying severity and corresponding medication use, treatment satisfaction, and patient-reported outcomes. Our analysis demonstrated that this large cohort of patients with SLE could be grouped into 4 distinct symptom clusters, despite the diverse presentation of this complex autoimmune disease. The clusters identified in our study had a pattern of worsening severity, disease burden, and progression moving from a very mild to a severe cluster. The patients comprising the severe cluster

experienced higher symptom burden, poorer health status, greater fatigue, and reduced productivity relative to patients in other clusters.

Previous studies have reported clusters of some clinical, hematologic, and immunologic features of SLE (5), reported 5 symptom clusters in patients with childhood-onset SLE (7), and identified 3 clusters based on laboratory data (6). Thus, our analysis focused on the burden of symptoms across organs or areas of the body, while other studies have identified clusters based on organ involvement (clinical and serologic laboratory findings). An analysis of data from the International Early Lupus Cohort and the Euro-Lupus Cohort studied the interrelationships between different manifestations of the same organ system (11,12). In that study, the clusters were derived based on the interaction between different manifestations of different organ systems. The scope/methodology of the current study and that from Touma et al (12) differ significantly, thus, comparing the results of the studies is not possible. However, this difference in methodologies and the findings thereof raises a very interesting research question that could be explored in a future study, and the interrelationships demonstrated in the International Early Lupus Cohort and the Euro-Lupus Cohort could be further validated in a large, international real-world data study collected from patients and physicians.

A published analysis of SLE manifestations at disease onset reported that the majority of patients have some form of skin rash (13). In our analysis, patients diagnosed for less than a year were predominantly in the very mild cluster in which skin involvement was the most common manifestation. However, our analysis showed patients in the very mild cluster to have no joint involvement, while the published study of newly diagnosed patients reported inflammatory arthritis in almost 60% of patients.

The EQ-5D-3L health utility index in our analysis ranged from 0.80 in the very mild cluster to 0.65 in the severe cluster, and the EQ-5D-3L VAS score from 76.0 in the very mild cluster to 58.8 in the severe cluster, indicating that patients in all clusters had poorer health status than the general population in the countries in which the research was conducted (14). Population norms for the FACIT-F scale have been published for Germany, with a mean score of 43.5 reported (15); our finding of 40.7 in the very mild cluster was similar to this result but findings in the mild, moderate, and severe clusters were 34.5, 34.7, and 27.9, respectively, suggesting that patients in these clusters experienced more fatigue than the general population.

The high symptom burden, occurrence of flares, and impact reported by patients speak to a substantial degree of unmet need across all patient clusters, although many patients were receiving multiple treatments. Treatment goals in SLE are achievement of remission, or at least low disease activity, and prevention of flares (16). Although the introduction of glucocorticoids and immunosuppressants to treat SLE in the 1990s resulted in markedly improved prognosis, some patients remain refractory to these treatments, which are also associated with risks of side effects, particularly with prolonged use (17,18). Treatment guidelines recommend hydroxychloroquine in almost all patients, with minimized chronic maintenance doses of glucocorticoids, immunomodulators as needed, and add-on biologics belimumab (in active or flaring extrarenal SLE) and rituximab (in organ-threatening, refractory disease, although this use is currently off-label) (16). Therefore, there remains a need for treatments that are safe even when administered repeatedly and effective in controlling symptoms (17). Overall, only 12% of patients in our study were receiving biologics, with <20% of patients even in the moderate and severe clusters on biologic treatment, despite >30% of the remaining patients in these clusters considered eligible for a biologic by their physician. Thus, these data highlight the fact that there is a large population (the moderate and severe clusters) of SLE patients who have a need for improved treatment options that can impact their HRQoL, treatment satisfaction, and ultimately their clinical outcome.

It is important to acknowledge some limitations of this research. Data were derived by a survey-based methodology, with physicians providing data relating to the next 5 consulting patients. This pragmatic approach identified a large patient population consisting predominantly of White, female patients and may have excluded relevant types of patients not currently consulting their rheumatologist. Although the sex distribution of our population reflected the finding that globally SLE is more prevalent in female patients than male patients, the preponderance of White patients did not reflect the fact that globally the highest prevalence of SLE is found in people of African ancestry and the lowest in those of White ethnicity (19,20). The under-representation of patients of African ancestry is difficult to explain, as our methodology aimed to recruit a geographically representative sample with a mix of rural/urban and private/public clinics involved. Care must therefore be taken in extending our findings to a broader SLE population.

Patients had a rheumatologist-confirmed diagnosis of SLE. The European Alliance of Associations for Rheumatology/American College of Rheumatology classification criteria for SLE were not collected; however, the patients included were reflective of the population managed in clinical practice as patients with SLE. Factors were derived from the signs, symptoms, and laboratory findings that physicians indicated patients were experiencing from a list of 39; although this list was developed with input from clinical experts and fine-tuned based on data reported in early waves of data collection, some signs/symptoms were possibly omitted from the list and not captured. Comparison of disease activity across the clusters was considered of interest, but few patients had scores from validated measures of disease activity, reflecting the focus of such measures in clinical research rather than real-world clinical practice (21). However, we believe that organ involvement and symptomatology, together with physician-reported disease severity and disease progression status, provide an acceptable proxy for disease activity.

This was a cross-sectional, rather than longitudinal, analysis, so that it allowed identification of associations, rather than causal

relationships, between variable factors and outcomes of interest. As with all observational research of this type, the quality of data depended largely on the accurate reporting of information by physicians (using available patient records) and patients with SLE. Finally, although minimal exclusion criteria governed the selection of physicians, physician inclusion depended upon willingness to participate in this type of detailed research.

In conclusion, this analysis provided evidence of the clustering of patients with similar manifestations in SLE. We identified 4 clusters based on the burden of SLE symptoms across multiple organs. Careful consideration of clinical features and review of laboratory markers throughout the disease course may be useful to inform patient management and treatment decisions. Additional analysis is needed to better understand treatment dynamics and the limited use of biologics, particularly in patients with more severe disease.

ACKNOWLEDGMENTS

Medical writing support under the guidance of the authors was provided by Carole Evans, PhD, on behalf of Adelphi Real World, and was funded by Janssen Inc., in accordance with Good Publication Practice guidelines. Proofreading and quality control were provided by Katerina Doslikova, an employee of Adelphi Real World.

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. Mr. Atkinson 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. Hoskin, Bell.

Acquisition of data. Hoskin, Bell.

Analysis and interpretation of data. Touma, Hoskin, Atkinson, Bell, Massey, Lofland, Berry, Karyekar, Costenbader.

ROLE OF THE STUDY SPONSOR

Data collection was undertaken by Adelphi Real World as part of an independent survey, entitled the Adelphi Lupus DSP. The analysis described here used data from the Adelphi Lupus DSP. The DSP is a wholly owned Adelphi product. Janssen Inc. is one of multiple subscribers to the DSP. Janssen Inc. did not influence the original survey through either contribution to the design of questionnaires or data collection. Publication of this article was not contingent upon approval by Janssen Inc.

REFERENCES

1. Mayo Clinic. Lupus. 2017. URL: https://www.mayoclinic.org/diseases-conditions/lupus/symptoms-causes/syc-20365789.
2. Kaul A, Gordon C, Crow MK, Touma Z, Urowitz MB, van Vollenhoven R, et al. Systemic lupus erythematosus. Nat Rev Dis Primers 2016;2:16039.
3. Lo MS, Tsokos GC. Recent developments in systemic lupus erythematosus pathogenesis and applications for therapy. Curr Opin Rheumatol 2018;30:222–8.
4. Choi J, Kim ST, Craft J. The pathogenesis of systemic lupus erythematosus: an update. Curr Opin Immunol 2012;24:651–7.
5. Font J, Cervera R, Ramos-Casals M, García-Carrasco M, Sents J, Herrero C, et al. Clusters of clinical and immunologic features in systemic lupus erythematosus: analysis of 600 patients from a single center. Semin Arthritis Rheum 2004;33:217–30.
6. Suh CH, Jung JY, Lee HY, Kim HA, Kim SS, Hong HA. Hierarchical cluster analysis of systemic lupus erythematosus. Lupus Sci Med 2017;4. doi: https://doi.org/10.1136/lupus-2017-000215.235.
7. Chiang YC, Huang JL, Wang CH, Lee HC, Lee MY, Hsiao YC. Symptom clustering in patients with childhood-onset systemic lupus erythematosus. J Adv Nurs 2019;75:5462.
8. Anderson P, Benford M, Harris N, Karavali M, Piercy J. Real-world physician and patient behaviour across countries: disease-specific programmes: a means to understand. Curr Med Res Opin 2008;24:3063–72.
9. Mardia KV, Kent JT, Bibby JM. Multivariate analysis. London: Academic Press; 1979.
10. StataCorp. 2017. Stata statistical software: release 15. College Station, TX: StataCorp LLC.
11. European Society for Opinion and Marketing Research. International code on market, opinion and social research and data analytics. 2016. URL: https://esomar.org/guidance.
12. Touma Z, Cervera R, Brinks R, Lorenzoni V, Tani C, Hoyer BF, et al. Associations between classification criteria items in systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2020;72:1820–6.
13. Mosca M, Costenbader KH, Johnson SR, Lorenzoni V, Sebastiani GD, Hoyer BF, et al. How do patients with newly diagnosed systemic lupus erythematosus present? A multicenter cohort of early systemic lupus erythematosus to inform the development of new classification criteria. Arthritis Rheumatol 2019;71:91–8.
14. Szende A, Oppe M, Devlin N. EQ-5D value sets: inventory, comparative review and user guide. New York: Springer; 2007.
15. Montan I, Löwe B, Cella D, Mehnert A, Hinz A. General population norms for the Functional Assessment of Chronic Illness Therapy (FACIT)-fatigue scale. Value Health 2018;21:1313–21.
16. Fanouriakis A, Kostopoulou M, Alunno A, Aringer M, Bajema I, Boletis JN, et al. 2019 update of the EULAR recommendations for the management of systemic lupus erythematosus. Ann Rheum Dis 2019;78:736–45.
17. Bakshi J, Segura BT, Wincup C, Rahman A. Unmet needs in the pathogenesis and treatment of systemic lupus erythematosus. Clin Rev Allergy Immunol 2018;55:35267.
18. Tan MK, Heng TY, Mak A. The potential use of metformin, dipyridamole, N-acetylcysteine and statins as adjunctive therapy for systemic lupus erythematosus. Cells 2019;8:E323.
19. Rees F, Doherty M, Grainge MJ, Lanyon P, Zhang W. The worldwide incidence and prevalence of systemic lupus erythematosus: a systematic review of epidemiological studies. Rheumatology (Oxford) 2017;56:1945–61.
20. Stojan G, Petri M. Epidemiology of systemic lupus erythematosus: an update. Curr Opin Rheumatol 2018;30:144–50.
21. Mikdashi J, Nived O. Measuring disease activity in adults with systemic lupus erythematosus: the challenges of administrative burden and responsiveness to patient concerns in clinical research. Arthritis Res Ther 2015;17:183.

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1089–1097
DOI 10.1002/acr.24550

American College of Rheumatology
Empowering Rheumatology Professionals

Impact of Antimalarial Adherence on Mortality Among Patients With Newly Diagnosed Systemic Lupus Erythematosus: A Population-Based Cohort Study

M. Rashedul Hoque,[1] J. Antonio Aviña-Zubieta,[2] Mary A. De Vera,[2] Yi Qian,[3] John M. Esdaile,[2] and Hui Xie[1]

Objective. To assess the association of antimalarial (AM) adherence with premature mortality among incident systemic lupus erythematosus (SLE) patients.

Methods. All patients with incident SLE and incident AM use in British Columbia, Canada, between January 1997 and March 2015 were identified using the provincial administrative databases. Follow-up started on the first day of having both SLE and AM. The outcome was all-cause mortality. An adherence measure, proportion of days covered (PDC), was calculated and categorized as adherent (PDC ≥ 0.90), nonadherent (0 < PDC < 0.90), and discontinuer (PDC = 0) during 30-day windows. We first used Cox models for time-to-death, adjusting for baseline and time-varying confounders on medication usages, health care utilization, and comorbidities. We then used marginal structural Cox models via inverse probability weighting designed for causal inference with time-varying confounders to assess the effect of AM adherence on premature mortality.

Results. We identified 3,062 individuals with incident SLE and incident AM use (mean age 46.9 years). Over the mean follow-up period of 6.4 years, 242 (7.9%) of those patients died. Adjusted hazard ratios (HR_{adj}) from the Cox model for AM adherent and nonadherent SLE patients were 0.20 (95% confidence interval [95% CI] 0.13–0.29) and 0.62 (95% CI 0.42–0.91), respectively, compared to discontinuers. The corresponding HR_{adj} from the marginal structural Cox model were 0.17 (95% CI 0.12–0.25) and 0.58 (95% CI 0.40–0.85), respectively. A significant trend in the HR_{adj} of mortality risk over the adherence levels was found ($P < 0.001$).

Conclusion. Patients with SLE adhering to AM therapy had a 71% and 83% lower risk of death than patients who do not adhere or who discontinued AMs, respectively.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a chronic inflammatory autoimmune disease that leads to excessive morbidity, including serious organ complications and premature mortality (1–3). A population-based UK study conducted from 1999 to 2014 found that survival in SLE relative to the general population had not improved in recent years (4). Adherence to prescribed medications is considered key in avoiding complications and death in SLE (5). A recent systematic review confirmed that SLE patients too often were not adhering to medication regimens (6).

Antimalarial (AM) medication is considered the first-line drug in SLE management for most patients (7). AM medications have been shown to improve SLE symptoms and to reduce inflammation of the lining of the heart and lung, the development of nephritis, central nervous system impairment, and flares in disease activity (8). The basis for the benefits of AMs includes antiinflammatory, antihyperlipidemic, antithrombotic, and antihyperglycemic effects (8–11). However, it is important that patients follow AM regimens as prescribed. A recent Canadian population-based study found a protective effect of AM adherence for SLE patients in preventing type 2 diabetes mellitus (9). Similarly, a US study

All inferences, opinions, and conclusions herein are those of the authors and do not reflect the opinions or policies of the Data Stewards of PopData BC.

Supported by Arthritis Research Canada (PRECISION grant), the CIHR (team grant THC-135235), and the Natural Sciences and Engineering Research Council of Canada (grant RGPIN-2018-04313).

[1]M. Rashedul Hoque, MSc, Hui Xie, PhD: Arthritis Research Canada, Vancouver, and Simon Fraser University, Burnaby, British Columbia, Canada; [2]J. Antonio Aviña-Zubieta, MD, PhD, Mary A. De Vera, PhD, John M. Esdaile, MD: Arthritis Research Canada, Vancouver, and University of British Columbia, Vancouver, British Columbia, Canada; [3]Yi Qian, PhD: University of British Columbia, Vancouver, British Columbia, Canada.

No potential conflicts of interest relevant to this article were reported.

Address correspondence to Hui Xie, PhD, Milan and Maureen Illich/Merck Frosst Chair in Biostatistics for Arthritis and Musculoskeletal Diseases, Arthritis Research Canada, 230 – 2238 Yukon Street, Vancouver, British Columbia V5Y 3P2, Canada. Email: hxie@arthritisresearch.ca.

Submitted for publication October 13, 2020; accepted in revised form December 22, 2020.

SIGNIFICANCE & INNOVATIONS

- Patients with newly diagnosed systemic lupus erythematosus adhering to antimalarial (AM) therapy had a 71% (relative to those who did not adhere to AM therapy) and 83% (relative to those who discontinued) lower risk of death.
- Our findings also show a dose-response relationship between AM adherence and premature mortality. As the adherence level increased by 1 level (e.g., discontinuation to nonadherence or nonadherence to adherence), on average, the risk of death decreased by 43%.
- The increased risk of death associated with poorer AM adherence persisted after adjusting for baseline confounders and time-varying confounders during follow-up, including medication use, hospitalizations, and comorbidities.
- Our results call for improved strategies to boost AM adherence among SLE patients for the sake of improving survival.

found an association between adherence to hydroxychloroquine and decreased use of acute care among Medicaid beneficiaries with SLE (12).

To our knowledge, there have been no population-based studies on the effects of time-varying AM adherence on mortality among incident SLE patients despite the documentation of the adverse effects of poor adherence on other health outcomes (9). This study aims to fill this research gap. Prior studies examining the effect of AM adherence on health outcomes used traditional time-dependent Cox proportional hazards (PH) models to control for time-varying confounders (9), which could produce biased effect estimates if those confounders were mediators (2). Our assessment used marginal structural models, the state of the art methodology designed for drawing a valid causal inference in the presence of time-varying confounders.

MATERIALS AND METHODS

Data source. Universal health care coverage is available for all residents of the province of British Columbia (BC), Canada (n = 4,683,139 in 2015 Statistics Canada) (13). Population Data BC provided all provincially funded health care service data, including all provincially funded health care professional visits (14), medically required registration services data (15), hospital separations (16), demographics (17), cancer registry (18), vital statistics (19), and comprehensive prescription drug database PharmaNet (20). PopData BC provided these data from January 1, 1990, to March 31, 2015, except for PharmaNet, where all dispensed medications are documented from 1996. Many recent population studies have successfully used PopData BC (2,9,21).

Study design and cohort definition. We employed a retrospective, longitudinal cohort study design to assess the independent effect of incident AM adherence on the risk of death among patients with incident SLE after adjusting for confounders. The case definition of incident SLE included the following: 1) age ≥18 years; 2) 2 principal International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) (710.0) or ICD-10-CM (M32.1, M32.8, M32.9) codes for SLE at least 2 months apart within 2 years from any physician or hospital visit, as previously used (2); and 3) no SLE diagnosis in a 7-year run-in period prior to the first ICD code for SLE to ensure incident SLE cases. In all, 99.4% of the SLE patients had at least 1 of the 2 ICD codes diagnosed by rheumatologists or from the hospitalization data set (2). This definition has 97.6% sensitivity and 97.5% positive predictive value in the Swedish registry data (22). SLE index date was defined as the date of the second SLE code. The study cohort (SLE-AM cohort) included all incident SLE patients with incident AM dispensation. Incident AM users were defined as those who had ≥1 AM dispensations between January 1997 and March 2015 and no prior AM dispensations since January 1996 (earliest date of available medication data). SLE patients with AM dispensations in 1996 were excluded to ensure incident AM users.

Exposure assessment. The primary exposure was adherence to AM medications (hydroxychloroquine, chloroquine, quinacrine). In the analysis, the follow-up time was divided into 30-day windows. For each window, a measure of adherence using the proportion of days covered (PDC) was calculated (23). We assumed that dispensed medications were taken, a common assumption when measuring medication adherence from administrative data (24). After computing PDC, we categorized PDC as adherent (PDC ≥ 0.90), nonadherent (0 < PDC < 0.90), and discontinuer (PDC = 0). PDC cutoff of 0.90 for the definition of AM adherence resulted from the recommendation of taking at least 90% AM medications in a recent study (9). Our sensitivity analysis used the alternative definition that considers PDC ≥ 0.80 as adherent (25). In our study, AM adherence is a time-varying variable that can change its status from one adherence level in one window to other levels in successive windows in the follow-up.

Ascertainment of outcome. The primary outcome was all-cause mortality, with death identified from the vital statistics registry. Follow-up for the outcome of this study started on the first day of having both SLE and AM, i.e., the SLE index date (second ICD code) for those whose first AM dispensation occurred before the SLE index date, or the date of the first AM dispensation if otherwise. Subjects were followed until death, leaving BC or the end-date of the study, March 31, 2015, whichever occurred first.

Covariate assessment. Baseline covariates were assessed within 12 months prior to the start of follow-up. Our baseline covariates included demographic variables (age, sex, location of residence, index calendar year, neighborhood income quintile), health resource utilization (hospital visits, length of stay during hospitalizations, number of outpatient visits, presence of nephrologist visits), medication usage (statins, other cardiovascular drugs including calcium-channel blocking agents, beta-blockers, nitrates, and antiarrhythmic agents, hormone replacement therapy, glucocorticoids, anticoagulant therapy, cyclooxygenase 2 inhibitors, immunosuppressive drugs), comorbidities (hypertension, chronic obstructive pulmonary disease, angina), and the Romano modification of the Charlson comorbidity index (26). Modified Charlson comorbidity index exhibited a weighted index of comorbid diseases with higher weights for more severe comorbid conditions. A weight of 1 was assigned to myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, peptic ulcer disease, connective tissue disease, mild liver disease, and diabetes mellitus without complications, whereas a weight of 2 was assigned to diabetes mellitus with complications, hemiplegia, renal disease, and cancer. Also, a weight of 3 was assigned to moderate or severe liver disease, and a weight of 6 was assigned to metastatic carcinoma and AIDS (26).

Time-varying covariates were updated in 30-day windows. They included the same set of variables regarding medication usage, health resource utilization, comorbidities, and Romano modification of the Charlson comorbidity index, as discussed above. As compared to baseline measures, these time-varying covariates, such as hospitalizations, comorbidities, and use of glucocorticoids and immunosuppressive drugs, provided far more instantaneous proxy measures for general health status and disease activity of SLE patients (2).

Statistical analysis. To determine the crude risk of mortality, we computed the crude incidence rates (IRs) of mortality per 1,000 person-years for each level of AM adherence in SLE patients. Poisson models with adherence levels as the only explanatory variable were used to estimate the crude IR ratios (IRRs). We obtained 95% confidence intervals (95% CIs) for the IRRs using empirical sandwich SEs to account for within-subject correlated observations.

We used a multivariable Cox proportional hazards model with time-varying covariates to assess the effect of AM adherence on the risk of death, adjusting for 1) baseline risk factors including demographic variables, health resource utilization, medication usage, comorbidities, and Charlson comorbidity index, and 2) time-varying health resource utilization, medication usage, comorbidities, and Charlson comorbidity index as covariates in the Cox model. The primary exposure, AM adherence, was a time-varying variable with updated adherence status for each 30-day window. The independent risk of death associated with AM adherence was assessed by the adjusted hazard ratio (HR_{adj}) estimates for the categorical exposure variable, AM adherence, with discontinuation as the reference level.

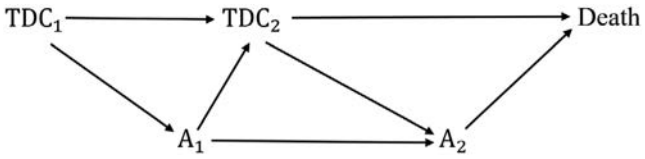


Figure 1. A directed acyclic graph for 2-period follow-up data illustrating time-dependent confounders also acting as mediators. A_t and TDC_t represent antimalarial (AM) adherence and time-varying confounders at period t (t = 1, t = 2, respectively). As illustrated above, a comorbidity (such as nephritis), initiation of immunosuppressive drugs, or hospitalizations may cause patients to discontinue or reduce AM use while affecting the mortality outcome. Thus, they are confounders of the effect of AM adherence on the mortality outcome. Furthermore, these confounders can be affected by the prior AM adherence behaviors (A_1 to TDC_2) and thus are mediators in the causal pathway AM adherence to death. The marginal structural model is a causal inference method designed to properly estimate the causal effect of AM adherence on mortality when there exist such time-varying confounders also acting as mediators. At any time point, the distributions of those time-varying confounders are balanced between all adherence groups via inverse probability weighting in marginal structural model analysis.

The above time-dependent Cox model could yield biased effect estimates of AM adherence when time-varying covariates are affected by the AM adherence in previous windows (27,28). To account for this possibility of certain time-varying covariates being mediators in the causal pathway from AM adherence to mortality (Figure 1), we conducted a marginal structural Cox analysis that can produce valid risk estimates for AM adherence–associated death wherein the distributions of time-varying confounders were balanced among the 3 adherence groups via inverse probability weighting (27,29). Marginal structural Cox models were implemented in 3 steps (2,27). First, the follow-up period was divided into 30-day windows, and the inverse probability of exposure weights was obtained for each person-window through a weighted, pooled, multinomial logistic regression model with AM adherence categories as the dependent variable and time-varying covariates, past AM adherence exposure status, and baseline covariates as the predictors. Second, the inverse probability of censoring weights was obtained using the same procedure with the censoring indicator at each window as the dependent variable to account for potential dependent censoring. At the final step, the risk of death was estimated using a generalized estimating equation approach with the inverse probability of exposure and censoring weights.

Sensitivity analyses. To examine the robustness of our results, we considered 4 sensitivity analyses. First, to assess the effect of increasing window size in the marginal structural Cox model, we conducted a window-size sensitivity analysis that used

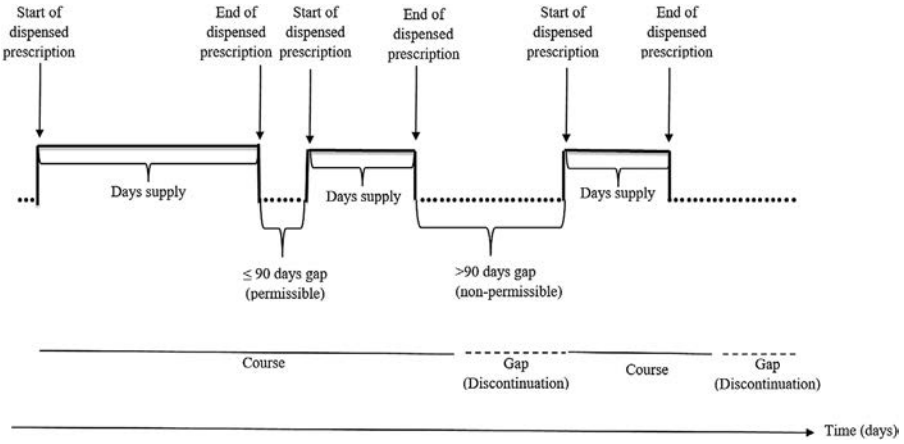


Figure 2. Schematic representation of antimalarial (AM) courses over follow-up time. AM courses with proportion of days covered (PDC) ≥ 0.90 show adherence, 0 < PDC < 0.9 show nonadherence, and gaps with >90 days show discontinuation.

90-day and 180-day windows in addition to the 30-day windows used in the primary analysis. The AM adherence exposure values were recalculated for 90-day and 180-day windows in these sensitivity analyses. Second, in an adherence cutoff sensitivity analysis, we used the standard definition of adherence, in which a patient is considered adherent if PDC ≥ 0.80, and nonadherent if PDC < 0.80 (25). To make results comparable with our analysis, we separated discontinuer (PDC = 0) from the nonadherent and defined nonadherent for 0 < PDC < 0.80. Third, we compared the results from primary analysis with the results from a course-based AM adherence design used in a recent article (see Figure 2 for the schematic representation of courses) (9). We established AM drug courses and gaps using PharmaNet data on dispensed prescription days and days of supply accordingly. Subsequent dispensed prescriptions with overlaps and permissible gaps of up to 90 days form 1 course. Here, AM adherence was a course-specific time-varying exposure with potentially different adherence statuses in different courses. Because the course durations varied among courses and across patients, and these courses did not coincide with the windows used in marginal structural Cox models, this design hindered performing marginal structural Cox analysis. The Cox PH model with time-dependent AM adherence status was then implemented to estimate how the risk of death is associated with course-specific AM adherence status. Fourth, in a model selection sensitivity analysis, we examined the possibility of overcorrection in our analysis that could arise from using all covariates. We compared our primary analyses with analyses that considered only significant covariates in the time-varying Cox model and all marginal structural Cox analysis stages, including weight calculations. SAS, version 9.4, was used for all analyses. All hazard ratios are presented with 95% CIs.

Ethics approval and study conduct. No personal identifying information was available in this study. British Columbia's

Table 1. Baseline characteristics of systemic lupus erythematosus antimalarial cohort patients (n = 3,062)*

Characteristic	Value
Demographic data	
Age, mean ± SD years	46.9 ± 15.2
Sex	
Female	2,718 (88.8)
Male	344 (11.2)
Neighborhood income quintile	
1	714 (23.3)
2	590 (19.3)
3	634 (20.7)
4	580 (18.9)
5	544 (17.8)
Location	
Urban	2,653 (86.6)
Rural	409 (13.4)
Health resource utilization	
No. of outpatient visits, mean ± SD	22.4 ± 16.4
No. of nephrologist visits, mean ± SD	
All patients	0.3 ± 3.4
Patients with at least 1 visit (n = 166)	5.9 ± 13.4
No. of psychiatrist visits, mean ± SD	
All patients	0.7 ± 4.1
Patients with at least 1 visit (n = 200)	10.2 ± 12.9
Any hospitalizations	1,099 (35.9)
Length of hospital stay, median ± SD days†	9.0 ± 92.5
Comorbidities	
Charlson comorbidity index, mean ± SD	0.5 ± 1.1
Hypertension	585 (19.1)
COPD	86 (2.8)
Angina	120 (3.9)
Medication usage	
Statin	196 (6.4)
Other CV drugs	793 (25.9)
Hormone therapy	252 (8.2)
Glucocorticoids	1,431 (46.7)
Immunosuppressive drugs	610 (19.9)
COX-2 inhibitors	343 (11.2)
Anticoagulant therapy	163 (5.3)

* Values are the number (%) unless indicated otherwise. All baseline characteristics are computed in the 12 months prior to the follow-up start date. COPD = chronic obstructive pulmonary disease; COX-2 = cyclooxygenase 2; CV = cardiovascular.
† Median used instead of the mean because the distribution of length of hospital stay was highly skewed.

Table 2. Follow-up assessment for different antimalarial adherence groups*

	Discontinuer (PDC = 0)	Nonadherent (0 < PDC < 0.90)	Adherent (PDC ≥ 0.90)
PDC, mean ± SD	0 ± 0	0.56 ± 0.24	0.98 ± 0.02
30-day windows in the follow-up, %	43.8	16.0	40.2
No. of deaths in the follow-up	173	35	34
Total person-years	9,250	3,373	8,486
Incidence rate of death (per 1,000 person-years)	18.7	10.4	4.0

* PDC = proportion of days covered.

Freedom of Information and Protection of Privacy Act was adhered to. Ethics approval was obtained from the University of British Columbia's Behavioral Research Ethics Board (H15-00887).

RESULTS

We identified 3,062 individuals with incident SLE and incident AM use. On average, there were 196 days (median 133 days) between the first ICD code and second ICD code for the patients in our SLE cohort. The interquartile range of the time difference between the 2 SLE codes was 155 days, whereas the SD was 152.8 days. The median number of 30-day windows for those SLE patients was 62 (mean ± SD 70.2 ± 47.9). In all, 96.4% of SLE patients had >1 AM adherence status during the follow-up. Length of stay during hospitalizations only accounted for 1% of the total follow-up time. Baseline characteristics of SLE-AM cohort patients are summarized in Table 1.

Risk of death among SLE-AM cohort patients. In terms of AM adherence, the 30-day windows were categorized as discontinuer (44%), adherent (40%), and nonadherent (16%). During a mean follow-up of 6.4 years, 242 (7.9%) SLE patients died (Table 2). Among those, 173, 35, and 34 died in windows in which the patients were AM discontinuers (PDC = 0), nonadherent (0 < PDC < 0.90), and adherent (PDC ≥ 0.90), respectively. The corresponding IRs of mortality for AM discontinuer, nonadherent, and adherent patients were 18.7, 10.4, and 4.0 per 1,000 person-years, respectively.

The Cox PH model with both the baseline and time-varying confounders was used to assess the risk of death due to time-varying AM adherence (Table 3). The corresponding HR_{adj} for AM nonadherent and adherent SLE patients were 0.62 (95% CI 0.42–0.91) and 0.20 (95% CI 0.13–0.29), respectively, compared to AM discontinuers.

The marginal structural Cox model was applied to account for the possibility that some time-varying confounders were mediators between AM adherence and premature mortality. The corresponding HR_{adj} for AM nonadherent and adherent incident SLE patients were 0.58 (95% CI 0.40–0.85) and 0.17 (95% CI 0.12–0.25), respectively, compared to AM discontinuers. The HR_{adj} for adherers compared to the nonadherers were 0.32 (95% CI 0.19–0.53) and 0.29 (95% CI 0.18–0.48) in the regular time-dependent Cox PH and the marginal structural Cox analyses.

The above HR_{adj} from the marginal structural Cox model meant that SLE patients adhering to AM therapy had a 71% (= 1.00 – 0.29) and 83% (= 1.00 – 0.17) lower risk of death than patients who did not adhere or who discontinued AMs, respectively. A statistically significant linear contrast test for trend in the HR_{adj} of mortality risk over the adherence levels was found (estimate 0.43, $P < 0.001$) (see Table 3 for linear trend).

Table 3. Overall risk of death in patients with incident systemic lupus erythematosus during follow-up*

Adherence levels	IR ratio (95% CI)	Cox PH, HR_{adj} (95% CI)	MSM, HR_{adj} (95% CI)
Discontinuer (PDC = 0) (ref.)	1.00	1.00	1.00
Nonadherent (0 < PDC < 0.90)	0.55 (0.39–0.80)	0.62 (0.42–0.91)	0.58 (0.40–0.85)
Adherent (PDC ≥ 0.90)	0.21 (0.15–0.31)	0.20 (0.13–0.29)	0.17 (0.12–0.25)
Contrast: adherent vs. nonadherent	0.39 (0.24–0.62)	0.32 (0.19–0.53)	0.29 (0.18–0.48)
Linear trend	–	0.46 (0.38–0.55)	0.43 (0.36–0.51)

* The multivariable models were adjusted for baseline covariates including demographic variables (age, sex, location of residence, index calendar year, neighborhood income quintile, health authority jurisdiction), health resource utilization (hospital visits, number of outpatient visits, nephrologist visits), medication usage (statins, other cardiovascular drugs, hormone replacement therapy, glucocorticoids, anticoagulant therapy, cyclooxygenase 2 inhibitors, immunosuppressive drugs), comorbidities (hypertension, chronic obstructive pulmonary disease, angina), and the Romano modification of the Charlson comorbidity index for administrative data. Also, the time-varying variables of health resource utilization, medication usage, Charlson comorbidity index, and comorbidities were included as control covariates in the Cox PH model for adjustment of risk estimates, whereas those are used to calculate weights in the marginal structural model. 95% CI = 95% confidence interval; HR_{adj} = adjusted hazard ratio; IR = incidence rate; MSM = marginal structural model; PDC = proportion of days covered; PH = proportional hazards; ref. = reference.

Table 4. Sensitivity analyses*

	Window size sensitivity analysis			Adherence cutoff sensitivity analysis†			Course-based sensitivity analysis‡	
Adherence levels§	Primary analysis (MSM 30-day window), HR_{adj} (95% CI)	Sensitivity analysis (MSM 90-day window), HR_{adj} (95% CI)	Sensitivity analysis (MSM 180-day window), HR_{adj} (95% CI)	IR ratio (95% CI)	Cox PH, HR_{adj} (95% CI)	MSM 30-day window, HR_{adj} (95% CI)	No. of events	Cox PH, HR_{adj} (95% CI)
Discontinuer (Ref.)	1.00	1.00	1.00	1.00	1.00	1.00	197	1.00
Nonadherent	0.58 (0.40–0.85)	0.94 (0.71–1.25)	0.98 (0.81–1.43)	0.68 (0.47–0.99)	0.77 (0.52–1.15)	0.73 (0.50–1.08)	15	0.18 (0.10–0.30)
Adherent	0.17 (0.12–0.25)	0.12 (0.07–0.20)	0.10 (0.05–0.17)	0.21 (0.15–0.30)	0.19 (0.13–0.28)	0.17 (0.11–0.24)	30	0.11 (0.07–0.16)
Contrast: adherent vs. nonadherent	0.29 (0.18–0.48)	0.12 (0.07–0.22)	0.09 (0.05–0.16)	0.33 (0.25–0.43)	0.25 (0.15–0.41)	0.23 (0.14–0.37)	–	–
Linear trend	0.43 (0.36–0.51)	0.45 (0.39–0.53)	0.46 (0.40–0.54)	–	0.46 (0.39–0.55)	0.44 (0.37–0.51)	–	–

* The multivariable models were adjusted for baseline covariates including demographic variables (age, sex, location of residence, index calendar year, neighborhood income quintile, health authority jurisdiction), health resource utilization (hospital visits, number of outpatient visits, nephrologist visits), medication usage (statins, other cardiovascular drugs, hormone replacement therapy, glucocorticoids, anticoagulant therapy, cyclooxygenase 2 inhibitors, immunosuppressive drugs), comorbidities (hypertension, chronic obstructive pulmonary disease, angina), and the Romano modification of the Charlson comorbidity index for administrative data. Also, the time-varying variables of health resource utilization, medication usage, Charlson comorbidity index, and comorbidities were included as control covariates in the Cox PH model for adjustment of risk estimates, whereas those are used to calculate weights in the marginal structural model. 95% CI = 95% confidence interval; AM = antimalarials; AMA = American Medical Association; HR_{adj} = adjusted hazard ratio; IR = incidence rate; MSM = marginal structural model; PDC = proportion of days covered; PH = proportional hazards; Ref. = reference.

† Cutoff at PDC ≥ 0.80 as per AMA definition.

‡ Cox PH analysis for AM adherence course exposure.

§ Discontinuer PDC = 0; nonadherent 0 < PDC < 0.90, and adherent PDC ≥ 0.90 for all analyses except adherence cutoff sensitivity analysis, in which nonadherent 0 < PDC < 0.80 and adherent PDC ≥ 0.80.

Results of sensitivity analyses. Results from the sensitivity analyses were similar to those of the primary analyses (Table 4). The sensitivity analysis with alternative window sizes (90-day and 180-day) for marginal structural analysis found that the protective effect of AM adherence compared to discontinuation remained statistically significant, whereas the protective effect of AM nonadherence compared to discontinuation became statistically nonsignificant. This suggests that the larger window size might not capture adherence exposure accurately, causing attenuation bias (30). Regardless of window sizes, AM adherence had a larger (reduction of 70% in death risk) (Table 4) and statistically significant protective effect on premature mortality compared with AM nonadherence. These findings emphasize the importance of AM adherence (at least 90% of the prescribed doses) in preventing premature mortality in SLE patients.

Similar findings were observed with an 80% cutoff for AM adherence. Results using a course-based adherence measure in third sensitivity analysis also supported the importance of AM adherence in preventing premature mortality. AM nonadherence was found to have a protective effect on premature mortality when compared with discontinuers, with an effect size close to but somewhat smaller than the protective effect of AM adherence relative to discontinuers. Courses were very long (the mean course duration was 594 days) and thus could include several AM adherent, nonadherent, and discontinued 30-day windows used in the primary marginal structural analyses in 1 course. Finally, the model selection sensitivity analysis including only significant covariates in the time-varying Cox PH and marginal structural Cox models provided almost the same risk estimates (thus, not reported in Table 4) as those in the primary analysis. As a result, we can rule out the possibility of overcorrection for the control variables in our study.

DISCUSSION

To our knowledge, this is the first large population-based study evaluating the effect of AM adherence on premature mortality in incident SLE patients after adjusting for baseline and time-varying confounders. This study found that AM adherence lowers the risk of death (risk reduction of 83% for adherent and 42% for nonadherent, respectively, compared to discontinuers). Our study also found a dose-response relationship: as adherence increased by one level (e.g., discontinuation to nonadherence or nonadherence to adherence), on average, the risk of death decreased by 43 percent.

Our findings of a large protective effect of AM medications (83% reduction of risk in premature mortality for AM adherent compared to discontinuer) is consistent with the large beneficial effects of taking AM reported in previous related studies. A 6-month, randomized, double-blind, placebo-controlled study on stable SLE patients found a 6-fold higher relative risk of severe disease flare-ups among hydroxychloroquine discontinuers compared to users (31). A multinational Latin American cohort study of 1,480 patients found that the risk of death decreased by 38% among the AM users (>6 months of continuous use) compared to the non-users (32). A US case–control study of 608 patients from a prevalent cohort found a large protective effect on survival (70% reduction in the mortality risk) of hydroxychloroquine use as compared with no use (33). However, neither of the latter 2 studies evaluated the effect of AM adherence on mortality. A key difference in our study is the explicit generation of AM adherence exposure, taking into account the dynamic nature of AM medication usage as well as the direct estimation of AM adherence on premature mortality. Furthermore, our study controlled for both the baseline and time-varying confounders of adherence, whereas the 2 prior studies (32,33) used baseline confounders only. Failing to include important time-varying confounders can introduce bias (2,27). Also, unlike the 2 prior studies (32,33), which could have generalizability issues as they used selected patients from selected centers, we used incident SLE samples from the entire BC population with a significantly larger sample size. True effect size could be underestimated in studies using prevalent SLE cohorts because of the associated survival bias: SLE patients who died before the study entry date were excluded from the prevalent cohort, making the selected nonuser group more likely to survive.

There are several limitations to our study. Misclassification of SLE cases could arise due to an inaccurate SLE case definition despite a strict algorithm for selection criteria. However, any misclassification is likely to underestimate the effect size. In all, 99.4% of patients in this study were diagnosed at visits with specialists or hospitalizations. Also, the sensitivity exceeds 97% for our SLE case definition (22). In our study, nonadherence and discontinuation of AM therapy could be either of patients' volition or suggested by doctors. We did not have data distinguishing these possible reasons, and hence, the overall effect of AM nonadherence and discontinuation for all reasons was evaluated. A detailed investigation could be done in future research with appropriate data regarding those reasons.

Although we have a large number of risk factors for death in our data for which risk estimates were adjusted for, some important risk factors of death, such as SLE disease activity, body mass index, and smoking were unavailable in the data. However, the large effect size observed in this study is expected to be robust to unmeasured weak-to-moderate confounders. Only very strong confounders could change the effect size. This is unlikely given the inclusion of important time-varying confounders related to poor clinical outcomes (e.g., comorbidities) and disease activity (e.g., glucocorticoids, immunosuppressives, number of medical visits) in the analysis that may correlate with these unmeasured confounders and capture some confounding effects of otherwise strong confounders. AM adherence was computed from the outpatient dispensation data, and during hospitalizations, we do not have medication records to assess adherence. However, the

length of stay during hospitalizations only accounted for 1% of our follow-up time. So, the AM adherence measure is unlikely to be affected by the hospitalization. Usually, patients admitted to hospitals have poor clinical outcomes, potentially attributable to SLE disease activity. We controlled for time-varying hospitalizations to offset this problem.

Our study has several strengths. First, the findings can be generalizable, as it used a large representative population-based sample of BC. Second, our study was confined to incident SLE defined with a 7-year washout period, which is a safeguard against selection and survival bias that could arise from a prevalent SLE cohort. Also, to our best knowledge, this is the first population-based study that accounts for the impact of time-varying confounders while assessing the risk of death in SLE patients for time-varying AM adherence. The traditional Cox model has been used commonly as a workhorse method for survival analysis despite its limitations for causal inference. Therefore, we first considered such Cox models and then used marginal structural models designed for causal inference to overcome traditional Cox models' limitations in handling time-varying confounders (27). The time-varying confounders of AM adherence could be affected by AM adherence in previous periods. Thus, these time-varying confounders (such as comorbidity) were likely mediators in the causal pathway from AM adherence to death in SLE patients (34–37). A statistical method designed for causal inference, marginal structural modeling, was used to account for time-varying confounders so that these confounders mediating the relationship between AM adherence and risk of death could be addressed (2,27). Inverse weighting in the marginal structural model balanced the distributions of these time-varying confounders, including comorbidities.

In contrast, despite their similarity with the estimates from the marginal structural models, the HR_{adj} estimates from the simpler time-dependent traditional Cox models generally do not have causal interpretations: these estimates are subject to bias of which the magnitude varies by data sets and depends on a number of factors (2,27). In particular, improper handling of the time-varying confounders in traditional Cox models can lead to biases in opposite directions (e.g., potentially upward and downward biases associated with unbalance of past AM adherence history and with adding time-varying comorbidities as covariates in Cox models, respectively) (27), which may result in overall HR_{adj} estimates similar to those from marginal structural models.

In conclusion, in this population-based study, we found that adherence to AM prevents premature mortality by 71% and 83% when compared to nonadherent or discontinuers, respectively. These findings are independent of baseline characteristics as well as time-varying confounders or risk factors. Therefore, strategies to support adherence to AM medications, including patients' perspectives, are urgently needed in the management of SLE.

ACKNOWLEDGMENT

We thank Shelby Marozoff for her editorial assistance in the preparation of the manuscript.

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. Xie 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. Hoque, De Vera, Qian, Esdaile, Xie.
Acquisition of data. Hoque, Aviña-Zubieta, Xie.
Analysis and interpretation of data. Hoque, Qian, Esdaile, Xie.

REFERENCES

1. Maidhof W, Hilas O. Lupus: an overview of the disease and management options. P T 2012;37:240.
2. Li L, Xie H, Lu N, Esdaile JM, Aviña-Zubieta JA. The impact of systemic lupus erythematosus on the risk of newly diagnosed hip fracture: a general population-based study. Arthritis Care Res (Hoboken) 2021;73:259–65.
3. Carter EE, Barr SG, Clarke AE. The global burden of SLE: prevalence, health disparities and socioeconomic impact. Nat Rev Rheumatol 2016;12:605.
4. Jorge AM, Lu N, Zhang Y, Rai SK, Choi HK. Unchanging premature mortality trends in systemic lupus erythematosus: a general population-based study (1999–2014). Rheumatology (Oxford) 2018; 57:337–44.
5. Costedoat-Chalumeau N, Pouchot J, Guettrot-Imbert G, Le Guern V, Leroux G, Marra D, et al. Adherence to treatment in systemic lupus erythematosus patients. Best Pract Res Clin Rheumatol 2013;27: 329–40.
6. Mehat P, Atiquzzaman M, Esdaile JM, Aviña-Zubieta JA, De Vera MA. Medication nonadherence in systemic lupus erythematosus: a systematic review. Arthritis Care Res (Hoboken) 2017;69:1706–13.
7. Dalle Vedove C, del Giglio M, Schena D, Girolomoni G. Drug-induced lupus erythematosus. Arch Dermatol Res 2009;301:99–105.
8. Aviña-Zubieta JA, Esdaile JM. Antimalarial medications. In: Wallace D, Hahn BH, editors. Dubois' lupus erythematosus and related syndromes. 9th edition. Elsevier; 2018. p. 650–60.
9. Salmasi S, Sayre EC, Aviña-Zubieta JA, Esdaile JM, De Vera MA. Adherence to antimalarial therapy and risk of type 2 diabetes mellitus among patients with systemic lupus erythematosus: a population-based study. Arthritis Care Res (Hoboken) 2021;73:703–6.
10. Petri M. Hydroxychloroquine use in the Baltimore Lupus Cohort: effects on lipids, glucose and thrombosis. Lupus 1996;5 Suppl 1: 16–22.
11. Canadian Hydroxychloroquine Study Group, Tsakonas E, Joseph L, Esdaile JM, Choquette D, Senécal JL, et al. A long-term study of hydroxychloroquine withdrawal on exacerbations in systemic lupus erythematosus. Lupus 1998;7:80–5.
12. Feldman CH, Yazdany J, Guan H, Solomon DH, Costenbader KH. Medication nonadherence is associated with increased subsequent acute care utilization among Medicaid beneficiaries with systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2015;67: 1712–21.
13. British Columbia Ministry of Health. Population Data BC. About PopData. URL: https://www.popdata.bc.ca/.
14. British Columbia Ministry of Health. Population Data BC. Medical Services Plan (MSP) payment information file. Data extract. URL: https://www.popdata.bc.ca/data.

15. British Columbia Ministry of Health. Population Data BC. Consolidation file (MSP Registration & Premium Billing). Data extract. URL: https://www.popdata.bc.ca/data.
16. Canadian Institute for Health Information. Population Data BC. Discharge abstract database (hospital separations). Data extract. URL: https://www.popdata.bc.ca/data.
17. Statistics Canada. Population Data BC. Statistics Canada income band data. Catalogue number: 13C0016. Data extract. URL: https://www.popdata.bc.ca/data.
18. British Columbia Cancer Agency. Population Data BC. BC Cancer Agency registry data. Data extract. URL: https://www.popdata.bc.ca/data.
19. British Columbia Vital Statistics Agency. Population Data BC. Vital statistics deaths. Data extract. URL: https://www.popdata.bc.ca/data.
20. British Columbia Ministry of Health. BC Ministry of Health. PharmaNet. Data Stewardship Committee. Data extract. URL: https://www.popdata.bc.ca/data.
21. Li L, McCormick N, Sayre EC, Esdaile JM, Lacaille D, Xie H, et al. Trends of venous thromboembolism risk before and after diagnosis of gout: a general population-based study. Rheumatology (Oxford) 2020;59:1099–107.
22. Arkema EV, Jönsen A, Rönnblom L, Svenungsson E, Sjöwall C, Simard JF. Case definitions in Swedish register data to identify systemic lupus erythematosus. BMJ Open 2016;6:e007769.
23. Hess LM, Raebel MA, Conner DA, Malone DC. Measurement of adherence in pharmacy administrative databases: a proposal for standard definitions and preferred measures. Ann Pharmacother 2006;40:1280–8.
24. Ho PM, Bryson CL, Rumsfeld JS. Medication adherence: its importance in cardiovascular outcomes. Circulation 2009;119:3028–35.
25. Cramer JA, Roy A, Burrell A, Fairchild CJ, Fuldeore MJ, Ollendorf DA, et al. Medication compliance and persistence: terminology and definitions. Value Health 2008;11:44–7.
26. Romano PS, Roos LL, Jollis JG. Presentation adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. J Clin Epidemiol 1993;46:1075–9.
27. Hernán MÁ, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology 2000;11:561–70.
28. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. Epidemiology 2009;20:488.
29. Xie D, Yang W, Jepson C, Roy J, Hsu JY, Shou H, et al. Statistical methods for modeling time-updated exposures in cohort studies of chronic kidney disease. Clin J Am Soc Nephrol 2017;12:1892–9.
30. Kyle RP, Moodie EE, Klein MB, Abrahamowicz M. Correcting for measurement error in time-varying covariates in marginal structural models. Am J Epidemiol 2016;184:249–58.
31. 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.
32. Shinjo SK, Bonfá E, Wojdyla D, Borba EF, Ramirez LA, Scherbarth HR, et al. Antimalarial treatment may have a time-dependent effect on lupus survival: data from a multinational Latin American inception cohort. Arthritis Rheum 2010;62:855–62.
33. Alarcón GS, McGwin G, Bertoli AM, Fessler BJ, Calvo-Alén J, Bastian HM, et al. Effect of hydroxychloroquine on the survival of patients with systemic lupus erythematosus: data from LUMINA, a multiethnic US cohort (LUMINA L). Ann Rheum Dis 2007;66:1168–72.
34. Ruiz-Irastorza G, Egurbide MV, Pijoan JI, Garmendia M, Villar I, Martinez-Berriotxoa A, et al. Effect of antimalarials on thrombosis and survival in patients with systemic lupus erythematosus. Lupus 2006;15:577–83.
35. Fessler BJ, Alarcón GS, McGwin G Jr, Roseman J, Bastian HM, Friedman AW, 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.
36. Nived O, Jönsen A, Bengtsson A, Bengtsson C, Sturfelt G. High predictive value of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology damage index for survival in systemic lupus erythematosus. J Rheumatol 2002;29:1398–400.
37. Rahman P, Gladman DD, Urowitz MB. Clinical predictors of fetal outcome in systemic lupus erythematosus. J Rheumatol 1998;25:1526.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1098–1104
DOI 10.1002/acr.24555

Physical Inactivity and Incident Depression in a Multiracial, Multiethnic Systemic Lupus Erythematosus Cohort

Sarah L. Patterson, Laura Trupin, Jinoos Yazdany, Maria Dall'Era, Cristina Lanata, Kimberly Dequattro, Wendy Hartogensis, and Patricia Katz

Objective. Physical activity is known to improve depressive symptoms. The present study was undertaken to examine physical inactivity as a predictor of incident depression in systemic lupus erythematosus (SLE).

Methods. Data derive from the California Lupus Epidemiology Study (CLUES), a longitudinal cohort with confirmed SLE diagnoses. Physical inactivity was assessed from a single item, "I rarely or never do any physical activities," and depressive symptoms by the 8-item Patient Health Questionnaire (PHQ-8). Analysis included those not depressed at baseline (PHQ-8 score <10) who completed an in-person baseline assessment and at least 1 follow-up visit (n = 225). Incident depression was defined as a PHQ-8 score of ≥10 at follow-up. Cox proportional hazards regression modeled incident depression over 2 years as a function of baseline physical inactivity, controlling for age, sex, race, income, comorbidities, disease activity, and disease damage.

Results. At baseline, the mean ± SD age of the participants was 45 ± 15 years, 88% were female, and 70% identified as non-White. Mean PHQ scores for those without depression at baseline did not differ by activity status, but those who were inactive at baseline were significantly more likely to develop depression over the next 2 years (hazard ratio [HR] 2.89 [95% confidence interval (95% CI) 1.46–5.71]). After adjusting for covariates, the association remained strong, including a >3-fold increased risk of incident depression among the sedentary group (HR 3.88 [95% CI 1.67–9.03]).

Conclusion. In this diverse SLE cohort, a simple question about physical inactivity was highly predictive of incident depression over the subsequent 2 years. Results suggest an urgent need for approaches to reduce sedentary behavior in this high-risk population.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a chronic systemic autoimmune condition characterized by autoantibody formation, multisystem organ involvement, and increased mortality. It also confers an increased risk of comorbid depression, and prior research has shown that the lifetime prevalence of major depressive disorder is 40–50% in the setting of SLE relative to 17% in the general population of US adults (1–3). The greater burden of depression experienced in this patient group is important due to the deleterious effects on quality of life, as well as associations with greater disability, interference with medication compliance, and worse patient-reported outcomes (4–6).

Although the higher prevalence of depression in lupus relative to the general population is well demonstrated, the psychosocial, biological, and lifestyle factors responsible (and measures that can be taken to mitigate them) are not yet well defined. Prior studies to better understand risk factors for depression in lupus suggest that disease activity (7) and treatment with glucocorticoids (8) may play a role, but the link between depression and disease activity has been inconsistent across studies. Moreover, indices of disease severity do not fully account for the relative burden of depression in this patient group (7–9). Physical inactivity confers an increased risk of incident depression in the general population (10–13) and may contribute to a higher incidence of mood disorders in SLE, but the link between inactivity and incident depression in this high-risk group has not been explored.

Supported by the Centers for Disease Control (grant 5U01DP005120), the Rheumatology Research Foundation, the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases grant P30-AR-070155), the Robert L. Kroc Chair in Rheumatic and Connective Tissue Diseases, and the Russell/Engleman Medical Research Center for Arthritis.

Sarah L. Patterson, MD, Laura Trupin, MPH, Jinoos Yazdany, MD, MPH, Maria Dall'Era, MD, Cristina Lanata, MD, Kimberly Dequattro, MD, Wendy Hartogensis, PhD, MPH, Patricia Katz, PhD (current address: Zuckerberg San Francisco General Hospital, San Francisco, California): University of California, San Francisco.

No potential conflicts of interest relevant to this article were reported.

Address correspondence to Patricia Katz, PhD, Zuckerberg San Francisco General Hospital, 1001 Potrero Avenue, Building 30, 3rd Floor, Room 3301, San Francisco, CA 94110. Email: patti.katz@ucsf.edu.

Submitted for publication July 30, 2020; accepted in revised form January 5, 2021.

SIGNIFICANCE & INNOVATIONS

- This is the first study examining the association between physical inactivity and risk of incident depression in systemic lupus erythematosus (SLE).
- After adjusting for potential confounding factors, physical inactivity conferred a >3-fold increased risk of developing depression over 2 years of follow-up among a diverse lupus cohort.
- Physical inactivity was the strongest independent predictor of new onset depression, even more than poverty-level income, racial-ethnic minority status, SLE disease activity, coexisting cardiovascular disease, or other comorbidities.
- Interventions to reduce sedentary behavior among SLE patients may reduce the disproportionate burden of depression experienced by this high-risk group.

In order to address this knowledge gap, we sought to determine whether physical inactivity predicts subsequent new onset depression among individuals with SLE, and if so, the magnitude of the associated risk. We used data from a racially and ethnically diverse lupus cohort to assess whether patients who were not depressed at baseline but reported low physical activity were at increased risk for developing depression over time.

PATIENTS AND METHODS

Study design and participants. Subjects were participants in the California Lupus Epidemiology Study (CLUES), a prospective longitudinal sample of individuals with SLE. Briefly, starting in 2015, participants in CLUES were recruited through the California Lupus Surveillance Project, which used outpatient, hospital, and laboratory records to identify all SLE patients residing in San Francisco County from 2007 to 2009 (14). Additional participants in the geographic region were identified through academic and community rheumatology clinics and from earlier studies of genetic risk factors for SLE outcomes (15,16). SLE diagnoses were confirmed by study physicians based on the following: 1) ≥4 of the 11 American College of Rheumatology (ACR) revised criteria for the classification of SLE (17,18); 2) meeting 3 of the 11 ACR criteria with a rheumatologist's documented diagnosis of SLE; or 3) a confirmed diagnosis of lupus nephritis. This combined definition of SLE has been used in prior population-based studies (16).

Participants were assessed annually either by telephone or in person. For the baseline assessment, the majority of participants (332 of 431) completed an in-person research clinic visit, whereas annual follow-up visits were conducted either in-person or by telephone. The in-person visits included collection and review of medical records prior to the visit; a history and physical examination conducted by a physician specializing in lupus; collection of biospecimens for clinical and research purposes; and completion of a structured interview administered by an experienced research assistant. CLUES specifically aimed to include a diverse patient sample, with representation from multiple racial and ethnic groups speaking multiple languages. Therefore, research clinic visits and interviews were conducted in 4 languages: English, Spanish, Mandarin, or Cantonese. The study was approved by the University of California, San Francisco Institutional Review Board, and all participants provided informed consent.

Given our objective to assess independent predictors of incident depression, participants were included in these analyses if they completed an in-person baseline assessment, at least 1 follow-up visit, and did not meet criteria for depression at baseline (see definition of depression below). There were 306 participants who completed an in-person assessment at baseline and had at least 1 follow-up assessment, of whom 81 met criteria for depression at baseline; the remaining 225 participants without depression at baseline were eligible for inclusion in this analysis.

Measures. *Sedentary behavior.* The primary predictor of interest was physical inactivity at the baseline assessment period. Inactivity was assessed using a single item from the Rapid Assessment of Physical Activity instrument; participants who agreed to the statement, "I rarely or never do any physical activities" were classified as inactive. We focused the analysis on endorsement of sedentary behavior rather than self-report of time spent exercising, as prior exercise studies indicate that self-report physical activity is frequently over-reported (19).

Incident depression. The primary outcome was incident depression, assessed by the 8-item Patient Health Questionnaire depression scale (PHQ-8), a validated screening measure for which scores of ≥10 have a high correspondence with clinical diagnoses of depressive disorders in large clinical studies (20). We use the term "depression," although we recognize that meeting the ≥10 cut point is not the equivalent of a clinical diagnosis of depression. Incident depression was defined as a change in PHQ-8 score from <10 at baseline to ≥10 during follow-up.

SLE-specific disease factors. Age of diagnosis was obtained by self-report. Disease damage was measured with the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index (SDI), a physician-completed assessment that provides a composite score for cumulative organ damage (21). Disease activity was measured with the Systemic Lupus Erythematosus Disease Activity Index (SLEDAI), a validated physician-completed instrument that consists of data from 24 weighted clinical and laboratory variables from 9 organ systems (22,23). Participants were also queried regarding current treatment with glucocorticoids (including dosage and frequency) as well as other immunomodulatory medications.

Other variables. Participants were asked about sociodemographic characteristics, including sex, age, race, educational

attainment (categorized as high-school graduate or less, versus those with additional education), and income (categorized for analysis as household income ≤ or >125% of the federal poverty level). Height and weight were measured during the baseline in-person visit or self-reported by telephone-only participants, and body mass index (BMI) was calculated as weight (kg) divided by height (m^2). Participants were also queried regarding smoking status and major comorbidities such as cardiovascular disease, diabetes mellitus, asthma, and cancer.

Statistical analysis. Differences in characteristics of participants who were inactive versus active at baseline were tested using *t*-tests and chi-square analyses. For the analysis of risk of onset depression, we defined follow-up time as the number of months from the baseline interview to the first interview with a PHQ-8 score of ≥10, or until the most recent interview date, for those whose PHQ-8 scores remained <10. Kaplan-Meier life table analysis was used to compare incident depression by physical activity level over time, and differences were tested using a log rank test. In bivariate analyses, we compared risk of depression onset based on physical activity status, sociodemographic factors, lupus disease characteristics such as disease activity (SLEDAI), and comorbidities using unadjusted Cox proportional hazards regression models. We also assessed for interaction, including for an interaction effect between physical inactivity and income, and physical inactivity and history of depression, to determine if the association between physical inactivity and incident depression differed by poverty status or prior depressive episodes. We fit a multivariable Cox model to evaluate the independent association of physical inactivity with risk of depression onset, adjusting for race, sex, age, poverty-level income, comorbidities (cardiovascular disease, diabetes mellitus, asthma, malignancy), disease activity (SLEDAI), and disease damage (SDI). The proportional hazards assumption was investigated by testing

Table 1. Characteristics of patients with systemic lupus erythematosus (SLE) according to physical activity category*

Characteristic	Overall (n = 225)	Active (n = 184)	Inactive (n = 41)	*P*
Sociodemographic factors				
Age, mean ± SD years	45.0 ± 14.2	44.5 ± 13.8	47.0 ± 15.4	0.306
Female	88.4	87.5	92.7	0.348
Race				0.031†
White	30.2	33.7	14.6	
Hispanic	22.2	19.6	34.2	
African American	10.2	8.7	17.1	
Asian	35.1	35.3	34.2	
Unspecified or other	2.2	2.7	0.0	
Poverty income‡	16.3	13.0	32.4	0.005†
High school education or less	18.7	13.6	41.5	<0.001†
Marital status				0.013†
Never married	35.4	36.8	29.3	
Married or living with partner	56.1	56.0	56.1	
Divorced	7.2	6.0	9.8	
Widowed	1.4	0.6	4.9	
Lupus-specific characteristics				
SLE disease duration, mean ± SD years	16.8 ± 10.5	16.9 ± 10.4	16.2 ± 11.0	0.696
Disease activity by SLEDAI, mean ± SD	2.9 ± 2.9	2.9 ± 3.0	2.6 ± 2.7	0.572
Disease damage by SDI, mean ± SD	1.8 ± 2.0	1.7 ± 1.9	2.2 ± 2.1	0.138
Lupus Severity Index, mean ± SD	6.9 ± 1.6	6.8 ± 1.6	7.1 ± 1.6	0.415
History of lupus nephritis	56.5	55.0	63.4	0.323
Prednisone ≥7.5 mg/day	21.8	21.7	22.0	0.976
Any glucocorticoid use over prior year	64.7	61.9	77.8	0.071
Current hydroxychloroquine use	95.7	96.2	93.6	0.523
Comorbidities and health status				
Cardiovascular disease§	12.0	9.2	24.4	0.007†
Diabetes mellitus	7.1	6.0	12.2	0.165
Asthma	9.3	8.2	14.6	0.197
History of malignancy	8.0	7.6	9.8	0.647
Body mass index, mean ± SD kg/m^2	25.6 ± 6.2	25.2 ± 5.2	27.8 ± 9.3	0.013†
Current smoker	3.6	3.9	2.5	0.679
History of depression	26.1	26.2	25.6	0.939

* Values are the percentage unless indicated otherwise. *P* values were calculated using chi-square tests for categorical measures and *t*-tests for continuous measures. SDI = Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index; SLEDAI = Systemic Lupus Erythematosus Disease Activity Index.
† Significant.
‡ Poverty income defined as ≤125% of the federal poverty level.
§ Cardiovascular disease: history of stroke, coronary artery disease, and/or myocardial infarction.

Table 2. Bivariate associations of physical inactivity and covariates with incident depression*

Characteristic	No depression (n = 188)	Incident depression (n = 37)	Unadjusted HR (95% CI)
Physical activity status			
Inactive†	14.4	37.8	2.89 (1.46–5.71)‡
At least minimal physical activity	85.6	62.2	Ref.
Sociodemographic factors			
Age, mean ± SD years	44.3 ± 14.1	48.2 ± 13.8	1.01 (0.99–1.04)
Female	89.4	83.8	0.64 (0.27–1.55)
Race/ethnicity			0.79 (0.61–1.03)
White	28.7	37.8	
Hispanic	21.3	27.0	
African American	10.6	8.1	
Asian	36.7	27.0	
Other	2.7	0.0	
Poverty income§	13.6	29.4	2.27 (1.08–4.77)‡
Education less than a bachelor's degree	43.6	59.5	1.56 (0.80–3.02)
Marital status			0.99 (0.61–1.61)
Never married	35.5	35.1	
Married or living with partner	55.9	56.8	
Divorced or separated	7.0	8.1	
Widowed	1.6	0.0	
Lupus-specific characteristics			
SLE disease duration, mean ± SD years	16.2 ± 10.1	19.7 ± 12.0	1.02 (0.99–1.05)
Age of diagnosis, mean ± SD years	28.2 ± 12.1	28.5 ± 12.2	1.00 (0.98–1.03)
Disease activity by SLEDAI, mean ± SD	2.9 ± 3.0	2.6 ± 2.8	0.97 (0.87–1.09)
Disease damage by SDI, mean ± SD	1.6 ± 1.8	2.8 ± 2.5	1.23 (1.08–1.40)‡
History of lupus nephritis	55.4	62.2	1.22 (0.63–2.40)
Prednisone ≥7.5 mg/day	22.3	18.9	0.77 (0.34–1.76)
Comorbidities			
Cardiovascular disease	8.5	29.7	3.46 (1.70–7.04)‡
Diabetes mellitus	5.3	16.7	2.27 (0.92–5.60)
Asthma	7.5	18.9	2.21 (0.95–5.12)
Obesity (BMI ≥30 kg/m^2)	18.1	24.3	1.36 (0.64–2.90)
History of malignancy	8.5	5.5	0.73 (0.17–3.02)
History of depression	24.7	33.3	0.73 (0.17–3.02)

* Values are the percentage unless indicated otherwise. "No depression" was defined as depressive symptoms by the 8-item Patient Health Questionnaire (PHQ-8) below the cutoff for depression (PHQ-8 score <10) throughout the study period. 95% CI = 95% confidence interval; BMI = body mass index; HR = hazard ratio; SDI = Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index; SLEDAI = Systemic Lupus Erythematosus Disease Activity Index; Ref. = reference.
† Inactive was defined as rare participation in physical activities by self-report.
‡ Significant.
§ Poverty income was defined as ≤125% of the federal poverty level.

the constancy of the log hazard ratio (HR) over time by means of the log-minus-log survival plots and interaction with time (log transformed); these tests revealed no violations of the proportional hazards assumption. All analyses were performed using Stata, version 14.

RESULTS

Sample characteristics. Table 1 shows the baseline characteristics of the study participants for the overall sample and according to physical activity status (physically inactive versus active). The cohort participants were racially and ethnically diverse; they were 35% Asian, 30% White, 22% Hispanic, 10% African American, and 2% percent other. Eighteen percent of participants reported doing no physical activity (sedentary), and people in the sedentary group were more likely to be Hispanic or African American, live on or below poverty income, and have less education. The participants in the inactive group were more likely to have a history of lupus nephritis and a higher BMI, but there was no significant association for physical inactivity with lupus disease activity or disease damage.

Bivariate associations of inactivity with incident depression. We included patients with a history of depression, which represented 26.1% of the cohort (Table 1), but no participants were depressed at baseline because participants meeting criteria for depression were excluded in order to assess for new-onset depression during follow-up. The inactive behavior was stable over the first year of the study; only 5% of participants provided a different response to the question regarding inactivity between the baseline assessment and study visit performed 1 year later. Importantly, the mean ± SD

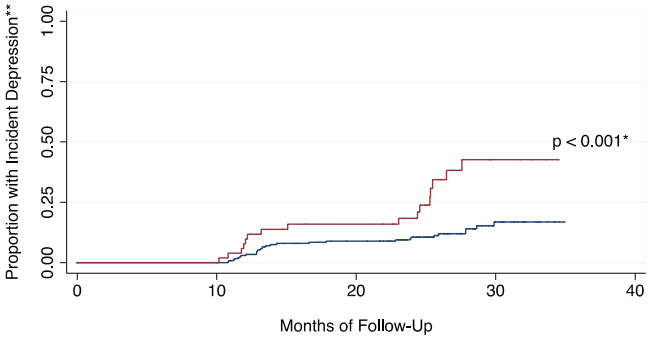


Figure 1. Cumulative proportion of systemic lupus erythematosus patients with depression by physical activity for the sedentary (red) and not sedentary (blue) groups. * = *P* by Kaplan-Meier life table analysis log rank test; ** = incident depression defined as a change in the 8-item Patient Health Questionnaire depression scale from <10 at baseline to ≥10 during follow-up.

PHQ-8 score at baseline did not differ by activity status (3.96 ± 2.78 among inactive, 3.43 ± 3.13 among active; $P = 0.23$, data not shown). In other words, scores of the participants in the inactive group were not hovering just below the PHQ-8 score threshold for depression during the baseline assessment.

There were 37 incident cases of depression (16% of the cohort) over a mean of 26 months of follow-up. Among participants who were inactive at baseline, the percent with incident depression was 38%, compared to 14% among the nonsedentary participants. In bivariate Cox proportional hazards regression analyses, several sociodemographic, behavior/lifestyle, and health factors were significantly associated with greater risk of incident depression. Physical inactivity showed a strong unadjusted association with incident depression (HR 2.89 [95% confidence interval (95% CI) 1.46–5.71]) (Table 2). Among the sociodemographic factors, poverty-level income was the only variable to significantly associate with depression (HR 2.27 [95% CI 1.08–4.77]). The factors related to health status that were significantly associated with incident depression included cardiovascular disease (HR 3.46 [95% CI 1.70–7.04]) and physician-assessed disease damage (SDI) (HR 1.23 [95% CI 1.08–1.40]). Kaplan-Meier analysis showed a significantly increased risk of depression onset among the inactive group in comparison to the active group (log-rank chi square = 12.4, $P < 0.001$) (Figure 1). There was no interaction between inactivity and poverty on risk of depression.

Multivariable analysis. In the multivariable Cox proportional hazards regression model, inactivity at baseline associated with more than a 3-fold increased risk of incident depression (HR 3.88 [95% CI 1.67–9.03]) during the follow-up period, adjusted for age, sex, race, income, self-report disease activity, self-report disease damage, and comorbidities (Table 3). The other variables in the multivariable model that significantly associated with elevated depression risk included male sex, White race, and higher lupus disease damage, but physical inactivity conferred the greatest and most statistically significant risk (data not shown).

DISCUSSION

This study is the first to investigate whether physical inactivity impacts new-onset depression in individuals with SLE. We found that low levels of physical activity were highly predictive of incident depression in this group. Among this cohort of individuals with lupus who were sedentary, there was a >3-fold increased risk of new-onset depression over the subsequent 2 years, even after adjusting for comorbidities, sociodemographic risk factors, and indices of disease severity and damage. Furthermore, physical inactivity was the strongest independent predictor of new-onset depression, even more than poverty-level income, racial and ethnic minority status, SLE disease activity, coexisting cardiovascular disease, or other comorbidities. Given the high burden of depression experienced by lupus patients relative to the general population (even among those with low disease activity and less severe disease), this finding is an important step toward understanding the contribution of lifestyle factors to mood symptoms in a uniquely vulnerable patient group.

Prior studies have demonstrated that exercise reduces the risk of incident depression in the general population, but this is the first study to our knowledge to investigate the relationship of physical inactivity to incident depression in patients with systemic lupus, a uniquely vulnerable group. Factors known to contribute to the higher burden of depression in SLE relative to the general population include reaction to chronic illness, fatigue, treatment side effects, and socioeconomic factors (4,8,9,24,25). In a minority of SLE patients, depression is immune mediated and associated with anti–ribosomal P antibodies and antibodies to *N*-methyl-D-aspartate receptors (26,27). This study builds on existing literature by establishing inactivity as a strong independent predictor of depression in SLE. Furthermore, we show that an affirmative response to the simple statement, "I rarely or never do any physical

Table 3. Adjusted risk of incident depression according to physical activity status among systemic lupus erythematosus patients*

	HR_{adj} (95% CI)†
Physically active	1 (–)
Physically inactive	3.88 (1.67–9.03)

* Among the 225 patients eligible for inclusion in the multivariable analysis, 22 had missing data for ≥1 of the covariates (e.g., income), resulting in n = 201 for the adjusted model. 95% CI = 95% confidence interval; HR_{adj} = adjusted hazard ratio.
† HR_{adj} obtained from the Cox proportional hazards model adjusted for age, sex, race, income, comorbidities (cardiovascular disease, diabetes mellitus, asthma, malignancy), disease activity by the Systemic Lupus Erythematosus Disease Activity Index, and disease damage by the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index.

activities" was the most predictive variable for subsequent depression, suggesting an important opportunity to reduce the burden of depression among lupus patients by screening and intervening on sedentary behavior as part of routine health care maintenance. For example, this question could be integrated during ambulatory rheumatology check-in procedures, and an affirmative response could trigger treating physicians to provide education, instruction, and prescription of exercise.

Even patients who did not meet public health guidelines for physical activity but participated in some amount of regular light activity were at significantly lower risk of incident depression relative to the sedentary group in our study. This finding is in keeping with the US Office of Disease Prevention and Health Promotion 2018 Physical Activity Guidelines Update, which asserts that people incur health benefits even with small increases in activity (28). Moore et al showed that there is no lower threshold for the amount of leisure time in physical activity that confers a benefit for all-cause mortality (any amount is helpful with an increasing magnitude of benefit up to 20 hours per week) (29), and our data suggest a similar relationship between any amount of physical activity and risk reduction for incident depression in SLE. The understanding that "any physical activity counts" toward reducing risk of worse health outcomes should be shared with people living with SLE who face physical, psychological, social, or environmental barriers to achieving recommended physical activity targets but can safely reduce sedentary behavior.

One limitation of this study is the use of patient-reported instruments to adjudicate the predictor and outcome variables. For example, incident depression was assessed using a depression screening measure as opposed to clinician-confirmed diagnosis, and therefore, depression may have been missed among participants who either did not feel comfortable, or who did not understand, all of the items included in the PHQ-8. However, multiple steps were taken to mitigate this limitation, including the use of a validated instrument with favorable psychometric properties (20), use of questionnaires in multiple languages administered by research staff with language concordance, and a script for study interviewers to increase participant comfort while answering sensitive questions. In addition, since we used a depression score cut point, there was a risk that the inactive participants were hovering just under the cutoff for depression during the baseline assessment. Given this concern, we examined the distribution of PHQ-8 scores among the study sample at baseline and found that they did not differ by physical activity status, indicating a meaningful change over time in PHQ-8 scores for the inactive group relative to the nonsedentary patients. Physical inactivity was also assessed by self-report, and some participants may not have responded accurately, but single-item self-report measures of physical inactivity have demonstrated similar accuracy compared to objectively measured inactivity (30). We intentionally evaluated self-reported absence of activity rather than self-reported levels of activity to mitigate the risk of activity overestimation.

We found a strong independent association between inactivity and incident depression, but as with all observational studies, there is a risk of unmeasured confounding, and we cannot definitively infer causation. However, we were able to leverage longitudinal data to exclude participants with depression at baseline and to prospectively assess whether physical activity relates to subsequent depressive symptoms. Additionally, we used detailed clinical and sociodemographic data provided by study participants, as well as physician-assessed measures of disease activity and damage completed by rheumatologists specializing in SLE, to build a comprehensive multivariable model that included covariates for each major factor with the potential to impact both physical activity and depression.

In conclusion, we found that physical inactivity, a modifiable lifestyle behavior, is common in SLE and confers a significant independent risk of incident depression among this patient group. Our findings have important clinical implications, as roughly 40% of people with lupus will experience depression during their lifetime (1), and strategies to prevent depression represent a major unmet need for those with this disease. Results support the importance of even low levels of physical activity and suggest an urgent need for approaches (such as health care providers' physical activity prescriptions and referrals to appropriate community-based exercise programs) to increase physical activity in this high-risk patient population. In addition to reducing the risk of important physical comorbidities such as cardiovascular disease, our data suggest that a small increase in physical activity may also reduce the risk of major mental health challenges experienced disproportionally in SLE.

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. Katz 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. Yazdany, Dall'Era, Katz.
Acquisition of data. Lanata, Dequattro.
Analysis and interpretation of data. Patterson, Trupin, Hartogensis.

REFERENCES

1. Bachen EA, Chesney MA, Criswell LA. Prevalence of mood and anxiety disorders in women with systemic lupus erythematosus. Arthritis Rheum 2009;61:822–9.
2. Palagini L, Mosca M, Tani C, Gemignani A, Mauri M, Bombardieri S. Depression and systemic lupus erythematosus: a systematic review. Lupus 2013;22:409–16.
3. Karol DE, Criscione-Schreiber LG, Lin M, Clowse ME. Depressive symptoms and associated factors in systemic lupus erythematosus. Psychosomatics 2013;54:443–50.
4. Julian LJ, Yelin E, Yazdany J, Panopalis P, Trupin L, Criswell LA, et al. Depression, medication adherence, and service utilization in systemic lupus erythematosus. Arthritis Rheum 2009;61:240–6.
5. Ward MM, Marx AS, Barry NN. Psychological distress and changes in the activity of systemic lupus erythematosus. Rheumatology (Oxford) 2002;41:184–8.

6. Kessler RC, Heeringa S, Lakoma MD, Petukhova M, Rupp AE, Schoenbaum M, et al. Individual and societal effects of mental disorders on earnings in the United States: results from the national comorbidity survey replication. Am J Psychiatry 2008;165:703–11.
7. Julian LJ, Tonner C, Yelin E, Yazdany J, Trupin L, Criswell LA, et al. Cardiovascular and disease-related predictors of depression in systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2011;63: 542–9.
8. Huang X, Magder LS, Petri M. Predictors of incident depression in systemic lupus erythematosus. J Rheumatol 2014;41:1823–33.
9. McCormick N, Trupin L, Yelin EH, Katz PP. Socioeconomic predictors of incident depression in systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2018;70:104–13.
10. Choi KW, Zheutlin AB, Karlson RA, Wang MJ, Dunn EC, Stein MB, et al. Physical activity offsets genetic risk for incident depression assessed via electronic health records in a biobank cohort study. Depress Anxiety 2020;37:106–14.
11. Mekary RA, Lucas M, Pan A, Okereke OI, Willett WC, Hu FB, et al. Isotemporal substitution analysis for physical activity, television watching, and risk of depression. Am J Epidemiol 2013;178:474–83.
12. Schuch FB, Vancampfort D, Firth J, Rosenbaum S, Ward PB, Silva ES, et al. Physical activity and incident depression: a meta-analysis of prospective cohort studies. Am J Psychiatry 2018;175: 631–48.
13. Strawbridge WJ, Deleger S, Roberts RE, Kaplan GA. Physical activity reduces the risk of subsequent depression for older adults. Am J Epidemiol 2002;156:328–34.
14. Dall'Era M, Cisternas MG, Snipes K, Herrinton LJ, Gordon C, Helmick CG. The incidence and prevalence of systemic lupus erythematosus in San Francisco County, California: the California Lupus Surveillance Project. Arthritis Rheumatol 2017;69:1996–2005.
15. Freemer MM, King TE Jr., Criswell LA. Association of smoking with dsDNA autoantibody production in systemic lupus erythematosus. Ann Rheum Dis 2006;65:581–4.
16. Parsa A, Lovett DH, Peden EA, Zhu L, Seldin MF, Criswell LA. Renin-angiotensin system gene polymorphisms predict the progression to renal insufficiency among Asians with lupus nephritis. Genes Immun 2005;6:217–24.
17. Hochberg MC, for the Diagnostic and Therapeutic Criteria Committee of the American College of Rheumatology. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus [letter]. Arthritis Rheum 1997;40:1725.
18. Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1982;25:1271–7.
19. Steene-Johannessen J, Anderssen SA, van der Ploeg HP, Hendriksen IJ, Donnelly AE, Brage S, et al. Are self-report measures able to define individuals as physically active or inactive? Med Sci Sports Exerc 2016;48:235–44.
20. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009;114:163–73.
21. Gladman D, Ginzler E, Goldsmith C, Fortin P, Liang M, Urowitz M, et al. The development and initial validation of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology damage index for systemic lupus erythematosus. Arthritis Rheum 1996;39:363–9.
22. Buyon JP, Petri MA, Kim MY, Kalunian KC, Grossman J, Hahn BH, et al. The effect of combined estrogen and progesterone hormone replacement therapy on disease activity in systemic lupus erythematosus: a randomized trial. Ann Intern Med 2005;142:953–62.
23. Romero-Diaz J, Isenberg D, Ramsey-Goldman R. Measures of adult systemic lupus erythematosus: updated Version of British Isles Lupus Assessment Group (BILAG 2004), European Consensus Lupus Activity Measurements (ECLAM), Systemic Lupus Activity Measure, Revised (SLAM-R), Systemic Lupus Activity Questionnaire for Population Studies (SLAQ), Systemic Lupus Erythematosus Disease Activity Index 2000 (SLEDAI-2K), and Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index (SDI). Arthritis Care Res (Hoboken) 2011;63 Suppl 11:S37–46.
24. Jump RL, Robinson ME, Armstrong AE, Barnes EV, Kilbourn KM, Richards HB. Fatigue in systemic lupus erythematosus: contributions of disease activity, pain, depression, and perceived social support. J Rheumatol 2005;32:1699–705.
25. Kozora E, Ellison MC, Waxmonsky JA, Wamboldt FS, Patterson TL. Major life stress, coping styles, and social support in relation to psychological distress in patients with systemic lupus erythematosus. Lupus 2005;14:363–72.
26. Lapteva L, Nowak M, Yarboro CH, Takada K, Roebuck-Spencer T, Weickert T, et al. Anti–*N*-methyl-D-aspartate receptor antibodies, cognitive dysfunction, and depression in systemic lupus erythematosus. Arthritis Rheum 2006;54:2505–14.
27. Schneebaum AB, Singleton JD, West SG, Blodgett JK, Allen LG, Cheronis JC, et al. Association of psychiatric manifestations with antibodies to ribosomal P proteins in systemic lupus erythematosus. Am J Med 1991;90:54–62.
28. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The Physical Activity Guidelines for Americans. JAMA 2018;320:2020–8.
29. Moore SC, Patel AV, Matthews CE, Berrington de Gonzalez A, Park Y, Katki HA, et al. Leisure time physical activity of moderate to vigorous intensity and mortality: a large pooled cohort analysis. PLoS Med 2012;9:e1001335.
30. Milton K, Bull FC, Bauman A. Reliability and validity testing of a single-item physical activity measure. Br J Sports Med 2011;45:203–8.

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1105–1112
DOI 10.1002/acr.24552

American College of Rheumatology
Empowering Rheumatology Professionals

Association of Renal Arteriosclerosis With Atherosclerotic Cardiovascular Disease Risk in Lupus Nephritis

Shivani Garg,[1] Amish N. Raval,[1] Karen E. Hansen,[1] Weixiong Zhong,[1] Yabing Huang,[2] Maureen Smith,[1] Sarah E. Panzer,[1] and Christie M. Bartels[1]

Objective. Lupus nephritis (LN) predicts a 9-fold higher atherosclerosis cardiovascular disease (ASCVD) risk, highlighting the urgent need to target ASCVD prevention. Studies in IgA nephropathy reported that severe renal arteriosclerosis (r-ASCL) in diagnostic biopsies strongly predicted ASCVD risk. We recently found that 50% of LN pathology reports overlooked r-ASCL reporting, which could explain prior negative LN ASCVD risk studies. The present study was undertaken to examine associations between a composite of reported and overread r-ASCL and ASCVD events in LN.

Methods. Data were abstracted from all LN patients who underwent diagnostic biopsy between 1994 and 2017, including demographic information, ASCVD risk factors, and pathology reports at the time of LN diagnosis. We manually validated all incident ASCVD events. We overread 25% of the biopsies to grade r-ASCL using the Banff criteria. We supplemented the overread r-ASCL grade, when available, to determine the composite of reported and overread r-ASCL grade.

Results. Among 189 incident LN patients, 78% were female, 73% White, and the median age was 25 years. Overall, 31% had any reported r-ASCL, and 7% had moderate-to-severe r-ASCL. After incorporating systematically re-examined r-ASCL grade, the prevalence of any and moderate-to-severe r-ASCL increased to 39% and 12%, respectively. We found 22 incident ASCVD events over 11 years of follow-up. Using a composite of reported and overread r-ASCL grade, we found that severe r-ASCL in diagnostic LN biopsies was associated with 9-fold higher odds of ASCVD.

Conclusion. Severe r-ASCL can predict ASCVD in LN; therefore, larger studies are required to systematically report r-ASCL and examine ASCVD associations.

INTRODUCTION

Systemic lupus erythematosus (SLE, or lupus) is a leading cause of mortality in young women, with a 3- to 5-fold higher standard mortality rate and accelerated atherosclerotic cardiovascular disease (ASCVD) risk compared to the general population (1,2). Furthermore, studies have shown that patients with lupus nephritis (LN) have a 9-fold higher risk of ASCVD (3) and a 2-fold higher risk of ASCVD or carotid plaques compared to lupus patients without nephritis (4). The historical concept of ASCVD as a late complication of SLE, 5–9 years after SLE diagnosis, was recently challenged by diverse SLE population-based studies (5–7). These studies reported a significant early and increased risk of ASCVD in SLE and LN patients around the time of SLE diagnosis (8–10). These studies suggest a role of subclinical autoimmunity in accelerating atherosclerosis early in the SLE disease course (8,9). Studies have also reported an interplay between inflammatory and traditional risk factors that predispose SLE and LN patients to ASCVD occurrence (11,12). Therefore, there is an urgent need to identify early risk factors of ASCVD to implement timely ASCVD prevention in SLE and LN patients.

Studies in IgA nephropathy and renal transplantation have shown that the presence of severe renal arteriosclerosis (r-ASCL) in kidney biopsies, performed at the time of disease diagnosis, is an early predictor of ASCVD in patients with IgA nephropathy and transplantation (13,14). Yet, a few studies including LN biopsies failed to report a similar association between the presence of severe r-ASCL and ASCVD occurrence (15,16). We previously reported an accelerated and higher burden of moderate-severe r-ASCL in LN patients, at the time of LN diagnosis, compared to

[1]Shivani Garg, MD, MS, Amish N. Raval, MD, Karen E. Hansen, MD, MS, Weixiong Zhong, MD, PhD, Maureen Smith, MD, PhD, Sarah E. Panzer, MD, Christie M. Bartels, MD, MS: University of Wisconsin, Madison; [2]Yabing Huang, MD, PhD: RenMin Hospital of Wuhan University, Hubei, China.

No potential conflicts of interest relevant to this article were reported.

Address correspondence to Shivani Garg, MD, MS, Department of Medicine-Rheumatology Division, MFCB, 4th Floor, 1685 Highland Avenue, Madison, WI 53705. Email: sgarg@medicine.wisc.edu.

Submitted for publication August 13, 2020; accepted in revised form January 5, 2021.

SIGNIFICANCE & INNOVATIONS

- The rate of incident atherosclerotic cardiovascular disease (ASCVD) in lupus nephritis (LN) patients in our cohort was 22%, with 12% moderate-to-severe renal arteriosclerosis (r-ASCL) at the time of LN diagnosis.
- Chronic kidney disease (CKD) stage ≥3 was associated with higher incident ASCVD occurrence in LN patients. Only 9% of eligible LN patients with CKD stage ≥3 or hyperlipidemia were prescribed statins, highlighting gaps in implementing ASCVD prevention in LN.
- After systematically re-examining biopsy results for r-ASCL using Banff criteria, we found that the presence of severe composite r-ASCL in diagnostic LN biopsies was associated with 9-fold higher odds of ASCVD. This interesting finding forms the basis to establish severe composite r-ASCL as a predictor of ASCVD events using a larger sample size in different study cohorts.

healthy peers (17). Furthermore, we reported an 80% positive predicted value of reported r-ASCL that indicated the accurate grading and reporting of r-ASCL by a pathologist when r-ASCL was present. However, we found that 50% of the pathology reports overlooked r-ASCL reporting. This oversight is because the current pathology reporting guidelines focus primarily on glomerular changes but place no emphasis on standard systematic r-ASCL and other vascular lesions grading in all LN biopsies. Therefore, we hypothesized that careful, systematic examination of biopsies for r-ASCL will yield a greater reported prevalence of r-ASCL, which, in turn, will be positively linked to clinical ASCVD in LN patients. Hence, we aimed to investigate whether the true relationship between r-ASCL and ASCVD in LN patients is underrecognized due to underreported r-ASCL. We also aimed to study the current gaps in implementing ASCVD prevention strategies based on 2012 American College of Rheumatology (ACR) LN guidelines, such as initiating statins in LN patients with chronic kidney disease (CKD) stage ≥3 or hyperlipidemia after LN diagnosis (18). We hypothesized that severe r-ASCL would be an early determinant of future ASCVD events in LN when a composite r-ASCL grade was determined using reported and overread r-ASCL grading, and that we would find significant gaps in statin use in eligible LN patients.

PATIENTS AND METHODS

Cohort. We identified all consecutive LN patients who underwent native renal biopsy between 1994 and 2017 at the University of Wisconsin Hospital and Clinics. We abstracted data on patient and disease characteristics from a comprehensive renal biopsy database and electronic health records. We used the 1997 updated ACR guidelines (19) and Systemic Lupus International Collaborating Clinics 2012 (20) criteria to validate the SLE diagnoses. We included the first native LN biopsies for all validated SLE patients in our cohort. LN diagnosis was validated using the 2003 International Society of Nephrology (ISN)/Renal Pathology Society (RPS) classification of LN (21). We excluded subsequent pathology reports after incident LN diagnosis, patients with transplant kidneys, and those who did not meet SLE diagnostic criteria and the ISN/RPS 2003 classification for LN. The University of Wisconsin Human Research Protection Program approved this study with a waiver of informed consent (IRB number 2016-1260).

Covariates (sociodemographic data and comorbidities). Using electronic health record and database information, we recorded sociodemographic and comorbidity data at the time of biopsy. Patient and disease characteristics included age, sex, race, smoking status, and comorbidities. Hypertension (HTN), hyperlipidemia, and diabetes mellitus (DM) were assessed using International Statistical Classification of Diseases and Related Health Problems, Tenth Revision codes, problem list diagnoses, or medication use. History of previous ASCVD events >1 year before LN diagnosis were manually identified by a physician author (SG), who was blinded to the pathology overread results, at the time of chart review to adjudicate ASCVD events. CKD stage was assessed using glomerular filtration rate at the time of biopsy, and the presence of nephrotic syndrome was defined as urine protein creatinine ratio or 24-hour urine protein ≥3 grams at the time of LN diagnosis. A modified risk score was calculated by adding the 7 risk factors used in ASCVD risk estimation: age, sex, race, smoking history, hyperlipidemia, HTN, and DM, and history of previous ASCVD events (>1 year before LN diagnosis) (22). We categorized LN patients in 2 categories: modified ASCVD score >1, and modified ASCVD score ≤1.

Renal histopathology. Renal biopsy was performed for clinical indication (edema, increase in serum creatinine, hematuria, and/or proteinuria), and pathologic assessments were performed using the 2003 ISN/RPS classification for LN. We abstracted the following data from renal pathology reports: 1) LN class (I–VI), which was further categorized into proliferative or nonproliferative LN, and 2) reported r-ASCL data were abstracted from renal pathology reports and were classified into the Banff r-ASCL grading categories as described in our previous manuscript (17).

Overreading of r-ASCL in LN biopsies using Banff criteria. As described in our previous manuscript, a blinded study pathologist (YH) overread a 25% random sample (n = 43 biopsies). This sample was randomly selected to overread ~50% with and without reported r-ASCL with oversampling of recent biopsies (2014–2017), which could have improved with new standards in transplant biopsy grading. Using the Banff criteria,

the r-ASCL grade was directly interpreted from the slides as none, mild (<25%), moderate (26–50%), and severe (>50%) luminal narrowing. A composite of reported and overread r-ASCL was identified by reviewing all original pathology reports and new findings for the overread sample of 43 biopsies (25% sample), which included oversampled recent and reported negative studies. When available, we supplemented the overread Banff r-ASCL grade to determine the composite of reported and overread r-ASCL grade.

Primary outcome (incident ASCVD event). ASCVD was defined using the following events: 1) ischemic heart disease (IHD) including myocardial infarction, coronary artery revascularization, abnormal stress test, abnormal angiogram, and events documented by a cardiologist; 2) stroke and transient ischemic attack (TIA), and 3) peripheral vascular disease (PVD) (such as abnormal ankle–brachial index, abnormal peripheral angiography, limb ischemia undergoing bypass or angioplasty or documented by vascular surgeon). Incident ASCVD was defined as the first ASCVD event that occurred between 1 year before LN diagnosis until 10 years after LN diagnosis. Patients with previous ASCVD events >1 year prior to LN diagnosis were excluded from the final analyses. In the case of multiple ASCVD events, the first ASCVD event was included in the analyses. All LN patient electronic health records were searched to manually identify events as defined above, which were then adjudicated by an author (SG) using the standard American College of Cardiology and American Heart Association guidelines (23–28).

Current care gaps in targeted ASCVD prevention in LN patients. The 2010 ACR management guidelines for LN recommend initiating statin therapy in all LN patients with hyperlipidemia. These guidelines also identify CKD stage ≥3 as a strong risk factor of ASCVD in LN (18). Likewise, other studies in the general population with CKD have reported the significant role of statins in reducing ASCVD risk in these patients (29). We thereby abstracted data on LN patients with CKD stage ≥3 and/or hyperlipidemia, defined as low-density lipoprotein >100 mg/dl, to determine rates of statin use. We manually searched and abstracted data on statin prescriptions written for LN patients who met these eligibility criteria for statin therapy.

Statistical analysis. Descriptive data were expressed as median and range for data that were not normally distributed or mean ± SD for normally distributed data. Data were also expressed as adjusted odds ratios (OR_{adj}) and 95% confidence intervals (95% CIs) where appropriate. We calculated the incidence of ASCVD events, starting 1 year before LN diagnosis until 10 years after LN diagnosis for follow-up.

We used similar methods as published in our recent paper: we analyzed key cutoffs for age and SLE duration before LN diagnosis and examined associations between ASCVD and a composite of reported and overread r-ASCL (17). Along with basic sociodemographic data, variables with a P value of <0.1 in univariable models and LN proliferative class were included in multivariable analyses. We used multivariable logistic regression to analyze the associations between incident ASCVD and reported r-ASCL, and incident ASCVD and a composite of reported and overread r-ASCL controlling for other covariates.

We compared the traditional atherosclerosis CVD risk factors, renal factors, reported r-ASCL, and composite r-ASCL at the time of LN diagnosis in patients who developed CVD and patients who did not develop CVD during the follow-up period. We used univariable logistic regression to analyze these groups. We also performed predictive modeling, using logistic regression, that incorporated key traditional, renal, and pathology-related risk factors, including HTN, CKD stage ≥3, and severe r-ASCL.

Finally, we calculated the percentage of LN patients eligible to start statin therapy per ACR and European Alliance of Associations for Rheumatology (EULAR) guidelines (18,30) and the percentage of eligible patients who started receiving statins. Statistical software R, version 3.4.1, was used for all analyses (31).

RESULTS

Among 189 incident validated LN patients with kidney biopsies, the median age at LN diagnosis was 25 years (40% were <19 years [range 2–79 years]), and other characteristics are shown in Table 1. Of the 189 patients, 78% were female, 73% were White, 21% were from other races, and 6% had missing race data. In our cohort at the time of LN diagnosis, 27% of the patients had CKD stage ≥3, 34% had >1 modified ASCVD risk score, and 49% of the patients were diagnosed with LN within 2 years of SLE diagnosis. We found that 26% of the patients had nephrotic syndrome at the time of LN diagnosis. Regarding renal histopathology at the time of LN diagnosis, 41% of patients were classified as having proliferative LN, 31% had any (mild, moderate, or severe) reported r-ASCL, and 7% had moderate-to-severe reported r-ASCL. After using the composite of reported and overread r-ASCL grade, we found the prevalence of any (mild, moderate, or severe) and moderate-to-severe r-ASCL increased to 39% and 12%, respectively (Table 1).

Incident ASCVD. Overall, we found 22 incident ASCVD events over an 11-year follow-up period, starting 1 year before LN diagnosis, and 2 of these events occurred 1 year before LN diagnosis. The ASCVD events were 54.5% stroke- or TIA-related events (n = 12), 31.9% IHD-related events (n = 7), and 13.6% PVD-related events (n = 3). In our cohort, the incident rate of ASCVD in LN patients was 12%.

Current gaps in initiating statin therapy in eligible LN patients. The ACR LN and EULAR guidelines define hyperlipidemia or CKD stage ≥3 as high-risk criteria for future ASCVD events in LN (18,30). These guidelines recommend initiating statin

Table 1. Demographic information on lupus nephritis (LN) patients (n = 189)*

Variable at LN diagnosis	Value
Age, median (range) years	25 (2–79)
Sex	
Female	148 (78)
Male	41 (22)
Race	
White	138 (73)
Black	17 (9)
Asian	15 (8)
Others	8 (4)
Unknown/missing	11 (6)
Hypertension	37 (20)
Diabetes mellitus	9 (5)
Hyperlipidemia	16 (9)
Smoking	44 (23)
ASCVD risk score >1	64 (34)
Chronic kidney disease stage ≥3	50 (27)
SLE duration <2 years	93 (49)
Nephrotic syndrome	49 (26)
Reported renal arteriosclerosis	
Mild	43 (24)
Moderate	13 (6)
Severe	2 (1)
Composite of reported and overread renal arteriosclerosis	
Mild	50 (27)
Moderate	17 (9)
Severe	5 (3)
ASCVD events over 11-year follow-up	22 (12)

* Values are the number (%) unless indicated otherwise. ASCVD = atherosclerotic cardiovascular disease; SLE = systemic lupus erythematosus.

therapy at the time of LN diagnosis in all patients who meet the high-risk criteria (hyperlipidemia and CKD stage ≥3). In our cohort, 22 total ASCVD events occurred starting 1 year before LN diagnosis through 11 years of follow-up; 20 events occurred after LN diagnosis, and 2 occurred 1 year before LN diagnosis. Among the 20 patients with incident ASCVD events after LN diagnosis in our cohort, none were receiving statin therapy at the time of LN diagnosis. Furthermore, 11 patients (55%) met high-risk criteria (hyperlipidemia and CKD stage ≥3) to implement statin therapy at the time of LN diagnosis, yet only 1 patient (9%) was initiated on statin therapy. All patients with severe r-ASCL met the high-risk criteria (hyperlipidemia and CKD stage ≥3) to initiate statin therapy, but none of them were started on statins for ASCVD prevention.

Determinants of incident ASCVD events among those with reported r-ASCL. Among the group with reported r-ASCL, we found a strong association with CKD stage ≥3, with a 5-fold higher odds of incident ASCVD in LN patients with CKD stage ≥3 (OR_{adj} 5.4 [95% CI 1.8–18.0], $P = 0.004$). However, on multivariable analysis, we found no association between ASCVD occurrence and a modified ASCVD risk score of >1, female sex, and age ≥30 years (Table 2).

Determinants of incident ASCVD among those with a composite of reported and overread r-ASCL. Upon using the composite of reported and overread r-ASCL, we found that the presence of composite of reported and overread severe r-ASCL was associated with a 9-fold higher odds of incident ASCVD compared to those without r-ASCL (OR_{adj} 9.1 [95% CI 1.1–94], $P = 0.04$) (Table 2). Furthermore, we found that CKD stage ≥3 was associated with a 4-fold higher odds of incident ASCVD in LN (OR_{adj} 4.1 [95% CI 1.4–13], $P = 0.01$) (Table 2). Greater than 1 ASCVD risk score, female sex, and age were not associated with ASCVD (Table 2).

Comparing risk factors in LN patients who developed ASCVD versus those who did not develop ASCVD. Table 3 shows the presence of traditional ASCVD risk factors, renal factors, reported r-ASCL, and composite r-ASCL at

Table 2. Predictors of atherosclerotic cardiovascular disease (ASCVD) in lupus nephritis patients using reported renal arteriosclerosis (r-ASCL) and a composite of reported and overread r-ASCL*

	Reported r-ASCL		Composite of reported and overread r-ASCL	
Variable	OR_{adj} (95% CI)	*P*	OR_{adj} (95% CI)	*P*
ASCL none	Ref.	Ref.	Ref.	Ref.
ASCL mild	0.5 (0.1–1.9)	0.4	0.7 (0.1–2.3)	0.5
ASCL moderate	1.3 (0.2–7.8)	0.7	0.8 (0.1–4.1)	0.8
ASCL severe	NA	NA	9.1 (1.1–94)†	0.04†
Age <30 years	Ref.	Ref.	Ref.	Ref.
Age ≥30 years	1.3 (0.3–4.6)	0.7	1.7 (0.5–5.7)	0.3
Female	0.4 (0.1–1.5)	0.2	0.5 (0.2–1.5)	0.2
ASCVD score ≤1	Ref.	Ref	Ref.	Ref.
ASCVD score >1	0.5 (0.2–1.6)	0.3	0.5 (0.2–1.7)	0.3
CKD stage <3	Ref.	Ref.	Ref.	Ref.
CKD stage ≥3	5.4 (1.8–18.0)†	0.004†	4.1 (1.4–13)†	0.01†

* 95% CI = 95% confidence interval; CKD = chronic kidney disease; NA = not applicable (unable to calculate 95% CIs, as fitted numerical probabilities reached 1); OR_{adj} = adjusted odds ratio; Ref. = reference.
† Significant ($P < 0.5$).

Table 3. Variables at lupus nephritis (LN) diagnosis in patients with cardiovascular disease (CVD) versus without CVD*

Variables at LN diagnosis	CVD, yes (n = 22)	CVD, no (n = 167)	OR (95% CI)†	*P*†
Non-White race	3 (14)	37 (22)	0.53 (0.12–1.7)	0.33
Smoking ever, yes	6 (27)	38 (23)	1.3 (0.42–3.6)	0.621
Hypertension, yes	3 (14)	34 (20)	0.54 (0.12–1.7)	0.34
Diabetes mellitus, yes	0	9 (5)	NA	NA
CKD stage ≥3, yes	12 (55)‡	38 (23)‡	4.3 (1.7–11.9)‡	0.0029‡
Hyperlipidemia, yes	1 (5)	15 (9)	0.43 (0.23–2.3)	0.428
Nephrotic syndrome, yes	5 (23)	44 (26)	1.6 (0.43–5.5)	0.475
Reported r-ASCL severe, yes	0	2 (1)	NA	NA
Composite r-ASCL severe, yes	3 (14)‡	2 (1)‡	12.0 (1.8–98)‡	0.0095‡

* 95% CI = 95% confidence interval; CKD = chronic kidney disease; NA = not applicable (unable to calculate, as fitted numerical probabilities reached 1); OR = odds ratio; r-ASCL = renal arteriosclerosis.
† Derived from univariable logistic regression.
‡ Significant ($P < 0.05$).

the time of LN diagnosis in patients who developed ASCVD and patients who did not develop ASCVD during the follow-up period. Interestingly, we found that LN patients who presented with CKD stage ≥3 had significantly higher odds of developing ASCVD compared to patients who did not have CKD stage ≥3. Likewise, patients with a composite severe r-ASCL grade had a 12-fold higher risk of ASCVD occurrence compared to patients who did not have a severe r-ASCL grade.

Furthermore, we performed predictive modeling (Table 4) that incorporated key traditional, renal, and pathology-related risk factors, including HTN, CKD stage ≥3, and severe r-ASCL. We found that the presence of CKD stage ≥3 or severe r-ASCL at the time of LN diagnosis were associated with a 6-fold higher odds of ASCVD occurrence (Table 4). Finally, we found that the presence of CKD stage ≥3 or severe r-ASCL along with HTN at the time of LN diagnosis did not determine ASCVD occurrence. In our cohort, no patients had all 3 risk factors (CKD stage ≥3, severe r-ASCL, and HTN); therefore, we were unable to elaborate on the association between the presence of all 3 risk factors and ASCVD occurrence.

Table 4. Predictors of atherosclerotic cardiovascular disease (ASCVD) using a composite risk score including 2 key predictors and a traditional CVD risk factor*

Variables at LN diagnosis	OR (95% CI)†	*P*†
No CKD or severe r-ASCL or HTN	Ref.	Ref.
CKD or severe r-ASCL, present	6.4 (2.2–19)‡	0.0006‡
CKD or severe r-ASCL and HTN, present	1.1 (0.06–6.8)	0.93
HTN, present	1.5 (0.21–7.0)	0.61
CKD, severe r-ASCL and HTN, present	NA	NA

* 95% CI = 95% confidence interval; CKD = chronic kidney disease (stage ≥3); HTN = hypertension; LN = lupus nephritis; NA = not applicable (unable to calculate, as no patients had all 3 variables); OR = odds ratio; r-ASCL renal arteriosclerosis; Ref. = reference.
† Derived from univariable logistic regression.
‡ Significant ($P < 0.05$).

DISCUSSION

Our findings are among the first to highlight that stage ≥3 CKD and presence of composite severe r-ASCL on diagnostic LN biopsy results were independently associated with incident ASCVD in LN. Moreover, we demonstrated that routine biopsy reports frequently overlooked r-ASCL in the absence of systematic Banff scoring, possibly explaining variation in prior reports. Studies have consistently shown that patients with LN have a 7-fold higher odds of IHD occurrence after LN diagnosis compared to healthy peers (32) and a 9-fold higher IHD odds compared to those with SLE without nephritis (3). Furthermore, studies have reported a 2 times higher prevalence of carotid plaques in patients with LN compared to healthy peers and age-matched SLE patients without nephritis (3,4). Therefore, LN is considered an independent risk factor of ASCVD in SLE (3,32,33). Previous studies examining predictors of ASCVD in LN have shown that smoking, elevated creatinine, and HTN are common risk factors (4,15,16,32,34,35). Our study further reports that CKD stage ≥3 at the time of LN diagnosis was a strong determinant of future ASCVD events. Conversely, we found no association between the presence of >1 traditional ASCVD risk factors at the time of LN diagnosis and future ASCVD risk.

The 2012 ACR and 2019 EULAR guidelines for LN and SLE recommend initiating statin therapy for all LN patients with hyperlipidemia or other risk factors such as CKD stage ≥3 (18,30). Despite these recommendations, our study found a significant gap in implementing statin therapy in LN patients with hyperlipidemia and/or CKD stage ≥3. Only 9% of patients with hyperlipidemia and/or CKD stage ≥3 started receiving statin therapy in our cohort. Despite the significant role of statin therapy in lowering ASCVD risk in patients with lupus, LN, and CKD (18,29,36), the optimal timing and thresholds for implementing ASCVD prevention are not clear (30,37). Therefore, there is a need to examine predictors at the time of LN diagnosis to prompt the initiation of statin therapy and other ASCVD preventive strategies.

Strikingly, previous studies reported that LN patients with ASCVD occurrence were significantly younger than peers, and

ASCVD risk was 42 times higher in LN patients ages 30–39 years (3,4,32). These studies underscore a possible role of immune-mediated and inflammatory risk factors, in addition to traditional risk factors, as contributors to significantly accelerated ASCVD risk in LN patients (3,4,32,38). In our previous study, we found that the burden of r-ASCL in diagnostic LN biopsies was significantly higher and earlier in LN patients, starting between 30 and 39 years of age (39). Therefore, these findings support a mechanistic association between the presence of r-ASCL in LN biopsies at the time of LN diagnosis and ASCVD occurrence in LN patients.

Studies in other renal diseases, such as IgA nephropathy and immune-mediated renal allograft changes, have reported that severe r-ASCL in renal biopsies is a strong predictor of ASCVD events (13,14,40). In IgA nephropathy, 37% of patients with severe r-ASCL experienced ASCVD events compared to 17% of patients without r-ASCL ($P < 0.05$) (13). Similarly, another study examining immune-mediated renal graft rejection reported that the presence of severe r-ASCL in transplant biopsies predicted a 4-fold higher ASCVD risk compared to those with minimal r-ASCL (14). Therefore, these and other studies have emphasized that there is a strong association between immune-mediated severe r-ASCL and intimal changes in systemic vessels leading to atherosclerosis and other ASCVD events (13,14,40). Yet, a similar relationship in LN patients, who are at a 9-fold higher risk of IHD and a 42-fold higher ASCVD risk, has not been established (15,16).

Two studies, one by Barber et al and another by Huang et al, reported >50% prevalence of any chronic (noninflammatory) renal arterial changes in initial diagnostic LN biopsies (15,16). Both studies concluded that the presence of r-ASCL in diagnostic LN biopsies was not a risk factor or early predictor of ASCVD occurrence in LN patients. However, these studies used only the 2003 ISN/RPS classification for LN to grade pathologic changes in renal biopsies and nonstandard semiquantitative methods to grade r-ASCL using pathology reports on renal arterial changes. For example, Huang et al calculated the mean r-ASCL score to examine the association between r-ASCL scores and ASCVD events, while Barber et al categorically examined composite renal arterial sclerosis, including both r-ASCL and hyalinosis in LN biopsy reports, as risk factors for ASCVD occurrence. Notably, the 2003 ISN/RPS classification for LN and the 2018 update provide no standard criteria, such as Banff, to systematically grade r-ASCL in all LN biopsy results (21,41). However, in our previous study, we reported that 50% of routine pathology reports using 2003 ISN/RPS guidelines overlooked r-ASCL that was identified and stratified by severity using Banff criteria review (39). Moreover, a previous study in other immune-mediated renal diseases highlighted that only severe r-ASCL was a risk factor for ASCVD occurrence, and no association between renal arteriolar hyalinosis and ASCVD occurrence was found (13). The failure to systematically use standard r-ASCL grading criteria, and significant underreporting of r-ASCL in LN biopsy reports, could explain prior negative findings in contrast to our strong association.

To overcome these limitations, we used biopsy reports supplemented with standard Banff criteria to grade r-ASCL on 25% of the sample. We further examined the association between all grades of the composite of reported and overread r-ASCL and ASCVD occurrence. In our study, when only routine pathology reports were used to grade r-ASCL, we found no association between mild, moderate, or severe r-ASCL and ASCVD occurrence, which is consistent with prior negative studies. However, in the composite of reported and overread r-ASCL scores, when available on 25% of the sample, we found a 9-fold higher odds of incident ASCVD occurrence in patients with severe r-ASCL at the time of LN diagnosis.

Finally, our analysis highlighted that the presence of CKD stage ≥3 and the presence of severe r-ASCL at the time of LN diagnosis were strongly associated with ASCVD occurrence. Due to our small sample size, and yet, this interesting finding, we suggest future studies to examine such associations in larger and more diverse study cohorts.

Despite the strengths of this study, such as the inclusion of a validated incident LN cohort, systematically using Banff criteria for r-ASCL grading, and manually adjudicating all incident ASCVD events, we also acknowledge limitations. First, 73% of patients at this Midwest center were White and may not represent the LN population in the US. Second, unlike in prior reports, we found no correlation between traditional ASCVD risk factors and ASCVD occurrence. This difference from prior research could be due to sample size limitations and using an ASCVD cumulative risk score instead of individual risk factors in our analyses. Third, we were only able to overread 25% of the sample, and further overread r-ASCL grading in the remaining sample was not feasible in this study. Fourth, we did not evaluate chronicity scoring or the presence of chronic lesions as determinants of ASCVD in our cohort. We will examine chronicity scoring in LN biopsy results and ASCVD risk in our future studies. Finally, using r-ASCL grading to determine future ASCVD occurrence may not apply broadly in clinical practice given that some LN patients, who are unstable or at high risk of bleeding, might not be eligible to undergo renal biopsy. To overcome this limitation, we plan to assess other biomarkers and imaging that correlates with r-ASCL and predicts CVD risk in such LN patients. Our future work will also focus on overreading all biopsy results to grade r-ASCL using standard Banff criteria in diverse LN cohorts and examine associations between Banff grades and ASCVD.

In conclusion, after systematically re-examining biopsy results for r-ASCL using Banff criteria, we found that the presence of severe composite r-ASCL in diagnostic LN biopsy reports was associated with a 9-fold higher odds of ASCVD. This interesting finding forms the basis of establishing severe composite r-ASCL as a predictor of ASCVD events using a larger sample size in

different cohorts. Our study also confirms CKD stage ≥3 at the time of LN diagnosis as an independent determinant of ASCVD occurrence. Despite these risks, we report that statin therapy was not used in any LN patients at baseline and was started in only 9% of eligible patients in our cohort.

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. Garg 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. Garg, Panzer, Bartels.

Acquisition of data. Garg, Hansen, Zhong, Huang, Smith, Panzer, Bartels.

Analysis and interpretation of data. Garg, Raval, Hansen, Zhong, Smith, Panzer, Bartels.

REFERENCES

1. Lim SS, Helmick CG, Bao G, Hootman J, Bayakly R, Gordon C, et al. Racial disparities in mortality associated with systemic lupus erythematosus: Fulton and DeKalb Counties, Georgia, 2002–2016. MMWR Morb Mortal Wkly Rep 2019;68:419–22.
2. Li H, Tong Q, Guo L, Yu S, Li Y, Cao Q, et al. Risk of coronary artery disease in patients with systemic lupus erythematosus: a systematic review and meta-analysis. Am J Med Sci 2018;356:451–63.
3. Hermansen ML, Lindhardsen J, Torp-Pedersen C, Faurschou M, Jacobsen S. The risk of cardiovascular morbidity and cardiovascular mortality in systemic lupus erythematosus and lupus nephritis: a Danish nationwide population-based cohort study. Rheumatology (Oxford) 2017;56:709–15.
4. Gustafsson JT, Herlitz Lindberg M, Gunnarsson I, Pettersson S, Elvin K, Ohrvik J, et al. Excess atherosclerosis in systemic lupus erythematosus: a matter of renal involvement: case control study of 281 SLE patients and 281 individually matched population controls. PLoS One 2017;12:e0174572.
5. Urowitz M, Gladman D, Bruce I. Atherosclerosis and systemic lupus erythematosus. Curr Rheumatol Rep 2000;2:19–23.
6. Urowitz MB, Bookman AA, Koehler BE, Gordon DA, Smythe HA, Ogryzlo MA. The bimodal mortality pattern of systemic lupus erythematosus. Am J Med 1976;60:221–5.
7. Manzi S, Meilahn EN, Rairie JE, Conte CG, Medsger TA Jr, Jansen-McWilliams L, et al. Age-specific incidence rates of myocardial infarction and angina in women with systemic lupus erythematosus: comparison with the Framingham Study. Am J Epidemiol 1997;145:408–15.
8. Bartels CM, Buhr KA, Goldberg JW, Bell CL, Visekruna M, Nekkanti S, et al. Mortality and cardiovascular burden of systemic lupus erythematosus in a US population-based cohort. J Rheumatol 2014;41:680–7.
9. Urowitz MB, Gladman DD, Anderson NM, Su J, Romero-Diaz J, Bae SC, et al. Cardiovascular events prior to or early after diagnosis of systemic lupus erythematosus in the systemic lupus international collaborating clinics cohort. Lupus Sci Med 2016;3:e000143.
10. Garg S, Bao G, Niyibizi M, Drenkard C, Lim SS. Differences between early and late cardiovascular disease in a population-based cohort of systemic lupus erythematosus patients [abstract]. Arthritis Rheumatol 2016; 68 Suppl 10. URL: https://acrabstracts.org/abstract/differences-between-early-and-late-cardiovascular-disease-in-a-population-based-cohort-of-systemic-lupus-erythematosus-patients/.
11. Tselios K, Sheane BJ, Gladman DD, Urowitz MB. Optimal monitoring for coronary heart disease risk in patients with systemic lupus erythematosus: a systematic review. J Rheumatol 2016;43:54–65.
12. Urowitz MB, Gladman D, Ibañez D, Fortin P, Sanchez-Guerrero J, Bae S, et al. Accumulation of coronary artery disease risk factors over three years: data from an international inception cohort. Arthritis Rheum 2008;59:176–80.
13. Myllymaki J, Syrjanen J, Helin H, Pasternack A, Kattainen A, Mustonen J. Vascular diseases and their risk factors in IgA nephropathy. Nephrol Dial Transplant 2006;21:1876–82.
14. Loupy A, Vernerey D, Viglietti D, Aubert O, Duong Van Huyen JP, Empana JP, et al. Determinants and outcomes of accelerated arteriosclerosis: major impact of circulating antibodies. Circ Res 2015;117:470–82.
15. Huang J, Han SS, Qin DD, Wu LH, Song Y, Yu F, et al. Renal interstitial arteriosclerotic lesions in lupus nephritis patients: a cohort study from China. PLoS One 2015;10:e0141547.
16. Barber C, Herzenberg A, Aghdassi E, Su J, Lou W, Qian G, et al. Evaluation of clinical outcomes and renal vascular pathology among patients with lupus. Clin J Am Soc Nephrol 2012;7:757–64.
17. Garg S, Bartels CM, Hansen KE, Zhong W, Huang Y, Semanik MG, et al. High burden of premature arteriosclerosis on renal biopsy results in incident lupus nephritis. Arthritis Care Res (Hoboken) 2021;73:394–401.
18. Hahn BH, McMahon MA, Wilkinson A, Wallace WD, Daikh DI, Fitzgerald JD, et al. American College of Rheumatology guidelines for screening, treatment, and management of lupus nephritis. Arthritis Care Res (Hoboken) 2012;64:797–808.
19. Hochberg MC, for the Diagnostic and Therapeutic Criteria Committee of the American College of Rheumatology. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus [letter]. Arthritis Rheum 1997;40:1725.
20. Petri M, Orbai AM, Alarcón GS, Gordon C, Merrill JT, Fortin PR, et al. Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum 2012;64:2677–86.
21. Weening JJ, D'Agati VD, Schwartz MM, Seshan SV, Alpers CE, Appel GB, et al. The classification of glomerulonephritis in systemic lupus erythematosus revisited. Kidney Int 2004;65:521–30.
22. Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol 2019;73:e285–350.
23. Alpert JS, Thygesen K, Antman E, Bassand JP. Myocardial infarction redefined: a consensus document of the Joint European Society of Cardiology/American College of Cardiology Committee for the redefinition of myocardial infarction. J Am Coll Cardiol 2000;36:959–69.
24. Hirsch AT, Haskal ZJ, Hertzer NR, Bakal CW, Creager MA, Halperin JL, et al. ACC/AHA 2005 practice guidelines for the management of patients with peripheral arterial disease (lower extremity, renal, mesenteric, and abdominal aortic): a collaborative report from the American Association for Vascular Surgery/Society for Vascular Surgery, Society for Cardiovascular Angiography and Interventions, Society for Vascular Medicine and Biology, Society of Interventional Radiology, and the ACC/AHA Task Force on Practice Guidelines (Writing Committee to Develop Guidelines for the Management of Patients With Peripheral Arterial Disease): endorsed by the American Association of Cardiovascular and Pulmonary Rehabilitation; National Heart, Lung, and Blood Institute; Society for Vascular Nursing; TransAtlantic Inter-Society Consensus; and Vascular Disease Foundation. Circulation 2006;113:e463–654.
25. Gibbons RJ, Chatterjee K, Daley J, Douglas JS, Fihn SD, Gardin JM, et al. ACC/AHA/ACP-ASIM guidelines for the management of patients with chronic stable angina: executive summary and recommendations. A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on

Management of Patients with Chronic Stable Angina). Circulation 1999;99:2829–48.

26. Sacco RL, Kasner SE, Broderick JP, Caplan LR, Connors JJ, Culebras A, et al. An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2013;44:2064–89.
27. Anderson JL, Adams CD, Antman EM, Bridges CR, Califf RM, Casey DE Jr, et al. ACC/AHA 2007 guidelines for the management of patients with unstable angina/non ST-elevation myocardial infarction: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the 2002 Guidelines for the Management of Patients With Unstable Angina/Non ST-Elevation Myocardial Infarction): developed in collaboration with the American College of Emergency Physicians, the Society for Cardiovascular Angiography and Interventions, and the Society of Thoracic Surgeons: endorsed by the American Association of Cardiovascular and Pulmonary Rehabilitation and the Society for Academic Emergency Medicine. Circulation 2007;116:e148–304.
28. Gerhard-Herman MD, Gornik HL, Barrett C, Barshes NR, Corriere MA, Drachman DE, et al. 2016 AHA/ACC guideline on the management of patients with lower extremity peripheral artery disease: executive summary. Vasc Med 2017;22:Np1-np43.
29. Obialo CI, Ofili EO, Norris KC. Statins and cardiovascular disease outcomes in chronic kidney disease: reaffirmation vs. repudiation. Int J Environ Res Public Health 2018;15.
30. Fanouriakis A, Kostopoulou M, Alunno A, Aringer M, Bajema I, Boletis JN, et al. 2019 update of the EULAR recommendations for the management of systemic lupus erythematosus. Ann Rheum Dis 2019; 78:736.
31. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. R version 3.4.1 ed; 2017.
32. Faurschou M, Mellemkjaer L, Starklint H, Kamper AL, Tarp U, Voss A, et al. High risk of ischemic heart disease in patients with lupus nephritis. J Rheumatol 2011;38:2400–5.
33. Wells DK, Ward MM. Nephritis and the risk of acute myocardial infarction in patients with systemic lupus erythematosus. Clin Exp Rheumatol 2010;28:223–9.
34. Doria A, Shoenfeld Y, Wu R, Gambari P, Puato M, Ghirardello A, et al. Risk factors for subclinical atherosclerosis in a prospective cohort of patients with systemic lupus erythematosus. Ann Rheum Dis 2003; 62:1071–7.
35. Selzer F, Sutton-Tyrrell K, Fitzgerald SG, Pratt JE, Tracy RP, Kuller LH, et al. Comparison of risk factors for vascular disease in the carotid artery and aorta in women with systemic lupus erythematosus. Arthritis Rheum 2004;50:151–9.
36. McMahon M, Hahn BH, Skaggs BJ. Systemic lupus erythematosus and cardiovascular disease: prediction and potential for therapeutic intervention. Expert Rev Clin Immunol 2011;7:227–41.
37. Yousef Yengej FA, Limper M, Leavis HL. Statins for prevention of cardiovascular disease in systemic lupus erythematosus. Neth J Med 2017;75:99–105.
38. Hansson GK, Robertson AK, Soderberg-Naucler C. Inflammation and atherosclerosis. Annu Rev Pathol 2006;1:297–329.
39. Garg S, Panzer S, Semanik M, Bartels CM. The burden of renal arteriosclerosis in lupus nephritis: a cohort study examining prevalence and predictors of renal arteriosclerosis [abstract]. Arthritis Rheumatol 2018; 70 Suppl 10. URL: https://acrabstracts.org/abstract/the-burden-of-renal-arteriosclerosis-in-lupus-nephritis-a-cohort-study-examining-prevalence-and-predictors-of-renal-arteriosclerosis/.
40. Ludewig B, Freigang S, Jaggi M, Kurrer MO, Pei YC, Vlk L, et al. Linking immune-mediated arterial inflammation and cholesterol-induced atherosclerosis in a transgenic mouse model. Proc Natl Acad Sci U S A 2000;97:12752–7.
41. Bajema IM, Wilhelmus S, Alpers CE, Bruijn JA, Colvin RB, Cook HT, et al. Revision of the International Society of Nephrology/Renal Pathology Society classification for lupus nephritis: clarification of definitions, and modified National Institutes of Health activity and chronicity indices. Kidney Int 2018;93:789–96.

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1113–1121
DOI 10.1002/acr.24542

American College of Rheumatology
Empowering Rheumatology Professionals

Challenges of Perceived Self-Management in Lupus

Paul R. Fortin,[1] Deborah Da Costa,[2] Carolyn Neville,[2] Anne-Sophie Julien,[1] Elham Rahme,[2] Vinita Haroun,[3] Wendy Singer,[3] Jodie Nimigon-Young,[3] Anna-Lisa Morrison,[3] Davy Eng,[1] Christine A. Peschken,[4] Evelyne Vinet,[2] Marie Hudson,[5] Doug Smith,[6] Mark Matsos,[7] Janet E. Pope,[8] Ann E. Clarke,[9] Stephanie Keeling,[10] J. Antonio Avina-Zubieta,[11] and Murray Rochon[12]

Objective. Systemic lupus erythematosus is a chronic autoimmune disease with varied and unpredictable levels of disease activity. The ability to self-manage lupus is important in controlling disease activity. Our objective was to determine levels of patient activation toward self-management in lupus.

Methods. We used baseline results from the MyLupusGuide study, which had recruited 541 lupus patients from 10 lupus centers. We used the Patient Activation Measure (PAM), a validated self-reported tool designed to measure activation toward self-management ability, as our primary variable and examined its association with demographic, disease-related, patient–provider communication and psychosocial variables captured in our study protocol. Univariable and multivariable linear regressions were performed using linear mixed models, with a random effect for centers.

Results. The mean ± SD age of participants was 50 ± 14 years, 93% were female, 74% were White, and the mean ± SD disease duration was 17 ± 12 years. The mean ± SD PAM score was 61.2 ± 13.5, with 36% of participants scoring in the 2 lower levels, indicating low activation. Variables associated with low activation included being single, having lower physical health status, lower self-reported disease activity, lower self-efficacy, use of more emotional coping and fewer distraction and instrumental coping strategies, and a perceived lack of clarity in patient–doctor communication.

Conclusion. Low patient activation was observed in more than one-third of lupus patients, indicating that a large proportion of patients perceived that they are lacking in lupus self-management skills. These results highlight a modifiable gap in perceived self-management ability among patients with lupus.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a chronic autoimmune, multiorgan, inflammatory disease that is characterized by numerous clinical manifestations and preferentially affects young women. Having SLE is associated with a 15% increase in mortality, high morbidity, and poor work outcomes. Despite an estimated prevalence of 1:1000 (1), SLE is mostly poorly understood by the general public, and information about the disease and access to specialized care remain limited. The chronic nature of the illness, unpredictability of the disease course, and complexity of treatment pose serious challenges to both patients and their treating physicians in disease management.

Activation refers to the ability and willingness to take on the role of managing one's own health and health care (2). This concept of activation may be used to evaluate preparedness and readiness to self-manage. Self-management is a crucial component of chronic disease management and is associated with

Supported by the Canadian Institute of Health Research (operating grant of the Knowledge to Action program) and in kind by Jack Digital Productions. Dr. Fortin is a tier 1 Canada Research Chair on Systemic Autoimmune Rheumatic Diseases. Dr. Avina-Zubieta is a BC Lupus Society Research Scholar and a Walter and Marilyn Booth Research Scholar.

[1]Paul R. Fortin, MD, MPH, FRCP(C), Anne-Sophie Julien, MSc, Davy Eng, PharmD, MSc: Université Laval, Quebec City, Quebec, Canada; [2]Deborah Da Costa, PhD, Carolyn Neville, BA, RN, Elham Rahme, PhD, Evelyne Vinet, MD, PhD, FRCPC: Centre for Outcomes Research & Evaluation, Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada; [3]Vinita Haroun, MSc, Wendy Singer, BA, Jodie Nimigon-Young MSW, Anna-Lisa Morrison BScN: MyLupusGuide Patient Advisory Committee, Université Laval, Quebec City, Quebec, Canada; [4]Christine A. Peschken, MD, MSc, FRCPC: University of Manitoba, Winnipeg, Manitoba, Canada; [5]Marie Hudson, MD, MPH, FRCPC: Jewish General Hospital, Lady Davis Institute for Medical Research and McGill University, Montreal, Quebec, Canada; [6]Doug Smith, MD, FRCPC: The Ottawa Hospital – University of Ottawa, Ottawa, Ontario, Canada; [7]Mark Matsos, MD, FRCPC: McMaster University, Hamilton, Ontario, Canada, [8]Janet E. Pope, MD, MPH, FRCPC: University of Western Ontario, London, Ontario, Canada; [9]Ann E. Clarke, MD, MSc, FRCPC: Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada; [10]Stephanie Keeling MD, MSc, FRCPC: University of Alberta, Edmonton, Alberta, Canada; [11]J. Antonio Avina-Zubieta, MD, PhD: University of British Columbia, Vancouver, British Columbia, Canada; [12]Murray Rochon, MArch: Jack Digital Productions, Montreal, Quebec and Toronto, Ontario, Canada.

No potential conflicts of interest relevant to this article were reported.

Address correspondence to Paul R. Fortin MD, MPH, FRCP(C): Centre de Recherche du CHU de Québec, 2705 Boulevard Laurier, Room TR-83, Quebec City, Quebec G1V 4G2, Canada. Email: paul.fortin@crchudequebec.ulaval.ca.

Submitted for publication February 21, 2020; accepted in revised form December 15, 2020.

SIGNIFICANCE & INNOVATIONS

- One-third of persons with systemic lupus erythematosus (SLE) report low activation and low confidence in self-managing their illness.
- Variables associated with low activation include being single, having lower physical health status, lower self-reported disease activity, lower self-efficacy, using more emotional coping and fewer distraction and instrumental coping strategies, and having a perceived lack of clarity in patient–doctor communication.
- There is a need for an intervention to provide support and solutions that will help persons with SLE develop confidence in self-managing their illness.

positive health outcomes (3); however, in order to manage one's illness, one requires the knowledge and skills necessary to promote the confidence needed to actively participate in decision-making about one's health care (4,5). Activation in patients with chronic conditions has been shown to be independently associated with several useful skills, including self-management behaviors (e.g., physical activity), use of self-management services, medication adherence, appropriate use of the health care system (e.g., having a regular source of care, not delaying care), consumeristic behaviors (e.g., preparing a list of questions for a doctor visit), improved chronic care self-management (e.g., keeping diary of blood pressure readings), control of chronic illness (e.g., better blood pressure reading, fewer hospitalizations), and health-related quality of life (6–9).

Previous studies examining factors associated with the degree to which patients with chronic conditions are activated for self-management have identified several associated factors under the following 4 categories: 1) patient sociodemographic characteristics, including age (10,11), sex (12), marital status (13), and education (11); 2) disease-related characteristics, including disease duration, disease severity, and health-related quality of life (11) (14); 3) patient–provider relational factors such as communication style (10); and 4) psychosocial factors including psychological distress (11), coping style (15), self-efficacy (16), and social support (11,14). Most studies have focused on sociodemographic and disease-related factors, and few studies have simultaneously examined a comprehensive set of factors, including modifiable psychosocial factors (see Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24542) that could be associated with level of patient activation for self-management (17). We had the opportunity to study patient activation in SLE using the data from a study entitled Measuring the Impact of MyLupusGuide in Canada (ClinicalTrials.gov identifier: NCT02950714). MyLupusGuide is a validated web-based program that was developed to facilitate engagement and self-management in patients with SLE (18). We report on the baseline findings of activation toward perceived health self-management in patients with SLE and the factors associated with lower activation.

PATIENTS AND METHODS

Patients for the MyLupusGuide study were recruited from 10 lupus clinics affiliated with the Canadian Network for Improved Outcomes in Systemic Lupus Erythematosus (CaNIOS). CaNIOS centers follow a unified protocol that uses a clinical diagnosis of SLE (19) to classify study participants as having SLE. Each CaNIOS center mailed invitations to all SLE patients who were at least 18 years of age and could read and write English or French. Patients who were willing to participate provided online consent and were asked to complete a series of online questionnaires, after which they were given access to the MyLupusGuide either immediately or 3 months later as part of a clinical trial. Analysis of Patient Activation Measure (PAM) data was preplanned as part of the MyLupusGuide study (ClinicalTrials.gov identifier: NCT02950714). We report here on the baseline data of the MyLupusGuide study.

Assessments: PAM. The widely used 13-item PAM evaluates the level of patient activation in patient health care (20,21). This measure is licensed and scored independently through Insignia Health (https://www.insigniahealth.com/products/pam-survey). This tool is designed to measure an individual's level of confidence, beliefs, knowledge, and skills about managing one's health. Respondents can answer with varying levels of agreement or disagreement (e.g., "I know how to prevent problems with my health"; "I am confident that I can tell a doctor my concerns, even when he or she does not ask") on a 4-point Likert scale. This instrument has been shown to have strong psychometric properties. The PAM has been used in observational and interventional studies as a patient-centered measure to monitor changes in patients' experiences over time, with higher scores related to greater activation and associated with greater self-management, healthy behaviors, medication adherence, better clinical outcomes, and higher levels of satisfaction with services (22, 23). The PAM can be used as a continuous score from 0 to 100 or be divided into 4 levels. Level 1 (score ≤47.0) indicates the lowest level of activation, e.g., the patient does not yet understand that an active role is important; level 2 (score 47.1–55.1) indicates, e.g., that the patient lacks knowledge and confidence to take action; level 3 (score 55.2–67.0) indicates that the patient is beginning to take action; and level 4 (score ≥67.1) indicates, e.g., that the patient is maintaining behaviors over time (24). Our primary outcome used the PAM score as a continuous variable. We classified levels 1 and 2 as low activation.

Sociodemographic questionnaires. Baseline assessment of sociodemographic characteristics were collected. These characteristics included age, sex, education, disease duration,

internet usage (including access and time spent searching online for health information).

Short Form 36 health survey version 1 (SF-36v1). This generic measure of health status has been recommended for use in SLE, as it is both valid and reliable (25,26) and it includes the important domain of vitality. In addition to scoring for each of the 8 domains, an algorithm also allows calculation of normalized scores for physical (the physical component summary [PCS]) and mental function (the mental component summary [MCS]) (27). Individual domain scores range 0–100, with higher scores indicating better function. Additional computations are required to transform the raw data into PCS and MCS scores normalized for a reference population at a score of 50. Any scores <48 on the PCS or MCS reflect a clinically relevant impairment in health status.

Systemic Lupus Activity Questionnaire. This 25-item self-reported questionnaire assesses disease activity in 9 organ systems and has demonstrated positive predictive values ranging from 56% to 89% for detecting clinically significant disease activity (28). It has been found to have adequate reliability and construct validity (29). Scores can range from 0 to 44, with higher scores indicating higher disease activity that correlate with the physician-completed Systemic Lupus Activity Measure (30).

Lupus Damage Questionnaire Index (LDIQ). This 56-item self-reported questionnaire assesses disease damage across 12 organ systems. It has been found to have construct validity with good correlations with the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index (SDI) for most organ systems. It may serve as a useful and reliable alternative to the SDI in assessing SLE-related damage in population studies (31). The LDIQ scores range from 0 to 43, with higher scores indicating greater damage.

Lupus Self-Efficacy Scale (LSES). The LSES, which was adapted for lupus from the Arthritis Self-Efficacy Scale (32), was used to measure patients' confidence related to lupus-specific domains. This 11-item version assesses level of confidence in managing or decreasing lupus-related symptoms (i.e., fatigue, mood). The construct and concurrent validity of this scale has been demonstrated (33). Self-efficacy has been found to be an important determinant in the adoption of self-management approaches among patients with arthritis (34). LSES scores range from 0 to 110, with higher scores indicating greater self-efficacy.

Center for Epidemiologic Studies Depression Scale (CES-D). The CES-D is a 20-item scale designed to measure symptoms of depression in the general population. The CES-D has been widely used in patients with chronic medical diseases and has been found to be more appropriate than other depression indices for patients with rheumatic disorders (35). CES-D scores range from 0 to 60, with higher scores indicating greater depression. A score of ≥16 has been used to identify individuals at risk for clinical depression, with good sensitivity and specificity and high internal consistency (36).

Morisky Levine scale (MLS-4). The MLS-4 is a valid and simple self-report 4-item scale that measures adherence to medication (37). It applies to all medications and diseases and does not measure any particular period. Scores range from 0 to 4, with low scores indicating better adherence.

Coping with Health Injuries and Problems (CHIP). This 32-item questionnaire assesses coping strategies typically used when dealing with health problems (38). Its 4 subscales include distraction (e.g., to dream of agreeable things rather than disease), palliative (e.g., to save his/her energy), instrumental (e.g., to look for efficient treatments), and emotional coping (e.g., to feel angry because of the disease). This instrument has been shown to have good psychometric properties and has been recommended for use with different medical populations (39). Scores for each of the 4 subscales range from 1 to 40. Higher values in each subscale are associated with greater use of the coping-related strategy.

Modified Medical Outcomes Study Social Support Survey (MMOS SSS). The MMOS SSS is a 7-item shortened version (40) of the original scale (41) that measures perceived support from one's social network related to emotional, tangible, and affectionate domains. Higher scores indicate higher perceived social support. Good internal consistency has been reported for the original version (41).

Interpersonal Processes of Care Survey Short Form (IPC-SF). The IPC-SF (42) assess 3 domains of patients' experiences with patient–doctor communication, including communication (lack of clarity, elicited concerns, explained results), patient-centered (decision-making), and interpersonal style (compassionate/respectful encounters with health care providers). For this study, we used the following 5 subscales: lack of clarity, elicited concerns, explained results, patient-centered decision-making, and compassionate interpersonal style. Scores for each scale range from 0 to 4. With the exception of the communication subscale "lack of clarity," higher scores indicate better experiences of the specific dimension, such as being provided with more explanations or more instances of being involved in treatment decisions. High scores in the communication subscale Lack of Clarity indicate worse experiences of receiving and interpreting disease-related information.

Analysis. Descriptive statistics are presented as mean ± SD or frequency with percentage. Univariable linear regressions were conducted using linear mixed models, with a random effect for centers. Results are presented as a regression coefficient (β) for continuous variables, where β represents a change in PAM score by a beta amount for a 1-unit increase in the variable, or as a mean PAM for categorical variables, with their SE and *P* value. Variables with a *P* value of <20% in the univariable models (that were not collinear) with a clinical relevance were included into a mixed-effect multivariable model with random center effect (like above). In order to understand counterintuitive results, where a coefficient sign was different between univariable and

Table 1. Descriptive and univariable analysis of demographic characteristics using PAM as principal variable*

Characteristics	No. (%) of patients	β or mean PAM†	SE	*P*
Age, mean ± SD years (n = 534)	50.07 ± 14.15	0.02	0.04	0.66
Sex (n = 528)				
Male	38 (7)	59.02	2.36	0.31
Female	490 (93)	61.32	1.11	
Marital status (n = 537)				
Single	106 (20)	59.04	1.54	0.07‡
Married	351 (65)	61.76	1.09	
Widowed	19 (4)	67.97	3.18	
Separated	22 (4)	62.30	2.97	
Divorced	39 (7)	59.73	2.29	
Education (n = 538)				
High school	79 (15)	60.27	1.71	0.03‡
Some post-secondary	74 (14)	63.80	1.78	
Some university	89 (17)	58.50	1.63	
University-complete	211 (39)	60.89	1.22	
Post-graduate studies	85 (16)	64.12	1.66	
Work disability (n = 535)				
Yes	109 (20)	57.29	1.54	<0.01‡
No	426 (80)	62.32	1.07	

* PAM = Patient Activation Measure.
† β is the regression coefficient of the model and represents a change in PAM score by a beta amount for a 1-unit increase in the variable studied.
‡ Significant.

multivariable models, several bivariable models were tested. In these bivariable model tests, the association of interest was adjusted for each of the other covariables separately.

RESULTS

Patients' characteristics. Of 1,916 patients who were mailed invitation letters, 541 (28%) provided consent and completed the first series of questionnaires. One patient was removed from the analyses due to unreliable response data, resulting in a total of 540 patients with complete data. Tables 1, 2, and 3 show the baseline demographic, disease, and psychosocial characteristics of 540 patients with SLE and their association with activation toward health self-management in univariable models.

Patient activation. The mean ± SD PAM score was 61.1 ± 13.5, with 16%, 20%, 42%, and 22% in levels 1, 2, 3, and 4, respectively. Thirty-six percent of patients scored in levels 1 or 2, indicating low activation toward perceived health self-management.

Demographic characteristics. The mean ± SD patient age was 50 ± 14 years, 93% were female, 39% had completed university-level education, and 16% had completed post-graduate

Table 2. Descriptive and univariable analysis of disease characteristics using PAM as principal variable*

Characteristics	No. of patients	Mean ± SD or no. (%)	β or mean PAM†	SE	*P*
Disease duration (years)	533	16.91 ± 11.93	0.12	0.05	0.01‡
SLAQ	540	14.01 ± 8.00	−0.31	0.07	<0.01‡
LDIQ	540	3.53 ± 3.06	−0.04	0.19	0.83
SF-36v1 PCS	535	39.05 ± 11.98	0.30	0.05	<0.01‡
SF-36v1 MCS	535	44.76 ± 11.65	0.23	0.05	<0.01‡
MLS-4 scores	540				<0.01‡
0		210 (39)	63.13	1.19	
1		202 (37)	61.50	1.22	
2		77 (14)	58.85	1.70	
3		30 (6)	59.93	2.54	
4		21(4)	52.02	3.01	

* LDIQ = Lupus Damage Index Questionnaire; MLS-4 = Morisky Levine Scale-4; PAM = Patient Activation Measure; SF-36v1 MCS = Short Form 36 health survey version 1 mental component summary score; SF-36v1 PCS = SF-36v1 physical component summary score; SLAQ = Systemic Lupus Activity Questionnaire.
† β is the regression coefficient of the model and represents a change in PAM score by a beta amount for a 1-unit increase in the variable studied.
‡ Significant.

Table 3. Descriptive and univariable analysis of patient–provider communication factors and psychosocial characteristics using PAM as principal variable (n = 540)*

Variable	Mean ± SD	β†	SE	*P*
CES-D (n = 538)	15.61 ± 10.68	−0.32	0.05	<0.01‡
CHIP (n = 539)				
Distraction	24.70 ± 6.16	0.56	0.09	<0.01‡
Palliative	24.06 ± 4.85	0.15	0.12	0.21
Instrumental	29.27 ± 5.26	0.78	0.11	<0.01‡
Emotional	19.84 ± 7.60	−0.45	0.07	<0.01‡
LSES (n = 539)	69.04 ± 23.60	0.26	0.02	<0.01‡
MMOS SSS (n = 539)	20.16 ± 6.68	0.46	0.08	<0.01‡
IPC-SF (n = 535)				
Lack of clarity	0.75 ± 0.77	−5.06	0.73	<0.01‡
Elicited concerns	3.24 ± 0.74	4.10	0.77	<0.01‡
Explained results	2.96 ± 0.99	3.19	0.58	<0.01‡
Decision-making§	2.58 ± 1.17	2.37	0.50	<0.01‡
Compassionate	3.20 ± 0.84	3.13	0.68	<0.01‡

* CES-D = Center for Epidemiologic Studies Depression Scale; CHIP = Coping with Health Injuries and Problems; LSES = Lupus Self-Efficacy Scale; MMOS SSS = Modified Medical Outcomes Study Social Support Survey; PAM = Patient Activation Measure.
† β is the regression coefficient of the model and represents a change in PAM score by a beta amount for a 1-unit increase in the variable studied.
‡ Significant.
§ For the decision-making component of the Interpersonal Processes of Care Survey Short Form (IPC-SF), n = 531.

studies, 65% were married, 54% were employed, and 20% were work-disabled. Seventy-four percent of participants were White, 11% were Asian, 5% were Black, and 10% reported being of other ethnicity. Participants reported a mean ± SD computer usage of 14.5 ± 13.7 hours per week, and 216 participants (41%) reported using the computer to search for health information frequently to very frequently.

Table 4. Demographic characteristics for the multivariable model using PAM as principal variable

Variable, levels	Mean PAM*	SE	*P*
Group†			
Now	62.62	1.82	0.93
Later	62.42	1.82	
Marital status			
Single	59.40	1.60	0.04‡
Married	60.52	1.37	
Widowed	68.39	2.96	
Separated	62.56	2.79	
Divorced	61.73	2.21	
Education			
High school	62.13	1.88	0.10
Some post-secondary	63.50	1.93	
University (incomplete)	60.70	1.83	
University (complete)	61.50	1.59	
Post-graduate studies	64.76	1.91	
Work disabled			
Yes	61.46	1.84	0.13
No	63.58	1.42	

* β is the regression coefficient of the model and represents a change in Patient Activation Measure (PAM) score by a beta amount for a 1-unit increase in the variable studied.
† The variable Group is included in the multivariable model, as the study consists of a randomized controlled trial (results of the intervention presented separately), and randomization was done by centers (1 group of centers invited to receive the intervention now and the other 3 months later). Although unlikely with randomization, belonging to 1 group or the other could have affected baseline characteristics, and we included the variable group in our multivariable analysis to adjust.
‡ Significant.

Disease characteristics. The mean ± SD disease duration was 17 ± 12 years and ranged between 0 and 63 years. The mean ± SD SF-36 PCS and MCS scores were 39 ± 12 and 45 ± 12, respectively, indicating poor physical and mental function. The mean ± SD self-reported lupus disease activity was 14.0 ± 8.0 (range 0–41), indicating low disease activity, and the mean ± SD lupus damage was moderate, at 3.5 ± 3.1.

Psychosocial characteristics. The mean ± SD CES-D score was 15.6 ± 10.7, with 232 participants (43.1%) scoring above the depressed mood cutoff score (≥16) indicating suspected depression. For CHIP, the use of instrumental coping was most frequent, and emotional preoccupation was the least frequently used coping mechanism.

Associations with lower PAM in patients with SLE. Results of univariable analyses are shown in Tables 1, 2, and 3. The univariable results demonstrate that lower PAM scores were associated with level of education, more disability, shorter disease duration, higher self-reported disease activity, lower physical and mental function (SF-36v1 PCS and MCS scores), and lower adherence to medication, reflected by a higher MLS-4 score (Tables 1 and 2). Among the psychosocial characteristics (Table 3), lower PAM scores were associated with more depression, more use of emotional coping (CHIP), lower self-efficacy, and more perceived lack of clarity in the patient–doctor

Table 5. Disease characteristics of the multivariable model using PAM as principal variable*

Variable	β or mean PAM†	SE	*P*
Disease duration (years)	0.06	0.05	0.21
SLAQ	0.28	0.10	<0.01‡
SF-36v1 PCS	0.16	0.07	0.02‡
SF-36v1 MCS	−0.03	0.06	0.68
MLS-4			
0	62.50	1.48	0.90
1	63.02	1.48	
2	62.38	1.80	
3	63.81	2.46	
4	60.88	2.97	

* MLS-4 = Morisky Levine Scale-4; SF-36v1 MCS = Short Form 36 health survey version 1 mental component summary score; SF-36v1 PCS = SF-36v1 physical component summary score; SLAQ = Systemic Lupus Activity Questionnaire.
† β is the regression coefficient of the model and represents a change in Patient Activation Measure (PAM) score by a beta amount for a 1-unit increase in the variable studied.
‡ Significant.

communication IPC subscale (Table 3). Higher PAM scores were associated with use of better coping strategies (distractive and instrumental) and better experiences with patient–physician communication (subscales elicited concerns, explained results, decision-making, and compassionate style) (Table 3).

Results of the overall multivariable model that included combined demographic, disease, and psychosocial characteristics in the same analytical model are shown in Tables 4, 5, and 6, respectively. Lower PAM scores were associated with single status, lower physical health, and lower self-efficacy. Of note, contrary to the result found in the univariable analysis, the multivariable model showed that lower PAM scores were associated with lower disease activity.

Among the CHIP coping strategies assessed, lower PAM scores were associated with higher use of emotional coping and less use of distractive and instrumental coping strategies. Among the interpersonal processes of care subscales, lower PAM scores were associated with more perceived "lack of clarity" on this IPC subscale.

DISCUSSION

Our primary outcome was to assess the perceived self-management ability in patients with SLE. The PAM is a valid and reliable tool to measure knowledge, skills, and confidence needed for self-management. Our findings showed that more than one-third of the lupus patients who participated in this study scored low for activation as measured by the PAM. Although a limitation of our study is that we did not measure self-management directly, these PAM scores suggest that a significant proportion of our participants reported a lack in knowledge, skills, or confidence in self-managing their disease. Interestingly, these results reflect what has been described in the general population of Northern Europe, with proportions of low activation ranging from 18% to 37% (43), which suggests that the level of activation in a population may not be related to being sick.

We were also interested in determining which factors were associated with lower activation in SLE. Lower self-assessed physical health, measured by the physical health component of the SF-36 health survey, was associated with a lower PAM score. This finding has been previously reported in other chronic disease populations (11,14) and may suggest that persons with poor physical health may have limited energy and may benefit from more tailored interventions with clear and specific instructions to take small steps to improve their engagement in self-care (44).

We found a positive association between self-reported lupus disease activity and the PAM scores in the multivariable model, while the association was negative in the univariable model. The result from the univariable model is in accordance with the notion that more patient activation is associated with better health outcomes (3). The reverse association in the multivariable model might be explained by the covariates we used in the model. For example, higher self-efficacy or physical and mental health status may give patients with more self-reported lupus activity an incentive to be more engaged in their disease management, resulting in a positive association with PAM.

Our results may suggest that when disease activity is low, people with SLE may be less motivated to continue to follow self-care strategies and increase their use when the disease is more active. However, the importance of maintaining self-care long-term to prevent future complications is essential. Providers caring for patients with lupus need to emphasize the importance of self-care as a long-term self-management strategy.

Table 6. Patient–provider communication factors and psychosocial characteristics sections of the multivariable model using PAM as principal variable*

Variable	β†	SE	*P*
CHIP			
Distraction	0.20	0.09	0.03‡
Palliative	−0.04	0.11	0.74
Instrumental	0.53	0.11	<0.01‡
Emotional	−0.17	0.08	0.04‡
LSES	0.15	0.03	<0.01‡
MMOS SSS	0.11	0.09	0.24
IPC-SF			
Lack of clarity	−3.38	0.71	<0.01‡
Elicited concerns	0.30	1.01	0.77
Explained results	1.08	0.65	0.10
Decision-making	0.70	0.54	0.20
Compassionate	−0.86	0.91	0.35

* CHIP = Coping with Health Injuries and Problems; IPC-SF Interpersonal Processes of Care survey Short Form; LSES = Lupus Self-Efficacy Scale; MMOS SSS = Modified Medical Outcomes Study Social Support Survey.
† β is the regression coefficient of the model and represents a change in Patient Activation Measurement (PAM) score by a beta amount for a 1-unit increase in the variable studied.
‡ Significant.

When viewed descriptively, the mean values on the CHIP subscales were similar to those previously reported by women in other medical populations (38). Among the health-related coping strategies, the use of more emotional-oriented coping strategies was associated with lower patient activation towards health self-management in this cross-sectional study. Focusing on the emotional aspects of the illness process has been associated with deficits in quality of life and increased psychological distress in chronic disease populations, including SLE (45,46). We extended these findings to SLE by demonstrating an association between higher levels of emotional preoccupation and less patient activation.

Distraction coping pertains to the attempts made to cope by focusing on more pleasant experiences or seeking the company of others. Participants who relied less on these strategies reported less activation toward self-management of their health. The use of distraction coping has been associated with avoidance and poorer health outcomes in the long-term, particularly when problems are manageable (47,48). Yet, items in this coping subscale include the use of social diversion to cope with health (such as inviting others to visit and enjoying the attention of friends and family) and more positive forms of distraction (such as thinking about the good times that one has experienced).

Consistent with prior studies that have examined more adaptive coping strategies used by persons with SLE (49,50), instrumental coping (which involves seeking knowledge about the illness and/or medical advice) and the use of methods to problem solve and set goals were associated with higher patient activation. Our findings suggest that interventions designed to increase patient activation should include these types of active coping strategies to empower patients with the skills needed to better self-manage their condition.

Patients who reported a lack of clarity in trying to understand their doctor due to the use of technical terms and to the speed with which the information was transmitted reported lower activation. In other chronic illness populations, poorer patient–provider communication has been associated with worse patient self-care behaviors and adherence to treatment regimens (51–53). Few studies have examined the relationship between patient–physician communications and activation for health self-management in SLE. Our findings suggest the importance of using nontechnical terms when conveying medical information to patients as well as efforts to ensure that patients have understood information. Digital technology platforms, such as the web-based interactive navigator MyLupusGuide that can be accessed 24/7 by the patient, may be an effective adjunctive modality to ensure that important information is conveyed and understood by the patient.

We acknowledge some limitations to our study. We used the baseline data from a clinical trial that will be published separately to describe patient activation in a sample of patients with lupus. As such, this is a cross-sectional study, and our analyses do not allow making inferences regarding cause–effect relationships on the associations that we are reporting. For example, although we observe an association between poor physical health, lack of clarity in physician–patient communication, or instrumental coping and low patient activation, we cannot conclude that low activation is a result of these other factors. Our study sample was recruited from the CaNIOS longitudinal observational study. In total, we recruited one-fourth (28%) of the CaNIOS cohort.

Sociodemographic and disease characteristics of our study sample, such as sex, race, marital status, education, work disability, self-reported disease activity, and damage and health status, reflected those of the CaNIOS cohort (see Supplementary Tables 1 and 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24542). We observed, however, that our sample was slightly older (mean ± SD age 50.07 ± 14.15 years for MyLupusGuide versus 47.97 ± 15.18 years for CaNIOS) with longer disease duration (mean ± SD disease duration 16.91 ± 11.93 years for MyLupusGuide versus 12.25 ± 12.18 years for CaNIOS) than that of the general CaNIOS cohort, which may have affected the attitude of our participants regarding self-management. In addition, unmeasured differences may have been present and undetected.

A selection bias of our participants is possible, and we must be cautious when generalizing our findings to that of all lupus patients. Our participants were recruited from lupus tertiary care centers and may not be entirely representative of the full spectrum of persons with SLE, as they may reflect a population that is a higher percentage White, better educated, and has greater access to more comprehensive specialized care and more information about lupus. Furthermore, our study was conducted online, requiring patients to have access to the internet via computer, tablet, or smartphone. Since we could not measure the ability to master the info-route, we could only assume that participants mastered the ability to use these devices and skills to complete the online tasks required for the study. Characteristics may differ between those who have computer devices and internet availability from those without; however, we were unable to collect such information because our study was conducted solely online and we were thus unable to survey patients without internet access.

Strengths of our study include a robust online methodology and a large sample. Moreover, our study included a comprehensive survey that captured several psychosocial characteristics that influence health care behaviors and patient activation.

Interventions to improve activation are becoming increasingly important as research shows that highly activated individuals tend to have a wide range of improved health-related outcomes (6). We observed low patient activation in more than one-third of a large sample of lupus patients participating in a study of MyLupusGuide. These findings suggest that this vulnerable population needs additional support resources to improve their ability in the self-management of SLE. We identified modifiable factors associated with low activation and recommend interventions that

focus on strategies to improve more adaptive coping and patient–doctor communication to help patients better self-manage their disease.

ACKNOWLEDGMENTS

We are thankful to our patient advisory committee, who provided time and effort toward the conduct of this study and the dissemination of MyLupusGuide.

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. Fortin 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. Fortin, Da Costa, Neville, Julien, Rahme, Haroun, Singer, Nimigon-Young, Morrison, Eng, Rochon.

Acquisition of data. Fortin, Da Costa, Neville, Eng, Peschken, Vinet, Hudson, Smith, Matsos, Pope, Clarke, Keeling, Avina-Zubieta, Rochon.

Analysis and interpretation of data. Fortin, Da Costa, Neville, Julien, Rahme, Rochon.

ROLE OF THE STUDY SPONSOR

Murray Rochon from Jack Digital Productions participated in study design and data collection, interpretation of the data, and approval of the content of the submitted manuscript. The publication of this article was not contingent upon approval by Jack Digital Productions.

ADDITIONAL DISCLOSURES

Author Murray Rochon is an employee of Jack Digital Productions.

REFERENCES

1. Bernatsky S, Joseph L, Pineau CA, Tamblyn R, Feldman DE, Clarke AE. A population-based assessment of systemic lupus erythematosus incidence and prevalence–results and implications of using administrative data for epidemiological studies. Rheumatology (Oxford) 2007;46:1814–8.
2. BMJ. Find resources to help you improve care. URL: http://personcentredcare.health.org.uk/sites/default/files/resources/patientactivation-1.pdf. 2019.
3. Bolen SD, Chandar A, Falck-Ytter C, Tyler C, Perzynski AT, Gertz AM, et al. Effectiveness and safety of patient activation interventions for adults with type 2 diabetes: systematic review, meta-analysis, and meta-regression. J Gen Intern Med 2014;29:1166–76.
4. Barry MJ, Edgman-Levitan S. Shared decision making–pinnacle of patient-centered care. N Engl J Med 2012;366:780–1.
5. Alegría M, Carson N, Flores M, Li X, Shi P, Lessios AS, et al. Activation, self-management, engagement, and retention in behavioral health care: a randomized clinical trial of the DECIDE intervention. JAMA Psychiatry 2014;71:557–65.
6. Greene J, Hibbard JH. Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. J Gen Intern Med 2012;27:520–6.
7. Hibbard JH, Cunningham PJ. How engaged are consumers in their health and health care, and why does it matter? Res Brief 2008:1–9.
8. Rask KJ, Ziemer DC, Kohler SA, Hawley JN, Arinde FJ, Barnes CS. Patient activation is associated with healthy behaviors and ease in managing diabetes in an indigent population. Diabetes Educ 2009;35:622–30.
9. Mosen DM, Schmittdiel J, Hibbard J, Sobel D, Remmers C, Bellows J. Is patient activation associated with outcomes of care for adults with chronic conditions? J Ambul Care Manage 2007;30:21–9.
10. Allen ML, Cook BL, Carson N, Interian A, La Roche M, Alegría M. Patient-provider therapeutic alliance contributes to patient activation in community mental health clinics. Adm Policy Ment Health 2017;44:431–40.
11. Bos-Touwen I, Schuurmans M, Monninkhof EM, Korpershoek Y, Spruit-Bentvelzen L, Ertugrul-van der Graaf I, et al. Patient and disease characteristics associated with activation for self-management in patients with diabetes, chronic obstructive pulmonary disease, chronic heart failure and chronic renal disease: a cross-sectional survey study. PLoS One 2015;10:e0126400.
12. Wetzstein MM, Shanta LL, Chlan LL. Patient activation among community-dwelling persons living with chronic obstructive pulmonary disease. Nurs Res. 2020;69:347–57.
13. O'Malley D, Dewan AA, Ohman-Strickland PA, Gundersen DA, Miller SM, Hudson SV. Determinants of patient activation in a community sample of breast and prostate cancer survivors. Psychooncology 2018;27:132–40.
14. Blakemore A, Hann M, Howells K, Panagioti M, Sidaway M, Reeves D, et al. Patient activation in older people with long-term conditions and multimorbidity: correlates and change in a cohort study in the United Kingdom. BMC Health Serv Res 2016;16:582.
15. Rapelli G, Donato S, Bertoni A, Spatola C, Pagani AF, Parise M, et al. The combined effect of psychological and relational aspects on cardiac patient activation. J Clin Psychol Med Settings 2020;27:783–94.
16. Goodworth MC, Stepleman L, Hibbard J, Johns L, Wright D, Hughes MD, et al. Variables associated with patient activation in persons with multiple sclerosis. J Health Psychol 2016;21:82–92.
17. Golubinski V, Oppel EM, Schreyögg J. A systematic scoping review of psychosocial and psychological factors associated with patient activation. Patient Educ Couns 2020;103:2061–8.
18. Neville C, Da Costa D, Rochon M, Peschken CA, Pineau CA, Bernatsky S, et al. Development of the lupus interactive navigator as an empowering web-based ehealth tool to facilitate lupus management: users perspectives on usability and acceptability. JMIR Res Protoc 2016;5:e44.
19. Peschken CA, Katz SJ, Silverman E, Pope JE, Fortin PR, Pineau C, et al. The 1000 Canadian faces of lupus: determinants of disease outcome in a large multiethnic cohort. J Rheumatol 2009;36:1200–8.
20. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004;39:1005–26.
21. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. Health Serv Res 2005;40:1918–30.
22. Adams RJ. Improving health outcomes with better patient understanding and education. Risk Manag Healthc Policy 2010;3:61–72.
23. Tzeng A, Tzeng TH, Vasdev S, Grindy A, Saleh JK, Saleh KJ. The role of patient activation in achieving better outcomes and cost-effectiveness in patient care. JBJS Rev 2015;3:e4.
24. Marshall R, Beach MC, Saha S, Mori T, Loveless MO, Hibbard JH, et al. Patient activation and improved outcomes in HIV-infected patients. J Gen Intern Med 2013;28:668–74.
25. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–83.
26. Stoll T, Gordon C, Seifert B, Richardson K, Malik J, Bacon PA, et al. Consistency and validity of patient administered assessment of quality

of life by the MOS SF-36; its association with disease activity and damage in patients with systemic lupus erythematosus. J Rheumatol 1997;24:1608–14.

27. McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 1993; 31:247–63.
28. Karlson EW, Daltroy LH, Rivest C, Ramsey-Goldman R, Wright EA, Partridge AJ, et al. Validation of a Systemic Lupus Activity Questionnaire (SLAQ) for population studies. Lupus 2003;12:280–6.
29. Yazdany J, Yelin EH, Panopalis P, Trupin L, Julian L, Katz PP. Validation of the systemic lupus erythematosus activity questionnaire in a large observational cohort. Arthritis Rheum 2008;59: 136–43.
30. Romero-Diaz J, Isenberg D, Ramsey-Goldman R. Measures of adult systemic lupus erythematosus: updated version of British Isles Lupus Assessment Group (BILAG 2004), European Consensus Lupus Activity Measurements (ECLAM), Systemic Lupus Activity Measure, Revised (SLAM-R), Systemic Lupus Activity Questionnaire for Population Studies (SLAQ), Systemic Lupus Erythematosus Disease Activity Index 2000 (SLEDAI-2K), and Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index (SDI). Arthritis Care Res (Hoboken) 2011;63 Suppl 11:S37–46.
31. Costenbader KH, Khamashta M, Ruiz-Garcia S, Perez-Rodriguez MT, Petri M, Elliott J, et al. Development and initial validation of a self-assessed lupus organ damage instrument. Arthritis Care Res (Hoboken) 2010;62:559–68.
32. Lorig K, Chastain RL, Ung E, Shoor S, Holman HR. Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis. Arthritis Rheum 1989;32:37–44.
33. Lorig K, Holman H. Arthritis self-efficacy scales measure self-efficacy. Arthritis Care Res 1998;11:155–7.
34. Keefe FJ, Lefebvre JC, Kerns RD, Rosenberg R, Beaupre P, Prochaska J, et al. Understanding the adoption of arthritis self-management: stages of change profiles among arthritis patients. Pain 2000;87:303–13.
35. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977;1: 385–401.
36. Lewinsohn PM, Seeley JR, Roberts RE, Allen NB. Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults. Psychol Aging 1997;12:277–87.
37. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. Med Care 1986; 24:67–74.
38. Endler NS, Parker JD, Summerfeldt LJ. Coping with health problems: developing a reliable and valid multidimensional measure. Psychol Assess 1998;10:195–205.
39. Endler NS, Parker JD. Coping with Health Injuries and Problems (CHIP). Toronto: Multi-Health Systems; 1998.
40. Czajkowski SM, Terrin M, Lindquist R, Hoogwerf B, Dupuis G, Shumaker SA, et al. Comparison of preoperative characteristics of men and women undergoing coronary artery bypass grafting (the Post Coronary Artery Bypass Graft [CABG] Biobehavioral Study). Am J Cardiol 1997;79:1017–24.
41. Sherbourne CD, Stewart AL. The MOS social support survey. Soc Sci Med 1991;32:705–14.
42. Stewart AL, Nápoles-Springer AM, Gregorich SE, Santoyo-Olsson J. Interpersonal processes of care survey: patient-reported measures for diverse groups. Health Serv Res 2007;42:1235–56.
43. Rademakers J, Maindal HT, Steinsbekk A, Gensichen J, Brenk-Franz K, Hendriks M. Patient activation in Europe: an international comparison of psychometric properties and patients' scores on the short form Patient Activation Measure (PAM-13). BMC Health Serv Res 2016; 16:570.
44. Magnezi R, Glasser S, Shalev H, Sheiber A, Reuveni H. Patient activation, depression and quality of life. Patient Educ Couns 2014;94: 432–7.
45. Kozora E, Ellison MC, Waxmonsky JA, Wamboldt FS, Patterson TL. Major life stress, coping styles, and social support in relation to psychological distress in patients with systemic lupus erythematosus. Lupus 2005;14:363–72.
46. Palominos PE, Gasparin AA, de Andrade NP, Xavier RM, da Silva Chakr RM, Igansi F, et al. Fears and beliefs of people living with rheumatoid arthritis: a systematic literature review. Adv Rheumatol 2018; 58:1.
47. Endler NS, Corace KM, Summerfeldt LJ, Johnson JM, Rothbart P. Coping with chronic pain. Pers Individ Dif 2003;34:323–46.
48. Penley JA, Tomaka J, Wiebe JS. The association of coping to physical and psychological health outcomes: a meta-analytic review. J Behav Med 2002;25:551–603.
49. Bricou O, Taïeb O, Baubet T, Gal B, Guillevin L, Moro MR. Stress and coping strategies in systemic lupus erythematosus: a review. Neuroimmunomodulation 2006;13:283–93.
50. Córdoba-Sánchez V, Limonero-García JT. Coping and quality of life in patients with systemic lupus erythematosus: a review. Pensando Psicología 2015;11:129–39.
51. Heisler M, Bouknight RR, Hayward RA, Smith DM, Kerr EA. The relative importance of physician communication, participatory decision making, and patient understanding in diabetes self-management. J Gen Intern Med 2002;17:243–52.
52. White RO, Eden S, Wallston KA, Kripalani S, Barto S, Shintani A, et al. Health communication, self-care, and treatment satisfaction among low-income diabetes patients in a public health setting. Patient Educ Couns 2015;98:144–9.
53. Zolnierek KB, Dimatteo MR. Physician communication and patient adherence to treatment: a meta-analysis. Med Care 2009;47:826–34.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1122–1132
DOI 10.1002/acr.24541

Predictors of Osteonecrosis in Systemic Lupus Erythematosus: A Prospective Cohort Study

Romy Kallas, Jessica Li, and Michelle Petri

Objective. We aimed at determining the predictors of osteonecrosis (ON) in a longitudinal lupus cohort.

Methods. Data were reviewed from the initiation of the cohort in 1987 until October 2019. In total, 2,428 patients were included in the analysis based on 224,295 person-months of follow-up. We used pooled logistic regression to assess the relationship between risk factors and rates of ON events. After identifying a set of variables related to ON incidence, we fit a final multivariable model to identify the most important risk factors for incident ON.

Results. In 18,691 person-years of follow-up after cohort entry, 122 incident ON events were observed (rate = 6.5/1,000 person-years). In the multivariable analysis, African American patients were at twice the risk for ON compared to White patients. Male patients and smokers had an increased risk for ON of ~80% and 50% compared to female patients and nonsmokers, respectively. For every 10-year increase in the age at diagnosis, there was a 20% reduced risk for ON. Patients diagnosed after the 1990s had a 50% reduced risk of ON compared to those diagnosed before the 1990s. A highest daily dosage of prednisone of 40 mg or higher, even when administered for a month or less, significantly increased the risk of ON. Use of pulse methylprednisolone or intramuscular triamcinolone was not associated with an increased risk of ON.

Conclusion. African American patients with systemic lupus erythematosus are at double the risk of experiencing ON compared to White patients. Oral prednisone at 20–39 mg for more than 1 month, or 40 mg daily for even 1 month, at any point in the disease course, remained the most important glucocorticoid predictor of ON.

INTRODUCTION

Osteonecrosis (ON) remains a serious complication in systemic lupus erythematosus (SLE), with prevalence ranging from 10–50% (1–4). The prevalence of asymptomatic ON has been estimated to be 40% (5). The femoral head is the most common site for ON (1,6–8) because of hemodynamics that make this area vulnerable to an ischemic state, with the potential to escalate to ON (9). Bilateral involvement is seen in 70–90% of individuals with ON (1,8). An early diagnosis of ON is challenging, as there is often a time lag between development of ON to symptomatic onset caused by collapse of the femoral head. Most patients will need joint arthroplasty (10), making it the costliest glucocorticoid-induced adverse event observed in SLE patients (11).

When Dubois and Cozen first described ON in patients with SLE, the role of glucocorticoids and other risk factors for ON was not clear (12). The increased incidence of ON in SLE patients compared to the general population (13), other autoimmune diseases, or other diseases requiring high doses of glucocorticoids (3,14) suggested the presence of SLE-specific risk factors (12,15). Over the years, multiple studies have evaluated predictors of ON. Of the known risk factors, glucocorticoid use remains the most important one. Whether the risk of ON is increased by the duration of glucocorticoid treatment, the initial dose during the first 3 months (15), the highest daily dose (16–18), the continuous high dose (19), or the mean or cumulative dose (5) still remains undetermined.

The possible association of ON with disease activity (1,18,20), lupus nephritis (1,21), neuropsychiatric manifestations (1,21), Raynaud's phenomenon (21,22), cutaneous vasculitis (20,21), Cushingoid features (16,21), and antiphospholipid (aPL) antibodies (4,6,7,15,23–25) has been reported in the literature. A recent study found that in African American patients, each additional risk allele of the *APOL1* gene increased the odds of prevalent ON, even after adjustment for prednisone dose (26).

The suspected mechanism for ON is vascular ischemia resulting in subchondral bone necrosis. Intrinsic risk factors related to SLE including vasculitis and vasculopathy have been proposed to be associated with the development of ON. Vasculitis was found on pathologic bone specimens from SLE patients with ON

The Hopkins Lupus Cohort was supported by NIH grant R01-AR069572.
Romy Kallas, MD, Jessica Li, MPH, Michelle Petri, MD, MPH: Johns Hopkins University School of Medicine, Baltimore, Maryland.
No potential conflicts of interest relevant to this article were reported.

Address correspondence to Romy Kallas, MD, 1830 East Monument Street, Suite 7500, Baltimore, MD 21205. Email: kallasr@mlhs.org.
Submitted for publication March 13, 2020; accepted in revised form December 15, 2020.

SIGNIFICANCE & INNOVATIONS

- Highest daily dosage of oral prednisone of >40 mg, even in individuals who receive this therapy for less than a month and irrespective of the time of administration during the disease course, is the most important predictor of osteonecrosis (ON). The data indicate that these doses should be avoided.
- Use of pulse methylprednisolone therapy for the treatment of major flares of disease activity does not increase the risk for ON.
- Use of intramuscular triamcinolone for the treatment of mild-to-moderate disease flares does not increase the risk for ON.
- African American patients are at an increased risk for ON.

who had not received therapy with any glucocorticoids, which is a scarce subpopulation (27). Patients with SLE may also have hemostatic abnormalities, such as the presence of aPL antibodies, resulting in vascular endothelial damage and intravascular coagulation leading to anoxia, hypoxia, and bone death.

Due to the longitudinal data available from The Hopkins Lupus Cohort, we were able to present the first prospective analysis of prednisone dose, coupled with duration, as risk factors for ON.

PATIENTS AND METHODS

The Hopkins Lupus Cohort. The Hopkins Lupus Cohort is a longitudinal cohort of patients diagnosed with SLE at Johns Hopkins Hospital. The cohort was established in 1987 and has been approved by the Johns Hopkins University School of Medicine Institutional Review Board on a yearly basis. All patients provided written informed consent. Data were collected prospectively during participation in The Hopkins Lupus Cohort. At cohort entry, a comprehensive medical history including year of diagnosis as well as clinical and laboratory information was obtained. During cohort participation, patients were followed up quarterly according to protocol, or more often as clinically indicated. At each quarterly clinic visit, clinical and laboratory assessments included the measurement of complement levels, anti–double-stranded DNA (anti-dsDNA), and lupus disease activity. In addition, other immunologic markers related to SLE, including anti-Sm, anti-RNP, anti-Ro, and anti-La, were collected at study enrollment. Multiple measures of aPL antibodies (lupus anticoagulant [LAC] by dilute Russell's viper venom time with confirmatory studies and anticardiolipin [aCL] antibody testing) were obtained at cohort entry and at follow-up visits. Information regarding each patient's glucocorticoid exposure before cohort entry was collected from review of complete patient medical records at cohort entry. From cohort entry onward, glucocorticoid doses were collected at each visit. The analysis for this study was based on cohort data collected through October 2019.

Subcohort for the present analysis. ON diagnosis was based on the Systemic Lupus International Collaborating Clinics (SLICC)/American College of Rheumatology (ACR) Damage Index, using imaging study reports including radiographs, computed tomography, or magnetic resonance imaging. Patients were followed up until the last cohort visit or until the time of their first ON event. Cohort members who had ON prior to cohort entry were excluded.

Definition of variables. Risk factors for ON used in the present analysis have been described in the literature (1,18,20). Variables were examined from cohort entry to either the first incident ON event or the last recorded visit in the database. Patient characteristics included age at SLE diagnosis, duration of SLE in years, year of diagnosis per 10-year difference, sex, socioeconomic status by education (less than 12 years and more than 12 years), and ethnicity (African American, White, Other). Social habits included smoking and alcohol abuse. Clinical characteristics included alopecia, oral ulcers, vasculitis, Raynaud's phenomenon, arthritis, serositis, neuropsychiatric, lupus nephritis, livedo reticularis, leukopenia, and thrombocytopenia. To evaluate the association between disease activity and the occurrence of ON, we noted the mean score and the most recent score on the Safety of Estrogens in Lupus Erythematosus National Assessment (SELENA) version of the Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) prior to ON occurrence. The antibody profile that was evaluated in the analysis included aCLs, anti–β_2-glycoprotein I, LAC, presence of low complement 3 (C3) or C4, anti-dsDNA, anti-Sm, anti-RNP, anti-Ro, and anti-La.

To evaluate the association between glucocorticoids and ON, we stratified the use of glucocorticoids as the number of months an individual received the highest dose of glucocorticoids (0–6 mg/day, 7–19 mg/day for 1 month, 7–19 mg/day for >1 month, 20–39 mg/day for 1 month, 20–39 mg/day for >1 month, 40–59 mg/day for 1 month, 40–59 mg/day for >1 month, ≥60 mg/day for 1 month, ≥60 mg/day for >1 month). We examined the number of months an individual received prednisone until first ON event, cumulative dose of prednisone, mean daily prednisone dose, number of triamcinolone intramuscular injections (used to treat mild/moderate SLE flares), and use of intravenous pulse methylprednisolone. Immunosuppressive treatments included cyclophosphamide, mycophenolate mofetil, tacrolimus, methotrexate, and azathioprine.

Statistical analysis. To facilitate the risk factor analysis, we constructed a data set with 1 record for each month of follow-up for each person. That record retained information on the patient's age at that month, their clinical and medication history up to that point, and their current disease activity. For each person-month, we created a variable indicating whether they were diagnosed with ON during that month.

To calculate the rate of ON in each demographic and clinical subgroup, we calculated the number of ON events divided by the

number of person-months at risk and converted this to rates per person-year. To assess the relationship between risk factors and rates of ON events, we used pooled logistic regression. Pooled logistic regression has been shown to be approximately equivalent to Cox regression with time dependent covariates and has practical advantages (28). Rate ratios (RRs) were adjusted for age for each variable. After identifying a set of variables related to ON incidence, we fit a final multivariable model to identify the most important risk factors for incident ON. Multicollinearity was eliminated by using collinearity diagnostics and excluding variables that were highly intercorrelated with other variables.

RESULTS

Person-month files were created from 2,602 cohort patients. The cumulative classification criteria were 48% malar rash, 19% discoid rash, 51% photosensitivity, 53% oral ulcer, 72% arthritis, 48% serositis, 45% renal disorder, 12% neurologic disorder, 67% hematologic disorder, 82% immunologic disorder, and 97% antinuclear antibody positivity, based on the revised ACR classification criteria for SLE (29). Additional SLICC classification criteria included 21% direct Coombs' test, 55% low C3, 48% low C4, and 16% low CH50 (30).

Among these 2,602 patients, 287 patients in our cohort had ON. Of these patients, 154 had ON in only 1 joint and 133 had ON in 2 or more joints. Incidences of ON that occurred before cohort entry were excluded. A total of 2,428 patients were eligible to be included in the analysis, based on a total of 224,295 person-months of follow-up (~18,691 person-years of follow-up). After cohort entry, 122 incident events of ON occurred and were included in the final analysis, with an incidence rate of 6.5 per 1,000 person-years. Only 3 patients developed ON without any glucocorticoid use. Demographic characteristics of the study cohort were as follows: 2,243 (92.4%) female, 1,298 (53.5%) White, 936 (38.6%) African American, and 194 (8.0%) other ethnicities (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24541). In this cohort, 49% were diagnosed as

Table 1. Rates of ON events by demographic and patient characteristics*

Subgroup	ON events	Person-years of follow-up	Rate of events per 1,000 person years	RR (95% CI)	*P*
All	122	18,691.3	6.52		
Sex					
Female	107	17,297.42	6.19	1.00 (Ref.)	–
Male	15	1,393.83	10.76	1.80 (1.05, 3.08)	0.0340
Ethnicity					
White	44	10,198.33	4.31	1.00 (Ref.)	–
African American	74	7,327.08	10.1	2.27 (1.56, 3.30)	<0.0001
Other	4	1,165.83	3.43	0.69 (0.25, 1.91)	0.4704
Year of SLE diagnosis					
<1980	15	1,109.83	13.52	1.00 (Ref.)	–
1980–1989	38	3,591.17	10.58	0.65 (0.36, 1.2)	0.1675
1990–1999	39	7,476.08	5.22	0.32 (0.17, 0.58)	0.0002
2000–2009	25	5,582.5	4.48	0.25 (0.13, 0.49)	<0.0001
2010–2019	5	925.08	5.4	0.27 (0.1, 0.77)	0.0136
Age at SLE diagnosis, years					
<30	72	9,013.0	7.99	1.00 (Ref.)	–
30 to <40	27	4,842.6	5.58	0.70 (0.45, 1.09)	0.1109
≥40	23	4,829.1	4.76	0.60 (0.37, 0.95)	0.0308
Duration of SLE, years					
<3	22	2,533.92	8.68	1.00 (Ref.)	–
3 to <5	21	2,022.75	10.38	1.24 (0.68, 2.26)	0.4786
5 to <10	27	4,713.67	5.73	0.72 (0.41, 1.27)	0.2548
10 to <15	25	3,710.92	6.74	0.91 (0.51, 1.63)	0.7413
≥15	27	5,622.17	4.80	0.75 (0.41, 1.37)	0.3497
Education, years					
≤12	55	6,697.25	8.21	1.00 (Ref.)	–
>12	65	11,828.75	5.50	0.66 (0.46, 0.94)	0.0215
Smoking status					
Never	65	11,759.5	5.53	1.00 (Ref.)	–
Ever	57	6,895	8.27	1.70 (1.18, 2.43)	0.0044
Alcohol abuse					
Never	109	17,323.17	6.29	1.00 (Ref.)	–
Ever	13	1,323.58	9.82	1.65 (0.93, 2.93)	0.0894

* Rate ratios (RR) and *P* values were adjusted for the age of the patient at each month of follow-up. 95% CI = 95% confidence interval; ON = osteonecrosis; Ref. = reference; SLE = systemic lupus erythematosus.

Table 2. Rates of ON events by clinical manifestations*

Subgroup	ON events	Person-years of follow-up	Rates of events per 1,000 person-years	RR (95% CI)	*P*
Alopecia					
No	47	8,426.33	5.58	1.00 (Ref.)	–
Yes	75	10,251.58	7.32	1.39 (0.96, 2)	0.0798
Oral ulcers					
No	48	8,514.08	5.64	1.00 (Ref.)	–
Yes	74	10,173.58	7.27	1.33 (0.93, 1.92)	0.1230
Vasculitis					
No	93	15,628.67	5.95	1.00 (Ref.)	–
Yes	29	3,055.83	9.49	1.62 (1.07, 2.46)	0.0236
Raynaud's phenomenon					
No	51	9,160.08	5.57	1.00 (Ref.)	–
Yes	71	9,525.75	7.45	1.37 (0.95, 1.96)	0.0902
Arthritis					
No	24	5,042.58	4.76	1.00 (Ref.)	–
Yes	98	13,623.33	7.19	1.67 (1.07, 2.62)	0.0249
Pleurisy					
No	54	10,730.5	5.03	1.00 (Ref.)	–
Yes	68	7,949	8.55	1.71 (1.19, 2.44)	0.0034
Pericarditis					
No	82	14,877.33	5.51	1.00 (Ref.)	–
Yes	40	3,781.58	10.58	1.92 (1.32, 2.8)	0.0007
Livedo					
No	94	13,688.92	6.87	1.00 (Ref.)	–
Yes	28	4,986.92	5.61	0.81 (0.53, 1.23)	0.3245
Leukopenia					
No	55	9,806.83	5.61	1.00 (Ref.)	–
Yes	67	8,876.5	7.55	1.31 (0.92, 1.87)	0.1415
Thrombocytopenia					
No	81	14,905.83	5.43	1.00 (Ref.)	–
Yes	40	3,773.17	10.6	1.85 (1.27, 2.71)	0.0015
Proteinuria					
No	41	10,815.33	3.79	1.00 (Ref.)	–
Yes	81	7,862.83	10.3	2.54 (1.74, 3.71)	<0.0001
Nephrotic syndrome					
No	84	15,744.08	5.34	1.00 (Ref.)	–
Yes	37	2,824.75	13.1	2.28 (1.54, 3.36)	<0.0001
Renal insufficiency					
No	88	14,532.58	6.06	1.00 (Ref.)	–
Yes	34	4,139.58	8.21	1.52 (1.02, 2.27)	0.0392
Renal failure					
No	112	17,803	6.29	1.00 (Ref.)	–
Yes	10	852.5	11.73	1.8 (0.94, 3.44)	0.0747
Seizure					
No	104	17,350.42	5.99	1.00 (Ref.)	–
Yes	18	1,339.83	13.43	2.18 (1.32, 3.59)	0.0023
Psychosis					
No	116	18,189.75	6.38	1.00 (Ref.)	–
Yes	6	500.75	11.98	1.9 (0.83, 4.31)	0.1268
OBS					
No	111	17,868	6.21	1.00 (Ref.)	–
Yes	11	796.33	13.81	2.49 (1.33, 4.64)	0.0042
Meningitis					
No	121	18,320.33	6.6	1.00 (Ref.)	–
Yes	1	364.67	2.74	0.42 (0.06, 3.01)	0.3883
Stroke					
No	114	17,856.33	6.38	1.00 (Ref.)	–
Yes	8	834.17	9.59	1.57 (0.77, 3.21)	0.219
Lupus headache					
No	103	17,053.67	6.04	1.00 (Ref.)	–
Yes	19	1,635.33	11.62	1.9 (1.17, 3.1)	0.0101

(Continued)

Table 2. (*Cont'd*)

Subgroup	ON events	Person-years of follow-up	Rates of events per 1,000 person-years	RR (95% CI)	*P*
Cognitive impairment					
No	114	17,199.08	6.63	1.00 (Ref.)	–
Yes	8	1,366.17	5.86	1.03 (0.5, 2.12)	0.9414
Optic neuritis					
No	121	18,423	6.57	1.00 (Ref.)	–
Yes	1	137.25	7.29	1.26 (0.18, 9.04)	0.8179
Peripheral neuropathy					
No	118	17,962	6.57	1.00 (Ref.)	–
Yes	4	598	6.69	1.27 (0.46, 3.46)	0.6432
Venous thrombosis					
No	96	15,793.7	6.08	1.00 (Ref.)	–
Yes	26	2,880.6	9.03	1.57 (1.01, 2.42)	0.0432
Arterial thrombosis					
No	97	16,223.8	5.98	1.00 (Ref.)	–
Yes	24	2,290.8	10.48	2.03 (1.29, 3.19)	0.0023
aCL antibody					
No	59	9,351.33	6.31	1.00 (Ref.)	–
Yes	60	9,111.08	6.59	1.09 (0.76, 1.56)	0.6549
Anti–β_2GPI					
No	76	12,151.75	6.25	1.00 (Ref.)	–
Yes	12	3,241.83	3.7	0.62 (0.34, 1.14)	0.1258
Low C3					
No	38	8,479	4.48	1.00 (Ref.)	–
Yes	84	10,209.5	8.23	1.7 (1.16, 2.51)	0.0068
Low C4					
No	50	9,935.58	5.03	1.00 (Ref.)	–
Yes	72	8,753.08	8.23	1.49 (1.03, 2.14)	0.034
Anti-dsDNA					
No	37	6,906.08	5.36	1.00 (Ref.)	–
Yes	85	11,776.5	7.22	1.27 (0.86, 1.87)	0.2236
Anti-RNP					
No	74	13,100.17	5.65	1.00 (Ref.)	–
Yes	48	5,436.25	8.83	1.43 (0.99, 2.06)	0.0577
Anti-Sm					
No	80	13,721.58	5.83	1.00 (Ref.)	–
Yes	41	4,813.17	8.52	1.34 (0.91, 1.95)	0.1359
Anti-Ro					
No	88	13,300.17	6.62	1.00 (Ref.)	–
Yes	34	5,261.17	6.46	0.96 (0.65, 1.43)	0.8554
Anti-La					
No	106	16,381.17	6.47	1.00 (Ref.)	–
Yes	16	2,177.92	7.35	1.14 (0.68, 1.93)	0.6207
Venous thrombosis					
No	96	15,793.7	6.08	1.00 (Ref.)	–
Yes	26	2,880.6	9.03	1.57 (1.01, 2.42)	0.0432
Arterial thrombosis					
No	97	16,223.8	5.98	1.00 (Ref.)	–
Yes	24	2,290.8	10.48	2.03 (1.29, 3.19)	0.0023
Mean SELENA–SLEDAI score					
0 to <1	18	5,200.9	3.46	(Ref.)	–
1 to <2.5	31	5,846.4	5.30	1.51 (0.85, 2.7)	0.1631
2.5 to <5	47	5,397.3	8.71	2.39 (1.39, 4.12)	0.0017
≥5	26	2,227.1	11.67	2.86 (1.55, 5.28)	0.0008
Immunosuppressant use					
Never	39	9,581.9	4.07	1.00 (Ref.)	–
Ever	83	9,096.7	9.12	2.16 (1.48, 3.16)	<0.0001

* Rate ratios (RRs) and *P* values were adjusted for the age of the patient at each month during follow-up. 95% CI = 95% confidence interval; aCL = anticardiolipin antibody; anti-β_2GPI = anti–β_2-glycoprotein I; anti-dsDNA = anti–double-stranded DNA; C3 = complement 3; OBS = organic brain syndrome; ON = osteonecrosis; Ref. = reference; SELENA-SLEDAI = Safety of Estrogens in Lupus Erythematosus National Assessment version of the Systemic Lupus Erythematosus Disease Activity Index; SLE = systemic lupus erythematosus.

Table 3. Rates of ON events by glucocorticoid use*

Subgroup	ON events	Person-years of follow-up	Rate of events per 1,000 patient-years	RR (95% CI)	*P*
Highest prednisone dose, mg/day					
0–6	5	4,427.7	1.13	1.00 (Ref.)	–
7–19	5	2,485.2	2.01	1.00 (Ref.)	–
20–39	16	3,616.9	4.42	2.95 (1.34, 6.51)	0.0073
40–59	25	3,209.2	7.79	5.19 (2.49, 10.81)	<0.0001
≥60	71	4,944.8	14.36	9.28 (4.78, 18.02)	<0.0001
Number of months receiving prednisone at highest dose†					
0–6 mg/day	5	4,427.7	1.13	1.00 (Ref.)	–
7–19 mg/day for 1 month	2	768.3	2.60	1.00 (Ref.)	–
7–19 mg/day for >1 month	3	1,716.9	1.75	1.00 (Ref.)	–
20–39 mg/day for 1 month	5	1,519.8	3.29	2.21 (0.76, 6.47)	0.1478
20–39 mg/day for >1 month	11	2,097.2	5.25	3.48 (1.48, 8.2)	0.0044
40–59 mg/day for 1 month	11	1,772.2	6.21	4.14 (1.76, 9.76)	0.0012
40–59 mg/day for >1 month	14	1,437.0	9.74	6.47 (2.87, 14.58)	<0.0001
≥60 mg/day for 1 month	41	2,638.9	15.54	10.12 (5.06, 20.23)	<0.0001
≥60 mg/day for >1 month	30	2,305.9	13.01	8.32 (4.06, 17.07)	<0.0001
Cumulative dose of prednisone, mg†					
0	3	3,389.8	0.89	1.00 (Ref.)	–
<3,650	9	2,883.1	3.12	1.00 (Ref.)	–
3,650–10,949	17	3,037.8	5.6	2.69 (1.29, 5.65)	0.0087
10,950–36,499	54	5,854	9.22	4.65 (2.49, 8.70)	<0.0001
≥36,500	39	3,519.1	11.08	5.79 (3.03, 11.06)	<0.0001
Mean daily prednisone dose, mg†					
0	3	3,389.8	0.89	1.00 (Ref.)	–
1–5	6	2,530.5	2.37	1.00 (Ref.)	–
6–9	19	4,194.6	4.53	2.97 (1.34, 6.56)	0.0072
≥10	94	8,568.8	10.97	6.68 (3.36, 13.28)	<0.0001
Number of prior months receiving prednisone†					
0–11	16	6,478.2	2.47	1.00 (Ref.)	–
12–23	14	1,834.6	7.63	2.82 (1.37, 5.78)	0.0048
≥24	92	10,378.5	8.86	3.55 (2.09, 6.04)	<0.0001
Recent prednisone dose, mg/day					
0	34	10,914.3	3.12	1.00 (Ref.)	–
1–9	35	4,345.4	8.05	2.57 (1.6, 4.11)	<0.0001
10–19	28	2,314.5	12.1	3.64 (2.2, 6.02)	<0.0001
≥20	25	1,098.7	22.75	6.49 (3.82, 11.03)	<0.0001
Number of triamcinolone injections during follow-up visits					
0	79	10,265.2	7.70	1.00 (Ref.)	–
1–5	32	6,217.5	5.15	0.68 (0.45, 1.04)	0.0742
≥6	11	2,208.6	4.98	0.65 (0.34, 1.24)	0.1887
Recent history of triamcinolone injection					
No	116	16,870.4	6.88	1.00 (Ref.)	–
Yes	6	1,820.8	3.30	0.52 (0.23, 1.17)	0.1146

* Rate ratios (RRs) and *P* values were adjusted for the age of the patient at each month during follow-up. 95% CI = 95% confidence interval; ON = osteonecrosis; Ref. = reference.
† Includes information on glucocorticoid use before cohort participation.

having SLE before they reached 30 years of age, 25% were diagnosed between the ages of 30 years and 40 years, and 26% were diagnosed at age 40 years and older. Further, 4.9% were diagnosed as having SLE prior to 1980, 15.8% in the 1980s, 33.7% in the 1990s, 33.2% in the 2000s, and 12.4% in the 2010s. Thirty-nine percent of patients joined the cohort within 1 year of SLE diagnosis, 27.2% joined 1–5 years after diagnosis, and 33.8% joined ≥5 years after diagnosis. Male sex, African American ethnicity, and smoking history were associated with an increased risk of ON (Table 1). Patients diagnosed after the 1990s and at an older age (more than 40 years old) had lower ON rates.

In Table 2, clinical manifestations necessitating higher doses of oral prednisone, such as thrombocytopenia and nephritis, were associated with a higher risk of ON. Other manifestations such as cutaneous vasculitis, arthritis, pleuritis, and pericarditis (flares of which were usually treated with intramuscular injections of triamcinolone in the cohort) were also associated with an increased risk of ON. There was no association between individual aPL

Table 4. Associations between ON rates and predictors using a multivariable model*

	RR (95% CI)	*P*
Maximum daily prednisone dose with duration of use, mg		
0–6	1.00 (Ref.)	–
7–19 for 1 month	1.00 (Ref.)	–
7–19 for >1 month	1.00 (Ref.)	–
20–39 for 1 month	2.03 (0.69, 5.94)	0.1973
20–39 for >1 month	2.74 (1.15, 6.49)	0.0222
40–59 for 1 month	3.54 (1.49, 8.37)	0.0041
40–59 for >1 month	4.98 (2.19, 11.32)	0.0001
≥60 mg/day for 1 month	8.47 (4.22, 16.99)	<0.0001
Sex		
Female	1.00 (Ref.)	–
Male	1.79 (1.03, 3.1)	0.0379
Ethnicity		
White	1.00 (Ref.)	–
African American	1.91 (1.31, 2.80)	0.0008
Other	0.74 (0.26, 2.08)	0.5655
Smoking history		
Never	1.00 (Ref.)	–
Ever	1.49 (1.03, 2.17)	0.0338
Year of SLE diagnosis		
Before 1990	1.00 (Ref.)	–
1990 and later	0.54 (0.37, 0.79)	0.0013
Age per 10-year increase	0.77 (0.67, 0.90)	0.0009

* 95% confidence interval; ON = osteonecrosis; Ref. = reference; RR = rate ratio; SLE = systemic lupus erythematosus.

antibodies and the development of ON. Univariate analysis showed that low complement levels, increased mean SELENA–SLEDAI score, and history of immunosuppressant use were associated with increased rates of ON. History of arterial and venous thrombosis was also associated with increased ON rates.

Glucocorticoid use, analyzed as maximal oral daily dose over a specified duration, oral cumulative dose, and mean daily dose of oral prednisone were associated with a higher risk of ON. Triamcinolone intramuscular injection, regardless of the number of injections, was not associated with ON (Table 3). In Table 4, multivariable analysis showed that African American patients were at twice the risk of experiencing ON compared to White patients. Male sex conferred an 80% increased risk of ON compared to female patients. Smokers had a 50% increased risk of ON compared to nonsmokers. With every 10-year increase in age, there was a 20% reduction in the risk of ON. Patients diagnosed after the 1990s had a 70% reduced risk of ON compared to patients diagnosed before the 1980s.

A daily prednisone dosage of 20–39 mg conferred an increased risk of ON when administered for longer than 1 month. A maximal daily dosage of prednisone of ≥40 mg, even when administered for 1 month, increased the risk of ON. As the dose and the duration of prednisone increased, the risk of ON increased. Compared to patients who received oral prednisone at a dosage of <20 mg/day, receiving prednisone at a dosage of 20–39 mg/day for >1 month was associated with a RR of 3.48 ($P = 0.0044$). A dosage of 40–59 mg/day, even if medication was received for just 1 month, was associated with an RR of 4.14 ($P = 0.0012$) compared to those who received prednisone at a daily dosage of <20 mg. A daily dosage of >60 mg, regardless of the duration, was associated with an even greater increase in RRs (RRs of 10.12 and 8.32 when administered for 1 month and >1 month, respectively [$P < 0.0001$]) compared to those who received a daily prednisone dosage of <20 mg. In Table 5 and 6, receiving pulse methylprednisolone did not increase the risk of developing ON after adjustment for maximal dose of oral prednisone received or after stratification by maximal dose of oral prednisone.

DISCUSSION

The present study evaluated highest daily prednisone dose based on duration, mean prednisone dose, cumulative prednisone dose, duration of prednisone therapy, number of intramuscular triamcinolone injections, and pulse methylprednisolone as possible predictors of ON occurrence in SLE. The present study highlighted that the risk of ON occurrence with oral prednisone use is both dose- and duration-dependent. A daily prednisone dosage of 20–39 mg increased the risk of ON when administered for >1 month. A daily dosage of ≥40 mg, even when administered for 1 month, predicted an even higher increased risk of ON. A mean daily dosage of <20 mg was safer to use in terms of risk of ON. Pulse methylprednisolone (after adjustment for oral prednisone use) and intramuscular triamcinolone injection did not increase ON risk.

Additional important findings unrelated to glucocorticoid-related risk factors should also be noted. First, we showed that the prevalence of ON has decreased over the past 2 decades. In 1995, we had reported the prevalence of ON in a study cohort as 14.5% (2). In the present analysis, the prevalence was 11%. Moreover, the rate of ON per decade has decreased gradually

Table 5. Association between ON rates and pulse methylprednisolone use*

	Number of ON events	Person-years	Rates per 1,000 person-years	Adjustment for age		Adjustment for age and maximum dose of prednisone	
				RR (95% CI)	*P*	RR (95% CI)	*P*
No	111	17,286.7	6.4	1.00 (Ref.)	–	1.00 (Ref.)	–
Yes	11	1,404.6	7.8	1.26 (0.68, 2.35)	0.4618	1.03 (0.55, 1.92)	0.9238

* 95% CI = 95% confidence interval; ON = osteonecrosis; Ref. = reference; RR = relative risk.

Table 6. Association between ON rates and methylprednisolone pulse therapy stratified by maximum prednisone use*

Subgroup	Number of ON events	Person-years	Rate per 1,000 person-years	RR (95% CI)	*P*
Pulse among those with maximum prednisone dose of 0–39 mg					
No	23	10,084.3	2.28	1.00 (Ref.)	–
Yes	3	445.4	6.74	3.02 (0.91, 10.09)	0.0719
Maximum prednisone dose of 0–19 mg					
No	9	6,691.5	1.34	1.00 (Ref.)	–
Yes	1	221.3	4.52	3.44 (0.43, 27.36)	0.2436
Maximum prednisone dose of 20–39 mg					
No	14	3,392.8	4.12	1.00 (Ref.)	
Yes	2	224.1	8.93	2.18 (0.50, 9.61)	0.3025
Pulse among those with maximum prednisone dose of ≥40 mg					
No	88	7,194.8	12.2	1.00 (Ref.)	–
Yes	8	959.2	8.3	0.71 (0.35, 1.48)	0.3618
Maximum prednisone dose of 40–59 mg					
No	24	2,872.4	8.4	1.00 (Ref.)	–
Yes	1	336.8	3.0	0.36 (0.05, 2.69)	0.3213
Maximum prednisone dose of ≥60 mg					
No	64	4,322.4	14.8	1.00 (Ref.)	–
Yes	7	622.4	11.2	0.79 (0.36, 1.72)	0.5477

* Rate ratios (RR) and *P* values were adjusted for the age of the patient at each month of follow-up. 95% CI = 95% confidence interval; ON = osteonecrosis; Ref. = reference.

since the 1990s. These findings are in agreement with a recent study by Gladman et al that showed the gradual decrease in ON incidence rate by decade (31). This reflects the decline in the use of high-dose oral prednisone in the management of SLE over the past several decades.

Second, the risk of developing ON was doubled in African American patients with SLE—an association that is well-known. African American SLE patients have more severe disease and more likely require higher doses of glucocorticoids. More recently, a genetic risk factor, *APOL1* risk alleles, was found to be more prevalent in African American patients with ON (26). Higher daily glucocorticoid requirements and genetic predisposition likely explain the increased risk conferred by ethnicity.

Third, a modifiable risk factor, smoking, remained associated with an increased risk of ON. The strong association between alcohol intake, cigarette smoking, and the role of heavy physical work and the occurrence of ON has been demonstrated previously (32). Current but not past smokers were at a higher risk for developing ON in a study (32). This was attributed to endothelial dysfunction compromising the blood supply to the bone. Fourth, we did not find an association between clinical or serologic manifestations and an increased risk of ON. This is in line with a Japanese study that confirmed that disease features such as Raynaud's phenomenon, hyperlipidemia, nephrotic syndrome, hypertension, and disease activity were not considered to be related to ON (25). An older study from our institution reported that patients with ON were more likely to have Raynaud's phenomenon, vasculitis, myositis, and hyperlipidemia (33). A meta-analysis by Zhu et al comprising 16 studies showed that arthritis, Cushingoid habitus, gastrointestinal involvement, hypertension, oral ulcers, pleuritis, renal disease, and vasculitis were associated with ON in SLE patients (34). The present study did not confirm the findings of the meta-analysis by Zhu et al.

Fifth, we did not find any association between hypercoagulable state or vasculopathy and increased risk of ON. We evaluated Raynaud's phenomenon, livedo reticularis, and cutaneous vasculitis as clinical indicators of vasculopathy and did not find an association with increased risk of ON. We also found that positivity for aPL antibodies was not a predictor of increased risk of ON. Our study is in agreement with a cohort study by Gladman et al (31), a cohort study by Petri (2), a case-control study of 265 patients by Mok et al (4), a prospective study on asymptomatic ON by Houssiau et al (5), and a cohort study of 500 patients by Alarcon-Segovia et al (35) that did not find an association between aCL or LAC and ON. The role of aPL antibodies in determining the risk of developing ON was initially reported by Alijotas and colleagues in 16 patients with Kienbock's disease (36). The development of ON in the absence of glucocorticoid administration (37–39) and as an initial presentation of primary aPL syndrome has been reported (40). In SLE patients, the prevalence of aCL antibody positivity in 37 patients with ON in a cohort of 800 patients was reported as 73% (7). Mont et al evaluated 103 SLE patients with ON and found higher levels of aCL IgG antibodies in these patients compared to SLE patients without ON. In

that study, antibodies were measured after the ON event (16). In the aforementioned studies, glucocorticoid use was a confounding factor. In a study by Campos et al evaluating 57 children with SLE, there was an association between the presence of aPL antibodies and ON (41). A high aPL score (odds ratio 5.12 [95% confidence interval 1.18–29.79]), which is also a predictor of thrombosis, but not individual aPL antibodies, was found to be an independent predictor of ON (42). Studies have also investigated other risk factors related to hypercoagulability and vasculopathy, including hyperlipidemia, smoking, Raynaud's phenomenon, superficial thrombosis, and preeclampsia, and have found an association between these factors and the risk of developing ON (6,16,39). The present study did not confirm these associations between risk factors and the risk of ON, except for smoking.

The present analysis showed that a highest daily dosage of prednisone of >40 mg is associated with an increased risk of ON even if prednisone is taken for only 1 month and irrespective of the time of glucocorticoid administration during the disease course. We also show that the daily dosage of 20–39 mg for >1 month increased the risk of ON. Our data indicate that these doses should be avoided. Many studies have evaluated the association between prednisone dose and ON, but ours is the first to address dose and duration jointly and prospectively. In a recent retrospective study, the use of more than 0.8 mg/kg/day of prednisone was associated with the development of ON (8). Our previous study, along with 3 other studies, highlighted the importance of a threshold of 40 mg (14,25,43) irrespective of body weight and duration of administration. In addition, our study showed that receiving 20–39 mg of prednisone for >1 month was sufficient to increase the risk of developing ON. In the meta-analysis by Mont et al that investigated the association between glucocorticoid use and occurrence of ON in 4 medical conditions (SLE, bone marrow transplant, renal transplant, and severe acute respiratory syndrome), every 10 mg–increase in prednisone above 40 mg was associated with a 3.6% increase in incidence of ON (43). A second meta-analysis by Zhu and colleagues assessed 16 studies on SLE patients and ON and concluded that cumulative dose, maximum daily dose, and mean daily dose of prednisone were all significantly higher in the ON group (34). Gladman et al reported that a mean glucocorticoid dosage as low as 10 mg/day may predispose an individual to developing ON (31). Previous work has outlined that the cumulative dose of glucocorticoids at 1 and 4 months is the most important predictor of ON (1,6). In 45 patients with newly diagnosed SLE who needed ≥40 mg per day of prednisone and 3 days of pulse methylprednisolone, 33% developed silent ON and 11% symptomatic ON of the femur (44). In 72 patients with active SLE who received high-dose glucocorticoids for the first time, 44% developed ON between 1 and 5 months after starting treatment with glucocorticoids (15).

The present study did not show an increased risk of ON with administration of pulse methylprednisolone, after adjustment for oral prednisone use. Results from previous works investigating the risk of ON conferred by pulse methylprednisolone are conflicting. While some studies reported no association between pulse methylprednisolone and ON (6,8), 1 meta-analysis did in fact show an association between this therapy and risk of ON (43).

In this analysis, a daily glucocorticoid dosage of <6 mg was not associated with an increased risk of ON. The results of the present study are in line with those of other studies that have tried to explain the mechanism leading to glucocorticoid-induced ON. In SLE patients, impaired femoral head blood supply occurred early after glucocorticoid administration compared to SLE patients and healthy adults not exposed to glucocorticoids. This phenomenon was not observed at mean daily dosages of prednisone <7.5 mg (45,46). In fact, bone repair was possible with chronic low doses of prednisone (47). Moreover, in vitro and animal studies have shown that glucocorticoid use causes differentiation of pluripotent mesenchymal cells to adipocytes and up-regulates the intracellular accumulation of fat (48,49). Fat conversion of the proximal femur hematopoietic marrow and its magnitude correlated with daily prednisone intake and with ON development. These changes were exclusively detected in those receiving a mean daily prednisolone dosage of ≥7.5 mg (50).

The present study is not without limitations. The patients were evaluated for ON after they became symptomatic. We had to exclude patients who developed ON before cohort entry. We did not have *APOL1* testing, which is a novel risk genetic factor in African American individuals (26). The results may not be completely generalizable, considering the demographic characteristics of the cohort. The cohort has a considerable number of African American participants, who are at a higher risk of developing ON.

A daily prednisone dosage of 20–39 mg for >1 month at any point during the disease course was a predictor of ON. Prednisone at 40 mg for just 1 month was an even stronger predictor of ON. More conservative use of oral prednisone, even in lupus nephritis, is now possible. The rituxilup pilot study (51) and the more recent voclosporin trial (52) have demonstrated that oral steroids can be safely limited even in patients with lupus nephritis.

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. Kallas had full access to all data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Kallas, Li, Petri.
Acquisition of data. Kallas, Li, Petri.
Analysis and interpretation of data. Kallas, Li, Petri.

REFERENCES

1. Mok CC, Lau CS, Wong RW. Risk factors for avascular bone necrosis in systemic lupus erythematosus. Br J Rheumatol 1998;37:895–900.

2. Petri M. Musculoskeletal complications of systemic lupus erythematosus in the Hopkins Lupus Cohort: an update. Arthritis Care Res 1995; 8:137–45.
3. Rascu A, Manger K, Kraetsch HG, Kalden JR, Manger B. Osteonecrosis in systemic lupus erythematosus, steroid-induced or a lupus-dependent manifestation? Lupus 1996;5:323–7.
4. Mok MY, Farewell VT, Isenberg DA. Risk factors for avascular necrosis of bone in patients with systemic lupus erythematosus: is there a role for antiphospholipid antibodies? Ann Rheum Dis 2000; 59:462–7.
5. Houssiau FA, N'Zeusseu Toukap A, Depresseux G, Maldague BE, Malghem J, Devogelaer JP, et al. Magnetic resonance imaging-detected avascular osteonecrosis in systemic lupus erythematosus: lack of correlation with antiphospholipid antibodies. Br J Rheumatol 1998;37:448–53.
6. Sayarlioglu M, Yuzbasioglu N, Inanc M, Kamali S, Cefle A, Karaman O, et al. Risk factors for avascular bone necrosis in patients with systemic lupus erythematosus. Rheumatol Int 2012;32:177–82.
7. Asherson RA, Liote F, Page B, Meyer O, Buchanan N, Khamashta MA, et al. Avascular necrosis of bone and antiphospholipid antibodies in systemic lupus erythematosus. J Rheumatol 1993;20:284–8.
8. Hisada R, Kato M, Ohnishi N, Sugawara E, Fujieda Y, Oku K, et al. Antiphospholipid score is a novel risk factor for idiopathic osteonecrosis of the femoral head in patients with systemic lupus erythematosus. Rheumatology (Oxford) 2019;58:645–9.
9. Nakamura F, Fujioka M, Takahashi KA, Ueshima K, Arai Y, Imahori Y, et al. Evaluation of the hemodynamics of the femoral head compared with the ilium, femoral neck and femoral intertrochanteric region in healthy adults: measurement with positron emission tomography (PET). Ann Nucl Med 2005;19:549–55.
10. Uea-areewongsa P, Chaiamnuay S, Narongroeknawin P, Asavatanabodee P. Factors associated with osteonecrosis in Thai lupus patients: a case control study. J Clin Rheumatol 2009;15: 345–9.
11. Petri M, Bechtel B, Dennis G, Shah M, McLaughlin T, Kan H, et al. Burden of corticosteroid use in patients with systemic lupus erythematosus: results from a Delphi panel. Lupus 2014;23:1006–13.
12. Dubois EL, Cozen L. Avascular (aseptic) bone necrosis associated with systemic lupus erythematosus. JAMA 1960;174:966–71.
13. Tse SM, Mok CC. Time trend and risk factors of avascular bone necrosis in patients with systemic lupus erythematosus. Lupus 2017;26:715–22.
14. Shigemura T, Nakamura J, Kishida S, Harada Y, Ohtori S, Kamikawa K, et al. Incidence of osteonecrosis associated with corticosteroid therapy among different underlying diseases: prospective MRI study. Rheumatology (Oxford) 2011;50:2023–8.
15. Oinuma K, Harada Y, Nawata Y, Takabayashi K, Abe I, Kamikawa K, et al. Osteonecrosis in patients with systemic lupus erythematosus develops very early after starting high dose corticosteroid treatment. Ann Rheum Dis 2001;60:1145–8.
16. Mont MA, Glueck CJ, Pacheco IH, Wang P, Hungerford DS, Petri M. Risk factors for osteonecrosis in systemic lupus erythematosus. J Rheumatol 1997;24:654–62.
17. Zonana-Nacach A, Barr SG, Magder LS, Petri M. Damage in systemic lupus erythematosus and its association with corticosteroids. Arthritis Rheum 2000;43:1801–8.
18. Nawata K, Nakamura J, Ikeda K, Furuta S, Nakajima H, Ohtori S, et al. Transitional changes in the incidence of osteonecrosis in systemic lupus erythematosus patients: focus on immunosuppressant agents and glucocorticoids. Rheumatology (Oxford) 2018;57:844–9.
19. Migliaresi S, Picillo U, Ambrosone L, Di Palma G, Mallozzi M, Tesone ER, et al. Avascular osteonecrosis in patients with SLE: relation to corticosteroid therapy and anticardiolipin antibodies. Lupus 1994;3: 37–41.
20. Fialho SC, Bonfa E, Vitule LF, D'Amico E, Caparbo V, Gualandro S, et al. Disease activity as a major risk factor for osteonecrosis in early systemic lupus erythematosus. Lupus 2007;16:239–44.
21. Hussein S, Suitner M, Beland-Bonenfant S, Baril-Dionne A, Vandermeer B, Santesso N, et al. Monitoring of osteonecrosis in systemic lupus erythematosus: a systematic review and metaanalysis. J Rheumatol 2018;45:1462–76.
22. Klipper AR, Stevens MB, Zizic TM, Hungerford DS. Ischemic necrosis of bone in systemic lupus erythematosus. Medicine (Baltimore) 1976; 55:251–7.
23. Sheikh JS, Retzinger GS, Hess EV. Association of osteonecrosis in systemic lupus erythematosus with abnormalities of fibrinolysis. Lupus 1998;7:42–8.
24. Cozen L, Wallace DJ. Risk factors for avascular necrosis in systemic lupus erythematosus. J Rheumatol 1998;25:188.
25. Nagasawa K, Ishii Y, Mayumi T, Tada Y, Ueda A, Yamauchi Y, et al. Avascular necrosis of bone in systemic lupus erythematosus: possible role of haemostatic abnormalities. Ann Rheum Dis 1989;48: 672–6.
26. Yip K, Efuni E, Qian Y, Clancy R, Buyon J, Blazer A. Avascular necrosis is associated with APOL1 variants in African Americans with systemic lupus erythematosus [abstract]. Arthritis Rheumatol 2019;71.
27. Leventhal GH, Dorfman HD. Aseptic necrosis of bone in systemic lupus erythematosus. Semin Arthritis Rheum 1974;4:73–93.
28. D'Agostino RB, Lee ML, Belanger AJ, Cupples LA, Anderson K, Kannel WB. Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. Stat Med 1990;9:1501–15.
29. Hochberg MC. Updating the American College of Rheumatology Revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1997;40:1725.
30. Petri M, Orbai AM, Alarcon GS, Gordon C, Merrill JT, Fortin PR, et al. Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum 2012;64:2677–86.
31. Gladman DD, Dhillon N, Su J, Urowitz MB. Osteonecrosis in SLE: prevalence, patterns, outcomes and predictors. Lupus 2018;27:76–81.
32. Hirota Y, Hirohata T, Fukuda K, Mori M, Yanagawa H, Ohno Y, et al. Association of alcohol intake, cigarette smoking, and occupational status with the risk of idiopathic osteonecrosis of the femoral head. Am J Epidemiol 1993;137:530–8.
33. Zizic TM, Hungerford DS, Stevens MB. Ischemic bone necrosis in systemic lupus erythematosus. II. The early diagnosis of ischemic necrosis of bone. Medicine (Baltimore) 1980;59:134–42.
34. Zhu KK, Xu WD, Pan HF, Zhang M, Ni J, Ge FY, et al. The risk factors of avascular necrosis in patients with systemic lupus erythematosus: a meta-analysis. Inflammation 2014;37:1852–64.
35. Alarcon-Segovia D, Deleze M, Oria CV, Sanchez-Guerrero J, Gomez-Pacheco L, Cabiedes J, et al. Antiphospholipid antibodies and the antiphospholipid syndrome in systemic lupus erythematosus: a prospective analysis of 500 consecutive patients. Medicine (Baltimore) 1989;68:353–65.
36. Alijotas J, Argemi M, Barquinero J. Kienbock's disease and antiphospholipid antibodies. Clin Exp Rheumatol 1990;8:297–8.
37. Tektonidou MG, Malagari K, Vlachoyiannopoulos PG, Kelekis DA, Moutsopoulos HM. Asymptomatic avascular necrosis in patients with primary antiphospholipid syndrome in the absence of corticosteroid use: a prospective study by magnetic resonance imaging. Arthritis Rheum 2003;48:732–6.
38. Vela P, Batlle E, Salas E, Marco P. Primary antiphospholipid syndrome and osteonecrosis. Clin Exp Rheumatol 1991;9:545–6.
39. Seleznick MJ, Silveira LH, Espinoza LR. Avascular necrosis associated with anticardiolipin antibodies. J Rheumatol 1991;18:1416–7.

40. Asherson RA, Khamashta MA, Ordi-Ros J, Derksen RH, Machin SJ, Barquinero J, et al. The "primary" antiphospholipid syndrome: major clinical and serological features. Medicine (Baltimore) 1989;68:366–74.
41. Campos LM, Kiss MH, D'Amico EA, Silva CA. Antiphospholipid antibodies and antiphospholipid syndrome in 57 children and adolescents with systemic lupus erythematosus. Lupus 2003;12:820–6.
42. Otomo K, Atsumi T, Amengual O, Fujieda Y, Kato M, Oku K, et al. Efficacy of the antiphospholipid score for the diagnosis of antiphospholipid syndrome and its predictive value for thrombotic events. Arthritis Rheum 2012;64:504–12.
43. Mont MA, Pivec R, Banerjee S, Issa K, Elmallah RK, Jones LC. High-dose corticosteroid use and risk of hip osteonecrosis: meta-analysis and systematic literature review. J Arthroplasty 2015;30:1506–12.
44. Nagasawa K, Tada Y, Koarada S, Horiuchi T, Tsukamoto H, Murai K, et al. Very early development of steroid-associated osteonecrosis of femoral head in systemic lupus erythematosus: prospective study by MRI. Lupus 2005;14:385–90.
45. Atsumi T, Kuroki Y. Role of impairment of blood supply of the femoral head in the pathogenesis of idiopathic osteonecrosis. Clin Orthop Relat Res 1992:22–30.
46. Nakamura J, Ohtori S, Watanabe A, Nakagawa K, Inoue G, Kishida S, et al. Recovery of the blood flow around the femoral head during early corticosteroid therapy: dynamic magnetic resonance imaging in systemic lupus erythematosus patients. Lupus 2012;21:264–70.
47. Nakamura J, Harada Y, Oinuma K, Iida S, Kishida S, Takahashi K. Spontaneous repair of asymptomatic osteonecrosis associated with corticosteroid therapy in systemic lupus erythematosus: 10-year minimum follow-up with MRI. Lupus 2010;19:1307–14.
48. Cui Q, Wang GJ, Balian G. Steroid-induced adipogenesis in a pluripotential cell line from bone marrow. J Bone Joint Surg Am 1997;79:1054–63.
49. Yamamoto T, Irisa T, Sugioka Y, Sueishi K. Effects of pulse methylprednisolone on bone and marrow tissues: corticosteroid-induced osteonecrosis in rabbits. Arthritis Rheum 1997;40:2055–64.
50. Vande Berg BC, Malghem J, Lecouvet FE, Devogelaer JP, Maldague B, Houssiau FA. Fat conversion of femoral marrow in glucocorticoid-treated patients: a cross-sectional and longitudinal study with magnetic resonance imaging. Arthritis Rheum 1999;42:1405–11.
51. Condon MB, Ashby D, Pepper RJ, Cook HT, Levy JB, Griffith M, et al. Prospective observational single-centre cohort study to evaluate the effectiveness of treating lupus nephritis with rituximab and mycophenolate mofetil but no oral steroids. Ann Rheum Dis 2013;72:1280–6.
52. Rovin BH, Solomons N, Pendergraft WF III, Dooley MA, Tumlin J, Romero-Diaz J, et al. A randomized, controlled double-blind study comparing the efficacy and safety of dose-ranging voclosporin with placebo in achieving remission in patients with active lupus nephritis. Kidney Int 2019;95:219–31.

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1133–1141
DOI 10.1002/acr.24553

American College of Rheumatology
Empowering Rheumatology Professionals

Impact or No Impact for Women With Mild Knee Osteoarthritis: A Bayesian Meta-Analysis of Two Randomized Controlled Trials With Contrasting Interventions

Risto Heikkinen,[1] Benjamin Waller,[2] Matti Munukka,[1] Juhani Multanen,[1] Ari Heinonen,[1] and Juha Karvanen[1]

Objective. To predict the probability of a benefit from 2 contrasting exercise programs for women with a new diagnosis of mild knee osteoarthritis, and to estimate the short- and long-term effects of aquatic resistance training (ART) and high-impact aerobic land training (HLT) compared with a control.

Methods. Original data sets from 2 previously conducted randomized controlled trials were combined and used in a Bayesian meta-analysis. Group differences in multiple response variables were estimated. Variables included cardiorespiratory fitness, dynamic maximum leg muscle power, maximal isometric knee extension and flexion force, pain, other symptoms, and quality of life. The statistical model included a latent commitment variable for each female participant.

Results. ART had a 55–71% probability of benefits in the outcome variables, and as the main effect, the intervention outperformed the control in cardiorespiratory fitness, with a probability of 71% immediately after the intervention period. HLT had a 46–63% probability of benefits after intervention with the outcome variables, but differently from ART; the positive effects of physical performance fade away during the follow-up period. Overall, the differences between groups were small, and the variation in the predictions between individuals was high.

Conclusion. Both interventions had benefits, but ART has a slightly higher probability of long-term benefits on physical performance. Because of high individual variation and no clear advantage of one training method over the other, personal preferences should be considered in the selection of the exercise program to ensure highest commitment to training.

INTRODUCTION

Exercise is one of the cornerstones in the management of osteoarthritis (OA) of the hip and knee (1). Since 2002 and unchanged with new evidence, it has been known that exercise is effective for the management of pain and impaired function in hip and knee OA (2,3). While there is strong evidence for positive effects from land-based neuromuscular exercise and muscular strength, aerobic, and aquatic exercise, there is no consensus on which type of exercise is superior. Recent systematic reviews have been unable to separate the different training environments (4,5), with only a consensus that the training should focus on a specific outcome and be completed 3 times a week (6).

In randomized controlled trials (RCTs) assessing the effect of exercises, the inference is often based on frequentist statistical analysis with group means and *P* values. A *P* value answers the question, "Under the null hypothesis, what is the probability to obtain this or more extreme result." A more important question not answered by *P* values (7) is, "What is the probability that each exercise program would be beneficial for each participant individually, and which exercise program should they choose?"

We use Bayesian analysis (8) to combine the relevant information from 2 different studies and calculate the probabilities that support the decision-making on exercise program recommendation. The first aim of the study is to calculate Bayesian posterior distributions and compare different exercise programs' probabilities

Supported by the Academy of Finland (grant 311877 under the Decision Analytics Utilizing Causal Models and Multiobjective Optimization [DEMO] research area), the Finnish Cultural Foundation, and the Central Finland Regional Fund.

[1]Risto Heikkinen, MSc, Matti Munukka, PhD, Juhani Multanen, PhD, Ari Heinonen, PhD, Juha Karvanen, DSc (Tech): University of Jyväskylä, Jyväskylä, Finland; [2]Benjamin Waller, PhD: University of Reykjavik, Reykjavik, Iceland.

Dr. Karvanen has received consulting fees and/or speaking fees from Biogen Finland and Tale (less than $10,000 each). No other disclosures relevant to this article were reported.

Address correspondence to Risto Heikkinen, MSc, University of Jyväskylä, Department of Mathematics and Statistics, PO Box 35 Jyvaskyla, Keski-Suomi 40014, Finland. Email: risto.heikkinen@statisti.fi.

Submitted for publication July 7, 2020; accepted in revised form January 5, 2021.

SIGNIFICANCE & INNOVATIONS

- We quantify the probability of a benefit from aquatic resistance training and high-impact land training in multiple measures of physical performance, symptoms, and quality of life for patients with mild knee osteoarthritis (OA).
- As the group differences are small compared to the variation between the individuals, patients should choose the exercise according to their preferences to ensure highest commitment to training.
- Medical professionals can make improved personal recommendations on training for individuals with knee OA based on predictive probability calculations that show tradeoffs between different outcomes.

of being beneficial to a new patient in physical performance, symptoms, and quality of life. The second aim is to improve the understanding about uncertainties and individual variation in predictions of exercise responses.

MATERIALS AND METHODS

Study design. This study utilized data from our 2 previous registered RCT studies, AquaRehab (ISRCTN: 65346593) and LuRu (ISRCTN: 58314639). The data sets were collected from January 2012 to May 2013, and March 2008 to April 2010 for AquaRehab and LuRu, respectively. Both studies had an exercise intervention group (aquatic resistance training [ART] in AquaRehab, high-impact aerobic land training [HLT] in LuRu), and both had a nonintervention control group. The study protocols of AquaRehab (9) and LuRu (10) can be found elsewhere and were followed without changes. Included participants were women ages 50–68 years with mild knee OA, body mass index of <35, and no medical reason preventing participation in intensive exercise. Mild knee OA was classified as experiencing knee pain on most days during the last 12 months, not exceeding 5 of 10 on a visual analog scale (0 = "no pain at all," and 10 = "worst pain imaginable"), with radiographic changes in tibiofemoral joint grades I (possible osteophytes) or II (definite osteophytes, possible joint space narrowing) according to the Kellgren/Lawrence classification (11). The design of both studies followed the Consolidated Standards of Reporting Trials (CONSORT) recommendations (12). Both AquaRehab (Dnro 19U/2011) and LuRu (Dnro1E/2008) study protocols were approved by the Ethics Committee of the Central Finland Health Care District and conform to the Declaration of Helsinki. Written informed consent was obtained from all participants prior to enrollment.

Subject recruitment and randomization. The recruitment methods and eligibility criteria for AquaRehab (13) and LuRu (10) are described elsewhere. Inclusion criteria in these 2 RCTs were otherwise similar except for age (AquaRehab: age range 60–68 years; LuRu age range: 50–66 years). The subjects in both studies were randomly allocated into 1 of the 2 arms of the study. Principal investigators were blinded to group allocation. The recruitment process is presented as a flow chart in Figure 1 of Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24553.

Interventions. Participants in the AquaRehab intervention group participated in ART lasting 1 hour, 3 times a week for 4 months. Variable resistance equipment was used to progress training intensity with 3 resistance levels: barefoot; small TheraBand resistance fins (Hygenic Corporation); and large Hydroboots resistance boots (Hydro-Tone Fitness Systems). Training intensity was set at "as hard and fast as possible." A full description of the training program, its progression, and daily training program can be found elsewhere (13).

Participants in the LuRu intervention group participated in supervised HLT, multidirectional aerobic and step-aerobic jumping lasting 55 minutes, 3 times a week for 12 months. The loading was gradually increased after 3 months by progressively raising the height of the fences from 5–20 cm in aerobic exercises, and the height of the step benches from 10–20 cm in jumping exercises. More detailed exercise protocol is provided elsewhere (10).

The control groups in both studies maintained usual care and were asked to continue their leisure time activities. The controls were offered 2 sessions consisting of 1 hour of light stretching, relaxation, and social interaction in AquaRehab, and a social group meeting every third month in LuRu.

Outcome measures. Measurement protocols were identical in both studies. In this study, we chose to use the secondary outcomes from both studies because these are more clinically applicable than the primary outcomes that required quantitative magnetic resonance imaging and dual-energy X-ray absorptiometry. Cardiorespiratory fitness (Vo_2 peak, ml/kg/minute) was calculated from the UKK 2 km walk test (14). Maximal isometric knee extension and flexion force (in newtons) of the affected knee was measured using an adjustable Good Strength dynamometer chair (Metitur) (15). Dynamic maximum leg muscle power (in watts) was examined by measuring peak instantaneous power production during the take-off phase of counter movement jump performed on a custom-made force plate (University of Jyväskylä, Finland). Self-reported pain, other symptoms, and quality of life were measured using the 3 domains of the Finnish version of the Knee Injury and Osteoarthritis Outcome Score (KOOS) (16). Scores were transformed into a score of 0–100, with a score of 0 indicating extreme knee problems, and 100 indicating no knee problems (17).

Statistical analysis. The changes in the response variables from baseline to the end of the intervention and from

Table 1. Group baseline measurements*

	AquaRehab		LuRu		Combined, control (n = 83)
	Control (n = 43)	Intervention (n = 42)	Control (n = 40)	Intervention (n = 35)	
Age, years	63.9 ± 2.2	63.7 ± 2.4	58.2 ± 4.3	57.4 ± 4.2	61.2 ± 4.4
Height, cm	163 ± 4.7	163 ± 5.3	162 ± 4.5	166 ± 5.9	163 ± 4.6
Body mass, kg	69.8 ± 10.6	68.6 ± 9.9	68.8 ± 11.3	72.6 ± 8.4	69.3 ± 10.9
BMI, kg/m^2	26.1 ± 3.2	25.7 ± 3.6	26.1 ± 4.0	26.4 ± 2.7	26.1 ± 3.6
Vo_2 peak, ml/kg/minute	24.9 ± 4.8	24.5 ± 5.6	29 ± 4.3	29.1 ± 3.9	26.9 ± 5.0
Power, W	1,663 ± 285	1,612 ± 260	1,798 ± 341	1,975 ± 382.9	1,728 ± 318
Force, N					
Extension	353 ± 78.9	333 ± 61.7	413 ± 74.6	408 ± 102	382 ± 82.2
Flexion	170 ± 43.1	165 ± 51.3	178 ± 54.5	189 ± 54.8	174 ± 48.8
KOOS score (range 0–100)					
Pain	82.2 ± 12.0	80.3 ± 10.3	86.9 ± 7.2	85.9 ± 10.3	84.5 ± 10.2
Symptoms	75 ± 14.3	74.1 ± 13.0	82.7 ± 10.1	78.3 ± 12.0	78.7 ± 13.0
QoL	70.6 ± 20.3	65.5 ± 17.4	78.5 ± 15.2	76.3 ± 15.1	74.4 ± 18.4

* Values are the mean ± SD. Only participants with at least 1 observed value in response variables have been included. BMI = body mass index; KOOS = Knee Injury and Osteoarthritis Outcome Score; N = newtons; QoL = quality of life; W = watts.

baseline to the end of the follow-up period were compared between the ART group, the HLT group, and the combined control group. In the Bayesian model, the 14-variate response variable (changes in the 7 secondary outcomes, at 2 time points, postintervention and follow-up) was explained by the exercise group effect and a personal effect modifier. The modifier is a latent variable that has not been measured but is estimated from the correlation structure of the data. This personal effect includes commitment, i.e., adherence and compliance with training, plus all other personal factors, for example, age and comorbidities that could cause systematic variation to one's exercise effects between individuals within the group. The modifier is defined so that the population average is 1, i.e., for a woman with average intervention effect, the latent coefficient has a value of 1. For example, if a woman has 5% higher than average differences in all outcome variables, her modifier has a value of 1.05. The minimum for the modifier is 0, indicating possible poor adherence and compliance with the intervention. It is assumed that other unmeasured personal factors do not change the sign of the intervention effect. A gamma distribution with mean 1 and variance 0.2 was chosen to describe our prior knowledge on this individual variation. This distribution has quantiles (Q) (Q[0.025] = 0.33 and Q[0.975] = 2.05) describing range of typical values based on our prior knowledge. The personal effect modifier was fixed to 0 for the control group, as there is no intervention effect. The error terms were modeled by the 14-variate normal distribution.

Missing values in the response variables were assumed to be missing at random and were handled as unknown parameters in the model, i.e., missing values were imputed parallel with estimation of the parameters of interest. The statistical modeling was carried out using R (18) and RStan (19). More information about the implemented Bayesian model, including mathematical

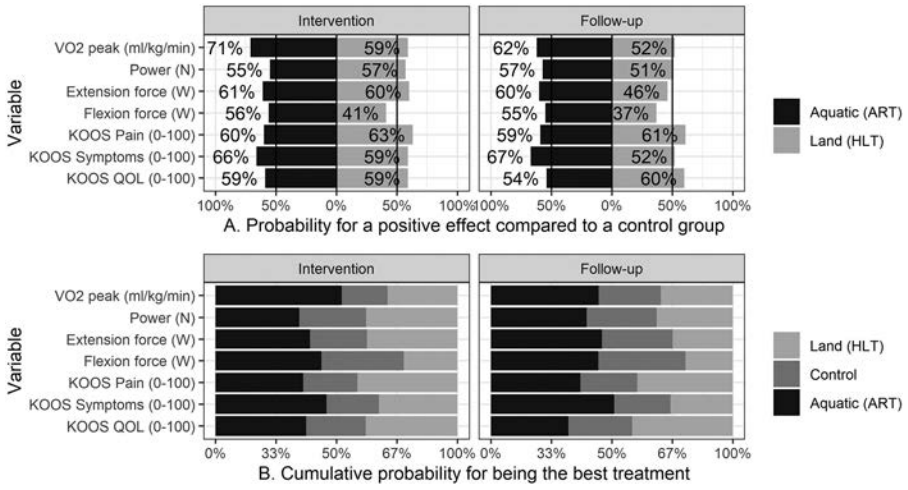
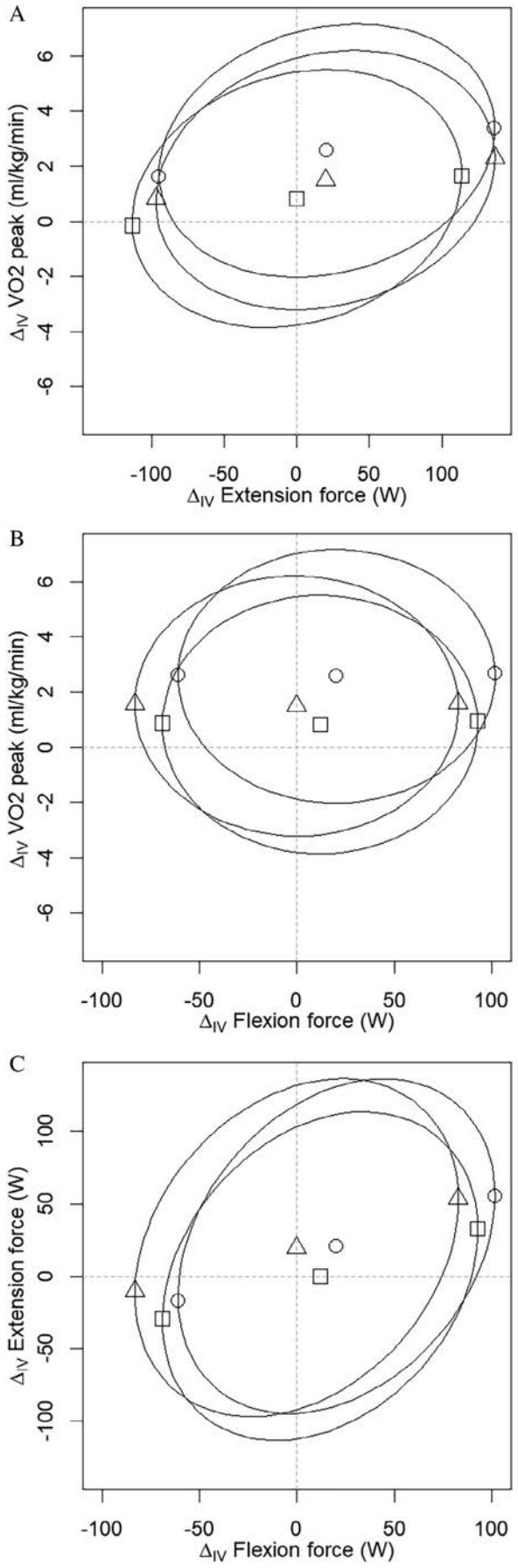


Figure 1. **A**, Probability for a positive effect compared to a control group. **B**, Cumulative probability for being the best treatment. ART = aquatic resistance training; HLT = high-impact aerobic land training; KOOS = Knee Injury and Osteoarthritis Outcome Score; QOL = quality of life.



formulas, is given in Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24553.

The Bayesian analysis results are posterior probability distributions (later posteriors) for the unknown parameters. These posteriors can be used to predict the response variables for a new female participant under different exercise programs. The predictive posterior distributions for pairs of the response variables are visualized by 2-dimensional ellipses that describe the 90% Bayes regions for estimated future values in different groups. With 90% probability, the response of a new patient (a randomly selected female patient with average commitment from the background population) will lie inside the 90% Bayes region. More details of these ellipses are given in Figure 2 of Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24553.

The predictive posterior distributions are summarized as probabilities of benefits. An intervention is considered beneficial if a randomly selected member of the exercise group with an average commitment has a larger change in response variable than a randomly selected member of the control group. Exact 50% probability of benefit is equal to throwing a coin when one predicts whether the exercise program will lead to a better result than the control treatment.

RESULTS

Baseline statistics. Participants who had at least 1 observed value in 7 outcome variables in either the postintervention or the follow-up measurement were included in the analysis. Within this population there were 10% of values missing in 14 response variables of 160 participants. Group sizes were 42 for the ART intervention group, 35 for the HLT intervention group, and 83 (AquaRehab n = 43, and LuRu n = 40) for the control group.

The average values and SDs of the background variables and response variables at baseline for the different study groups are summarized in Table 1. Participants in the AquaRehab study were on average ~6 years older and less active compared to participants in the LuRu study. For the response variables in the statistical model, differences from baseline, the group averages together with SDs, are presented in Tables 1–3 of Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24553.

Figure 2. The mean change between baseline and postintervention and 2-dimensional 90% Bayes region for the prediction of extension force and Vo_2 peak (**A**), flexion force and Vo_2 peak (**B**), and flexion force and extension force (**C**). The values in the top right indicate a positive outcome in both variables. There is 1 mean point and Bayes region for each group. Circles represent the aquatic resistance training group, triangles represent the high-impact aerobic land training group, and squares represent the control group. W = watts.

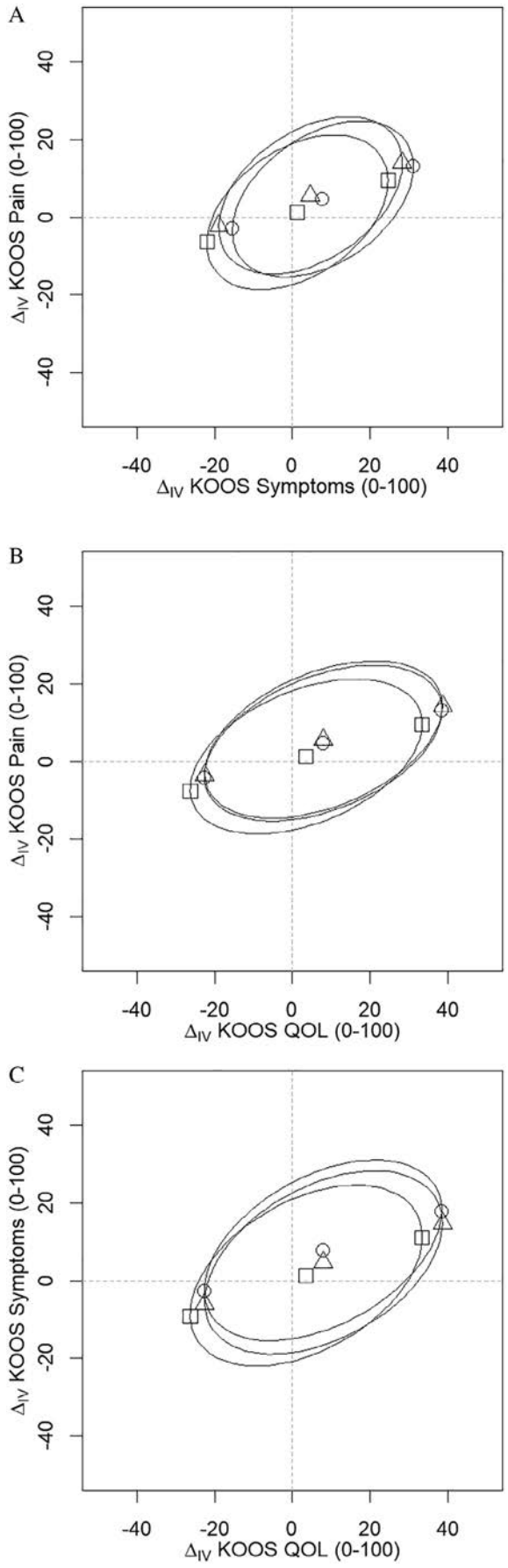


Figure 3. The mean change between baseline and postintervention and 2-dimensional 90% Bayes region for the prediction of Knee Injury and Osteoarthritis Outcome Score (KOOS) symptoms and KOOS pain (**A**), KOOS quality of life (QoL) and KOOS pain (**B**), and KOOS QoL and KOOS symptoms (**C**). The values in the top right indicate a positive outcome in both variables. There is 1 mean point and Bayes region for each group. Circles represent the aquatic resistance training group, triangles represent the high-impact aerobic land training group, and squares represent the control group.

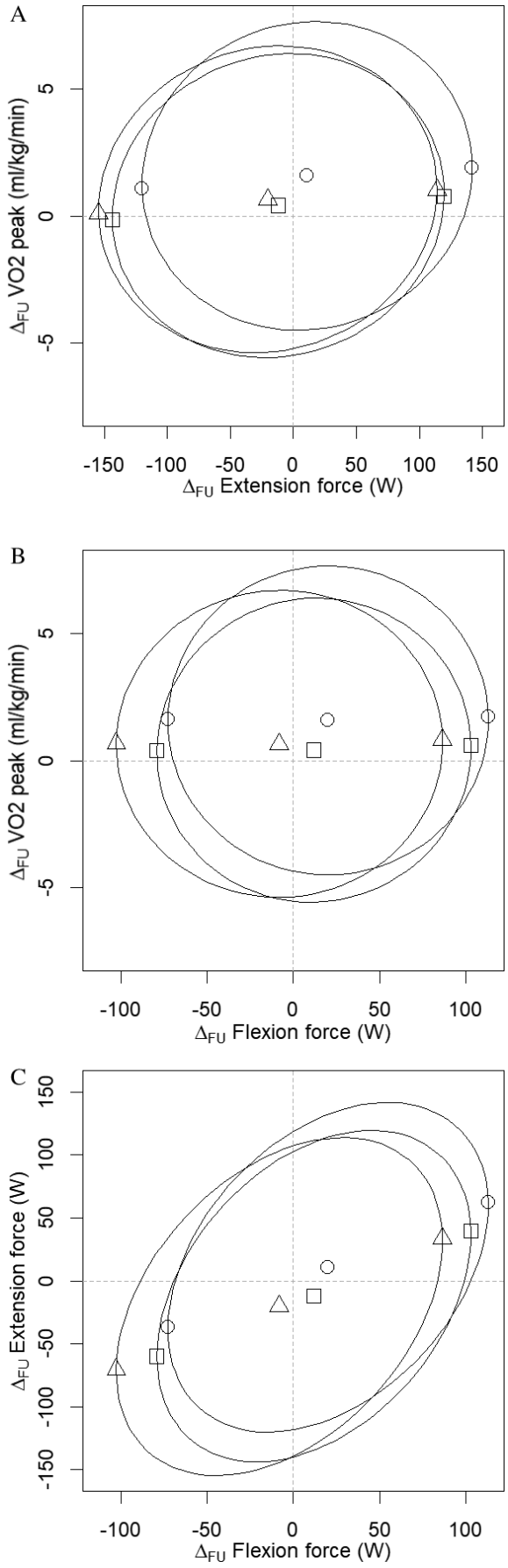
Probabilities of benefits. The estimated probabilities of benefits are summarized in Figure 1A. These probabilities have been calculated for both time periods, from baseline to postintervention and from baseline to the end of the follow-up. The highest probabilities for benefits after intervention are seen in V_{O_2} peak and symptoms in favor of ART. From the posteriors, it was calculated that ART intervention led to higher V_{O_2} peak than the control, with a probability of 71%, and to better KOOS symptoms score, with a probability of 66%. All probability calculations are based on the parameter estimates of the Bayesian model and presented in Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24553.

All variables other than flexion force have close to a 60% probability to have benefits after intervention with the HLT exercise program. However, the benefits in physical performance variables faded away during the follow-up period, ending up close to a 50% probability with most of the variables. The long-term benefits in flexion strength have only a 37% probability with HLT, meaning that the intervention is likely to be ineffective.

The effects for ART remained during the follow-up. As the highest long-term effect, with a probability of 67%, the ART led to fewer symptoms after the follow-up compared to controls. In addition, it was calculated that with a probability of 65%, ART is better than HLT for reducing symptoms in the long term (not shown in Figure 1). In all of the 7 variables, ART is likely to be beneficial even though the probabilities are not much higher than 50%.

In addition, Figure 1B reports the calculated probabilities for the most effective treatment from posteriors for 3 female participants with typical commitment. For instance, on average, the best result in reducing pain right after intervention is achieved by ART, with a probability of 36%, by HLT, with a probability of 41%, and by the control treatment, with a probability of 23%.

Personal effect modifiers. The estimated expected values for the latent personal effect modifiers were distributed between 0.58 and 1.53. This result indicates that the most committed patient had 53% more beneficial intervention than a patient with average commitment, and the least committed got 42% less benefits compared to average commitment. Even though we are interpreting this latent variable primarily as commitment, it also includes all other causes for the individual variation in benefits of



the exercise program. For ART, the commitment had an average of 0.93 and SD of 0.19. For HLT, the average was 0.92 and the SD 0.15, thus there were no major differences between the groups regarding to the variability of commitment.

Pairwise predictions of variables. The ellipses in the Figures visualize the posteriors. The major axis of the ellipse is tilted for predictions that are correlated, such as the flexion force and the extension force. The ellipses also show the amount of individual variation in predictions of differences in the outcome variables. For a clear visualization, the variable power, which has the least differences between groups among physical performance variables, has been dropped out of these ellipse figures.

The ellipses in Figure 2 show the predictions for the differences after the intervention period in the physical performance variables. The overlapping ellipses indicate that the groupwise predictions have a lot of individual variation. The largest differences between the groups are seen on the variables Vo_2 peak and flexion force, where the ART group has the highest predicted improvement.

The predictions after the intervention period in the KOOS domains in Figure 3 indicate that all differences are correlated with each other. Both exercise groups have higher predictions than the control group on average on all variables, but there are a lot of individual variations.

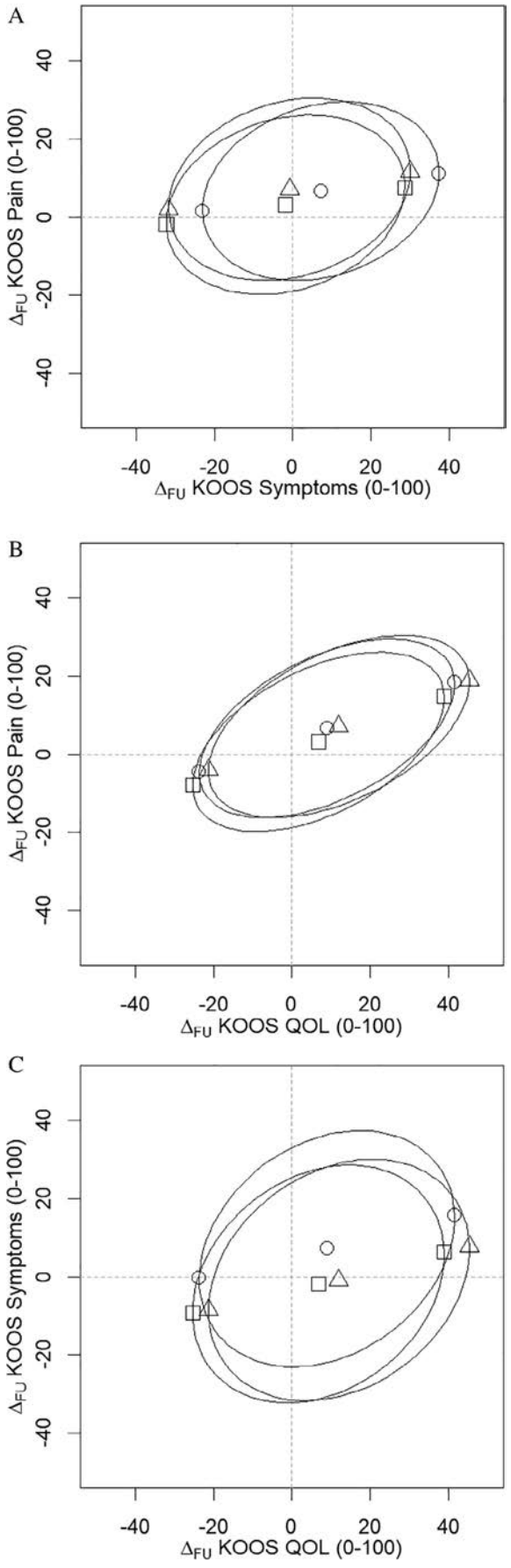
The main finding in the follow-up predictions in Figure 4 is that there are not any signs of benefits of HLT exercise compared to the control group in physical performance variables. Vo_2 peak and extension force show the biggest differences in the ART group compared to other groups.

Figure 5 presents the follow-up predictions in the KOOS domains. The main difference compared to postintervention results is the lack of correlation between symptoms and other variables. Both exercise groups have slightly higher predictions than the control group on average, but there are a lot of individual differences.

DISCUSSION

This was the first Bayesian meta-analysis that estimates posterior probability distributions for exercise interventions in the management of OA. Differently from *P* values, posteriors allow us to calculate the probability of a hypothesis given the data.

Figure 4. The mean change between baseline and post–follow-up and 2-dimensional 90% Bayes region for the prediction of extension force and Vo_2 peak (**A**), flexion force and Vo_2 peak (**B**), and flexion force and extension force (**C**). The values in the top right indicate a positive outcome in both variables. There is 1 mean point and Bayes region for each group. Circles represent the aquatic resistance training group, triangles represent the high-impact aerobic land training group, and squares represent the control group. W = watts.



The Bayesian model combining information from 2 different studies offered a possibility to calculate probabilities for benefits in multiple outcome measurements. Overall, ART seems to be slightly more beneficial than HLT.

The main short-term (from baseline to postintervention) effect was in cardiorespiratory fitness (Vo_2 peak), which had a 71% probability to have a positive short-term benefit with ART intervention. Other short-term probabilities varied between 55% and 66% in ART, and HLT had a >50% probability with all other variables except flexion force. Thus, neither of these interventions predicted overall harm to the patients.

Interestingly, ART is considered an aerobic training intervention that would not be specific enough to improve muscle strength (20). However, results indicate that ART had an equal probability of a positive short-term outcome on muscle strength and power compared to HLT, which showed a lower probability of improvement in Vo_2 peak (Figure 2).

As the main long-term (from baseline to follow-up) effect, ART led to benefits in self-reported OA-related symptoms, with a probability of 67%. The benefits on physical performance in the HLT group faded out during the follow-up period. This can be explained by the fact that subjects' exercise and physical activity intensity decreased after the intervention, thus leading to a detraining effect (21). Short duration interventions (up to a year) have only a short-term effect. Therefore, other interventions, including lifestyle chances and education, may be necessary in these populations to maintain training effects. Interestingly, based on participants' feedback, some found it difficult to find suitable training options after the cessation of the intervention; for example, the ART group found available aquatic exercise groups easy and not effective enough.

Unsurprisingly, the long-term probabilities for pain and quality of life stayed at the same ~60% level with both exercise programs, i.e., there was only a slightly better than coin toss chance to get a better result than with control treatment. It is important to notice that patients with severe knee OA were not included, and therefore the opportunity for change in the KOOS domains was limited (ceiling effect).

The ART program took 4 months compared to the HLT, which took 12 months, which for many participants is a significant difference for commitment. A longer period of exercise could be predicted to produce larger improvements in

Figure 5. The mean change between baseline and post–follow-up and 2-dimensional 90% Bayes region for the prediction of Knee Injury and Osteoarthritis Outcome Score (KOOS) symptoms and KOOS pain (**A**), KOOS quality of life (QOL) and KOOS pain (**B**), and KOOS QOL and KOOS symptoms (**C**). The values in the top right indicate a positive outcome in both variables. There is 1 mean point and Bayes region for each group. Circles represent the aquatic resistance training group, triangles represent the high-impact aerobic land training group, and squares represent the control group.

outcomes that are easier to maintain, but this was not the case in the current study. We found that the 4-month intensive ART had a slightly higher probability of improvement, which is in line with previous recommendations for intervention durations (6). However, one primary aim for the HLT intervention was to impart an effect on bone traits, and thus a longer intervention would be more suitable. Therefore, careful match of desired outcome should be considered when choosing which intervention to recommend.

Our results show that both ART and HLT had a moderately high probability of benefits, but we cannot strongly differentiate between the 2, which is in line with comparisons between land and aquatic exercise (22). The Bayes regions for different treatments were largely overlapping (Figures 2–5), which indicates that uncertainty on the individual level is large compared to the average treatment effects. Approximately 3% of the uncertainty in predictions was due to uncertainty in the estimated model parameters, and ~97% due to variation between individuals (see Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24553), and we were unable to perform counterfactual reasoning on the individual level. There are many interacting individual factors that have an impact on the positive or negative outcome of an exercise intervention, which is shown by in the biopsychosocial model (23). Given that commitment interacts with outcome, we can recommend that the type of the intervention is not as important as the willingness and ability to commit to the intervention.

Based on these results, a woman with mild knee OA who especially wants to improve muscle power in long term, can decide to take ART exercise. On the other hand, there could be another woman, to whom quality of life variables are more important than muscle power and who does not feel that comfortable in a swimming hall. She can decide to take HLT exercise because quality of life variables are not expected to be major differences between the exercises. A third woman could be especially interested in improving V_{O_2} peak and extension force. Her exercise decision could be supported with predictions in Figure 4A. These examples demonstrate the multicriteria nature of decision-making, and Bayesian meta-analysis offers support for comparing tradeoffs between personally important criteria.

The analysis has some limitations. The sizes of the exercise groups are not large enough to allow for reliable estimation of the interaction with individual level background variables such as age and body mass index. The control group was a combination of 2 control groups in different studies that had some differences: patients in the AquaRehab control group were ~6 years older than in the LuRu group, and the AquaRehab's intervention took 4 months compared to LuRu's 12 months of intervention. However, the length of the follow-up period was the same: 12 months in both studies.

In conclusion, Bayesian meta-analysis shows potential as a decision-making tool that can help with choosing between different exercise programs. As common advice for a random female patient without further information, this analysis suggests a slight preference of ART because it is likely to cause benefits in the short term and the long term for every variable. From the quality of life point of view, there does not seem to be any difference as to which exercise program to choose. In addition, in physical performance variables, even the highest probability (71%) leaves a lot of room for individual differences, and it is not necessary to force patients to exercise against their preferences. Instead, this research could be continued within this Bayesian framework toward more personalized recommendations. In a sufficiently large study, this could be done by taking background variables into account and developing a multicriteria decision-making procedure that recommends an exercise program based on the background information of a new patient and her motivation for different kinds of exercises.

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. Mr. Heikkinen 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. Heikkinen, Waller, Munukka, Heinonen, Karvanen.

Acquisition of data. Waller, Munukka, Multanen, Heinonen.

Analysis and interpretation of data. Heikkinen, Karvanen.

REFERENCES

1. Fransen M, Mcconnell S, Harmer AR, van der Esch M, Simic M, Bennell KL. Exercise for osteoarthritis of the knee. Cochrane Database Syst Rev 2015;2015.
2. Uthman OA, van Der Windt DA, Jordan JL, Dziedzic KS, Healey EL, Peat GM, et al. Exercise for lower limb osteoarthritis: systematic review incorporating trial sequential analysis and network meta-analysis. Br J Sports Med 2014;48:1579.
3. Goh SL, Persson MS, Stocks J, Hou Y, Welton NJ, Lin J, et al. Relative efficacy of different exercises for pain, function, performance and quality of life in knee and hip osteoarthritis: systematic review and network meta-analysis. Sport Med 2019;49:743–61.
4. Batterham SI, Heywood S, Keating JL. Systematic review and meta-analysis comparing land and aquatic exercise for people with hip or knee arthritis on function, mobility and other health outcomes. BMC Musculoskelet Disord 2011;12:123.
5. Dong R, Wu Y, Xu S, Zhang L, Ying J, Jin H, et al. Is aquatic exercise more effective than land-based exercise for knee osteoarthritis? Medicine (Baltimore) 2018;97:e13823.
6. Juhl C, Christensen R, Roos EM, Zhang W, Lund H. Impact of exercise type and dose on pain and disability in knee osteoarthritis: a systematic review and meta-regression analysis of randomized controlled trials. Arthritis Rheumatol 2014;66:622–36.
7. Wasserstein RL, Lazar NA. The ASA's statement on P-values: context, process, and purpose. Am Stat 2016;70:129–33.
8. Gelman A, Carling JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis, 3rd ed. Chapman & Hall / CRC; 2013.
9. Waller B, Munukka M, Multanen J, Rantalainen T, Pöyhönen T, Nieminen MT, et al. Effects of a progressive aquatic resistance

exercise program on the biochemical composition and morphology of cartilage in women with mild knee osteoarthritis: protocol for a randomised controlled trial. BMC Musculoskelet Disord 2013;14:1–14.

10. Multanen J, Nieminen MT, Häkkinen A, Kujala UM, Jämsä T, Kautiainen H, et al. Effects of high-impact training on bone and articular cartilage: 12-month randomized controlled quantitative MRI study. J Bone Miner Res 2014;29:192–201.
11. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. Arthritis Rheum 1986;29:1039–49.
12. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surg 2012;10:28–55.
13. Munukka M, Waller B, Rantalainen T, Häkkinen A, Nieminen MT, Lammentausta E, et al. Efficacy of progressive aquatic resistance training for tibiofemoral cartilage in postmenopausal women with mild knee osteoarthritis: a randomised controlled trial. Osteoarthr Cartil 2016;24:1708–17.
14. Laukkanen RM, Oja P, Pasanen ME, Vuori IM. A two-kilometer walking test: effect of walking speed on the prediction of maximal oxygen uptake. Scand J Med Sci Sports 2007;3:263–6.
15. Sipilä S, Multanen J, Kallinen M, Era P, Suominen H. Effects of strength and endurance training on isometric muscle strength and walking speed in elderly women. Acta Physiol Scand 1996;156: 457–64.
16. Multanen J, Honkanen M, Häkkinen A, Kiviranta I. Construct validity and reliability of the Finnish version of the Knee Injury and Osteoarthritis Outcome Score. BMC Musculoskelet Disord 2018;19:155.
17. Bekkers JE, de Windt TS, Raijmakers NJ, Dhert WJ, Saris DB. Validation of the Knee Injury and Osteoarthritis Outcome Score (KOOS) for the treatment of focal cartilage lesions. Osteoarthritis Cartilage 2009; 17:1434–9.
18. R Core Team. R: a language and environment for statistical computing. 2020. URL: https://www.r-project.org/.
19. Stan Development Team. {RStan}: the {R} interface to {Stan}. 2020. URL: http://mc-stan.org.
20. Lund H, Weile U, Christensen R, Rostock B, Downey A, Bartels EM, et al. A randomized controlled trial of aquatic and land-based exercise in patients with knee osteoarthritis. J Rehabil Med 2008;40:137–44.
21. Waller B, Munukka M, Kujala UM, Heinonen AO. Response to the comments on "Effects of high intensity aquatic resistance training on body composition and walking speed in women with mild knee osteoarthritis: a 4-month RCT with 12-month follow-up." Osteoarthritis Cartilage 2017;25:e19–20.
22. Heywood S, McClelland J, Mentiplay B, Geigle P, Rahmann A, Clark R. Effectiveness of aquatic exercise in improving lower limb strength in musculoskeletal conditions: a systematic review and meta-analysis. Arch Phys Med Rehabil 2017;98:173–86.
23. Kanavaki AM, Rushton A, Efstathiou N, Alrushud A, Klocke R, Abhishek A, et al. Barriers and facilitators of physical activity in knee and hip osteoarthritis: a systematic review of qualitative evidence. BMJ Open 2017;7:e017042.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1142–1153
DOI 10.1002/acr.24557

Multivariable Modeling of Biomarker Data From the Phase I Foundation for the National Institutes of Health Osteoarthritis Biomarkers Consortium

David J. Hunter,[1] Leticia A. Deveza,[1] Jamie E. Collins,[2] Elena Losina,[2] Jeffrey N. Katz,[2] Michael C. Nevitt,[3] John A. Lynch,[3] Frank W. Roemer,[4] Ali Guermazi,[5] Michael A. Bowes,[6] Erik B. Dam,[7] Felix Eckstein,[8] C. Kent Kwoh,[9] Steve Hoffmann,[10] and Virginia B. Kraus[11]

Objective. To determine the optimal combination of imaging and biochemical biomarkers for use in the prediction of knee osteoarthritis (OA) progression.

Methods. The present study was a nested case–control trial from the Foundation of the National Institutes of Health OA Biomarkers Consortium that assessed study participants with a Kellgren/Lawrence grade of 1–3 who had complete biomarker data available (n = 539 to 550). Cases were participants' knees that had radiographic and pain progression between 24 and 48 months compared to baseline. Radiographic progression only was assessed in secondary analyses. Biomarkers (baseline and 24-month changes) that had a *P* value of <0.10 in univariate analysis were selected, including quantitative cartilage thickness and volume on magnetic resonance imaging (MRI), semiquantitative MRI markers, bone shape and area, quantitative meniscal volume, radiographic progression (trabecular bone texture [TBT]), and serum and/or urine biochemical markers. Multivariable logistic regression models were built using 3 different stepwise selection methods (complex models versus parsimonious models).

Results. Among baseline biomarkers, the number of locations affected by osteophytes (semiquantitative), quantitative central medial femoral and central lateral femoral cartilage thickness, patellar bone shape, and semiquantitative Hoffa-synovitis predicted OA progression in most models (C statistic 0.641–0.671). In most models, 24-month changes in semiquantitative MRI markers (effusion-synovitis, meniscal morphologic changes, and cartilage damage), quantitative central medial femoral cartilage thickness, quantitative medial tibial cartilage volume, quantitative lateral patellofemoral bone area, horizontal TBT (intercept term), and urine N-telopeptide of type I collagen predicted OA progression (C statistic 0.680–0.724). A different combination of imaging and biochemical biomarkers (baseline and 24-month change) predicted radiographic progression only, which had a higher C statistic of 0.716–0.832.

Conclusion. The present study highlights the combination of biomarkers with potential prognostic utility in OA disease-modifying trials. Properly qualified, these biomarkers could be used to enrich future trials with participants likely to experience progression of knee OA.

INTRODUCTION

There are currently no pharmacologic therapies approved by regulatory agencies to prevent or stop the progression of knee osteoarthritis (OA) (1), although some therapies have recently been found to beneficially modify structural progression (2,3). Half of patients with knee OA are estimated to progress to end-stage disease requiring total knee replacement (TKR) (4). Improvements

The Foundation of the NIH Osteoarthritis (FINH OA) Biomarkers Consortium and the present study were supported by the Arthritis Foundation, AbbVie, Amgen, Bioibérica SA, Johnson & Johnson (DePuy Mitek), Flexion Therapeutics, GlaxoSmithKline Australia, Merck Serono, Rottapharm (Madaus), Sanofi, Stryker Corporation, the Pivotal Osteoarthritis Initiative (OAI) MRI Analyses (POMA) study, and NIH grant HH-SN-2682010000. Additional funding partners include Merck Research Laboratories, Novartis Pharmaceuticals Corporation, GlaxoSmithKline, and Pfizer. Private sector funding for the Consortium and OAI is managed by the FNIH. The OAI is a public-private partnership comprised of 5 contracts funded by the NIH (N01-AR-22258, N01-AR-22259, N01-AR-22260, N01-AR-22261, and N01-AR-22262). Dr. Hunter's work is supported by a National Health and Medical Research Council of Australia Practitioner Fellowship and a grant from the FNIH OA Biomarkers Consortium. Dr. Deveza's work is supported by the Royal Australasian College of Physicians and Australian Rheumatology Association and a DEV Starr Research Establishment Fellowship. Dr. Collins's work is supported by the Rheumatology Research Foundation Investigator Award. Dr. Losina's work is supported by a grant from the FNIH OA Biomarkers Consortium and NIH grants R01-AR-064320, K24-AR-057827, and P30-AR-072577 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS). Dr. Katz's work is supported by NIH grants P30-AR-072577, U01-AR-071658, and R21-AR-076156 from NIAMS. Dr. Eckstein's work is supported by the NIH, the FNIH, the European Union, the Paracelsus Medical University Research Fund, and the Federal Ministry of Education and Research of

SIGNIFICANCE & INNOVATIONS

- Several imaging and biochemical markers have been shown to have prognostic validity for measuring the progression of knee osteoarthritis (OA). The present study evaluated biomarkers from all biomarker domains (i.e., magnetic resonance imaging, radiographic imaging, and biochemical analysis) in multivariable models and demonstrated which biomarkers (measured at baseline and change over 24 months) had prognostic value for knee OA progression.
- Findings from the present study show the most promising biomarkers that can be used in future structure-modifying OA trials, specifically in the selection of study participants who were more likely to have progression of knee OA (baseline biomarkers) as well as biomarkers that could be used as structural end points (longitudinal change biomarkers), if properly qualified.

in clinical trial design are critically needed to overcome barriers to the development of disease-modifying treatments for the improvement of OA care. Biomarkers may enhance the success of every phase of the drug development process, as they can improve predictability by identifying patients more likely to benefit from certain therapies, those most likely to incur adverse events, or by contributing to a better understanding of drug mechanisms and actions (5,6).

Further refinement and improvement of measures of joint structural change based on imaging and/or biochemical markers are needed to identify individuals with knee OA that is likely to progress radiographically and symptomatically and to overcome the limited responsiveness of existing imaging biomarkers (e.g., radiographic joint space width [JSW] loss) (7). To overcome these obstacles, the Foundation for the National Institutes of Health (FNIH) OA Biomarkers Consortium carried out an extensive phase I biomarker validation study from 2012 to 2015 using a nested case–control sample of symptomatic and/or radiographic knee OA progression within the Osteoarthritis Initiative (OAI) (8). The overarching project objective was to establish the prognostic validity of several imaging and biochemical biomarkers for knee OA progression. Some results of this study have been published in work that focused on individual biomarker domains (9–13).

As some of these biomarkers may be highly correlated with each other, the specific purpose of the current work and ultimate aim of the FNIH phase I study was to determine the optimal combination of imaging and biochemical biomarkers in multivariable analyses. This final step will allow for the development of a multifactorial model of biomarkers that will best predict the risk of OA progression for further validation in the phase II trial of the OA Biomarkers Consortium. To this end, we evaluated the association and prognostic validity between biomarkers (assessed either at baseline or change over 24 months) with radiographic and pain progression over the longer term (baseline to 48 months) in knees affected by mild to moderate tibiofemoral OA.

PATIENTS AND METHODS

Study design. Six hundred participants in the OAI were selected for participation in the FNIH Biomarkers Consortium based on a participant having at least 1 knee with frequent pain and a Kellgren/Lawrence (K/L) grade of 1, 2, or 3 on knee radiographs at baseline (8). Selected participants were required to have baseline and 24 months of radiographic minimum medial tibiofemoral JSW data

Germany (BMBF). Dr. Kraus's work is supported by the FNIH, the Claude D. Pepper Older American Independence Centers program at the NIH (National Institutes of Aging grant 5P30-AG-028716), and the NIH (grant R01-AR-071450).

[1]David J. Hunter, PhD, Leticia A. Deveza, PhD: Royal North Shore Hospital and University of Sydney, Sydney, Australia; [2]Jamie E. Collins, PhD, Elena Losina, PhD, Jeffrey N. Katz, MD, MSc: Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; [3]Michael C. Nevitt, PhD, John A. Lynch, PhD: University of California, San Francisco; [4]Frank W. Roemer, MD: Boston University School of Medicine, Boston, Massachusetts, and University of Erlangen, Nuremberg, Erlangen, Germany; [5]Ali Guermazi, PhD: Boston University School of Medicine, Boston, Massachusetts; [6]Michael A. Bowes, PhD: Imorphics Ltd., Manchester, UK; [7]Erik B. Dam, PhD: University of Copenhagen and BiomedIQ, Copenhagen, Denmark; [8]Felix Eckstein, PhD: Paracelsus Medical University, Salzburg, Austria, and Chondrometrics GmbH, Ainring, Germany; [9]C. Kent Kwoh, PhD: University of Arizona College of Medicine, Tucson; [10]Steve Hoffmann, MS: Foundation for the National Institutes of Health, North Bethesda, Maryland; [11]Virginia B. Kraus, PhD: Duke University School of Medicine, Durham, North Carolina.

Drs. Hunter and Deveza contributed equally to this work.

Dr. Hunter has received consulting fees from Merck Serono, TLC Biopharmaceuticals, Pfizer, and Eli Lilly and Company (less than $10,000 each). Dr. Deveza has received royalties from UptoDate and partial reimbursement of conference registration costs from Pfizer (less than $10,00). Dr. Collins has received consulting fees from Boston Imaging Core Labs (less than $10,000). Dr. Losina has received research support from Samumed and Flexion Therapeutics and consulting fees from Pfizer (less than $10,000). Dr. Katz has received research support from Samumed and Flexion Therapeutics. Dr. Lynch has received consulting fees from Boston Imaging Core Lab (BICL) (less than $10,000). Dr. Roemer is a co-owner of BICL and has received consulting fees from the California Institute of Biomedical Research (less than $10,000). Dr. Guermazi is a shareholder in BICL and has received consulting fees from Roche, Galapagos, and AstraZeneca (less than $10,000 each) and Pfizer, Merck Serono, and TissueGene (more than $10,000 each). Dr. Bowes owns stock or stock options in Imorphics, Ltd., a wholly owned subsidiary of Stryker Corporation. Dr. Dam is a co-owner of BiomedIQ, a company that holds the patent for the knee imaging quantification software used in meniscus quantification. Dr. Eckstein is CEO/CMO of Chondrometrics GmbH, owns stock or stock options in Chondrometrics GmbH, and has received consulting fees from Merck KGaA, AbbVie, Galapagos, Novartis, Kolon-TissueGene, Servier Laboratories, and Roche (less than $10,000 each) and research support from OrthoTrophix, MerckK GaA, Samumed, TissueGene, BICL, Galapagos, and Novartis. Dr. Kwoh has received research support from Merck Serono and consulting fees from Thuasne, TLC Biopharmaceuticals, GlaxoSmithKline, Regeneron Pharmaceuticals, Amzell BV, Astellas Pharma, Regulus Therapeutics, Focus Medical Communications, Prime Education LLC (less than $10,000 each), and Express Scripts (more than $10,000 each). Dr. Kraus holds a patent related to the methods used in the reduction of the complex fractal data collected from trabecular bone texture parameters.

Address correspondence to David J. Hunter, PhD, Rheumatology Department, Royal North Shore Hospital, Reserve Road, St. Leonards, New South Wales 2065, Australia. Email: David.Hunter@sydney.edu.au.

Submitted for publication May 12, 2020; accepted in revised form January 5, 2021.

(measured using automated software [14]), magnetic resonance imaging (MRI) of the knee, stored serum and urine specimens, and clinical data.

A predetermined number of index knees were selected based on outcome assessment at 48 months (1 knee per participant) in the following 4 mutually exclusive groups: 1) knees with both radiographic and pain progression (n = 194), 2) knees with radiographic progression but not pain progression (n = 103), 3) knees with pain but not radiographic progression (n = 103), and 4) knees with neither radiographic nor pain progression (n = 200). The main analysis compared knees with both radiographic and pain progression (n = 194) to all other knees (n = 406). We took this approach to ensure the 2 main OA outcome domains (structural and symptomatic) were represented in the main progression definition. Radiographic and pain progression were determined as previously described (9). Briefly, radiographic progression was defined as a minimum JSW loss of ≥0.7 mm, and pain progression was defined as a persistent (sustained at ≥2 time points) increase of ≥9 points on the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain subscale (0–100 scale) (8,15,16). Knees were excluded from analysis under the following circumstances: if progression criteria were met by 12 months so as to enable the study of change in biomarkers before the progression definition was met, if radiographic lateral joint space narrowing grade 2 or 3 was present at baseline (17), or if TKR or total hip replacement had occurred prior to 24 months due to possible effects on biochemical markers. The complete flow diagram of the study is shown in Figure 1. Knees and participants were frequency matched for baseline K/L grade and body mass index (BMI) categories, respectively (10).

Knee MRI acquisition. MRI acquisition was performed using a Trio 3T MRI system (Siemens Healthcare) at the 4 OAI clinical sites. Additional parameters of the full OAI pulse sequence protocol and sequence parameters have been previously

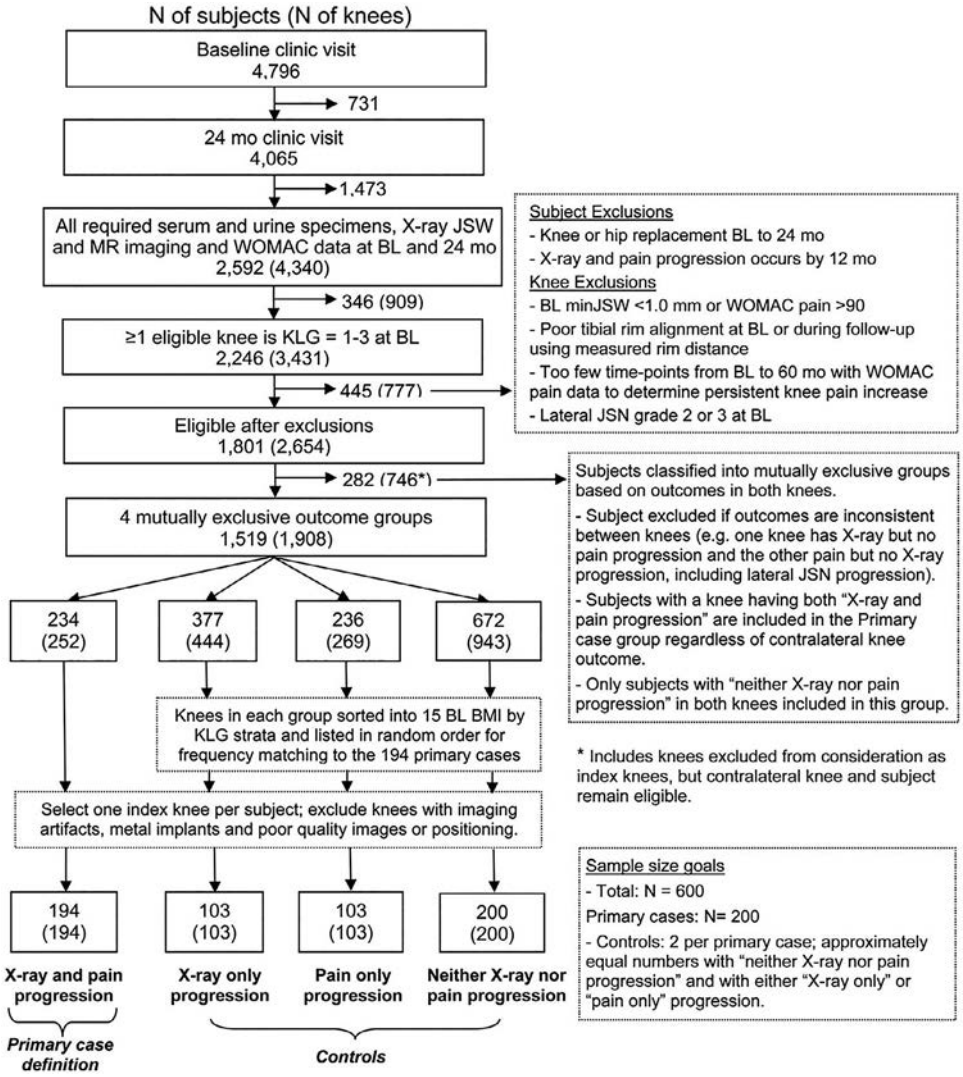


Figure 1. Flow diagram of the study participants. BL = baseline; BMI = body mass index; JSN = joint space narrowing; minJSW = minimum joint space width; MR = magnetic resonance; KLG = Kellgren/Lawrence grade; WOMAC = Western Ontario and McMaster Universities Arthritis Index.

published in detail (18) (see Supplementary Methods, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24557).

Biomarkers. Biomarkers included MRI (quantitative cartilage thickness and volume, semiquantitative MRI markers, bone shape and area, and quantitative meniscal volume), radiographic markers (trabecular bone texture [TBT]), and serum and/or urine biochemical markers, which has been previously described (10–13) (see the Supplementary Methods). The reproducibility of the biomarker measurements was overall satisfactory and has been previously reported (10,12,13,19).

Semiquantitative analyses. Semiquantitative scoring of MRI findings included the assessment of cartilage and meniscal damage, bone marrow lesions, osteophytes and effusion/synovitis using water-sensitive conventional MRI acquisitions (20–23). MRIs were read according to the MRI Osteoarthritis Knee Score system (24) in sequential order and without blinding to the time point of acquisition. The readers (AG and FWR) were blinded to clinical characteristics and case/control status.

Quantitative cartilage morphometry. Cartilage thickness analysis relied on sagittal double-echo steady-state (DESS) imaging (9). Segmentation of the femorotibial cartilage surfaces at the medial and lateral tibia and weight-bearing femur were processed as triplets by the same reader. The readers were blinded to case/control status and image acquisition order.

Bone shape and area. Femur, tibia, and patella bone surfaces were automatically segmented from DESS with water excitation images using active appearance models (10). Two measures were used: 1) subchondral bone area on the medial and lateral femur, tibia and patella and 2) position on 3-dimensional shape vectors for the femur, tibia, and patella (Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24557). Shape measures were normalized to a Z-scale with the mean non-OA shape represented as +1 and the mean OA shape as –1.

Meniscal volume and radiographic TBT. Medial and lateral meniscus volumes were automatically quantified using the computer-based Knee Imaging Quantification framework. The framework combined multi-atlas registration and supervised classification to segment the knee tissues (25). TBT is a way of representing the state of the vertical and horizontal bone trabeculae. Quantification of TBT is a two-step process (Supplementary Methods, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24557) using a semiautomated software (12,26).

Biochemical markers. Biochemical markers were quantified in both serum and/or urine (13). All urinary markers were normalized to urinary creatinine concentration. Interassay coefficients of variation ranged from 3.0% to 12.3% (13).

Patient and public involvement. Consumers are part of the steering committee guiding the design and ongoing conduct of the present study. Once published, the results of the present work will be disseminated through advocacy groups, Twitter, and other mainstream media to engage with the wider public.

Statistical analysis. All variables that had a P value of <0.10 in univariate analysis were advanced to multivariable modeling. In total, 27 and 43 biomarkers were tested in the baseline analysis and change in biomarkers over 24 months analysis, respectively. Models were fit separately for baseline and change in biomarkers. For both sets of models, we first considered models with imaging parameters only (models 1–3) and then added the biochemical markers in a second step (models 4–6) in order to assess the additional prognostic value of adding biochemical markers to imaging parameters only. Three different stepwise selection methods were used to determine the best subset of predictors: 1) Akaike's information criterion (AIC) (models 1 and 4), 2) Schwarz Bayesian Criterion (models 2 and 5), and 3) P value (models 3 and 6) ($P = 0.2$ for entry and $P = 0.1$ for retention). Results were compared across the 3 types of selection procedures in order to assess the robustness of the results. Multivariable logistic regression was used for the analysis that included participants who had complete data for all biomarker parameters.

To assess the prognostic ability of each multivariable model, we presented the area under the curve (AUC), the receiver operating characteristic curve (C statistic), the integrated discrimination improvement (IDI) and the category-less net reclassification (NRI) for each model (27,28). The AUCs are presented for results that were not adjusted for covariates, results that were adjusted for covariates (sex, race, and the following baseline measures: minimum JSW, WOMAC pain score, age, BMI, K/L grade, and use of pain medications), and results that were adjusted with a 10-fold cross-validation. The IDI and NRI are calculated as improvement versus the model with covariates only and are calculated with a 10-fold cross-validation (28) (Supplementary Methods, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24557). For the TBT and biochemical markers, change over 24 months was quantified as time-integrated concentration (TIC). TICs are equivalent to the AUC defined by the individual values for the specific time interval (13).

Sensitivity analysis. We used structural (i.e., radiographic) progression, irrespective of pain progression (n = 297), as the definition of progression in secondary analyses, with all radiographic nonprogressors considered controls (n = 303). As a sensitivity analysis, we ran models using absolute change

of biomarkers over 24 months (24-month value minus baseline value) for TBT and biochemical markers. Because most missing data were in TBT parameters, we ran a sensitivity analysis excluding the TBT parameters (n = 600 and n = 596 in the baseline and 24-month change analysis, respectively).

RESULTS

Study sample. Of the 600 participants included in the FNIH study, 46 participants were missing TBT data. Initial univariate analyses were run in the cohort of 554 study participants with TBT data. Results of the univariate analysis that used baseline and 24-month change in biomarkers are provided in Supplementary Tables 1 and 2 (available at http://onlinelibrary.wiley.com/doi/10.1002/acr.2455), respectively. After further excluding participants that did not have complete data on all selected biomarkers, 550 participants (92%; 173 cases and 377 controls) and 539 participants (90%; 171 cases and 368 controls) were included in the baseline and 24-month change multivariable analysis, respectively. Demographic characteristics of the study sample included in the baseline and 24-month change analysis are provided in Table 1 and Supplementary Table 3 (available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24557), respectively.

Baseline biomarkers predicting pain and radiographic progression over 48 months. For the imaging biomarkers, the adjusted AUCs with 10-fold cross-validation ranged from 0.641 to 0.669, with inclusion of the number of locations affected by osteophyte (semiquantitative) and patella shape in all models (Table 2). Quantitative central medial femoral cartilage thickness (external), central lateral femoral cartilage thickness (internal), and Hoffa-synovitis (semiquantitative) were associated with case status in 2 of the 3 models. When biochemical markers were added to the imaging biomarkers, no biochemical markers were selected using the TIC or *P* value–based approaches, whereas the AIC-based approach additionally selected serum N-telopeptide of type I collagen (NTX-1) (AUC of 0.671).

Change in biomarkers over 24 months predicting pain and radiographic progression over 48 months. In models including only imaging markers, worsening in semiquantitative effusion-synovitis and semiquantitative meniscal damage were predictive of progression in all 3 models, with the addition of the intercept (horizontal) TBT parameter (Table 3). Other markers were significantly associated with case status in 2 of the 3 models: increase in the number of areas with worsening semiquantitative cartilage morphology, loss of quantitative cartilage thickness in the central medial femur (center), loss of quantitative cartilage volume in the medial tibia, and change in the quantitative lateral patellofemoral bone area. AUCs ranged from 0.680 to 0.713. Increases in serum or urine NTX-I were associated with outcome in at least 1 model. The AUCs of the models including biochemical markers ranged from 0.683 to 0.724.

Table 1. Characteristics of analytic samples included in the analysis at baseline for the prediction of radiographic and pain progression (n = 550)*

Characteristic	Cases (knees with radiographic and pain progression)	Control knees			
		All 3 groups	Radiographic progression only	Pain progression only	No radiographic or pain progression
Age, mean ± SD years	61.9 ± 8.8	61.6 ± 8.8	63.5 ± 8.2	59.3 ± 8.7	61.9 ± 9.0
Sex					
Male	73 (42)				
Female	100 (58)				
BMI, mean ± SD kg/m^2	30.8 ± 4.9	30.7 ± 4.8	30.7 ± 4.7	31.1 ± 5.0	30.5 ± 4.7
Kellgren/Lawrence grade at baseline					
1	22 (13)	227 (60)	43 (46)	64 (65)	120 (65)
2	76 (44)	30.7 (4.8)	30.7 (4.7)	31.1 (5.0)	30.5 (4.7)
3	75 (43)	50 (13)	14 (15)	13 (13)	23 (12)
Race		209 (55)	44 (47)	59 (60)	106 (57)
White	36 (21)	118 (31)	35 (38)	26 (27)	57 (31)
Not White	137 (79)	76 (20)	10 (11)	28 (29)	38 (20)
Pain medications at baseline		301 (80)	83 (89)	70 (71)	148 (80)
No	115 (66)	275 (73)	74 (80)	63 (64)	138 (74)
Yes	58 (34)	102 (27)	19 (20)	35 (36)	48 (26)
WOMAC pain at baseline, mean ± SD	10.2 ± 12.7	12.5 ± 16.3	15.1 ± 18.6	9.3 ± 13.1	13.0 ± 16.3
Minimum JSW at baseline, mean ± SD	3.8 ± 1.4	3.9 ± 1.0	3.9 ± 1.1	4.0 ± 1.0	3.9 ± 1.0

* Except where indicated otherwise, values are the number (%). BMI = body mass index; JSW = joint space width; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index.

Table 2. Results of multivariable modeling, with baseline biomarkers for the prediction of radiographic and pain progression (n = 550)*

	Imaging biomarkers only			Imaging + biochemical markers		
	Model 1 (Selection method: stepwise; AIC)	Model 2 (Selection method: stepwise; SBC)	Model 3† (Selection method: stepwise; *P*)	Model 4 (Selection method: stepwise; AIC)	Model 5‡ (Selection method: stepwise; SBC)	Model 6† (Selection method: stepwise; *P*)
Model characteristics						
Unadjusted AUC	0.684	0.646	0.684	0.688	0.646	0.684
Adjusted AUC§	0.718	0.688	0.718	0.722	0.688	0.718
Adjusted AUC with 10-fold cross-validation	0.669	0.641	0.669	0.671	0.641	0.669
IDI	0.0832	0.0548	0.0832	0.0861	0.0548	0.0832
NRI	0.4746	0.4048	0.4746	0.5007	0.4048	0.4746
% of cases correctly reclassified¶	28	29	28	28	29	28
% of controls correctly reclassified#	19	12	19	22	12	19
Biomarkers included						
Semiquantitative Hoffa-synovitis score	0.0467	–	0.0467	0.0682	–	0.0467
0	Ref.	–	Ref.	Ref.	–	Ref.
1	1.67 (1.09, 2.56)	–	1.67 (1.09, 2.56)	1.61 (1.05, 2.48)	–	1.67 (1.09, 2.56)
2–3	1.87 (0.86, 4.05)	–	1.87 (0.86, 4.05)	1.83 (0.84, 3.95)	–	1.87 (0.86, 4.05)
Quantitative mean cartilage thickness of the central lateral femur (internal), Z score	0.0374	–	0.0374	0.0374	–	0.0374
OR for each 1-unit SD increase	1.29 (1.02, 1.65)	–	1.29 (1.02, 1.65)	1.29 (1.02, 1.65)	–	1.29 (1.02, 1.65)
Quantitative mean cartilage thickness of the central medial femur (external), Z score	0.0013	–	0.0013	0.0010	–	0.0013
OR for each 1-unit SD increase	0.65 (0.50, 0.85)	–	0.65 (0.50, 0.85)	0.65 (0.50, 0.84)	–	0.65 (0.50, 0.85)
Number of locations affected by any osteophyte (semiquantitative)	0.0008	<0.0001	0.0008	0.0008	<0.0001	0.0008
0–2	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
3–5	1.92 (0.92, 4.02)	1.92 (0.93, 3.95)	1.92 (0.92, 4.02)	1.99 (0.95, 4.18)	1.92 (0.93, 3.95)	1.92 (0.92, 4.02)
6+	3.47 (1.72, 7.02)	3.94 (1.99, 7.78)	3.47 (1.72, 7.02)	3.55 (1.75, 7.22)	3.94 (1.99, 7.78)	3.47 (1.72, 7.02)
Patella shape (vector), Z score	0.0028	0.0015	0.0028	0.0029	0.0015	0.0028
OR for each 1-unit SD increase	0.72 (0.58, 0.89)	0.71 (0.58, 0.88)	0.72 (0.58, 0.89)	0.72 (0.59, 0.90)	0.71 (0.58, 0.88)	0.72 (0.58, 0.89)
Serum NTX-I**	–	–	–	0.1197	–	–
OR for each 1-unit SD increase	–	–	–	1.16 (0.96, 1.41)	–	–

* Except where indicated otherwise, values are the odds ratios (ORs) and 95% confidence intervals. *P* values are shown for individual biomarkers. AIC = Akaike's information criterion; AUC = area under the curve; IDI = integrated discrimination improvement; NRI = category-less net reclassification; NTX-I = N-telopeptide of type I collagen; SBC = Schwarz Bayesian Criterion.
† Biomarkers selected were the same as model 1.
‡ Biomarkers selected were the same as model 2.
§ Adjusted for age, body mass index, sex, race, minimum joint space width at baseline, Western Ontario and McMaster Universities Osteoarthritis Index pain score at baseline, Kellgren/Lawrence score at baseline, and use of pain medications.
¶ Percentage of cases correctly reclassified = percentage of cases with a higher probability of being a case in a new model versus an old model.
# Percentage of controls correctly reclassified = percentage of controls with a lower probability of being a case in a new model versus an old model.
** Interpolated research value if below lower limit.

Table 3. Results of multivariable modeling, with 24-month change in biomarkers to predict radiographic and pain progression (n = 539)*

	Imaging biomarkers only			Imaging + biochemical markers		
	Model 1 (Selection method: stepwise; AIC)	Model 2 (Selection method: stepwise; SBC)	Model 3 (Selection method: stepwise; *P*)	Model 4 (Selection method: stepwise; AIC)	Model 5 (Selection method: stepwise; SBC)	Model 6 (Selection method: stepwise; *P*)
Model characteristics						
Unadjusted AUC	0.765	0.713	0.755	0.774	0.713	0.764
Adjusted AUC†	0.777	0.730	0.770	0.788	0.731	0.781
Adjusted AUC with 10-fold cross-validation	0.713	0.680	0.713	0.724	0.683	0.724
IDI‡	0.1612	0.1129	0.1510	0.1699	0.1145	0.1616
NRI‡	0.6611	0.5164	0.6951	0.6498	0.5482	0.6819
% of cases correctly reclassified§	23	10	26	25	15	26
% of controls correctly reclassified¶	43	41	43	40	40	42
Biomarkers included						
Change in SQ effusion-synovitis	0.0686	0.0026	0.0424	0.0481	0.0012	0.0310
No change versus improvement	1.17 (0.59, 2.32)	1.41 (0.74, 2.68)	1.33 (0.68, 2.62)	1.07 (0.54, 2.15)	1.36 (0.71, 2.60)	1.22 (0.62, 2.41)
Worsening versus improvement	2.01 (0.95, 4.27)	2.86 (1.41, 5.80)	2.28 (1.09, 4.78)	1.97 (0.93, 4.19)	2.96 (1.46, 6.02)	2.23 (1.06, 4.69)
Number of subregions with worsening in SQ cartilage thickness	0.0534	–	0.0391	0.0462	–	0.0357
1 versus 0	1.38 (0.81, 2.32)	–	1.48 (0.89, 2.47)	1.36 (0.80, 2.30)	–	1.45 (0.87, 2.43)
2 versus 0	1.44 (0.73, 2.84)	–	1.65 (0.86, 3.15)	1.42 (0.72, 2.79)	–	1.60 (0.84, 3.05)
3+ versus 0	3.36 (1.39, 8.13)	–	3.15 (1.35, 7.36)	3.53 (1.45, 8.61)	–	3.32 (1.41, 7.82)
Any regions with worsening SQ meniscal morphology	0.0084	0.0014	0.0098	0.0060	0.0027	0.0067
Yes versus no	2.30 (1.24, 4.27)	2.47 (1.42, 4.32)	2.18 (1.21, 3.94)	2.41 (1.29, 4.50)	2.33 (1.34, 4.05)	2.29 (1.26, 4.17)
Worsening in SQ Hoffa-synovitis score	0.0620	–	–	0.0694	–	–
Yes versus no	1.92 (0.97, 3.81)	–	–	1.90 (0.95, 3.80)	–	–
Number of subregions with worsening in SQ cartilage surface area across entire knee (include within-grade change)	0.1063	–	0.0330	0.0694	–	–
1 versus 0	1.35 (0.80, 2.28)	–	1.33 (0.79, 2.23)	1.90 (0.95, 3.80)	–	–
2 versus 0	1.77 (0.92, 3.43)	–	2.00 (1.05, 3.82)	0.1287	–	0.0512
3+ versus 0	2.27 (1.13, 4.56)	–	2.50 (1.27, 4.91)	1.41 (0.83, 2.38)	–	1.39 (0.82, 2.34)
Maximum change in osteophyte score of ≥1 across all areas of the knee	0.0423	–	–	0.0523	–	–
Yes versus no	0.51 (0.27, 0.98)	–	–	0.52 (0.27, 1.01)	–	–
Change in mean quantitative cartilage thickness of the central medial femur (center), Z score	0.0581	0.0001	–	0.0711	0.0001	–
OR for each 1–unit increase in SD#	1.30 (0.99, 1.71)	1.56 (1.25, 1.96)	–	1.29 (0.98, 1.70)	1.56 (1.24, 1.95)	–
Change in quantitative medial tibial cartilage volume, Z score	0.0278	–	0.0071	0.0152	–	0.0034

(Continued)

Table 3. (*Cont'd*)

	Imaging biomarkers only			Imaging + biochemical markers		
	Model 1 (Selection method: stepwise; AIC)	Model 2 (Selection method: stepwise; SBC)	Model 3 (Selection method: stepwise; *P*)	Model 4 (Selection method: stepwise; AIC)	Model 5 (Selection method: stepwise; SBC)	Model 6 (Selection method: stepwise; *P*)
OR for each 1-unit increase in SD#	1.29 (1.03, 1.63)	–	1.36 (1.09, 1.69)	1.33 (1.06, 1.68)	–	1.40 (1.12, 1.75)
Change in quantitative lateral patellofemoral region on femur area, Z score	0.0785	–	0.0164	0.0896	–	0.0209
OR for each 1-unit increase in SD	1.25 (0.98, 1.59)	–	1.33 (1.05, 1.68)	1.24 (0.97, 1.59)	–	1.32 (1.04, 1.67)
Urine NTX-I, Z score**	–	–	–	0.0094	–	0.0063
OR for each 1-unit increase in SD	–	–	–	1.34 (1.07, 1.67)	–	1.35 (1.09, 1.68)
Serum NTX-I, Z score**	–	–	–	–	0.0057	–
OR for each 1-unit increase in SD	–	–	–	–	1.32 (1.08, 1.61)	–
Intercept (horizontal) TBT parameter, Z score	0.0021	0.0161	0.0011	0.0020	–	0.0013
OR for each 1-unit increase in SD	1.48 (1.15, 1.89)	1.30 (1.05, 1.62)	1.50 (1.18, 1.92)	1.48 (1.15, 1.90)	–	1.50 (1.17, 1.92)
Slope (horizontal) TBT parameter, Z score	0.0080	–	0.0070	0.0036	–	0.0031
OR for each 1-unit increase in SD	0.73 (0.58, 0.92)	–	0.73 (0.58, 0.92)	0.70 (0.55, 0.89)	–	0.70 (0.56, 0.89)

* Except where indicated otherwise, values are the odds ratios (ORs) and 95% confidence intervals. *P* values are shown for individual biomarkers. Time-integrated concentrations were used for biochemical and trabecular bone texture biomarkers. AIC = Akaike's information criterion; AUC = area under the curve; IDI = integrated discrimination improvement; NRI = category-less net reclassification; NTX-I = N-telopeptide of type I collagen; SBC = Schwarz Bayesian Criterion; SQ = semiquantitative; TBT = trabecular bone texture.
† Adjusted for age, body mass index, sex, race, minimum joint space width at baseline, Western Ontario and McMaster Universities Osteoarthritis Index pain score at baseline, Kellgren/Lawrence score at baseline, and use of pain medications.
‡ Compared to covariates-only model.
§ Percentage of cases correctly reclassified = percentage of cases with a higher probability of being a case in a new model versus an old model.
¶ % controls correctly reclassified = % of controls with a lower probability of being a case in a new model versus an old model.
# Coded such that increasing OR = increasing change.
** Interpolated research value if below lower limit.

Sensitivity analyses. *Change in biomarkers over 24 months predicting pain and radiographic progression over 48 months (absolute change used for biochemical markers and TBT).* Compared to the model using TICs for biochemical markers and TBT, the same selection of imaging markers was associated with case status, with the main difference being that no biochemical marker or TBT parameter was selected when absolute change in these markers was used (Supplementary Table 4, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24557). The adjusted 10-fold cross-validated AUCs were slightly lower, ranging from 0.668 to 0.700.

Baseline biomarkers predicting radiographic progression over 48 months. The number of locations affected by semiquantitative osteophytes, medial meniscus volume and quantitative cartilage thickness at the central lateral femur (internal), medial tibia (external), and lateral tibia (posterior) were associated with case status in all 3 models (Supplementary Table 5, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24557). Semiquantitative cartilage morphology (maximum full-thickness cartilage loss score) and semiquantitative Hoffa-synovitis score were included in 2 of the 3 models. The adjusted 10-fold cross-validated AUCs, using imaging markers only, ranged from 0.716 to 0.723. When biochemical markers were added, AUCs were 0.716 to 0.732. The same imaging markers were selected, with the addition of urinary C-propeptide of type II collagen (CTX-II) and serum N-propeptide of type IIA collagen (PIIANP), in 2 of the 3 models.

Change in biomarkers over 24 months predicting radiographic progression over 48 months. The adjusted 10-fold cross-validated AUCs were higher in the models predicting radiographic progression only (AUCs OF 0.793 to 0.832) compared to the models using pain and radiographic progression as the outcome (Supplementary Table 6, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24557). A different set of biomarkers was associated with progression in all 3 models that included imaging and biochemical markers: the number of areas of cartilage damage with worsening in surface area (semiquantitative MRI), worsening meniscus extrusion (semiquantitative), reduction in mean cartilage thickness at the central medial femur (quantitative MRI), and decrease in serum PIIANP. Several other markers were found to be significant in models 1 (AIC) and/or 3 (*P* value), including measures of bone shape and area, quantitative cartilage thickness and volume, semiquantitative effusion-synovitis, semiquantitative cartilage and meniscal damage, the number of locations with osteophytes (semiquantitative), and serum NTX-1 and CTX-1.

Baseline and change in biomarkers predicting pain and radiographic progression, excluding TBT parameters. The results of the analysis using baseline biomarkers as predictors were consistent overall with the main analysis including TBT parameters, with 3 main exceptions as follows: 1) Hoffa-synovitis score was not significant in any model, 2) medial meniscus volume was significant in all models, and 3) urinary CTX-II was associated with case status in the models that used AICs and *P* values. The AUCs ranged from 0.668 to 0.694 (Supplementary Table 7). In the 24-month change analysis, the imaging markers were overall consistent with the original analysis; however, a different biochemical marker was significant in all models: serum CTX-I (Supplementary Table 8, available at *the Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24557). AUCs were similar compared to the main analysis.

DISCUSSION

The baseline biomarkers that predicted subsequent pain and radiographic progression in most models were the number of locations affected by osteophyte (semiquantitative MRI), quantitative central medial femoral and central lateral femoral cartilage thickness, patellar bone shape, and semiquantitative Hoffa-synovitis score. Only the number of locations affected by semiquantitative osteophytes and patella shape were significantly associated with case status in all models. The 24-month change in biomarkers that predicted pain and radiographic progression in all models were as follows: worsening in semiquantitative effusion-synovitis (versus improvement), increase in the number of knee regions that had worsening in semiquantitative meniscal damage, and the horizontal TBT (intercept term). An increase in the number of areas with worsening semiquantitative cartilage morphology, loss of quantitative cartilage thickness in the central medial femur (center), loss of cartilage volume in the medial tibial, and change in quantitative lateral patellofemoral bone area were significantly associated with case status in 2 of the models. For TBT parameters and biochemical markers, 24-month TIC values performed better than change scores. The fact that the strongest biochemical predictor in univariate analysis in this cohort, urinary CTX-II, did not contribute to model predictions containing the core set of cartilage MRI markers, suggests its collinearity with these imaging parameters, which is in line with previous studies (29,30). The overall AUCs were similar with the addition of the biochemical markers as compared to the earlier models with the core set of MRI markers only (adjusted AUCs with 10-fold cross-validation of 0.669 versus 0.671 in the baseline analysis and 0.713 versus 0.724 in the 24-month change analysis for models 1 and 4, respectively).

Higher AUCs yielded by a different set of imaging and biochemical markers were found in the secondary analysis to predict radiographic progression only. This is important since surrogate end points such as radiographic progression might in theory be accepted by the US Food and Drug Administration for the initial drug approval of a disease-modifying OA agent, although post-marketing studies showing benefits on clinically important outcomes would be needed (31). Imaging and biochemical markers of structural progression are objective and more fully developed than biomarkers of pain, which to date are largely subjective and

self-reported measures. A recent genome-wide association study of knee pain identified *GDF5* as the primary locus (32). *GDF5* is the same gene most strongly and repeatedly associated with OA based on structural diagnoses. Therefore, it may not be a different underlying pathologic process driving symptom and structural progression, but instead our ability to measure them with adequate sensitivity.

Although synovitis (Hoffa-synovitis or effusion-synovitis) was consistently selected in all models, this study demonstrates that the other biomarkers that predict progression vary dependent on whether baseline biomarkers or changes in biomarkers over 24 months are evaluated for their ability to predict longer-term (48 month) outcomes of radiographic and symptomatic progression. Both biomarker types may be useful for the same clinical trial, but with different purposes, namely 1) participant selection for inclusion (baseline biomarkers) and 2) structural end point (longitudinal change). These could be particularly important in enhancing the efficiency and shortening the duration of phase II and III clinical trials, thereby reducing the cost of conducting these trials, increasing the likelihood of drug approval (6), and improving the time to market (33).

Other studies have also developed models to predict OA progression using baseline and/or longitudinal biomarker data. A recent study has used a machine-learning approach in the same FNIH data set to identify differences in a variety of baseline characteristics between progressors and nonprogressors (34). Similar to the present study, the number of locations with osteophytes was a strong contributor to the progressor phenotype, which supports previous findings showing the role of osteophytes in OA progression (35). Bone marrow lesions and urine CTX-II were also highlighted as prognostic biomarkers, which is in line with our other findings, although bone marrow lesions were not included in the final multivariable model. However, synovitis did not differentiate progressors and nonprogressors in that study despite robust evidence indicating that inflammation plays an important role in OA progression (36). It is of note that their control definition was different (knees with neither clinical nor radiographic progression).

In the present work, we utilized logistic regression because our focus was not only on the variable selection, but also on computing interpretable effect estimates (i.e., odds ratios) for each parameter. Another study tested different models to predict moderate-to-severe OA (clinical and/or radiographic) over 8 years and found that adding MRI biomarkers (cartilage morphology and T2 and meniscal tear) significantly improved the prognostic ability of the model compared to clinical and radiographic characteristics only (37). The AUCs were similar to this study (0.71–0.72 for the models including biomarkers). We have used a shorter follow-up period of 4 years in order to make the results more informative for clinical trial design. Although OA progression is typically slow, a large epidemiologic study has shown that radiographic progression over a 5-year interval occurs in 12–23% of knees with radiographic OA (38). Other more sophisticated methodologic approaches have also been tested, such as different machine learning and regression algorithms, but to date, no prediction model has been sufficiently validated and qualified for use in trials (39,40).

There are a few limitations of the present study. First, these analyses were performed on a subsample of the FNIH cohort for which all biomarkers were available; missing data were largely related to missing TBT biomarkers, mostly secondary to poor radiographic positioning. Sensitivity analyses excluding TBT showed similar results for the imaging markers and a different selection of biochemical markers. Second, the results may not be generalizable to race/ethnicities not represented in the OAI, which mostly included White participants. Third, there are no reproducibility data for meniscal volume on scan-rescan images. Fourth, the analyses were conducted first with imaging parameters only, with subsequent addition of biochemical biomarkers; as the order of addition can affect the incremental explanatory power of the variable, results could vary with a different approach. In addition, participants with radiographic and pain progression by 12 months were excluded from analysis, which may have excluded a small number of cases with very fast progression. It is also worth noting that the control definition used in the main analysis included knees with pain only and radiographic only progression, which may have reduced the strength of the association between biomarkers and case status. The approach we used has been predefined for the overall FNIH project and used in previous work studying individual biomarker domains (9–13). Finally, we did not explicitly control for multiple testing. Instead, we sought to examine the robustness of the models by comparing the variables selected across the different selection methods. Machine learning methods that can assess enormous numbers of predictors could be an alternative strategy to variable selection and model fitting.

In conclusion, our study highlights the combination of biomarkers that could provide prognostic utility in the context of OA disease-modifying trials. At baseline, semiquantitative imaging markers (osteophytes and Hoffa-synovitis score) and quantitative (cartilage thickness and patella shape) imaging markers were selected. Different biomarkers were selected in the 24-month change analysis including semiquantitative measures (effusion-synovitis, meniscal morphology, and cartilage morphology) and quantitative measures of cartilage thickness and volume, radiographic TBT, and urinary or serum NTX-I. Phase II of the OA Biomarkers Consortium is currently underway to externally validate the present findings and enable the submission of these biomarkers for regulatory review and formal qualification for use as prognostic biomarkers in disease-modifying OA trials.

ACKNOWLEDGMENTS

We thank the Osteoarthritis Research Society International organization for their leadership and expertise on the FNIH OA Biomarker Consortium project.

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. Drs. Hunter and Collins had full access to all data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Hunter, Collins, Losina, Nevitt, Katz, Kraus.

Acquisition of data. Collins, Nevitt, Roemer, Guermazi, Bowes, Dam, Eckstein, Lynch, Kraus.

Analysis and interpretation of data. Hunter, Deveza, Collins, Losina, Nevitt, Roemer, Guermazi, Bowes, Dam, Eckstein, Lynch, Katz, Kwoh, Hoffmann.

REFERENCES

1. Matthews GL, Hunter DJ. Emerging drugs for osteoarthritis. Expert Opin Emerg Drugs 2011;16:479–91.
2. Hochberg MC, Guermazi A, Guehring H, Aydemir A, Wax S, Fleuranceau-Morel P, et al. Effect of intra-articular sprifermin vs placebo on femorotibial joint cartilage thickness in patients with osteoarthritis: the FORWARD randomized clinical trial. JAMA 2019;322:1360–70.
3. Conaghan PG, Bowes MA, Kingsbury SR, Brett A, Guillard G, Rizoska B, et al. Disease-modifying effects of a novel cathepsin K inhibitor in osteoarthritis: a randomized, placebo-controlled study. Ann Intern Med 2020 21;172:86–95.
4. Weinstein AM, Rome BN, Reichmann WM, Collins JE, Burbine SA, Thornhill TS, et al. Estimating the burden of total knee replacement in the United States. J Bone Joint Surg Am 2013;95:385–92.
5. Lotz M, Martel-Pelletier J, Christiansen C, Brandi ML, Bruyere O, Chapurlat R, et al. Value of biomarkers in osteoarthritis: current status and perspectives. Ann Rheum Dis 2013;72:1756–63.
6. Thomas DW, Burns J, Audette J, Carroll A, Dow-Hygelund C, Hay M. Clinical development success rates 2006–2015. BIO Industry Analysis 2016;1:1–26.
7. Hunter DJ, Losina E, Guermazi A, Burstein D, Lassere MN, Kraus V. A pathway and approach to biomarker validation and qualification for osteoarthritis clinical trials. Curr Drug Targets 2010;11:536–45.
8. Hunter DJ, Nevitt M, Losina E, Kraus V. Biomarkers for osteoarthritis: current position and steps towards further validation. Best Pract Res Clin Rheumatol 2014;28:61–71.
9. Eckstein F, Collins JE, Nevitt MC, Lynch JA, Kraus VB, Katz JN, et al. Cartilage thickness change as an imaging biomarker of knee osteoarthritis progression: data from the Foundation for the National Institutes of Health Osteoarthritis Biomarkers Consortium. Arthritis Rheumatol 2015;67:3184–9.
10. Hunter D, Nevitt M, Lynch J, Kraus VB, Katz JN, Collins JE, et al. Longitudinal validation of periarticular bone area and 3D shape as biomarkers for knee OA progression? Data from the FNIH OA Biomarkers Consortium. Ann Rheum Dis 2016;75:1607–14.
11. Collins JE, Losina E, Nevitt MC, Roemer FW, Guermazi A, Lynch JA, et al. Semiquantitative imaging biomarkers of knee osteoarthritis progression: data from the Foundation for the National Institutes of Health Osteoarthritis Biomarkers Consortium. Arthritis Rheumatol 2016;68: 2422–31.
12. Kraus VB, Collins JE, Charles HC, Pieper CF, Whitley L, Losina E, et al. Predictive validity of radiographic trabecular bone texture in knee osteoarthritis: the Osteoarthritis Research Society International/ Foundation for the National Institutes of Health Osteoarthritis Biomarkers Consortium. Arthritis Rheumatol 2018;70:80–7.
13. Kraus VB, Collins JE, Hargrove D, Losina E, Nevitt M, Katz JN, et al. Predictive validity of biochemical biomarkers in knee osteoarthritis: data from the FNIH OA Biomarkers Consortium. Ann Rheum Dis 2017;76:186–95.
14. Neumann G, Hunter D, Nevitt M, Chibnik LB, Kwoh K, Chen H, et al. Location specific radiographic joint space width for osteoarthritis progression. Osteoarthritis Cartilage 2009;17:761–5.
15. Ornetti P, Brandt K, Hellio-Le Graverand MP, Hochberg M, Hunter DJ, Kloppenburg M, et al. OARSI-OMERACT definition of relevant radiological progression in hip/knee osteoarthritis. Osteoarthritis Cartilage 2009;17:856–63.
16. Angst F, Aeschlimann A, Michel BA, Stucki G. Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. J Rheumatol 2002;29:131–8.
17. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. Osteoarthritis Cartilage 2007;15 Suppl A:A1–56.
18. Peterfy CG, Schneider E, Nevitt M. The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee. Osteoarthritis Cartilage 2008;16:1433–41.
19. Roemer FW, Guermazi A, Collins JE, Losina E, Nevitt MC, Lynch JA, et al. Semi-quantitative MRI biomarkers of knee osteoarthritis progression in the FNIH biomarkers consortium cohort – Methodologic aspects and definition of change. BMC Musculoskelet Disord 2016;17:466.
20. Peterfy CG, Guermazi A, Zaim S, Tirman PF, Miaux Y, White D, et al. Whole-Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis. Osteoarthritis Cartilage 2004;12:177–90.
21. Biswal S, Hastie T, Andriacchi TP, Bergman GA, Dillingham MF, Lang P. Risk factors for progressive cartilage loss in the knee: a longitudinal magnetic resonance imaging study in forty-three patients. Arthritis Rheumatol 2002;46:2884–92.
22. Hunter D, Gale D, Grainger G, Lo G, Conaghan P. The reliability of a new scoring system for knee osteoarthritis MRI and the validity of bone marrow lesion assessment: BLOKS (Boston Leeds Osteoarthritis Knee Score). Ann Rheum Dis 2008;67:206–11.
23. Kornaat PR, Ceulemans RY, Kroon HM, Riyazi N, Kloppenburg M, Carter WO, et al. MRI assessment of knee osteoarthritis: Knee Osteoarthritis Scoring System (KOSS)—inter-observer and intra-observer reproducibility of a compartment-based scoring system. Skeletal Radiology 2005;34:95–102.
24. Hunter DJ, Guermazi A, Lo GH, Grainger AJ, Conaghan PG, Boudreau RM, et al. Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). Osteoarthritis Cartilage 2011;19:990–1002.
25. Dam EB, Lillholm M, Marques J, Nielsen M. Automatic segmentation of high- and low-field knee MRIs using knee image quantification with data from the osteoarthritis initiative. J Med Imaging (Bellingham) 2015;2:024001.
26. Kraus VB, Feng S, Wang S, White S, Ainslie M, Brett A, et al. Trabecular morphometry by fractal signature analysis is a novel marker of osteoarthritis progression. Arthritis Rheumatol 2009;60:3711–22.
27. Gu W, Pepe M. Measures to summarize and compare the predictive capacity of markers. Int J Biostat 2009;5:27.
28. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 2008;27:157–72.
29. Eckstein F, Le Graverand MP, Charles HC, Hunter DJ, Kraus VB, Sunyer T, et al. Clinical, radiographic, molecular and MRI-based predictors of cartilage loss in knee osteoarthritis. Ann Rheum Dis 2011; 70:1223–30.
30. Deveza LA, Kraus VB, Collins JE, Guermazi A, Roemer FW, Bowes M, et al. Association between biochemical markers of bone turnover and bone changes on imaging: data from the Osteoarthritis Initiative. Arthritis Care Res (Hoboken) 2017;69:1179–91.
31. Kraus VB, Simon LS, Katz JN, Neogi T, Hunter D, Guermazi A, et al. Proposed study designs for approval based on a surrogate endpoint and a post-marketing confirmatory study under FDA's accelerated approval regulations for disease modifying osteoarthritis drugs. Osteoarthritis Cartilage 2019;27:571–9.

32. Meng W, Adams MJ, Palmer CNA, Agee M, Alipanahi B, Bell RK, et al. Genome-wide association study of knee pain identifies associations with *GDF5* and *COL27A1* in UK Biobank. Commun Biol 2019;2:321.
33. Berndt ER, Gottschalk AH, Strobeck MW. Opportunities for improving the drug development process: results from a survey of industry and the FDA. URL: https://www.journals.uchicago.edu/doi/abs/10.1086/ipe.6.25056181.
34. Nelson AE, Fang F, Arbeeva L, Cleveland RJ, Schwartz TA, Callahan LF, et al. A machine learning approach to knee osteoarthritis phenotyping: data from the FNIH Biomarkers Consortium. Osteoarthritis Cartilage 2019;27:994–1001.
35. van der Kraan PM, van den Berg WB. Osteophytes: relevance and biology. Osteoarthritis Cartilage 2007;15:237–44.
36. Wang X, Hunter DJ, Jin X, Ding C. The importance of synovial inflammation in osteoarthritis: current evidence from imaging assessments and clinical trials. Osteoarthritis Cartilage 2018;26:165–74.
37. Joseph GB, McCulloch CE, Nevitt MC, Neumann J, Gersing AS, Kretzschmar M, et al. Tool for osteoarthritis risk prediction (TOARP) over 8 years using baseline clinical data, X-ray, and MRI: data from the Osteoarthritis Initiative. J Magn Reson Imaging 2018;47:1517–26.
38. Leyland KM, Hart DJ, Javaid MK, Judge A, Kiran A, Soni A, et al. The natural history of radiographic knee osteoarthritis: a fourteen-year population-based cohort study. Arthritis Rheumatol 2012;64:2243–51.
39. Widera P, Welsing PM, Ladel C, Loughlin J, Lafeber FP, Petit Dop F, et al. Multi-classifier prediction of knee osteoarthritis progression from incomplete imbalanced longitudinal data. Sci Rep 2020;10:8427.
40. Halilaj E, Le Y, Hicks JL, Hastie TJ, Delp SL. Modeling and predicting osteoarthritis progression: data from the osteoarthritis initiative. Osteoarthritis Cartilage 2018;26:1643–50.

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1154–1162
DOI 10.1002/acr.24549

American College of Rheumatology
Empowering Rheumatology Professionals

Concept End Points Informing Design Considerations for Confirmatory Clinical Trials in Osteoarthritis

Yura Kim, Gregory Levin, Nikolay P. Nikolov, Robert Abugov, and Rebecca Rothwell

Objective. There is an unmet need for therapies that target the underlying pathophysiology of osteoarthritis (OA). However, defining appropriate measures for clinical trials of such therapies is challenging. Our objective was to propose concept clinical end points that directly capture clinical benefit in this setting and evaluate the feasibility of their use.

Methods. This analysis used the multicenter, longitudinal, observational Osteoarthritis Initiative (OAI) database. OAI participants primarily had knee OA, with follow-up of up to 9 years and assessments of joints, surgical interventions, performance outcomes, and patient-reported outcomes. We examined this data set to identify existing outcome measures of direct clinical benefit. We evaluated the feasibility of conducting trials using these candidate end points by estimating incidence rates and resulting required sample sizes and study durations in time-to-event analyses.

Results. We identified candidate end points based on total knee replacement (TKR) and composite end points defined by TKR and conservative thresholds of patient-reported outcomes of pain and function. Using time to TKR as an end point, a study with an average follow-up time of 3 years requires approximately 3,000 to 18,000 subjects, depending on effect size. Alternatively, for a composite end point, such as "time to TKR or severe pain or severely impaired functioning," the required sample sizes ranged from approximately 2,000 to 11,000 for a 3-year study.

Conclusion. The proposed concept end points can reliably and feasibly evaluate the effectiveness of therapies for this unmet need. In particular, the composite end point approach can substantially reduce sample sizes (up to approximately 40%) compared to the use of TKR alone.

INTRODUCTION

Osteoarthritis (OA) is a joint disease, affecting over 250 million people worldwide (1), with currently rising prevalence rates. Knee OA is the most prevalent form, representing approximately one-third of OA cases in the United States (2).

As the disease progresses, many patients can have substantially impaired functioning, often ultimately leading to disability and the need for joint replacement surgery. However, currently available approved drugs treat only short-term symptoms of OA, primarily pain and function, and do not target the underlying causes or long-term progression of the disease. Therefore, there is an unmet need for drugs to alter the underlying disease process. One approach to development of such drugs discussed in the literature is to use an assessment of structural changes in the joint based on imaging as a primary end point (e.g., radiograph and/or magnetic resonance imaging measures of cartilage thickness/catabolism/anabolism, pathologic remodeling of subchondral bone, or synovial inflammation) (3–5). However, as noted in the recent US Food and Drug Administration draft guidance (6), "the ability of treatment effects on common measures of structural progression to reliably predict treatment effects

This article was prepared using an Osteoarthritis Initiative (OAI) public-use data set, and its contents do not necessarily reflect the opinions or views of the OAI Study Investigators, the NIH, or the private funding partners of the OAI. The OAI is a public–private partnership between the NIH (contracts N01-AR-2-2258, N01-AR-2-2259, N01-AR-2-2260, N01-AR-2-2261, and N01-AR-2-2262) and private funding partners (Merck Research Laboratories, Novartis Pharmaceuticals, GlaxoSmithKline, and Pfizer, Inc.) and is conducted by the OAI Study Investigators. Private sector funding for the OAI is managed by the Foundation for the NIH. The authors of this article are not part of the OAI investigative team.

This article reflects the views of the authors and should not be construed to represent views or policies of the US Food and Drug Administration.

Supported by an appointment to the Research Participation Program at the US Food and Drug Administration, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the US Department of Energy and the US Food and Drug Administration.

Yura Kim, PhD, Gregory Levin, PhD, Nikolay P. Nikolov, MD, Robert Abugov, PhD, Rebecca Rothwell, PhD: US Food and Drug Administration, Silver Spring, Maryland.

No potential conflicts of interest relevant to this article were reported.

Address correspondence to Yura Kim, PhD, 10903 New Hampshire Avenue, Silver Spring, MD 20993. Email: Yura.Kim@fda.hhs.gov.

Submitted for publication March 11, 2020; accepted in revised form December 17, 2020.

SIGNIFICANCE & INNOVATIONS

- Clinical studies for products intended to treat the underlying disease process of osteoarthritis (OA) will require a larger sample size and longer duration than historical OA clinical trials.
- Use of a composite end point (e.g., time to total knee replacement and surpassing a threshold on patient-reported outcomes of pain or disability) can improve feasibility of osteoarthritis drug trials of products intended to alter the underlying pathophysiology of the disease by increasing the background incidence rate.
- A composite end point based on these direct measures of how a patient functions, feels, or survives, provides a more clinically relevant approach to OA drug development while mitigating issues of external factors affecting the incidence of total knee replacement surgery.

on direct measures of how patients function and feel has not been established." An alternative, and more clinically relevant, approach to development of drugs targeting the underlying pathophysiology is to use a long-term clinical end point, i.e., a direct measure of how patients function, feel, or survive (7), as the primary end point. While the term "survive" is referring to literal survival of the person in the definition of a clinical end point, in the case of OA, the term may also be interpreted as the survival of the joint.

However, there have been concerns that using such end points would not be feasible in a clinical trial to support drug approval. In this article, we consider several concept clinical end points for a long-term OA trial and calculate sample sizes needed to detect clinically relevant treatment effects, in order to inform considerations regarding the appropriate design of future OA drug trials of products intended to alter the underlying pathophysiology of the disease.

To explore potential end points, we used data from the publicly available Osteoarthritis Initiative (OAI) (8), a multicenter, longitudinal, observational study of 4,796 subjects, primarily focusing on knee OA. Participants were followed for up to 9 years, with at least annual evaluations of patient-reported symptoms, physician assessments, surgical status (e.g., total knee replacement [TKR]), imaging measures, and biochemical markers.

After reviewing the measurements collected in the OAI, we focused specifically on surgical status (i.e., TKR) and long-term patient-reported outcomes of disability and pain (i.e., Western Ontario and McMaster Universities Osteoarthritis Index [WOMAC] subscales) because these provide direct measures of how a patient functions, feels, or survives. More specifically, we evaluated the feasibility of using a composite end point: time to first TKR or severe pain or severely impaired functioning, defined by passing a clinically appropriate threshold in appropriate patient-reported outcomes, such as WOMAC pain or disability.

PATIENTS AND METHODS

Study population. The OAI is a multicenter, longitudinal, observational study intended to improve understanding on how to prevent and treat knee OA. The OAI study population included 4,796 women and men, ages 45–79 years. Patients were categorized into two primary cohorts, referred to as the progression cohort and the incidence cohort. The progression cohort included subjects who had frequent knee symptoms and radiographic evidence of tibiofemoral knee OA (definite tibiofemoral osteophytes measured as Osteoarthritis Research Society International atlas grades 1–3, equivalent to Kellgren/Lawrence grade ≥2 on the fixed flexion radiograph). The incidence cohort included subjects who had eligibility risk factors of knee OA, which included knee symptoms, frequent use of medications, being overweight, knee injury/surgery, and family history. The number of risk factors required for inclusion in this cohort depended on the patient's age. The study enrolled 3,284 participants in the incidence cohort and 1,390 participants in the progression cohort. An additional 122 participants were enrolled as healthy controls.

We primarily focused on the progression cohort participants, who had symptomatic knee OA with radiographic evidence of OA at baseline, because this cohort is most representative of subjects who would be expected to be included in clinical trials to evaluate potentially disease-modifying OA drugs. Among the 1,390 progression cohort participants, 474 had OA in the right knee but not the left knee at baseline (right-knee-affected), 427 had OA in the left knee but not the right knee at baseline (left-knee-affected), and 489 had OA in both knees at baseline (both-knees-affected).

Although participants in the OAI were followed for up to 9 years, for the analyses in this article, we focused on each subject's initial 5 years of follow-up data, which is considered a more reasonable study duration for a potential confirmatory clinical trial. Approximately 80% of progression cohort participants had follow-up of at least 60 months (Table 1).

Selected candidate outcome measures. Among the measurements collected in the OAI, we selected TKR and WOMAC subscales for further investigation. Total replacement of the joint, including TKR, has been previously proposed as a primary outcome in OA randomized trials (9–11) because, as

Table 1. Proportion of participants who completed the follow-up*

Month of follow-up visit								
12	24	36	48	60	72	84	96	108
95.8	92.2	88.9	86.5	79.8	77.7	76.5	73.3	64.3

* Values are the percentage.

opposed to some proposed imaging-based candidate surrogate end points, TKR is a direct measure of the survival of the joint. In addition, as a major surgery, it involves some risk and considerable patient recovery time and effort. Total replacement is generally considered when all available nonsurgical treatments have been exhausted and is recommended to patients with severe symptoms (11). Total replacement has also been shown to effectively reduce joint pain and disability in most cases (11). However, there have been questions regarding the feasibility of conducting trials with this end point due to the perceived low incidence rate (11). Furthermore, there are additional factors beyond pain and function (e.g., race, sex, socioeconomic status, access to care, surgeon preference, and health care systems) that may limit access to TKR (12–19), reducing the interpretability of the estimated treatment effect.

In the OAI, the date of joint replacement surgery and the type of surgery (partial or total) were provided. We also considered patient-reported outcomes, specifically the WOMAC index (20). The WOMAC index was developed for hip and knee OA and has 3 subscales of pain, stiffness, and functional disability. WOMAC pain and disability have been used as both primary end points (21–24) and secondary end points (25,26) in previous OA trials. Questionnaire items are scored on a scale of 0–4, which correspond to: none = 0, mild = 1, moderate = 2, severe = 3, and extreme = 4, with the response based on the past 7-day period. In the OAI, the questionnaire items evaluate each knee separately, asking whether the subject has pain, stiffness, or functional difficulty performing an activity with "the target knee." The scores for each subscale are summed, and a total WOMAC score is obtained by summing across subscales. Higher scores on each WOMAC subscale indicate worse pain, stiffness, and functional limitations.

In the OAI database, all 3 WOMAC subscales were provided. For the purposes of this article, based on clinical considerations, we focused on the disability and pain subscales. WOMAC disability consists of 17 questionnaire items, with a possible total score range of 0–68. Each item asks whether a person has difficulty doing a specific activity (Table 2). Thus, a threshold of 51 (a score of 3 representing "severe difficulty" × 17 items), represents a subject with severe disability in all categories or extreme disability in some categories.

WOMAC pain consists of 5 questionnaire items, with a possible score range of 0–20. Each item asks whether a person has pain doing a specific activity (Table 3). Thus, a threshold of 15 (a score of 3 representing "severe pain" × 5 items), represents a subject with severe pain in all categories or extreme pain in some categories.

Table 2. Items in the WOMAC disability score*

Down stairs	Walking	Lying down
Up stairs	In car/out of car	Get in/out of bathtub
Stand from sitting	Shopping	On/off toilet
Standing	Socks on	Heavy chores
Sitting	Socks off	Light chores
Bending	Get out of bed	–

* WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index.

Table 3. Items in the WOMAC pain score*

Walking	In bed	Standing
Stairs	Sit or lie down	–

* WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index.

In the OAI, TKR status and WOMAC measurements were available for each knee. For our analyses, we used the subject-specific measurement based on the knee that was affected at baseline if only 1 knee was affected or based on the worst outcome if both knees were affected at baseline. More specifically, for the right-knee-affected/left-knee-affected participants, we used the time to right/left knee TKR, respectively. For the both-knees-affected participants, we used the time to first knee (either knee) TKR. Similarly, for WOMAC subscales, in the right-knee-affected/left-knee-affected participants, we used the right/left knee scores, respectively. For the both-knees-affected participants, we used the worst knee scores at each follow-up time point. We chose this approach to mirror potential use in clinical trials. In such trials, only the disease-affected knee (i.e., the target knee) is likely to be treated, and its clinical status will be of primary interest. Under other scenarios, different definitions of target knees for patient-reported outcomes or TKR may be more appropriate.

Clinical end points. In this study, we considered 7 concept clinical end points, each defined as time to a certain event: 1) time to TKR, 2) time to TKR or WOMAC disability of ≥51, 3) time to TKR or WOMAC pain of ≥15, 4) time to TKR or WOMAC disability of ≥51 in 2 consecutive visits, 5) time to TKR or WOMAC pain of ≥15 in 2 consecutive visits, 6) time to TKR or WOMAC disability of ≥51 *or* WOMAC pain of ≥15, and 7) time to TKR or WOMAC disability of ≥43 *and* WOMAC pain of ≥13.

We evaluated the potential use of a threshold with the patient-reported outcomes to identify severe pain or disability. Thresholds have been proposed previously when considering patient-reported outcomes for OA (27–29). In particular, Gossec et al (2011) and Dougados et al (2009) attempted to identify thresholds that would correspond to the orthopedic surgeon's recommendation for joint replacement surgery. Patients meeting their criteria were considered as having a "virtual joint replacement." Our goal and focus of using thresholds differ from these studies. The thresholds we considered in this article were selected to identify severe levels of pain and functional debilitation, not to predict future knee replacement surgeries. Therefore, the thresholds used were chosen by examining the items and scorings of these patient-reported outcomes.

The thresholds of 51 for WOMAC disability and 15 for WOMAC pain were based on aforementioned thresholds

corresponding to severe disability and severe pain. Therefore, end points 2 and 3 correspond to time to first TKR or severe disability and time to first TKR or severe pain, respectively.

End points 4 and 5 require repeated pain or function measurements beyond a certain threshold, to ensure that severe pain or loss of function is stable and represents chronic pain or disability. We note that the incidence of such an event might be greater in a study with more frequent follow-up visits than the OAI (which had only annual WOMAC assessments). For example, Manno et al (2012) considered the duration of symptoms with consecutive visits, where consecutive visits were 3–6 months apart.

End point 6 takes into consideration impacts on either severe disability *or* pain. Having no event with respect to this end point would suggest that the subject had no TKR, had no severe disability, and had no severe pain within the follow-up period. End point 7 allows for requiring *both* disability and pain. Because of this combined requirement, we considered a lower threshold for each subscale to be reasonable. Furthermore, noting that some item tasks (such as sitting and lying down) may not be expected to be impacted as directly by knee OA, a less stringent threshold may also be appropriate. Therefore, rather than using a threshold of 51 for disability and 15 for the pain, here we considered 43 and 13, respectively. These thresholds would correspond to half severe and half moderate responses (or some combination with extreme responses).

Evaluation of feasibility and application in clinical trials. For each candidate end point, we calculated the survival probability function and cumulative hazard function from the OAI data set. In each case, we calculated the incidence rate of each end point–defined event. We also examined the distribution of event times to determine whether the hazard rate was approximately constant to justify basing a sample size calculation on an assumed constant incidence rate.

Incidence rates were calculated based on 1,819 days (5-year follow-up), by dividing the total number of each event by the sum of person-time at risk. Person-time at risk for each individual was either the time when censored (for those who dropped out before an event happened), the time of an event (for those who had an event within 1,819 days), or 1,819 days (for those who remained through the follow-up period without an event).

Using these incidence rates, for each end point we calculated the sample size that would be required in future clinical trials to detect different magnitudes of treatment effects with 80% power and a 5% 2-sided type 1 error probability, assuming a parallel-group design with 1:1 randomization to the treatment group and control group. We further assumed: 1) a constant hazard rate over time for each group, which was justified through the estimated hazard functions, and 2) equal average follow-up time for the control group and treatment group.

RESULTS

TKR. In 1,332 participants in the progression cohort of the OAI who had postbaseline observations, there was a total of 138 TKRs within 5 years. Among 456 right-knee-affected participants who had postbaseline observations, 44 had a right knee TKR within 5 years. Among 413 left-knee-affected participants who had postbaseline observations, 39 had a left knee TKR within 5 years. Among 463 both-knees-affected participants who had postbaseline observations, 55 had a least 1 TKR (either knee) within 5 years.

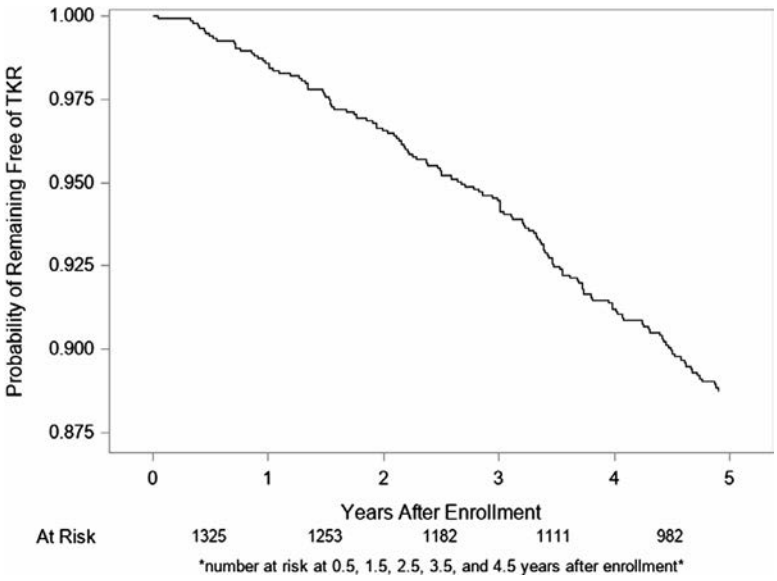


Figure 1. Kaplan-Meier plot for total knee replacement, progression cohort. TKR = total knee replacement.

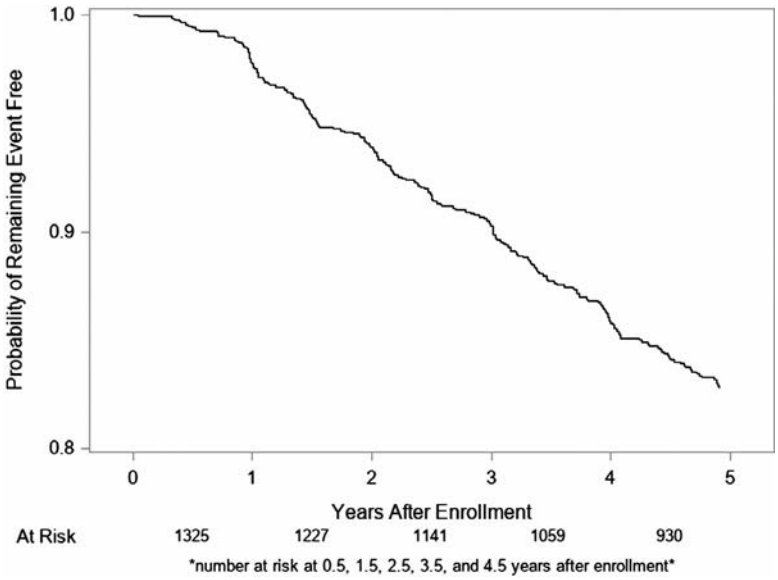


Figure 2. Kaplan-Meier plot for time to total knee replacement or Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) disability of ≥51 or WOMAC pain of ≥15, progression cohort.

Figure 1 shows the survival probability function estimated for the progression cohort. The cumulative hazard function estimated for the progression cohort indicates that the hazard rate was relatively constant over the 5-year period (Supplementary Figures 3–18, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24549; note Supplementary Figure 12). The estimated incidence rate of a TKR in an affected knee based on 5 years (1,819 days) of follow-up was approximately 24 cases per 1,000 person-years.

Feasibility and application in clinical trials. For each of the 7 clinical end points proposed, we estimated and examined the survival probability function and the cumulative hazard function as discussed above for the TKR (Figure 2 and Supplementary Figures 7–18, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24549). For each end point considered, the hazard rate was relatively constant over the 5-year period for the overall progression cohort.

Table 4 shows the incidence rate (λ_c) per 1,000 person-years, based on 5 years of follow-up (1,819 days) for each end point–defined event and the corresponding sample size calculation results. For the time to TKR (end point 1), the estimated incidence rate was approximately 24 cases per 1,000 person-years. This rate corresponds to 3,675 subjects required for a study with an average follow-up time of 5 years, with an assumed hazard ratio of 0.75, i.e., a 25% reduction in the rate of TKR while receiving the drug compared to the control. For different effect sizes (assumed hazard ratio varies between 0.67 and 0.85), the required sample sizes ranged from 1,987 to 10,886 for a study with an average follow-up time of 5 years. For a shorter study, the required sample sizes would be increased. For example, for a study with an average follow-up time of 3 years with an assumed hazard ratio of 0.75, we would need 6,105 subjects.

With composite end points (end points 2–7), we calculated higher estimated incidence rates. For end points 4 and 5, where we required the patient-reported outcomes scores to be consistently high for 2 consecutive years, we obtained estimated incident rates of approximately 26 and 27 cases per 1,000 person-years, respectively. For the other end points we assessed, estimated incidence rates ranged from 32 to 39. Specifically, end point 6, time to TKR or WOMAC disability of ≥51 or WOMAC pain of ≥15, resulted in an incidence rate of approximately 39 cases per 1,000 person-years, which corresponds to 2,251 subjects required for a study with an average follow-up time of 5 years, with an assumed hazard ratio of 0.75. Among the end points considered, this composite end point is the most feasible in terms of sample size, reducing the required number of total subjects by approximately 40% compared to an end point based only on TKR. For different effect sizes (assumed hazard ratio varies between 0.67 and 0.85), the required sample sizes ranged from 1,217 to 6,666 for a study with an average follow-up time of 5 years. For a shorter study, with an average follow-up time of 3 years with an assumed hazard ratio of 0.75, the required sample size would be approximately 3,700 subjects. These analyses are based on average follow-up time, and thus total study duration would be greater and would depend on the enrollment rate.

DISCUSSION

In this article, we examined the feasibility of using clinical end points in OA trials derived from analyses of the OAI database. We first considered the use of TKR alone, which has been previously

Table 4. Sample sizes with clinical end points of total knee replacement and WOMAC subscales*

HR, end point	λ_t†	Events required	Sample size required at follow-up		
			3-year‡	4-year	5-year
1) λ_c = 23.71§					
0.85	20.16	1,190	18,083	13,563	10,886
0.75	17.79	380	6,105	4,579	3,675
0.67	15.89	196	3,300	2,475	1,987
2) λ_c = 32.59¶					
0.85	27.70	1,190	1,3158	9,869	7,921
0.75	24.44	380	4,442	3,332	2,674
0.67	21.84	196	2,401	1,801	1,446
3) λ_c = 37.07#					
0.85	31.51	1,190	11,570	8,677	6,965
0.75	27.80	380	3,906	2,930	2,352
0.67	24.83	196	2,111	1,584	1,271
4) λ_c = 25.68**					
0.85	21.83	1,190	16,699	12,524	10,053
0.75	19.26	380	5,638	4,228	3,394
0.67	17.21	196	3,047	2,286	1,835
5) λ_c = 27.34††					
0.85	23.24	1,190	15,688	11,766	9,444
0.75	20.50	380	5,296	3,972	3,188
0.67	18.31	196	2,863	2,147	1,724
6) λ_c = 38.73‡‡					
0.85	32.92	1,190	11,073	8,305	6,666
0.75	29.05	380	3,738	2,804	2,251
0.67	25.95	196	2,021	1,516	1,217
7) λ_c = 37.82§§					
0.85	32.15	1,190	11,340	8,505	6,826
0.75	28.36	380	3,828	2,871	2,305
0.67	25.34	196	2,069	1,552	1,246

* HR = hazard ratio; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index.
† HR (or incidence rate) of treatment group, rate per 1,000 person-years.
‡ Average follow-up. Total study duration would be greater and would depend on the enrollment rate.
§ Time to total knee replacement (TKR); HR (or incidence rate) of control group based on estimates from the Osteoarthritis Initiative, rate per 1,000 person-years.
¶ Time to TKR or WOMAC disability of ≥51.
# Time to TKR or WOMAC pain of ≥15.
** Time to TKR or WOMAC disability of ≥51 twice in consecutive visits.
†† Time to TKR or WOMAC pain of ≥5 twice in consecutive visits.
‡‡ Time to TKR or WOMAC disability of ≥51 *or* WOMAC pain of ≥15.
§§ Time to TKR or WOMAC disability of ≥43 *and* WOMAC pain of ≥13.

proposed as a primary outcome in long-term OA randomized trials (9–11). However, there has been reluctance to use this end point due to the perceived low incidence rate and potential biases associated with the influence of external variables (e.g., socioeconomic factors). We estimated that for a study with time to TKR alone as an end point with an average follow-up time of 5 years and an assumed hazard rate of 0.75, the required total sample size would be approximately 4,000 subjects. A shorter study with an average follow-up of 3 years would require approximately 6,100 subjects.

In considering these limitations, we also evaluated the feasibility of using a composite end point: time to first TKR or time to severe pain or severely impaired functioning, defined by passing a clinically appropriate threshold in appropriate patient-reported outcomes, such as WOMAC pain or disability. Our approach has several advantages. First, it mitigates the potential issues of external factors affecting the incidence of TKR surgery alone. By including a measure of severe pain or severely impaired functioning, the composite end point can appropriately capture subjects who would be candidates for TKR based on severe symptoms and physical function but who may not receive the surgery for nondisease related reasons (e.g., socioeconomic status, access to care, surgeon preference, and health care systems).

Furthermore, the composite end point approach has the ability to address feasibility concerns for clinical studies by increasing the incidence rate of the composite end points, while still providing direct measures of clinical benefit in how a patient functions, feels, and survives. The use of composite end points will substantially reduce required sample sizes (up to approximately 40%) compared to the use of TKR alone. For example, with the composite end point of time to TKR or WOMAC disability of ≥51 or WOMAC pain of ≥15, a study with an average follow-up time of 3 years and an assumed hazard rate of 0.75 would require approximately 3,800 subjects, rather than 6,100 subjects if only

TKR was used. For a study with an average follow-up of 5 years, the total sample size would be approximately 2,300 subjects for the composite end point, as opposed to 3,700 subjects if only TKR was used.

Our estimates suggest that clinical studies for products intended to treat the underlying disease process of OA will require a larger sample size and longer duration than historical OA clinical trials. However, trials of even larger size and duration are not unprecedented in drug development in other common diseases with potential significant public health implications (30–37). Importantly, such OA studies will provide the evidence needed to confirm a direct long-term patient benefit and to ensure that the magnitude of benefit outweighs any risks of new OA interventions. The feasibility of such studies could be further improved by employing enrichment strategies, innovative trial designs, or use of models of accelerated OA, such as posttraumatic OA. In particular, studies in patients with accelerated OA are likely to accrue events at a much higher rate, resulting in smaller/shorter studies, and could provide data needed to establish structural or other biomarkers as surrogates to significantly advance the field of OA drug development. Notably, the OAI database represents a broad OA population where the background rates of progression to TKR are relatively low. Future clinical trials enriching for patients who are more likely to progress would likely increase study feasibility relative to the estimates based on OAI described in this article.

Based on our analyses, we propose the concept of using a composite end point (e.g., time to TKR or surpassing a threshold on WOMAC pain, time to TKR or surpassing a threshold on WOMAC disability, or time to TKR or surpassing a threshold on WOMAC pain and/or surpassing a threshold on WOMAC disability) to improve feasibility and clinical relevance. This approach uses an end point based on direct measures of how a patient functions, feels, or survives. In addition, the incorporation of thresholds for pain and function in the end point improves feasibility by increasing the background incidence rate. It also mitigates issues of external factors affecting incidence of TKR surgery by ensuring that patients with substantial chronic pain and disability who may not have access to or may not choose to undergo a TKR are still captured as having an undesirable outcome. Repeat measurements, i.e., more frequent visits than once a year (as in end points 4 or 5), may ensure that chronic impairment is detected earlier and may increase the background event rate and trial feasibility.

We note that this article considered a single patient-reported outcomes instrument (WOMAC), based on its frequent use in current OA trials; however, other appropriate patient-reported outcomes (including other versions of the WOMAC) may be considered using a similar approach. In the composite end points we proposed, the thresholds for pain and disability were chosen by examining the items and scorings of the patient-reported outcomes (WOMAC subscales) to ensure they represent meaningful impacts of disease. The choice of thresholds for PROs would require further discussion to ensure that the definitions capture chronic, substantial pain and/or disability, but our examples can be considered as proof of concept that such composite end points can improve feasibility and capture additional relevant patient outcomes. We have used high cutoffs that ensure the composite outcomes capture only patients with substantial pain and/or disability for the examples, but lower thresholds may also be considered (for example, responses of "severe" on half of the items of pain and disability would result in thresholds of 9 for pain and 27 for disability). Lowering the thresholds would result in smaller sample sizes and further increase feasibility (for example, with the composite end point of time to TKR or WOMAC disability of ≥27 or WOMAC pain of ≥9, a study with an average follow-up time of 3 years and an assumed hazard rate of 0.75 would require approximately 1,000 subjects).

In addition, the end points presented here focus on a target knee. This approach of following the knee initially affected may be appropriate for local treatments, but for systemic therapy, the end point could also be time to first TKR in any knee, which would further improve the feasibility of a trial. We found that among 456 right-knee-affected participants who had postbaseline observations, 42 had their first TKR on the right knee, while 9 had their first TKR on the left knee within 5 years (total of 51 TKR). Among 413 left-knee-affected participants who had postbaseline observations, 36 had their first TKR on the left knee; 8 had their first TKR on the right knee within 5 years (total of 44 TKR).

One limitation of the current study is the simple method used to calculate the sample sizes based on a constant hazard rate. This approach did not consider an accrual period or a delayed treatment effect. Also, our calculations did not incorporate the potential for missing data due to subject dropout, which could motivate an increase in sample size. All these factors could be considered for a more specific calculation within a specific drug development program and may affect the actual sample size requirements.

We note, however, that this OAI database reflected a relatively low dropout rate of approximately 20% across 5 years. Considering that this is an observational study, we are optimistic that high follow-up rates can be achieved in long-term interventional studies in this patient population. Additional measures could be implemented to prevent dropout, e.g., using a large, simple outcomes trial design with limited site visits and taking steps to ensure that patients who discontinue treatment continue to be followed.

In summary, for products being developed to treat the underlying pathophysiology or structural progression of OA, confirmatory studies to demonstrate direct clinical benefit to patients will be needed. Recognizing that experience in designing such studies has been limited, we examined the feasibility of using clinical end points in OA trials using the OAI database, comprising a cohort of patients with OA, representative of the general OA

population. The data from our analyses indicate that studies using composite clinical end points could be feasible and should be explored further to advance the field of OA drug development.

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. Kim 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. Kim, Levin, Nikolov, Abugov, Rothwell.

Analysis and interpretation of data. Kim, Levin, Nikolov, Abugov, Rothwell.

REFERENCES

1. Ondrésik M, Azevedo Maia FR, da Silva Morais A, Gertrudes AC, Dias Bacelar AH, Correia C, et al. Management of knee osteoarthritis: current status and future trends. Biotechnol Bioeng 2017;114:717–39.
2. Lawrence RC, Felson DT, Helmick CG, Arnold LM, Choi H, Deyo RA, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. Arthritis Rheum 2008;58:26–35.
3. Burstein D. MRI for development of disease-modifying osteoarthritis drugs. NMR Biomed 2006;19:669–80.
4. Karsdal MA, Michaelis M, Ladel C, Siebuhr AS, Bihlet AR, Andersen JR, et al. Disease-modifying treatments for osteoarthritis (DMOADs) of the knee and hip: lessons learned from failures and opportunities for the future. Osteoarthritis Cartilage 2016;24:2013–21.
5. Oo WM, Yu SP, Daniel MS, Hunter DJ. Disease-modifying drugs in osteoarthritis: current understanding and future therapeutics. Expert Opin Emerg Drugs 2018;23:331–47.
6. Food and Drug Administration. Osteoarthritis: structural end points for the development of drugs. 2018. URL: https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm071577.pdf.
7. Fleming TR, Powers JH. Biomarkers and surrogate end points in clinical trials. Stat Med 2012;31:2973–84.
8. Osteoarthritis Initiative. URL: https://oai.nih.gov.
9. Abadie E, Ethgen D, Avouac B, Bouvenot G, Branco J, Bruyere O, et al. Recommendations for the use of new methods to assess the efficacy of disease-modifying drugs in the treatment of osteoarthritis. Osteoarthritis Cartilage 2004;12:263–8.
10. Dougados M, Gueguen A, Nguyen M, Berdah L, Lequesne M, Mazieres B, et al. Requirement for total hip arthroplasty: an outcome measure of hip osteoarthritis? J Rheumatol 1999;26:855–61.
11. Maillefert JF, Hawker GA, Gossec L, Mahomed NN, Lohmander S, Dieppe PA, et al. Concomitant therapy: an outcome variable for musculoskeletal disorders? Part 2: total joint replacement in osteoarthritis trials. J Rheumatol 2005;32:2449–51.
12. Boutron I, Rannou F, Jardinaud-Lopez M, Meric G, Revel M, Poiraudeau S. Disability and quality of life of patients with knee or hip osteoarthritis in the primary care setting and factors associated with general practitioners' indication for prosthetic replacement within 1 year. Osteoarthritis Cartilage 2008;16:1024–31.
13. Brennan SL, Lane SE, Lorimer M, Buchbinder R, Wluka AE, Page RS, et al. Associations between socioeconomic status and primary total knee joint replacements performed for osteoarthritis across Australia 2003-10: data from the Australian Orthopaedic Association National Joint Replacement Registry. BMC Musculoskelet Disord 2014;15:356
14. Feldman CH, Dong Y, Katz JN, Donnell-Fink LA, Losina E. Association between socioeconomic status and pain, function and pain catastrophizing at presentation for total knee arthroplasty. BMC Musculoskelet Disord 2015;16:18.
15. Gossec L, Tubach F, Baron G, Ravaud P, Logeart I, Dougados M. Predictive factors of total hip replacement due to primary osteoarthritis: a prospective 2 year study of 505 patients. Ann Rheum Dis 2005;64:1028–32.
16. Hannan MT, Felson DT, Pincus T. Analysis of the discordance between radiographic changes and knee pain in osteoarthritis of the knee. J Rheumatol 2000;27:1513–7.
17. Hawker GA, Guan J, Croxford R, Coyte PC, Glazier RH, Harvey BJ, et al. A prospective population-based study of the predictors of undergoing total joint arthroplasty. Arthritis Rheum 2006;54:3212–20.
18. Merx H, Dreinhofer K, Schrader P, Sturmer T, Puhl W, Gunther KP, et al. International variation in hip replacement rates. Ann Rheum Dis 2003;62:222–6.
19. Wetterholm M, Turkiewicz A, Stigmar K, Hubertsson J, Englund M. The rate of joint replacement in osteoarthritis depends on the patient's socioeconomic status. Acta Orthop 2016;87:245–51.
20. Collins NJ, Misra D, Felson DT, Crossley KM, Roos EM. Measures of knee function: International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form, Knee Injury and Osteoarthritis Outcome Score (KOOS), Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS), Knee Outcome Survey Activities of Daily Living Scale (KOS-ADL), Lysholm Knee Scoring Scale, Oxford Knee Score (OKS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Activity Rating Scale (ARS), and Tegner Activity Score (TAS). Arthritis Care Res (Hoboken) 2011;63 Suppl 11:S208–28.
21. Hochberg MC, Fort JG, Svensson O, Hwang C, Sostek M. Fixed-dose combination of enteric-coated naproxen and immediate-release esomeprazole has comparable efficacy to celecoxib for knee osteoarthritis: two randomized trials. Curr Med Res Opin 2011;27:1243–53.
22. McAlindon TE, LaValley MP, Harvey WF, Price LL, Driban JB, Zhang M, et al. Effect of intra-articular triamcinolone vs saline on knee cartilage volume and pain in patients with knee osteoarthritis: a randomized clinical trial. JAMA 2017;317:1967–75.
23. Gibofsky A, Hochberg MC, Jaros MJ, Young CL. Efficacy and safety of low-dose submicron diclofenac for the treatment of osteoarthritis pain: a 12-week, phase 3 study. Curr Med Res Opin 2014;30:1883–93.
24. Altman R, Hochberg M, Gibofsky A, Jaros M, Young C. Efficacy and safety of low-dose SoluMatrix meloxicam in the treatment of osteoarthritis pain: a 12-week, phase 3 study. Curr Med Res Opin 2015;31:2331–43.
25. Conaghan PG, Cohen SB, Berenbaum F, Lufkin J, Johnson JR, Bodick N. A phase IIb trial of a novel extended-release microsphere formulation of triamcinolone acetonide for intraarticular injection in knee osteoarthritis. Arthritis Rheumatol 2018;70:204–11.
26. Hellio le Graverand MP, Clemmer RS, Redifer P, Brunell RM, Hayes CW, Brandt KD, et al. A 2-year randomised, double-blind, placebo-controlled, multicentre study of oral selective iNOS inhibitor, cindunistat (SD-6010), in patients with symptomatic osteoarthritis of the knee. Ann Rheum Dis 2013;72:187–95.
27. Dougados M, Hawker G, Lohmander S, Davis AM, Dieppe P, Maillefert JF, et al. OARSI/OMERACT criteria of being considered a candidate for total joint replacement in knee/hip osteoarthritis as an end point in clinical trials evaluating potential disease modifying osteoarthritic drugs. J Rheumatol 2009;36:2097–9.
28. Gossec L, Paternotte S, Bingham CO, Clegg DO, Coste P, Conaghan PG, et al. OARSI/OMERACT initiative to define state of severity and indication for joint replacement in hip and knee osteoarthritis. J Rheumatol 2011;38:1765–9.
29. Manno RL, Bingham CO, Paternotte S, Gossec L, Halhol H, Giacovelli G, et al. OARSI-OMERACT initiative: defining thresholds for symptomatic severity and structural changes in disease modifying

osteoarthritis drug (DMOAD) clinical trials. Osteoarthritis Cartilage 2012;20:93–101.

30. Scirica BM, Bhatt DL, Braunwald E, Steg PG, Davidson J, Hirshberg B, et al. Saxagliptin and cardiovascular outcomes in patients with type 2 diabetes mellitus. N Engl J Med 2013;369:1317–26.
31. Green JB, Bethel MA, Armstrong PW, Buse JB, Engel SS, Garg J, et al. Effect of sitagliptin on cardiovascular outcomes in type 2 diabetes. N Engl J Med 2015;373:232–42.
32. Marso SP, Daniels GH, Brown-Frandsen K, Kristensen P, Mann JF, Nauck MA, et al. Liraglutide and cardiovascular outcomes in type 2 diabetes. N Engl J Med 2016;375:311–22.
33. Schwartz GG, Steg PG, Szarek M, Bhatt DL, Bittner VA, Diaz R, et al. Alirocumab and cardiovascular outcomes after acute coronary syndrome. N Engl J Med 2018;379:2097–107.
34. Sabatine MS, Giugliano RP, Keech AC, Honarpour N, Wiviott SD, Murphy SA, et al. Evolocumab and clinical outcomes in patients with cardiovascular disease. N Engl J Med 2017;376:1713–22.
35. Action to Control Cardiovascular Risk in Diabetes Study Group, Gerstein HC, Miller ME, Byington RP, Goff DC Jr, Bigger JT, et al. Effects of intensive glucose lowering in type 2 diabetes. N Engl J Med 2008;358:2545–59.
36. Jamerson K, Weber MA, Bakris GL, Dahlöf B, Pitt B, Shi V, et al. Benazepril plus amlodipine or hydrochlorothiazide for hypertension in high-risk patients. N Engl J Med 2008;359:2417–28.
37. ACTIVE Investigators, Connolly SJ, Pogue J, Hart RG, Hohnloser SH, Pfeffer M, et al. Effect of clopidogrel added to aspirin in patients with atrial fibrillation. N Engl J Med 2009;360:2066–78.

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1163–1171
DOI 10.1002/acr.24554

American College of Rheumatology
Empowering Rheumatology Professionals

Longitudinal Relationship Between Physical Activity and Joint Space Narrowing: Forty-Eight–Month Follow-Up Data From the Osteoarthritis Initiative

Bo Hu,[1] DongBai Han,[2] Michael C. Nevitt,[3] Barton L. Wise,[4] and Neil A. Segal[5]

Objective. To determine whether the amount of physical activity (PA) is a determinant of joint space narrowing (JSN) worsening over 48 months in participants with knee osteoarthritis.

Methods. Data were obtained from the Osteoarthritis Initiative. PA, measured using the Physical Activity Scale for the Elderly (PASE), was defined as the mean value of the annual measurements conducted prior to development of worsening JSN. Worsening JSN was defined as at least a partial grade increase in the Osteoarthritis Research Society International JSN score over 48 months, in comparison with baseline. Restricted cubic spline function was used to group participants based on the linear association between PA and JSN worsening. A pooled logistic regression model was used to evaluate the association between PA and JSN worsening adjusted for confounders.

Results. A total of 2,167 participants were included. In total, 625 participants (28.8%) had JSN worsening over 48 months. Compared with a PASE score of 141–180, PASE scores of 101–140 and >220 were associated with an increased risk of JSN worsening in men, with odds ratios (ORs) of 1.73 (95% confidence interval [95% CI] 1.07–2.81) and 1.83 (95% CI 1.14–2.93), respectively. Similarly, in participants with Kellgren/Lawrence (K/L) grade 2, compared with a PASE score of 141–180, PASE scores of ≤100 and >220 were associated with increased risks of JSN worsening, with an OR of 1.69 (95% CI 1.13–2.54) and 1.64 (95% CI 1.05–2.56), respectively.

Conclusion. Compared to moderate PA, higher or lower amounts of PA are associated with an elevated risk for JSN worsening in men and in participants with K/L grade 2 knees.

INTRODUCTION

The prevalence of knee osteoarthritis (OA) and knee replacements is increasing each year (1). Approximately 40 million Americans are living with OA, and the number will increase by 50% over the next decade (2). Knee OA is a primary cause of disability among the elderly (3). Disability due to OA is associated with an extremely high economic burden and elevated risks of hospitalization, institutionalization, and mortality (4,5).

Physical activity (PA) is defined as "any bodily movement produced by skeletal muscles that results in energy expenditure" and is a broad classification of movement that encompasses sport and exercise in addition to other activities (6). The broad components of PA are occupational, transportation-related, domestic, and leisure time (which consists of exercise and recreational or competitive sport). Evidence supports the idea that PA is a modifiable factor with beneficial effects on overall health (6), although either too little or too much PA could negatively influence joint health (7,8).

ClinicalTrials.gov identifier: NCT00080171.

This article does not necessarily reflect the opinions or views of the OAI investigators, the NIH, or the private funding partners.

This article was prepared using an Osteoarthritis Initiative (OAI) public-use data set, and its contents do not necessarily reflect the opinions or views of the OAI Study Investigators, the NIH, or the private funding partners of the OAI. The OAI is a public–private partnership between the NIH (contracts N01-AR-2-2258, N01-AR-2-2259, N01-AR-2-2260, N01-AR-2-2261, and N01-AR-2-2262) and private funding partners (Merck Research Laboratories, Novartis Pharmaceuticals, GlaxoSmithKline, and Pfizer, Inc.) and is conducted by the OAI Study Investigators. Private sector funding for the OAI is managed by the Foundation for the NIH. Dr. Nevitt was a part of the OAI investigative team.

[1]Bo Hu, PhD: Fuwai Hospital, Peking Union Medical College, and Chinese Academy of Medical Sciences, Beijing, China; [2]DongBai Han, MS: Yantai Center for Disease Control and Prevention, Yantai, Shandong Province, China; [3]Michael C. Nevitt, PhD, MPH: University of California, San Francisco; [4]Barton L. Wise, MD, MSc, FACP: University of California, Davis School of Medicine, Sacramento; [5]Neil A. Segal, MD, MS: University of Kansas, Kansas City.

No potential conflicts of interest relevant to this article were reported.

Address correspondence to Neil A. Segal, MD, MS, Department of Rehabilitation Medicine, University of Kansas Medical Center, 3901 Rainbow Boulevard–MS 1046, Kansas City, KS 66160. Email: segal-research@kumc.edu.

Submitted for publication June 17, 2020; accepted in revised form January 5, 2021.

SIGNIFICANCE & INNOVATIONS

- Evidence of an association between physical activity amount and the risk for worsening of knee joint space narrowing could guide recommendations for the amount of physical activity that can optimize positive health outcomes.
- The finding that men with the highest or lower amounts of physical activity had a greater risk for worsening of joint space narrowing over 48 months follow-up suggests a need to moderate physical activity in men at elevated risk for knee osteoarthritis worsening.
- The finding that participants with Kellgren/Lawrence grade 2 knees with the highest and lowest amounts of physical activity had elevated risk for worsening of joint space narrowing over 48 months suggests a need to moderate physical activity in these participants at elevated risk for knee osteoarthritis worsening.

Several studies have explored the association of PA and progression of knee OA. However, whether PA may be harmful or protective on progression of knee OA remains unclear, possibly due to the populations and methods used in prior studies. Cooper et al (9) found that leisure activities such as sports, walking, cycling, gardening, and dancing were not associated with progression of knee OA. The negative result may have been related to a smaller number of subjects in whom progression was detected or by the limited definition of assessed PA. In a study of middle-aged men and women, leisure time PA, assessed by a questionnaire adapted from the Minnesota Leisure Time Physical Activity Questionnaire (10), showed no consistent overall relationship with incidence of severe knee OA over 11 years, defined as joint replacement due to OA (11). Therefore, there is uncertainty regarding whether PA is associated with OA structural progression.

Structural progression of knee OA can be defined by quantitative radiographic joint space narrowing (JSN) (12) or by semiquantitative measurement (13). Semiquantitative assessment has been used extensively in epidemiologic studies and has been the primary measure of structural progression accepted by regulatory agencies for clinical trials (14).

In this study, we tested the hypothesis that higher amounts of PA are longitudinally associated with an increased risk for worsening of JSN over 48 months of follow-up. Considering that there are different determinants of risk for knee OA progression in women and men (15–17), we tested this hypothesis separately in men and women.

PATIENTS AND METHODS

Subjects. Data were obtained from the Osteoarthritis Initiative (OAI) database, which is available for public access at https://oai.epi-ucsf.org/datarelease/. The OAI offers high-quality longitudinal data with detailed information on 4,796 participants with, or at elevated risk of developing, symptomatic knee OA (e.g., daily knee symptoms, overweight, history of knee injury/surgery, family history of knee replacement, or repetitive knee flexion [https://oai.epi-ucsf.org/datarelease/docs/StudyDesignProtocol.pdf]). This study of progression of knee OA structural worsening included participants with both preradiographic (Kellgren/Lawrence [K/L] grade 1 [18]) and radiographic knee OA (K/L grade ≥2), who did not have end-stage JSN (<3) at baseline. We excluded healthy reference subjects without knee OA or risk factors for knee OA (n = 122), participants who had knee replacements at baseline (n = 64), participants who had rheumatoid arthritis or some other type of inflammatory arthritis at baseline or follow-up (n = 408), participants with missing main information (n = 546), and participants with K/L grade of 0 at baseline (n = 1,489).

Assessment of PA. General PA was assessed using the Physical Activity Scale for the Elderly (PASE), an instrument that quantifies multiple domains of activity in older adults that has been validated for use in persons with knee OA (19). Questions included on the PASE were designed to evaluate both occupational and nonoccupational walking, recreational activities, exercise, housework, yard work, and caring for others in the past 7 days, adapted from a widely used instrument that has shown associations of these activities with knee OA in multiple studies (19,20). The PASE score correlated with performance on the 6-minute walk test, the isokinetic thigh strength test, and the perceived difficulty with physical functioning test, which supports both convergent and construct validity (19).

Image acquisition and assessment of JSN. Bilateral posteroanterior fixed-flexed knee radiographs were acquired using a SynaFlexer frame to position participants' feet reproducibly (21). In the OAI data sets, a key measure of structural worsening of knee OA is Osteoarthritis Research Society International (OARSI) JSN (13). This individual grading scale uses an atlas to compare radiographs to representative images and assign a grade for the severity of JSN from 0 to 3 in the medial or lateral tibiofemoral compartments (13). Radiographs were centrally read for this measurement as part of the OAI (22).

Version 8 of the OAI data release provided summary radiograph outcome variables for the first visit at which JSN had progressed compared to baseline. These radiographic measurements of JSN worsening were assessed at 12, 24, 36, and 48 months. Worsening was defined by at a least partial grade increase in OARSI JSN score between baseline and follow-up. Bilateral knee radiographs were acquired on each OAI participant at each time point and used for assessing JSN. JSN was treated as a person-based outcome. Participants were defined as demonstrating progression if at least 1 knee demonstrated JSN worsening.

Follow-up assessments. The OAI protocol included measurements of PA, JSN, and other covariates, including knee pain and knee injury every 12 months. The time point at

which worsening occurred was defined as the follow-up visit at which JSN worsening was detected. The earlier follow-up time point was selected if both of a participant's knees demonstrated worsening. Participants who had undergone knee replacement by follow-up were included in the definition of JSN worsening.

Assessment of potential confounders. Age, sex (23), race (24), and body mass index (BMI) (25) have been found to be related to knee OA and were therefore entered as potential confounders in multivariable analyses for exploring the relationship between PA and JSN worsening. BMI was defined as body mass divided by the square of height (kg/m^2). Race was collapsed into 2 categories, White and non-White. K/L grades were scored by 2 radiograph readers (26). Knee pain is an important risk factor for knee OA. The pain subscale (range 0–20) of The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) was used, as described in the statistical methods below, to characterize participants' knee pain. The WOMAC, a survey based on self-report, is the most extensively validated instrument used to assess change in persons with OA and is both recommended for and widely used in studies of participants with OA. The WOMAC pain score was used as a potential confounder in multivariate analyses. In our study, K/L grade and WOMAC pain scores were collected separately for each knee within participants. Thus, K/L grade and pain scores were included for the knee in which JSN worsening occurred. K/L grade and WOMAC pain score of the knee with earlier worsening was selected if both of a participant's knees met the definition for worsening. When there was no difference in the time to either worsening or lack of worsening of a participant's knees, the K/L grade and pain score were selected for the leg with which the participant preferred to kick a ball.

Because of the likelihood that previously injured knees would be at high risk of OA (27), knee injury was also included in analyses as a potential confounder. Knee injury was defined as a knee having been injured badly enough to limit a participant's ability to walk for at least 2 days since their last annual visit for OAI. Similar to knee pain, knee injury was also included for the knee in which JSN progressed.

Statistical analysis. Baseline differences in characteristics between different PASE scores were compared using a chi-square test and analysis of variance. Exact time for JSN worsening could not be identified definitively because JSN was measured and recorded annually as interval censored data in data sets. Since PASE scores and all confounders were measured annually in the OAI, a pooled logistic regression model, validated by the Framingham Heart Study (28), was selected to estimate the risk for knee JSN worsening associated with the PASE score. Every follow-up interval was treated as a mini–follow-up study, and risk factors at the interval start to predict relative events in the same interval (e.g., for the 12–24-month interval, values of confounders collected at 12-month follow-up were used in the logistic model, for the 24–36-month interval, values of confounders collected by 24-month follow-up were used in the logistic model, etc.). In our study, confounders (age, sex, race, BMI, knee pain, and history of knee injury) at relevant time intervals were adjusted in pooled logistic regression models. In addition to sex-stratified analyses, a subgroup analysis was conducted by K/L grade. The strength and direction of the associations were assessed with odds ratios (ORs) and corresponding 95% confidence intervals (95% CIs). Sensitivity analyses were conducted to assess the impact of including versus excluding 24 knees that, at baseline, had a JSN score of 2 and also were graded K/L 4.

Based on an unclear association between the PASE score and JSN worsening, we initially analyzed the PASE score as a continuous variable, fitting a restricted cubic spline (RCS) function (29) with 4 knots (located at the 5th, 35th, 65th, and 95th percentiles) using pooled data sets, to make a proper grouping by finding the participants with the lowest risk for JSN worsening. The median of the PASE score was chosen to be the reference value for all spline plots. The RCS function was used to estimate and list all hazard ratios for JSN worsening for each PASE score by comparing it with the median of the PASE score. Age, sex, race, BMI, knee injury, and WOMAC pain score over the observation period were included as adjustment variables in logistic regression models. An SAS macro (29) was used to build the RCS function. A *P* value of 0.05 or less was considered to indicate statistical significance. All statistical analyses were performed using SAS software, version 9.4.

RESULTS

Of the 4,796 participants, the following were ineligible for inclusion: healthy reference participants (n = 122), participants who had knee replacement at baseline (n = 64), participants who had inflammatory arthritis at baseline (n = 408), participants with missing main information at baseline (n = 546), and participants with K/L grade score of 0 at baseline (n = 1,489). Thus, 2,167 participants were included in statistical analyses. The mean ± SD age was 62.2 ± 9.0 years, 38.4% of participants were male, and 79.0% were White (Table 1). After pooling 4 time intervals (baseline–12-month follow-up, 12–24-month follow-up, 24–36-month follow-up, and 36–48-month follow-up), a total of 7,407 data items were included in the analytic data set (Figure 1). Chronologically at the 4 respective time points, there were 287 cases of JSN worsening in 2,167 participants, 145 cases of JSN worsening in 1,880, 110 cases of JSN worsening in 1,735 participants, and 83 cases of JSN worsening in 1,625 participants (Table 2), providing a total of 625 participants with JSN worsening, 235 men and 390 women. Over 48 months, in participants with baseline K/L grades 1, 2, 3, and 4, respectively,

Table 1. Baseline characteristics of participants*

Characteristic	Overall (n = 2,167)	Men (n = 834)	Women (n = 1,333)
Age, mean ± SD years	62.2 ± 9.0	62.3 ± 9.3	62.2 ± 8.9
White	1,712 (79.0)	711 (85.3)	1,001 (75.1)
BMI, mean ± SD kg/m^2	29.4 ± 4.8	29.2 ± 4.1	29.5 ± 5.2
Knee pain, median (IQR)†	1.0 (0.0–4.0)	1.0 (0.0–4.0)	2.0 (0.0–5.0)
Knee injury	659 (30.4)	309 (37.1)	350 (26.3)
K/L grade			
1	689 (31.8)	267 (32.0)	422 (31.7)
2	965 (44.5)	336 (40.3)	629 (47.2)
3	489 (22.6)	219 (26.3)	270 (20.3)
4	24 (1.1)	12 (1.4)	12 (0.9)

* Values are the number (%) unless indicated otherwise. BMI = body mass index; IQR = interquartile range; K/L = Kellgren/Lawrence grading scale.
† Measured with the Western Ontario and McMaster Universities Osteoarthritis Index scale.

4.7%, 7.9%, 14.0%, and 46.8% demonstrated worsening JSN (Table 3).

In both men and women, no linear (*P* values for linear trend in Table 2) or nonlinear (Figure 2) associations were found between the PASE score and JSN worsening. Based on the results of RCS analyses, the risk for JSN worsening plateaus when PASE scores are between 100 and 200 in the total cohort (test for overall association $P = 0.4721$; test for nonlinear association $P = 0.6813$ with the median PASE reference value of 143), as well as in men (Figure 2A) and in women (Figure 2B). We selected the range of PASE score of 141–180 as the reference category because this was the range associated with the lowest risk (OR range from 1.00 to 1.00) for JSN worsening. The OR ranges in men and women corresponding to PASE score of 141–180 are 1.01–1.01 and 0.96–1.00, respectively. Given the number of participants with PASE scores between 141–180 and the plateaus of the distribution for the risk of JSN worsening on RCS curves, participants were divided into 5 groups by PASE score: ≤100, 101–140, 141–180, 181–220, and >220.

In men, after adjusting for potential confounders, compared with a PASE score of 141–180, a PASE score of >220 was associated with an increased risk of JSN worsening, with an OR of 1.83 (95% CI 1.14–2.93), and a PASE score of 101–140 was associated with an increased risk of JSN worsening, with an OR of 1.73 (95% CI 1.07–2.81). However, in women, compared with a PASE score of 141–180, PASE scores >220 were not associated with an increased risk (OR 0.99 [95% CI 0.68–1.45]) of JSN worsening. Compared with PASE scores of 141–180, we found no association between PASE score groups other than those

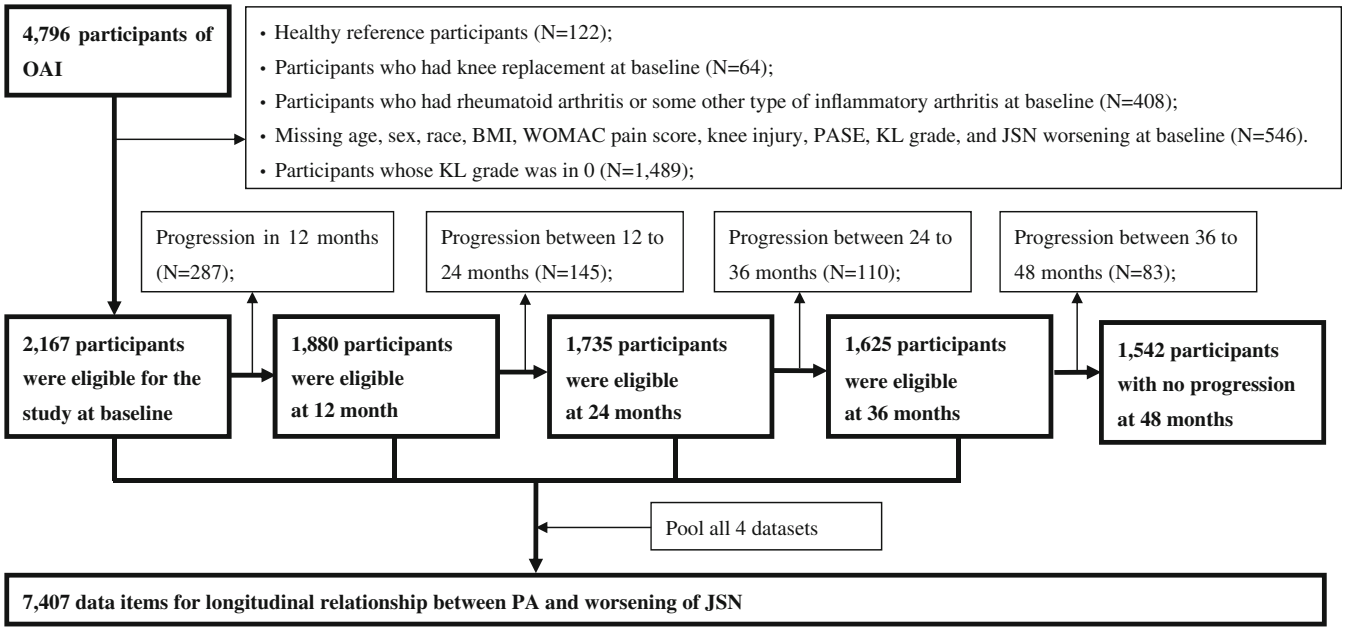


Figure 1. Flow chart of the identification of participants for study inclusion. Missing main information included age, sex, race, body mass index, Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), knee injury, Physical Activity Scale for the Elderly (PASE) score, Kellgren/Lawrence (KL) grading scale, or joint space narrowing (JSN) worsening. BMI = body mass index; OAI = Osteoarthritis Initiative.

Table 2. Participants with joint space narrowing progression by annual PASE score*

Time, months	Overall	≤100	101–140	141–180	181–220	>220	*P* for trend
Total							
Baseline–12 (n = 2,167)	287	87 (14.8)	54 (13.2)	52 (12.9)	29 (10.4)	65 (13.4)	0.3143
12–24 (n = 1,880)	145	2 (18.2)	76 (8.5)	14 (4.2)	21 (7.8)	32 (8.5)	0.8146
24–36 (n = 1,735)	110	39 (7.6)	24 (6.5)	20 (6.3)	12 (5.6)	15 (4.7)	0.0830
36–48 (n = 1,625)	83	29 (5.9)	14 (4.1)	11 (3.7)	12 (6.0)	17 (5.8)	0.8693
Men							
Baseline–12 (n = 834)	101	20 (10.8)	18 (12.9)	16 (11.0)	11 (9.7)	36 (14.3)	0.3979
12–24 (n = 733)	55	2 (33.3)	29 (9.5)	0 (0.0)	7 (6.5)	17 (8.7)	0.5368
24–36 (n = 678)	43	14 (8.2)	10 (7.5)	8 (5.8)	4 (5.3)	7 (4.4)	0.1140
36–48 (n = 635)	36	8 (5.1)	5 (4.0)	5 (4.4)	7 (9.0)	11 (6.8)	0.2524
Women							
Baseline–12 (n = 1,333)	186	67 (16.5)	36 (13.3)	36 (14.0)	18 (10.9)	29 (12.3)	0.0844
12–24 (n = 1,147)	90	0 (0.0)	47 (8.0)	14 (6.5)	14 (8.6)	15 (8.3)	0.7864
24–36 (n = 1,057)	67	25 (7.4)	14 (5.9)	12 (6.6)	8 (5.7)	8 (5.1)	0.3421
36–48 (n = 990)	47	21 (6.2)	9 (4.1)	6 (3.4)	5 (4.1)	6 (4.6)	0.3042

* Values are the number (%) unless indicated otherwise. PASE = Physical Activity Scale for the Elderly.

reported above and JSN worsening in either men or women (Table 4). In participants with K/L grade 2, compared with a PASE score of 141–180, PASE scores ≤100 and >220 were associated with increased risks of JSN worsening after adjusting for potential confounders; ORs were 1.69 (95% CI 1.13–2.54) and 1.64 (95% CI 1.05–2.56), respectively. However, no associations were found in participants with K/L grade 1 and 3 (Table 4).

There were 24 knees that were graded K/L 4 but had a JSN score of 2 at baseline and therefore had the potential to demonstrate JSN worsening. Of these, 4 knees worsened in the 0–12-month, 10 knees worsened in 12–24-month, and 10 knees worsened in 24–36-month assessment periods. Sensitivity analyses revealed that ORs were unchanged based on inclusion or exclusion of these knees. An RCS plot for these knees is presented in the supplementary materials (see Supplementary Figures 1–5, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24554).

DISCUSSION

Our results revealed that, in men, both higher amounts of PA (PASE score >220) and lower amounts of PA (PASE scores 101–

Table 3. Participants with joint space narrowing progression by annual PASE score*

Time (months), K/L grade	Overall	≤100	101–140	141–180	181–220	>220
Pooled time points						
1 (n = 2,383)	113 (4.7)	30 (4.6)	24 (5.0)	21 (4.7)	11 (3.8)	27 (5.4)
2 (n = 3,466)	274 (7.9)	96 (9.5)	53 (7.8)	35 (5.6)	30 (6.5)	60 (8.7)
3 (n = 1,496)	209 (14.0)	64 (14.3)	41 (13.0)	37 (14.1)	30 (14.6)	37 (13.9)
4 (n = 62)	29 (46.8)	8 (42.1)	9 (64.3)	4 (28.6)	3 (75.0)	5 (45.5)
Baseline–12 (n = 2,167)						
1 (n = 689)	45 (6.5)	13 (7.6)	8 (5.8)	9 (6.7)	5 (5.8)	10 (6.3)
2 (n = 965)	132 (13.7)	36 (13.7)	30 (16.3)	22 (12.9)	13 (10.3)	31 (13.9)
3 (n = 489)	106 (21.7)	36 (24.7)	16 (18.6)	21 (22.8)	9 (14.1)	24 (23.8)
4 (n = 24)	4 (16.7)	2 (20.0)	0 (0.0)	0 (0.0)	2 (100.0)	0 (0.0)
12–24 (n = 1,880)						
1 (n = 601)	27 (4.5)	0 (0.0)	16 (5.8)	4 (3.9)	2 (2.3)	5 (2.8)
2 (n = 873)	53 (6.1)	1 (25.0)	29 (7.1)	4 (2.4)	5 (4.2)	14 (8.0)
3 (n = 382)	54 (14.1)	1 (20.0)	23 (12.1)	6 (9.8)	13 (20.6)	11 (17.5)
4 (n = 24)	11 (45.8)	0 (0.0)	8 (57.1)	0 (0.0)	1 (50.0)	2 (40.0)
24–36 (n = 1,721)						
1 (n = 564)	20 (3.5)	3 (1.8)	6 (5.0)	5 (4.4)	1 (1.8)	5 (4.6)
2 (n = 829)	44 (5.3)	22 (8.9)	5 (3.0)	4 (2.9)	7 (6.0)	6 (3.8)
3 (n = 328)	32 (9.8)	11 (11.7)	9 (11.3)	7 (11.1)	4 (9.8)	1 (2.0)
4 (n = 14)	14 (100.0)	3 (100.0)	4 (100.0)	4 (100.0)	0 (0.0)	3 (100.0)
36–48 (n = 1,625)						
1 (n = 529)	21 (2.6)	5 (3.2)	3 (2.8)	3 (3.1)	3 (4.8)	7 (6.8)
2 (n = 799)	45 (5.6)	16 (6.6)	10 (5.8)	5 (3.3)	5 (5.1)	9 (6.6)
3 (n = 297)	17 (5.7)	8 (8.4)	1 (1.5)	3 (6.5)	4 (10.8)	1 (1.9)
4 (n = 0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)

* Values are the number (%). K/L = Kellgren/Lawrence grading scale; PASE = Physical Activity Scale for the Elderly.

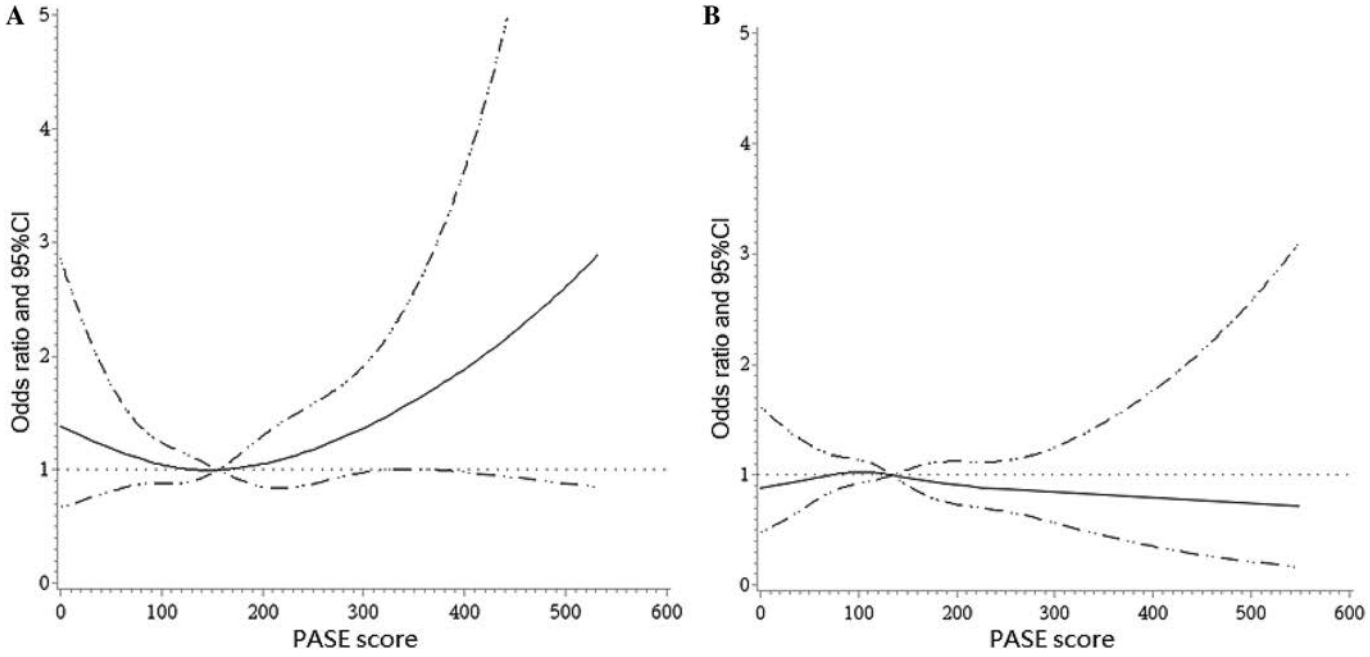


Figure 2. Association between Physical Activity Scale for the Elderly (PASE) score and joint space narrowing progression over 48 months using the restricted cubic spline method (4 nodes), spline plot for adjusted logistic models, adjusted for age, sex, race, body mass index, knee pain, and knee injury. **A**, Men (test for overall association $P = 0.1103$; test for nonlinear association $P = 0.1507$, with median PASE score reference value of 155.5); **B**, Women (test for overall association $P = 0.4339$; test for nonlinear association $P = 0.4869$, with median PASE score reference value of 134). 95% CI = 95% confidence interval. Solid gray line shows adjusted odds ratios of joint space narrowing progression. Broken dash/dot line shows 95% confidence interval. Broken dotted line corresponds to 1.

140) were associated with an increased risk of JSN worsening compared with a PASE score of 141–180 over a 48-month follow-up. In addition, in participants with baseline K/L grade 2, both higher and lower amounts of PA were associated with an elevated risk for JSN worsening (PASE scores ≤100 or >220 compared to 141–180).

While the current study focused on worsening of radiographic JSN in knees with preexisting radiographic OA, these

Table 4. Association between PASE score and JSN worsening over 48 months by sex and K/L grade*

	≤100	101–140	141–180	181–220	>220
JSN total (pooled no. = 7,407)					
Worsening, pooled no. (%)	198 (9.3)	127 (8.5)	97 (7.2)	74 (7.7)	129 (8.8)
Pooled logistic regression†	1.23 (0.95–1.59)	1.20 (0.91–1.59)	1 (Ref.)	1.08 (0.78–1.49)	1.27 (0.95–1.69)
JSN men (pooled no. = 2,880)					
Worsening, pooled no. (%)	57 (8.3)	49 (9.2)	29 (5.6)	29 (7.8)	71 (9.2)
Pooled logistic regression‡	1.40 (0.87–2.25)	1.73 (1.07–2.81)	1 (Ref.)	1.34 (0.77–2.32)	1.83 (1.14–2.93)
JSN women (pooled no. = 4,527)					
Worsening, pooled no. (%)	141 (9.8)	78 (8.2)	68 (8.2)	45 (7.7)	58 (8.2)
Pooled logistic regression‡	1.12 (0.82–1.53)	0.98 (0.70–1.39)	1 (Ref.)	0.95 (0.64–1.42)	0.99 (0.68–1.45)
K/L 1 (pooled no. = 2,383)					
JSN worsening, pooled no. (%)	30 (4.6)	24 (5.0)	21 (4.7)	11 (3.8)	27 (5.4)
Pooled logistic regression†	0.88 (0.49–1.57)	1.03 (0.56–1.88)	1 (Ref.)	0.73 (0.34–1.55)	1.20 (0.65–2.21)
K/L 2 (pooled no. = 3,466)					
JSN worsening, pooled no. (%)	96 (9.5)	53 (7.8)	35 (5.6)	30 (6.5)	60 (8.7)
Pooled logistic regression†	1.69 (1.13–2.54)	1.43 (0.91–2.23)	1 (Ref.)	1.20 (0.72–1.99)	1.64 (1.05–2.56)
K/L 3 (pooled no. = 1,496)					
JSN worsening, pooled no. (%)	64 (14.3)	41 (13.0)	37 (14.1)	30 (14.6)	37 (13.9)
Pooled logistic regression†	1.00 (0.63–1.57)	0.98 (0.60–1.61)	1 (Ref.)	1.06 (0.62–1.80)	0.85 (0.51–1.43)
K/L 4 (pooled no. = 62)					
JSN worsening, pooled no. (%)	8 (42.1)	9 (64.3)	4 (28.6)	3 (75.0)	5 (45.5)
Pooled logistic regression†	1.93 (0.41–9.03)	3.25 (0.62–16.98)	1 (Ref.)	11.13 (0.60–204.92)	0.92 (0.07–12.67)

* Values are the odds ratio (95% confidence interval) unless indicated otherwise. JSN = joint space narrowing; K/L = Kellgren/Lawrence grading scale; PASE = Physical Activity Scale for the Elderly; Ref. = reference.
† Adjusted for age, sex, race, body mass index (BMI), knee pain, and knee injury.
‡ Adjusted for age, race, BMI, knee pain, and knee injury.

findings are consistent with those of a study of knees without symptomatic or radiographic OA (WOMAC pain of 0 and K/L grade <2) (8). In that study of T2 relaxation time worsening in OAI participants without pain or knee OA, PASE scores in the highest 33% and lowest 15% were associated with T2 progression, indicative of more rapid cartilage damage in comparison with those in the middle tertile of PASE scores (8).

Overall, studies of the association between PA and worsening of radiographic OA, defined by JSN, have had inconsistent results. Felson et al (26) combined data from the Multicenter Osteoarthritis Study and OAI and determined the effect of PA on knee OA development in 3,542 knees without radiographic knee OA at baseline. In that incidence study, JSN occurred in 3.41% of knees in the active group (highest quartile of PASE score) versus 4.04% in the other groups (OR 0.9 [95% CI 0.5–1.5]). Øiestad et al (30) investigated the association between objectively measured daily walking and knee structural change, defined either as radiographic worsening or as cartilage loss by magnetic resonance imaging, in 1,179 participants age ≥60 years in the Multicenter Osteoarthritis Study. They found no significant associations between daily walking and radiographic worsening or cartilage loss after adjusting for confounders. More recently, Qin et al investigated the association between moderate-to-vigorous PA detected by uniaxial accelerometry and development of incident knee OA in OAI participants without knee OA at baseline (31). No association was detected between PA and risk for developing incident knee OA or JSN over 48 months of follow-up. With regard to cartilage morphology, Racunica et al (32) reported that vigorous PA appeared to have a beneficial effect on knee articular cartilage in 297 healthy, community-based adults with no history of knee injury or disease.

In contrast, consistent with the results of our current study in knees with preexisting pathology, some studies have found associations between higher amounts of PA and JSN worsening. Doré et al (33) reported that walking ≥10,000 steps/day was associated with a greater risk of an increasing cartilage defect score in those with prevalent bone marrow lesions at baseline (risk ratio 1.36 [95% CI 1.03–1.69]) in 405 community-dwelling adults ages 51–81 years. Doré et al suggested that individuals with knee abnormalities should avoid walking ≥10,000 steps/day.

One explanation for conflicting results in these studies may be thresholds for the association between PA and JSN worsening. We found a plateau section in the RCS curves for the associations of PA and JSN worsening. This finding would suggest that a significantly increased risk for JSN worsening may be found in participants with the higher or lower PA amounts. Thus, the seemingly disparate results may be due to different classification methods for PA used in these studies. In prior studies, PA was classified as the lowest versus highest quartile (26), as <5,859 steps/day and >7,846 steps/day versus 5,859–7,846 steps/day (30), as vigorous PA versus less vigorous PA (32), as sedentary activity and moderate-to-vigorous versus light PA (34), or as <10,000 steps/day versus >10,000 steps/day (33). Stratifying into fewer groups (2–3 in these prior studies) may have reduced the ability to detect differences between groups by pooling heterogeneity within the broader categories, thereby potentially failing to detect associations between PA and radiographic worsening or cartilage loss in those studies. Additionally, the studies of incident JSN in people without radiographic OA seem to have different outcomes than the current study of progression of JSN in people with radiographic OA.

Other reasons for differences in findings could include the use of different measures of PA. Jayabalan et al reported on a substudy of OAI that collected accelerometry data at the 48-month visit, finding that worsening of K/L grade between 48 and 96 months was not associated with either light or moderate-to-vigorous PA (34). The participants in that study differed from our current study of worsening OA severity in that 57% of the knees included in the accelerometry study had K/L grade 0 at baseline. In addition, the number of minutes classified as moderate-to-vigorous PA in that study sample was very low (~2.2% of all PA minutes) (34). The focus on mixed incidence and progression of knee OA, as well as the relative lack of higher-level PA in that study, may contribute to divergent findings with our study.

Accelerometers may reflect PA differently or may reflect different types of PA than self-report and also may reflect the fact that survey instruments may suffer from recall bias. The PASE instrument has been validated for adults age ≥55 years (35) over a range of age and health status as well as for older adults with knee pain (19). On a similar instrument (e.g., IPAQ), unlike accelerometers, respondents tended to underestimate sedentary time and overestimate higher intensities of PA, a bias that was most pronounced in men (36). In our study, assessing the magnitude of recall bias on measured PASE scores is difficult. However, based on the RCS results, recall bias seems unlikely to result in 2 opposite trends on the association between PA and JSN worsening, given that both lower reported PA and higher reported PA were associated with a greater risk for worsening JSN in men.

Through subgroup analyses, we found that the associations between PA and JSN worsening were observed in men, but not in women. Very few published studies have explored sex differences when assessing associations between PA and worsening of knee OA. Doré et al (33) did not find any sex differences when examining PA and tibial bone area change. However, Wise et al found that at every level of functional limitation, the risk ratio for total knee arthroplasty was higher for men than women (37). Reasons for sex differences are complex and multifactorial. Bone and muscle strength, alignment, ligamentous laxity, pregnancy history, and neuromuscular activation could contribute to the sex differences (38). Srikanth et al (16) found that women tended to have more severe knee OA, particularly in those age >55 years (after menopause). The age range of 45–79 for participants in our study is in the perimenopausal period. However, we did not find any clear association between PA and JSN worsening in women.

Finding that higher PA was associated with an increased risk of JSN worsening in men, we feel a possible explanation for this result may be a lower amount of PA among women in the study sample.

We found a PASE score >220 was associated with an increased risk of JSN worsening compared with a PASE score of 141–180, after adjusting for other known risk factors for knee OA progression. The 2018 Physical Activity Guidelines for Americans from the Department of Health and Human Services (DHHS) recommend that all adults accumulate at least 150–300 minutes/week of moderate intensity PA in at least 10-minute bouts (39). Furthermore, similar guidelines were issued by other organizations (39–41). In addition, some researchers have reported that 10,000 steps/day was more effective in increasing the PA amount than the DHHS recommendation in low-active overweight and obese populations (42,43). As the PASE score cannot be directly translated into PA level (low, moderate, and high) or steps/day, we cannot directly compare our findings with these recommendations. However, these recommendations focused on the health risks of low PA, specifically to improve cardiorespiratory and muscular fitness, and to reduce the risk of chronic diseases, depression, and cognitive decline. The results of our study suggest that ~15% of the OAI cohort, who have knee OA, may increase the risk for worsening of JSN over 4 years by following recommendations at the upper range of PA. This finding suggests that there may be a need to examine PA recommendations for people with knee OA.

One limitation of our study was the inclusion of OAI participants with radiographic evidence of knee OA at baseline, so the findings are generalizable to similar people, rather than generalizable to the overall population. Another potential limitation was the investigation method of PA in our study. The PASE score is a comprehensive scoring method involving leisure, household, and occupational PA but does not distinguish between the weight-bearing and nonweight-bearing impact of PA. Most likely, only weight-bearing PA aggravates symptoms such as pain and inflammation (44,45). Thus, evaluating the effects for different types of PA on JSN is difficult in our study. This study included knees with radiographic knee OA (K/L grade ≥1) and potential for JSN to worsen (JSN <3) at baseline. Due to different radiographic reading methods, in the OAI data set, 24 knees that were scored K/L grade 4 at baseline were not rated as JSN score of 3 at baseline. While this study excluded JSN grade 3 knees at baseline due to inability for JSN to worsen, it did not exclude these 24 knees that had the potential to demonstrate worsening (4 knees worsened in the 0–12-month, 10 knees worsened in the 12–24-month, and 10 knees worsened in the 24–36-month assessment periods). Sensitivity analyses were completed to determine the effect of including versus excluding these knees, and ORs were unchanged.

Strengths of this study include the large number of participants and outcomes, and the availability of detailed covariates to adjust for a broad range of potential confounders. In summary, compared to moderate PA, higher and lower PA appears to increase the risk for JSN worsening over a 48-month follow-up period in men and in people with a baseline K/L grade of 2.

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. Drs. Hu and Han 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. Hu, Nevitt, Segal.

Acquisition of data. Nevitt.

Analysis and interpretation of data. Hu, Han, Nevitt, Wise, Segal.

REFERENCES

1. Yu D, Jordan KP, Bedson J, Englund M, Blyth F, Turkiewicz A, 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.
2. Garstang SV, Stitik TP. Osteoarthritis: epidemiology, risk factors, and pathophysiology. Am J Phys Med Rehabil 2006;85 Suppl:S2–11.
3. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380:2163–96.
4. Cutler DM. Disability and the future of Medicare. N Engl J Med 2003; 349:1084–5.
5. Lubitz J, Cai L, Kramarow E, Lentzner H. Health, life expectancy, and health care spending among the elderly. N Engl J Med 2003;349:1048–55.
6. Khan KM, Thompson AM, Blair SN, Sallis JF, Powell KE, Bull FC, et al. Sport and exercise as contributors to the health of nations. Lancet 2012;380:59–64.
7. Eckstein F, Hudelmaier M, Putz R. The effects of exercise on human articular cartilage. J Anat 2006;208:491–512.
8. Lin W, Alizai H, Joseph GB, Srikhum W, Nevitt MC, Lynch JA, et al. Physical activity in relation to knee cartilage T2 progression measured with 3 T MRI over a period of 4 years: data from the Osteoarthritis Initiative. Osteoarthritis Cartilage 2013;21:1558–66.
9. Cooper C, Snow S, McAlindon TE, Kellingray S, Stuart B, Coggon D, et al. Risk factors for the incidence and progression of radiographic knee osteoarthritis. Arthritis Rheum 2000;43:995–1000.
10. Calling S, Hedblad B, Engstrom G, Berglund G, Janzon L. Effects of body fatness and physical activity on cardiovascular risk: risk prediction using the bioelectrical impedance method. Scand J Public Health 2006;34:568–75.
11. Ageberg E, Engstrom G, Gerhardsson de Verdier M, Rollof J, Roos EM, Lohmander LS. Effect of leisure time physical activity on severe knee or hip osteoarthritis leading to total joint replacement: a population-based prospective cohort study. BMC Musculoskelet Disord 2012;13:73.
12. Ornetti P, Brandt K, Hellio-Le Graverand MP, Hochberg M, Hunter DJ, Kloppenburg M, et al. OARSI-OMERACT definition of relevant radiological progression in hip/knee osteoarthritis. Osteoarthritis Cartilage 2009;17:856–63.
13. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. Osteoarthritis Cartilage 2007;15 Suppl A:A1–56.
14. Laslett LL, Pelletier JP, Cicuttini FM, Jones G, Martel-Pelletier J. Measuring disease progression in osteoarthritis. Curr Treatm Opt Rheumatol 2016;2:97–110.

15. Glass N, Segal NA, Sluka KA, Torner JC, Nevitt MC, Felson DT, et al. Examining sex differences in knee pain: the multicenter osteoarthritis study. Osteoarthritis Cartilage 2014;22:1100–6.
16. Srikanth VK, Fryer JL, Zhai G, Winzenberg TM, Hosmer D, Jones G. A meta-analysis of sex differences prevalence, incidence and severity of osteoarthritis. Osteoarthritis Cartilage 2005;13:769–81.
17. Segal NA, Glass NA, Torner J, Yang M, Felson DT, Sharma L, et al. Quadriceps weakness predicts risk for knee joint space narrowing in women in the MOST cohort. Osteoarthritis Cartilage 2010;18:769–75.
18. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957;16:494–502.
19. Martin KA, Rejeski WJ, Miller ME, James MK, Ettinger WH Jr, Messier SP. Validation of the PASE in older adults with knee pain and physical disability. Med Sci Sports Exerc 1999;31:627–33.
20. Coggon D, Croft P, Kellingray S, Barrett D, McLaren M, Cooper C. Occupational physical activities and osteoarthritis of the knee. Arthritis Rheum 2000;43:1443–9.
21. Kothari M, Guermazi A, von Ingersleben G, Miaux Y, Sieffert M, Block JE, et al. Fixed-flexion radiography of the knee provides reproducible joint space width measurements in osteoarthritis. Eur Radiol 2004;14:1568–73.
22. The Osteoarthritis Initiative. URL: https://nda.nih.gov/oai/.
23. Felson DT, Zhang Y, Hannan MT, Naimark A, Weissman B, Aliabadi P, et al. Risk factors for incident radiographic knee osteoarthritis in the elderly: the Framingham Study. Arthritis Rheum 1997;40:728–33.
24. Allen KD. Racial and ethnic disparities in osteoarthritis phenotypes. Curr Opin Rheumatol 2010;22:528–32.
25. Zheng H, Chen C. Body mass index and risk of knee osteoarthritis: systematic review and meta-analysis of prospective studies. BMJ Open 2015;5:e007568.
26. Felson DT, Niu J, Yang T, Torner J, Lewis CE, Aliabadi P, et al. Physical activity, alignment and knee osteoarthritis: data from MOST and the OAI. Osteoarthritis Cartilage 2013;21:789–95.
27. Wilder FV, Hall BJ, Barrett JP Jr, Lemrow NB. History of acute knee injury and osteoarthritis of the knee: a prospective epidemiological assessment: the Clearwater Osteoarthritis Study. Osteoarthritis Cartilage 2002;10:611–6.
28. D'Agostino RB, Lee ML, Belanger AJ, Cupples LA, Anderson K, Kannel WB. Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. Stat Med 1990;9:1501–15.
29. Desquilbet L, Mariotti F. Dose-response analyses using restricted cubic spline functions in public health research. Stat Med 2010;29:1037–57.
30. Øiestad BE, Quinn E, White D, Roemer F, Guermazi A, Nevitt M, et al. No association between daily walking and knee structural changes in people at risk of or with mild knee osteoarthritis: prospective data from the Multicenter Osteoarthritis Study. J Rheumatol 2015;42:1685–93.
31. Qin J, Barbour KE, Nevitt MC, Helmick CG, Hootman JM, Murphy LB, et al. Objectively measured physical activity and risk of knee osteoarthritis. Med Sci Sports Exerc 2018;50:277–83.
32. Racunica TL, Teichtahl AJ, Wang Y, Wluka AE, English DR, Giles GG, et al. Effect of physical activity on articular knee joint structures in community-based adults. Arthritis Rheum 2007;57:1261–8.
33. Doré DA, Winzenberg TM, Ding C, Otahal P, Pelletier JP, Martel-Pelletier J, et al. The association between objectively measured physical activity and knee structural change using MRI. Ann Rheum Dis 2013;72:1170–5.
34. Jayabalan P, Kocherginsky M, Chang AH, Rouleau GW, Koloms KL, Lee J, et al. Physical activity and worsening of radiographic findings in persons with or at higher risk of knee osteoarthritis. Arthritis Care Res (Hoboken) 2019;71:198–206.
35. Washburn RA, McAuley E, Katula J, Mihalko SL, Boileau RA. The physical activity scale for the elderly (PASE): evidence for validity. J Clin Epidemiol 1999;52:643–51.
36. Dyrstad SM, Hansen BH, Holme IM, Anderssen SA. Comparison of self-reported versus accelerometer-measured physical activity. Med Sci Sports Exerc 2014;46:99–106.
37. Wise BL, Niu J, Felson DT, Hietpas J, Sadosky A, Torner J, et al. Functional impairment is a risk factor for knee replacement in the Multicenter Osteoarthritis Study. Clin Orthop Relat Res 2015;473:2505–13.
38. Johnson VL, Hunter DJ. The epidemiology of osteoarthritis. Best Pract Res Clin Rheumatol 2014;28:5–15.
39. Physical activity guidelines for Americans. 2nd edition. Washington (DC): US Department of Health and Human Services; 2018.
40. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee IM, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc 2011;43:1334–59.
41. World Health Organization. Global recommendations on physical activity for health. Geneva: World Health Organization; 2010. URL: https://www.ncbi.nlm.nih.gov/books/NBK305057/.
42. Pal S, Cheng C, Ho S. The effect of two different health messages on physical activity levels and health in sedentary overweight, middle-aged women. BMC Public Health 2011;11:204.
43. Jones DB, Richeson NE, Croteau KA, Farmer BC. Focus groups to explore the perceptions of older adults on a pedometer-based intervention. Res Q Exerc Sport 2009;80:710–7.
44. Baliunas AJ, Hurwitz DE, Ryals AB, Karrar A, Case JP, Block JA, et al. Increased knee joint loads during walking are present in subjects with knee osteoarthritis. Osteoarthritis Cartilage 2002;10:573–9.
45. Lephart SM, Pincivero DM, Rozzi SL. Proprioception of the ankle and knee. Sports Med 1998;25:149–55.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1172–1178
DOI 10.1002/acr.24561

In Vivo Compositional Changes in the Articular Cartilage of the Patellofemoral Joint Following Anterior Cruciate Ligament Reconstruction

Michelle C. Boling,[1] Matthew Dupell,[1] Steven J. Pfeiffer,[2] Kyle Wallace,[2] David Lalush,[3] Jeffrey T. Spang,[2] Daniel Nissman,[2] and Brian Pietrosimone[2]

Objective. To compare T1ρ relaxation times of the medial and lateral regions of the patella and femoral trochlea at 6 and 12 months following anterior cruciate ligament reconstruction (ACLR) on the ACLR and contralateral extremity. Greater T1ρ relaxation times are associated with a lower proteoglycan density of articular cartilage.

Methods. This study involved 20 individuals (11 males, 9 females; mean ± SD age 22 ± 3.9 years, weight 76.11 ± 13.48 kg, and height 178.32 ± 12.32 cm) who underwent a previous unilateral ACLR using a patellar tendon autograft. Magnetic resonance images from both extremities were acquired at 6 and 12 months post-ACLR. Voxel by voxel T1ρ relaxation times were calculated using a 5-image sequence. The medial and lateral regions of the femoral trochlea and patellar articular cartilage were manually segmented on both extremities. Separate extremity (ACLR and contralateral extremity) by time (6 months and 12 months) analysis of variance tests were performed for each region ($P < 0.05$).

Results. For the medial patella and lateral trochlea, T1ρ relaxation times increased in both extremities between 6 and 12 months post-ACLR (medial patella $P = 0.012$; lateral trochlea $P = 0.043$). For the lateral patella, T1ρ relaxation times were significantly greater on the contralateral extremity compared to the ACLR extremity ($P = 0.001$). The T1ρ relaxation times of the medial trochlea on the ACLR extremity were significantly greater at 6 ($P = 0.005$) and 12 months ($P < 0.001$) compared to the contralateral extremity. T1ρ relaxation times of the medial trochlea significantly increased from 6 to 12 months on the ACLR extremity ($P = 0.003$).

Conclusion. Changes in T1ρ relaxation times occur within the first 12 months following ACLR in specific regions of the patellofemoral joint on the ACLR and contralateral extremity.

INTRODUCTION

The incidence of anterior cruciate ligament (ACL) injury is highest between ages 14 and 25 years (1) when individuals are more likely to engage in dynamic physical activity. Unfortunately, those who sustain an ACL injury are at increased risk of developing posttraumatic osteoarthritis (OA) regardless of ACL reconstruction (ACLR) and rehabilitation (2,3). Approximately 50% of individuals are reported to have signs of radiographic OA in the tibiofemoral (TF) and/or patellofemoral (PF) joints within 12 years following ACLR (4,5). Younger individuals who desire to be engaged in a high level of physical activity and who develop posttraumatic OA early in life may experience a disproportionately high level of disability compared to individuals who develop idiopathic OA later in life (6). Therefore, a better understanding of the deleterious joint tissue changes that occur following ACL injury may improve the capacity to detect and manage the early development of posttraumatic OA in young physically active individuals.

Compositional changes to articular cartilage often predate morphologic alterations in the early phases of OA

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, the North Carolina Translational and Clinical Sciences Institute, or the National Athletic Trainers Association Research and Education Foundation.

Supported by the NIH/National Institute of Arthritis and Musculoskeletal and Skin Diseases (1R03-AR-066840-01A1), the North Carolina Translational and Clinical Sciences Institute, and the National Athletic Trainers Association Research and Education Foundation (14NewInv001).

[1]Michelle C. Boling, PhD, Matthew Dupell, DPT: University of North Florida, Jacksonville; [2]Steven J. Pfeiffer, MS, Kyle Wallace, BA, Jeffrey T. Spang, MD, Daniel Nissman, MD, Brian Pietrosimone, PhD: University of North Carolina at Chapel Hill; [3]David Lalush, PhD: North Carolina State University and University of North Carolina at Chapel Hill.

No potential conflicts of interest relevant to this article were reported.

Address correspondence to Michelle C. Boling, PhD, University of North Florida, 1 UNF Drive, Jacksonville, FL 32224. Email: m.boling@unf.edu.

Submitted for publication May 26, 2020; accepted in revised form January 12, 2021.

SIGNIFICANCE & INNOVATIONS

- Following unilateral anterior cruciate ligament reconstruction (ACLR), the involved and uninvolved extremities display compositional changes of the articular cartilage of the patellofemoral (PF) joint.
- The findings from this study provide novel evidence that deleterious tissue changes associated with posttraumatic osteoarthritis are occurring within specific regions of the PF joint post-ACLR.

development (7,8). Proteoglycans are macromolecules embedded in the extracellular matrix of articular cartilage and they play a critical role in force attenuation in the tissue (9). Proteoglycan depletion within TF articular cartilage is an early marker associated with OA development (10,11). T1ρ magnetic resonance imaging (MRI) relaxation times are associated with the proteoglycan density of articular cartilage and have been used to evaluate early in vivo compositional changes in articular cartilage (12,13). Increased T1ρ MRI relaxation times, interpreted as decreased proteoglycan density, have been reported in the TF cartilage of the ACLR extremity compared to uninjured extremities within the first 2 years following ACL injury (14–16). Although the prevalence of PF joint OA is reported to be similar to that of TF joint OA following ACLR (17), only a few studies have investigated longitudinal changes in T1ρ relaxation times at the PF joint following ACLR (18,19). The results from these studies demonstrate significant increases in T1ρ relaxation times from baseline to 6 months post-ACLR occurring in the femoral trochlea articular cartilage but no changes in the entire patellar articular cartilage on the injured extremity (18,19). These previous studies assessed T1ρ relaxation times of the articular cartilage averaged across the entire PF cartilage and did not evaluate changes to different regions of the patella or femoral trochlea (18,19), which could allow for a more sensitive analysis.

The PF joint is complex, as different regions of the patella articular cartilage contact multiple regions of the femoral trochlea during normal knee movements required for activities of daily living. Changes to joint tissue metabolism (20) and PF joint biomechanics (21,22) following ACLR may lead to alterations in cartilage composition in different regions of the PF joint. Furthermore, there is evidence to support a higher prevalence of medial PF cartilage damage as compared to lateral PF cartilage damage among individuals with PF OA (23). Due to the high incidence (47%) of PF joint degenerative changes between 5 and 9 years following ACLR (4), more research is needed to understand the factors contributing to the long-term articular cartilage changes within specific regions of the PF joint. Subsectioning the patella and femoral trochlea into medial and lateral regions of interest is important for understanding the nature of compositional changes following ACLR and how these changes may be related to long-term damage at the PF joint. No studies to date have assessed the T1ρ relaxation times for the medial or lateral regions of the PF joint following ACLR.

Therefore, the objective of this investigation was to determine whether changes in T1ρ relaxation times occur in the medial and lateral regions of the patella and femoral trochlea on the ACLR extremity and uninvolved contralateral extremity from 6 to 12 months post-ACLR. We hypothesized that significant longitudinal increases in T1ρ relaxation times would be observed over time in specific regions (medial and lateral) of the PF joint on the ACLR extremity but not on the uninvolved contralateral extremity. A secondary purpose of this study was to determine whether T1ρ relaxation times are greater in specific regions of the PF joint on the ACLR extremity compared to the uninvolved contralateral extremity at both 6 and 12 months post-ACLR. We hypothesized that the T1ρ relaxation times for all regions of the PF joint on the ACLR extremity would be significantly greater than the uninvolved contralateral extremity at both 6 and 12 months post-ACLR.

PATIENTS AND METHODS

Study design. We conducted a longitudinal cohort study from a subset of individuals who underwent MRI analysis as part of a larger prospective longitudinal cohort study. The subset included all individuals with a primary ACLR who had completed the T1ρ MRI collections in both extremities for the 6 and 12 month follow-up examinations at the time of this study. No statistically significant differences were found between the sample used in the current study and the overall cohort for participant age ($P = 0.21$), height ($P = 0.09$), or weight ($P = 0.90$). All participants were initially identified upon presentation in the orthopedic clinic within 14 days of ACL injury and prior to ACLR. Participants who attended both the 6-month (mean ± SD 198.5 ± 23.0 days postsurgery) and 12-month (mean ± SD 369.2 ± 18.6 days postsurgery) follow-up examination after their ACLR were included in the current study.

While formal outpatient rehabilitation services were prescribed following ACLR by each surgeon, the rehabilitation protocol was not standardized across the cohort (24–26). During the rehabilitation process, participants were prescribed therapeutic exercise by their physician to be supervised by an athletic trainer or physical therapist. Standardized evidence-based rehabilitation guidelines were provided to the participants and clinicians to help progress the rehabilitation process (27). The Knee injury and Osteoarthritis Outcomes Score (28) was collected from each patient at the 6- and 12-month follow-up examinations in order to describe the self-reported function of the cohort (Table 1). All participants provided informed consent that was approved by Institutional Review Board at the University of North Carolina at Chapel Hill (13-2385) prior to participating in any research-related procedures.

Participants. Individuals age 18–35 years with a history of a unilateral primary ACL injury were included. We excluded those with a previous history of ACL injury on either extremity,

Table 1. Participant demographic information*

Characteristic	Value
Men/women, no.	11/9
Age, years	22 ± 3.9
Height, cm	178.32 ± 12.32
Mass, kg	76.11 ± 13.48
Concomitant medial meniscus injury, no. (%)	6 (30)
Concomitant lateral meniscus injury, no. (%)	15 (75)
Concomitant chondral injury, no. (%)	8 (40)
KOOS at 6 months	
Symptoms	80.15 ± 11.91
Pain	86.50 ± 8.26
Activities of daily living	96.85 ± 3.66
Sport	69.50 ± 14.95
Quality of life	54.75 ± 18.36
KOOS at 12 months	
Symptoms	84.85 ± 9.46
Pain	91.70 ± 7.41
Activities of daily living	97.35 ± 4.18
Sport	84.00 ± 14.47
Quality of life	74.50 ± 17.96

* Values are the mean ± SD unless indicated otherwise. KOOS = Knee injury and Osteoarthritis Outcome Score.

as well as those who sustained a second ACL injury at any point during the observation period. We did not exclude individuals with a concomitant meniscal or chondral injury. Those who were pregnant at the time of consent or planned to become pregnant during the 12-month observation period, had been previously diagnosed with any form of arthritis, needed a multiligament reconstruction, or were not planning to undergo ACLR were excluded.

All participants underwent a unilateral arthroscopically assisted single incision ACLR (mean ± SD 31 ± 16 days following ACL injury) using a patellar tendon autograft performed by 1 of 3 participating orthopedic surgeons as previously reported (29). Briefly, the middle third of the patellar tendon was harvested via an anterior longitudinal incision. Next, a target was determined on the lateral wall of the intercondylar notch of the femur and a femoral tunnel was drilled through the inframedial arthroscopic portal with the knee in 120° of flexion. A pin was drilled and over-reamed into the ACL footprint from the inframedial tibia to create a tibial tunnel. The proximal bone-plug of the patellar tendon graft was affixed to the femur with a metal interference screw. Finally, a metal interference screw was used to affix the distal bone-plug of the patellar tendon graft to the tibia. The attending orthopedic surgeon recorded data regarding meniscal and articular cartilage injury at the time of surgery.

Previous work has demonstrated interextremity effects that range from moderate to strong for articular cartilage T1ρ relaxation times at 12 months following ACLR in different TF regions of interest (30). Based on these data, we estimated (G*Power software, version 3.1.9.2) that 20 individuals would be needed to demonstrate a statistically significant difference between extremities and over time if a moderate effect was found (d = 0.65, 1-β = 0.8, α = 0.05).

MRI acquisition. MRI images from both extremities were acquired using either a Siemens Magnetom TIM Trio 3T scanner with a 4-channel Siemens large flex coil (516 mm × 224 mm) or a Siemens Magnetom Prisma 3T PowerPack scanner with an XR 80/200 gradient coil (60 cm × 213 cm). Strong interscanner reliability for absolute agreement has been previously determined for T1ρ relaxation times in the entire medial (intraclass coefficient $[ICC]_{2,1}$ 0.99) and lateral ($ICC_{2,1}$ 0.96) weight-bearing regions of the femoral condyle in a separate cohort of 6 knees assessed in both scanners approximately 45 days apart (25). Upon arrival to the imaging center, participants remained seated for 30 minutes to unload the knee cartilage (31). We used a T1ρ prepared 3-dimensional fast low-angle shot with a spin-lock power at 500 Hz, 5 different spin-lock durations (40, 30, 20, 10, and 0 msec) and a voxel size of 0.8 mm × 0.4 mm × 3 mm (field of view 288 mm, slice thickness 3.0 mm, time to recovery 9.2 msec, 160 × 320 matrix, gap 0 mm, flip angle 10°, echo-train duration time 443 msec, phase encode direction of anterior/posterior) (30,32).

T1ρ relaxation time quantification, registration, and segmentation. Prior to segmentation, an affine technique was performed to register the ACLR extremity image to the uninjured extremity image using the 0 msec spin-lock image with 3D Slicer software (http://www.slicer.org) (33). Following the affine registration, a nonrigid deformable, voxel-by-voxel intensity-based registration technique was applied to accurately align the ACLR femur and tibia to that of the uninvolved contralateral extremity at each time point. The articular cartilage of the femoral trochlea and the patella acquired during the 0 msec spin-lock duration were manually segmented using ITK-SNAP software (version 3.6; http://www.itksnap.org) (34) for both the ACLR and uninvolved contralateral extremities. Following the initial segmentation, the articular cartilage of the femoral trochlea and patella were evenly divided into medial and lateral regions of interest. We separately determined the total number of image slices that included the patella and femur and divided each bone in half to derive the medial and lateral regions of interest for the patella and femoral trochlea. The medial and lateral regions of interest of the femoral trochlea and the patella were included in the data analysis. Voxel-by-voxel T1ρ relaxation times were calculated using a 5-image sequence created with a MatLab program (R2014b [8.4.0]) with the following equation: $S(TSL) = S_0 \exp(-TSL/T1\rho)$ (14), where TSL is the duration of the spin-lock time, S_0 is signal intensity when TSL equals zero, S corresponds to signal intensity, and T1ρ is the T1 relaxation time in the rotating frame, as previously reported (30,32).

Statistical analysis. Means ± SDs were calculated for all continuous demographic variables and T1ρ relaxation times for all regions of interest, while frequencies were counted for all noncontinuous demographic variables. Data distributions were assessed using the Shapiro-Wilk test for normality, and stem

and leaf plots were visually inspected for potential outliers. Separate extremity (ACLR extremity and uninvolved contralateral extremity) by time (6 months and 12 months) analysis of variance tests were performed for each region of interest (medial patella, lateral patella, medial trochlea, and lateral trochlea). There were no covariates included in this analysis. An a priori level of significance for all analyses was set at a P value less than 0.05, and all analyses were performed using SPSS software, version 21.0.

RESULTS

Demographic information. Table 1 shows the demographic characteristics of the participants. If an individual sustained a concomitant meniscal injury or chondral injury, this injury was addressed during the ACLR procedure (medial meniscal tear: 6 repairs; lateral meniscal tear: 4 repairs, 7 meniscectomy, 3 repairs and meniscectomy, 1 tear did not require surgical intervention; 8 chondral injuries: 1 chondroplasty, 1 microfracture, 6 did not require surgical intervention). Table 2 shows the average T1ρ relaxation times in all regions of interest on the ACLR and uninvolved contralateral extremity at 6 and 12 months post-ACLR. All outcome measures were normally distributed.

Patella articular cartilage. For the medial patella, T1ρ relaxation times increased in both extremities from 6 to 12 months post-ACLR ($F[1,19] = 7.79$, $P = 0.012$; $\eta_\rho^2 = 0.29$; mean difference 1.76 msec [95% confidence interval (95% CI) –0.04, 3.55]). For the lateral patella, T1ρ relaxation times were significantly greater on the uninvolved contralateral extremity compared to the ACLR extremity ($F[1,19] = 14.156$, $P = 0.001$; $\eta_\rho^2 = 0.43$; mean difference 2.47 msec [95% CI 0.13, 4.82]).

Trochlear articular cartilage. For the lateral trochlea, T1ρ relaxation times increased in both extremities from 6 to 12 months post-ACLR ($F[1,19] = 4.698$, $P = 0.043$; $\eta_\rho^2 = 0.20$; mean difference 1.68 msec [95% CI –0.32, 3.67]) and the T1ρ relaxation times were significantly greater on the ACLR extremity compared to the uninvolved contralateral extremity ($F[1,19] = 11.311$, $P = 0.003$; $\eta_\rho^2 = 0.37$; mean difference 2.47 msec [95% CI 0.51, 4.43]). A significant extremity by time interaction was found for T1ρ relaxation times in the medial trochlea ($F[1,19] = 6.136$, $P = 0.023$; $\eta_\rho^2 = 0.24$). T1ρ relaxation times in the medial trochlea on the ACLR extremity were significantly greater at 6 ($P = 0.005$; mean difference 2.62 msec [95% CI 0.87, 4.37]) and 12 months ($P < 0.001$; mean difference 4.90 msec [95% CI 3.52, 6.28]) post-ACLR compared to the uninvolved contralateral extremity. T1ρ relaxation times in the medial trochlea on the ACLR extremity significantly increased from 6 to 12 months on the ACLR extremity ($P = 0.003$; mean difference –3.41 msec [95% CI –5.48, –1.34]). There were no significant changes in T1ρ relaxation times of the medial trochlea from 6 to 12 months on the uninvolved contralateral extremity ($P = 0.163$; mean difference –1.13 msec [95% CI –2.77, 0.50]).

DISCUSSION

In agreement with our hypotheses, T1ρ relaxation times in the medial femoral trochlea were greater in the ACLR extremity compared to the uninvolved contralateral extremity at both time points and increased in the ACLR extremity between 6 and 12 months post-ACLR. In partial agreement with our hypotheses, we found that T1ρ relaxation times in the medial patella and lateral trochlea increased bilaterally between 6 and 12 months following

Table 2. T1ρ relaxation times (msec) at 6 months and 12 months post-ACLR for each region of interest*

	6 months	12 months	Mean difference (95% CI)
Medial patella†			
ACLR	53.67 ± 3.74	54.62 ± 4.37	–0.95 (–2.90, 1.01)
Uninvolved	52.94 ± 2.77	55.51 ± 2.08	–2.57 (–4.01, –1.12)
Mean difference (95% CI)	0.73 (–0.92, 2.38)	–0.89 (–2.6, 0.89)	–
Lateral patella‡			
ACLR	55.24 ± 4.53	54.52 ± 5.62	0.71 (–1.82, 3.25)
Uninvolved	56.96 ± 3.26	57.73 ± 3.59	–0.77 (–2.51, 0.97)
Mean difference (95% CI)	–1.73 (–3.42, –0.03)	–3.21 (–4.97, –1.45)	–
Medial trochlea			
ACLR	57.69 ± 4.50	61.10 ± 4.22	–3.41 (–5.48, –1.34)†
Uninvolved	55.07 ± 3.01	56.20 ± 3.88	–1.13 (–2.77, 0.50)
Mean difference (95% CI)	2.62 (0.87, 4.37)‡	4.90 (3.52, 6.28)‡	–
Lateral trochlea§			
ACLR	56.10 ± 3.23	57.80 ± 4.69	–1.70 (–3.62, 0.22)
Uninvolved	53.65 ± 3.22	55.30 ± 3.03	–1.66 (–3.28, –0.03)
Mean difference (95% CI)	2.45 (1.04, 3.86)	2.50 (0.54, 4.46)	–

* Values are the mean ± SD unless indicated otherwise. 95% CI = 95% confidence interval; ACLR = anterior cruciate ligament reconstruction.
† Significant increase from 6 to 12 months for main effect or interaction effect ($P < 0.05$).
‡ Significant difference between extremities for main effect or interaction effect ($P < 0.05$).
§ Significant increase from 6 to 12 months for main effect or interaction effect, and significant difference between extremities for main effect or interaction effect ($P < 0.05$).

unilateral ACLR; yet T1ρ relaxation times were always higher in the lateral trochlea on the ACLR extremity compared to the uninvolved contralateral extremity. Contrary to our hypotheses, we found that T1ρ relaxation times were greater in the lateral patellar cartilage on the uninvolved contralateral extremity compared to the ACLR extremity. Overall, these findings provide novel evidence that specific regions of the PF joint may be more susceptible to deleterious tissue changes associated with posttraumatic OA and that these compositional changes in specific portions of the PF joint occur bilaterally following unilateral ACLR.

Previous studies have reported longitudinal T1ρ relaxation time changes in the femoral trochlea of the ACLR extremity and patella on the uninvolved contralateral extremity (18,19). Both Amano et al (19) and Pedoia et al (18) reported a significant increase in T1ρ relaxation times of the articular cartilage in the femoral trochlea between ACL injury (prior to ACLR) and 6 months post-ACLR. In addition, Pedoia et al (18) reported increased T1ρ relaxation times at the patella on the uninvolved contralateral extremity from ACL injury (prior to ACLR) to 6 months. The findings from our study build on previous work, by defining more specific regions of the PF joint that demonstrate changes in T1ρ relaxation times.

The underlying mechanisms leading to these compositional changes in PF joint articular cartilage composition remain unclear, yet both biochemical and biomechanical changes following ACLR may be important factors related to these degenerative joint tissue changes. Increased proinflammatory cytokines and degenerative enzymes associated with cartilage breakdown have been reported within the first 12 months following ACL injury and ACLR (20). In addition to biochemical changes, both overloading and underloading of the lower extremity during dynamic tasks have been associated with deleterious cartilage compositional changes following ACLR (25,35,36).

Aberrant gait biomechanics, which impact knee joint loading, are common following ACLR (37). Individuals with an ACLR demonstrate altered joint loading (decreased vertical ground reaction force [vGRF] in early stance and increased vGRF in midstance) of the ACLR extremity and uninvolved contralateral extremity during the stance phase of gait in the first 12 months following ACLR compared to uninjured controls (38). During more dynamic tasks, such as squatting and jumping, individuals with an ACLR display decreased vGRF on the ACLR extremity compared to the uninvolved contralateral extremity 1 to 2 years post-ACLR (39–42). The changes in vGRF likely lead to alterations in loads placed across the PF joint. In a recent study investigating PF joint contact forces during running 12–24 months post-ACLR, peak PF joint contact forces were decreased on the ACLR extremity as compared to the uninvolved contralateral extremity (22). Decreased loading on the ACLR extremity has been associated with altered TF cartilage composition (25) but whether decreased loading leads to changes in PF joint cartilage composition is not clear. Future research needs to investigate how these changes in joint loading contribute to articular cartilage changes at the PF joint.

There is also evidence to suggest that specific kinematic changes occur at the PF joint following ACLR. Lin et al (43) demonstrated a significant increase in patellar external rotation, lateral patellar tilt, and lateral translation following ACLR. These changes in PF kinematics could be expected to lead to increased loads placed across the lateral patella and femoral trochlea and decreased loads across the medial patella and femoral trochlea. Our study demonstrated that the T1ρ relaxation times in the articular cartilage of medial femoral trochlea significantly increased between 6 and 12 months post-ACLR in the injured extremity, which could be a sign of altered loading across this joint. Further investigation into changes in PF kinematics on the ACLR extremity following ACLR and their effect on T1ρ relaxation times is needed to better understand how these factors influence articular cartilage compositional changes at the PF joint.

While underloading the ACLR extremity following ACLR appears to be a general tendency, we should be aware of the increased loads placed on the uninvolved contralateral extremity during dynamic tasks. The alterations in T1ρ relaxation times of the medial and lateral compartments of the patella and lateral trochlea on the uninvolved contralateral extremity in this study could be attributed to increased loading on the contralateral extremity during dynamic activities. Evidence supports increased vGRF during squatting and jumping on the uninvolved contralateral extremity compared to the ACLR extremity (39–42). The higher vGRF on the uninvolved contralateral extremity could lead to increased loads placed on the PF joint and compositional alterations from overloading the articular cartilage (44).

The finding of compositional cartilage changes at the PF joint on both the ACLR and uninvolved contralateral extremity highlights the need for clinicians to focus on both extremities following ACLR to ensure that symmetrical and sufficient loading is restored upon return to sport or activity. Gaining an understanding of optimal loading of articular cartilage and the ability to counteract these deleterious articular cartilage changes through interventions should help to inform the development of effective interventions post-ACLR. Increases in T1ρ relaxation times of the knee articular cartilage following a period of non–weight-bearing have been shown to be transient and return to baseline levels upon return to normal loading of the joint (45), although how adjusting loading in those with knee injury may impact proteoglycan density over time is unknown. More research is needed to understand whether there are intervention strategies that can be implemented to improve proteoglycan concentrations in articular cartilage at the PF joint post-ACLR.

When interpreting the results of this study, awareness that all participants in this investigation underwent a bone patellar tendon bone autograft is important. Increased degenerative changes have been reported at the PF joint as compared to the TF joint 7 years post-ACLR in individuals undergoing a bone patellar tendon bone

autograft (46). Whether these same changes in T1ρ relaxation times within specific regions of the PF joint would also occur following other ACL graft procedures is unknown. Furthermore, changes in T1ρ relaxation times in the PF joint do not clearly predispose individuals to the development of posttraumatic OA. The articular cartilage in the femoral trochlea has been reported to thin 24 months post-ACLR (47). Yet additional research is needed to determine whether an increase in T1ρ relaxation times between 6 and 12 months post-ACLR will result in the eventual thinning of articular cartilage in the femoral trochlea or patella.

While this is the first study to evaluate T1ρ relaxation times within specific regions of the PF joint post-ACLR, there are some limitations that can inform future investigations. We did not collect baseline T1ρ relaxation times in this cohort; therefore, we cannot determine preinjury T1ρ relaxation times or how cartilage composition may have changed within the first 6 months following ACLR on the ACLR extremity and uninvolved contralateral extremity. Furthermore, proteoglycan density changes in the articular cartilage may be reversable, and we do not know whether the changes in T1ρ relaxation times are transient or if these changes will continue. Longer follow-up times are needed to determine if the changes in T1ρ relaxation times at the PF joint following ACLR lead to chronic symptoms and radiographic PF OA on the ACLR or uninvolved contralateral extremity.

We also did not assess the influence of patient function, rehabilitation progression, or their physical activity levels at the time of the 6- and 12-month MRI acquisitions. An assessment of the progression of the participant through their rehabilitation program and their physical activity levels at each time point could help to inform the interpretation of the biologic changes occurring within the PF joint post-ACLR. We recommend the inclusion of an assessment of physical activity levels and rehabilitation progress in future investigations on articular cartilage changes at the PF joint following ACLR. While all participants were prescribed therapeutic exercise and provided standardized evidence-based rehabilitation guidelines, we did not standardize the rehabilitation protocol in this study. Finally, this is a relatively small subset of individuals from a larger investigation. Due to the sample size, we were unable to determine how other covariates may have impacted the change in the T1ρ relaxation times (i.e., sex, age, concomitant meniscal/chondral injury).

In conclusion, compositional changes in articular cartilage occur within the first 12 months following ACLR in specific regions of the PF joint on the ACLR and uninvolved contralateral extremity. These compositional changes may increase the risk for the development of posttraumatic OA early in the PF joint. Continued research is needed to understand whether these compositional changes at the PF joint are associated with biochemical and/or biomechanical changes that occur following ACLR and whether these changes lead to long-term damage at the PF joint. Additional research on compositional changes in the articular cartilage of the PF joint following ACLR could help to inform the development of effective treatment strategies aimed at preventing the development of posttraumatic OA at the PF joint.

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. Boling 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. Boling, Pfeiffer, Wallace, Lalush, Pietrosimone.

Acquisition of data. Pfeiffer, Wallace, Spang, Nissman, Pietrosimone.

Analysis and interpretation of data. Boling, Dupell, Pfeiffer, Wallace, Pietrosimone.

REFERENCES

1. Sanders TL, Maradit Kremers H, Bryan AJ, Larson DR, Dahm DL, Levy BA, et al. Incidence of anterior cruciate ligament tears and reconstruction: a 21-year population-based study. Am J Sports Med 2016; 44:1502–7.
2. Daniel DM, Stone ML, Dobson BE, Fithian DC, Rossman DJ, Kaufman KR. Fate of the ACL-injured patient: a prospective outcome study. Am J Sports Med 1994;22:632–44.
3. Luc B, Gribble PA, Pietrosimone BG. Osteoarthritis prevalence following anterior cruciate ligament reconstruction: a systematic review and numbers-needed-to-treat analysis. J Athl Train 2014;49:806–19.
4. Jarvela T, Paakkala T, Kannus P, Jarvinen M. The incidence of patellofemoral osteoarthritis and associated findings 7 years after anterior cruciate ligament reconstruction with a bone-patellar tendon-bone autograft. Am J Sports Med 2001;29:18–24.
5. Lohmander LS, Östenberg A, Englund M, Roos H. High prevalence of knee osteoarthritis, pain, and functional limitations in female soccer players twelve years after anterior cruciate ligament injury. Arthritis Rheum 2004;50:3145–52.
6. Ackerman IN, Bucknill A, Page RS, Broughton NS, Roberts C, Cavka B, et al. The substantial personal burden experienced by younger people with hip or knee osteoarthritis. Osteoarthritis Cartilage 2015;23:1276–84.
7. Dijkgraaf LC, de Bont LG, Boering G, Liem RS. The structure, biochemistry, and metabolism of osteoarthritic cartilage: a review of the literature. J Oral Maxillofac Surg 1995;53:1182–92.
8. Li X, Majumdar S. Quantitative MRI of articular cartilage and its clinical applications. J Magn Reson Imaging 2013;38:991–1008.
9. Lohmander S. Proteoglycans of joint cartilage: structure, function, turnover and role as markers of joint disease. Baillieres Clin Rheumatol 1988;2:37–62.
10. Moskowitz RW, Howell DS, Goldberg VM, Muniz O, Pita JC. Cartilage proteoglycan alterations in an experimentally induced model of rabbit osteoarthritis. Arthritis Rheum 1979;22:155–63.
11. Young AA, McLennan S, Smith MM, Smith SM, Cake MA, Read RA, et al. Proteoglycan 4 downregulation in a sheep meniscectomy model of early osteoarthritis. Arthritis Res Ther 2006;8:R41.
12. Hatcher CC, Collins AT, Kim SY, Michel LC, Mostertz WC III, Ziemian SN, et al. Relationship between T1rho magnetic resonance imaging, synovial fluid biomarkers, and the biochemical and biomechanical properties of cartilage. J Biomech 2017;55:18–26.
13. Regatte RR, Akella SV, Lonner JH, Kneeland JB, Reddy R. T1rho relaxation mapping in human osteoarthritis (OA) cartilage: comparison of T1rho with T2. J Magn Reson Imaging 2006;23:547–53.
14. Theologis AA, Haughom B, Liang F, Zhang Y, Majumdar S, Link TM, et al. Comparison of T1rho relaxation times between ACL-

reconstructed knees and contralateral uninjured knees. Knee Surg Sports Traumatol Arthrosc 2014;22:298–307.

15. Su F, Hilton JF, Nardo L, Wu S, Liang F, Link TM, et al. Cartilage morphology and T1rho and T2 quantification in ACL-reconstructed knees: a 2-year follow-up. Osteoarthritis Cartilage 2013;21: 1058–67.
16. Li AK, Pedoia V, Tanaka M, Souza RB, Ma CB, Li X. Six-month post-surgical elevations in cartilage T1rho relaxation times are associated with functional performance 2 years after ACL reconstruction. J Orthop Res 2020;38:1132–40.
17. Culvenor AG, Cook JL, Collins NJ, Crossley KM. Is patellofemoral joint osteoarthritis an under-recognised outcome of anterior cruciate ligament reconstruction? A narrative literature review. Br J Sports Med 2013;47:66–70.
18. Pedoia V, Su F, Amano K, Li Q, McCulloch CE, Souza RB, et al. Analysis of the articular cartilage T1rho and T2 relaxation times changes after ACL reconstruction in injured and contralateral knees and relationships with bone shape. J Orthop Res 2017;35: 707–17.
19. Amano K, Li AK, Pedoia V, Koff MF, Krych AJ, Link TM, et al. Effects of surgical factors on cartilage can be detected using quantitative magnetic resonance imaging after anterior cruciate ligament reconstruction. Am J Sports Med 2017;45:1075–84.
20. Harkey MS, Luc BA, Golightly YM, Thomas AC, Driban JB, Hackney AC, et al. Osteoarthritis-related biomarkers following anterior cruciate ligament injury and reconstruction: a systematic review. Osteoarthritis Cartilage 2015;23:1–12.
21. Van de Velde SK, Gill TJ, DeFrate LE, Papannagari R, Li G. The effect of anterior cruciate ligament deficiency and reconstruction on the patellofemoral joint. Am J Sports Med 2008;36:1150–9.
22. Sritharan P, Schache AG, Culvenor AG, Perraton LG, Bryant AL, Crossley KM. Between-extremity differences in patellofemoral joint forces during running at 12 to 24 months after unilateral anterior cruciate ligament reconstruction. Am J Sports Med 2020;48:1711–9.
23. Gross KD, Niu J, Stefanik JJ, Guermazi A, Roemer FW, Sharma L, et al. Breaking the law of valgus: the surprising and unexplained prevalence of medial patellofemoral cartilage damage. Ann Rheum Dis 2012;71:1827–32.
24. Pietrosimone B, Pfeiffer SJ, Harkey MS, Wallace K, Hunt C, Blackburn JT, et al. Quadriceps weakness associates with greater T1rho relaxation time in the medial femoral articular cartilage 6 months following anterior cruciate ligament reconstruction. Knee Surg Sports Traumatol Arthrosc 2019;27:2632–42.
25. Pfeiffer SJ, Spang J, Nissman D, Lalush D, Wallace K, Harkey MS, et al. Gait mechanics and T1rho MRI of tibiofemoral cartilage 6 months after ACL reconstruction. Med Sci Sports Exerc 2019;51:630–9.
26. Pietrosimone B, Blackburn JT, Padua DA, Pfeiffer SJ, Davis HC, Luc-Harkey BA, et al. Walking gait asymmetries 6 months following anterior cruciate ligament reconstruction predict 12-month patient-reported outcomes. J Orthop Res 2018;36:2932–40.
27. Adams D, Logerstedt DS, Hunter-Giordano A, Axe MJ, Snyder-Mackler L. Current concepts for anterior cruciate ligament reconstruction: a criterion-based rehabilitation progression. J Orthop Sports Phys Ther 2012;42:601–14.
28. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS): development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998;28:88–96.
29. Pietrosimone B, Loeser RF, Blackburn JT, Padua DA, Harkey MS, Stanley LE, et al. Biochemical markers of cartilage metabolism are associated with walking biomechanics 6-months following anterior cruciate ligament reconstruction. J Orthop Res 2017;35:2288–97.
30. Pietrosimone B, Nissman D, Padua DA, Blackburn JT, Harkey MS, Creighton RA, et al. Associations between cartilage proteoglycan density and patient outcomes 12 months following anterior cruciate ligament reconstruction. Knee 2018;25:118–29.
31. Souza RB, Stehling C, Wyman BT, Hellio Le Graverand MP, Li X, Link TM, et al. The effects of acute loading on T1rho and T2 relaxation times of tibiofemoral articular cartilage. Osteoarthritis Cartilage 2010; 18:1557–63.
32. Pfeiffer S, Harkey MS, Stanley LE, Blackburn JT, Padua DA, Spang JT, et al. Associations between slower walking speed and T1ρ magnetic resonance imaging of femoral cartilage following anterior cruciate ligament reconstruction. Arthritis Care Research (Hoboken) 2018;70:1132–40.
33. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging 2012;30:1323–41.
34. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 2006;31:1116–28.
35. Teng HL, Wu D, Su F, Pedoia V, Souza RB, Ma CB, et al. Gait characteristics associated with a greater increase in medial knee cartilage T1rho and T2 relaxation times in patients undergoing anterior cruciate ligament reconstruction. Am J Sports Med 2017;45:3262–71.
36. Teng HL, Pedoia V, Link TM, Majumdar S, Souza RB. Local associations between knee cartilage T1rho and T2 relaxation times and patellofemoral joint stress during walking: a voxel-based relaxometry analysis. Knee 2018;25:406–16.
37. Slater LV, Hart JM, Kelly AR, Kuenze CM. Progressive changes in walking kinematics and kinetics after anterior cruciate ligament injury and reconstruction: a review and meta-analysis. J Athl Train 2017; 52:847–60.
38. Davis-Wilson HC, Pfeiffer SJ, Johnston CD, Seeley MK, Harkey MS, Blackburn JT, et al. Bilateral gait 6 and 12 months post-anterior cruciate ligament reconstruction compared with controls. Med Sci Sports Exerc 2020;52:785–94.
39. Chmielewski TL, Wilk KE, Snyder-Mackler L. Changes in weight-bearing following injury or surgical reconstruction of the ACL: relationship to quadriceps strength and function. Gait Posture 2002;16: 87–95.
40. Neitzel JA, Kernozek TW, Davies GJ. Loading response following anterior cruciate ligament reconstruction during the parallel squat exercise. Clin Biomech (Bristol, Avon) 2002;17:551–4.
41. Paterno MV, Ford KR, Myer GD, Heyl R, Hewett TE. Extremity asymmetries in landing and jumping 2 years following anterior cruciate ligament reconstruction. Clin J Sport Med 2007;17:258–62.
42. Renner KE, Franck CT, Miller TK, Queen RM. Extremity asymmetry during recovery from anterior cruciate ligament reconstruction. J Orthop Res 2018;36:1887–93.
43. Lin Z, Tang Y, Tan H, Cai D. Patellofemoral kinematic characteristics in anterior cruciate ligament deficiency and reconstruction. BMC Musculoskelet Disord 2019;20:82.
44. Martin JA, Buckwalter JA. Post-traumatic osteoarthritis: the role of stress induced chondrocyte damage. Biorheology 2006;43:517–21.
45. Souza RB, Baum T, Wu S, Feeley BT, Kadel N, Li X, et al. Effects of unloading on knee articular cartilage T1rho and T2 magnetic resonance imaging relaxation times: a case series. J Orthop Sports Phys Ther 2012;42:511–20.
46. Jarvela T, Kannus P, Jarvinen M. Anterior cruciate ligament reconstruction in patients with or without accompanying injuries: a re-examination of subjects 5 to 9 years after reconstruction. Arthroscopy 2001;17:818–25.
47. Frobell RB. Change in cartilage thickness, posttraumatic bone marrow lesions, and joint fluid volumes after acute ACL disruption: a two-year prospective MRI study of sixty-one subjects. J Bone Joint Surg Am 2011;93:1096–103.

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1179–1187
DOI 10.1002/acr.24563

American College of Rheumatology
Empowering Rheumatology Professionals

Kellgren/Lawrence Grading in Cohort Studies: Methodological Update and Implications Illustrated Using Data From a Dutch Hip and Knee Cohort

Erin M. Macri, Jos Runhaar, Jurgen Damen, Edwin H. G. Oei, and Sita M. A. Bierma-Zeinstra

Objective. The Cohort Hip and Cohort Knee (CHECK) is a cohort of middle-aged individuals with hip or knee pain. Radiographs were assigned Kellgren/Lawrence (K/L) scores under different conditions at each follow-up visit for 10 years. We aimed to describe and consolidate each scoring approach, then illustrate implications of their use by comparing baseline K/L scores assigned using 2 of these approaches, and evaluating their respective associations with joint replacement and incident radiographic osteoarthritis (ROA).

Methods. We compared baseline K/L scores assigned to hips and knees using 2 scoring approaches: 1) assigned by senior researchers to baseline images alone and 2) assigned by trained readers, with images read paired and in known sequence with up to 10 years of follow-up radiographs (Poisson regression). We evaluated the associations of baseline ROA (any: K/L grade ≥1; established: K/L ≥2) with joint replacement, and of K/L 1 joints with incident established ROA (survival analysis).

Results. Of 1,002 participants (79% women, mean ± SD age 55.9 ± 5.2 years, body mass index 26.2 ± 4.0 kg/m^2), the second scoring approach had 2.4 times (95% confidence interval [95% CI] 1.8–3.1 for knees) and 2.9 times (95% CI 2.3–3.7 for hips) higher prevalence of established ROA than the first approach. Established hip ROA had a higher risk of joint replacement using the first approach (hazard ratio [HR] 24.2 [95% CI 15.0–39.8] versus second approach HR 7.7 [95% CI 4.9–12.1]), as did knees (HR 19.3 [95% CI 10.3–36.1] versus second approach HR 4.8 [95% CI 2.4–9.6]). The risk of incident ROA did not differ by approach.

Conclusion. This study demonstrates that evaluating ROA prevalence and predicting outcomes depends on the scoring approach.

INTRODUCTION

The presence and severity of knee or hip radiographic osteoarthritis (ROA) is commonly graded using the Kellgren/Lawrence (K/L) method (1). This semiquantitative approach primarily evaluates osteophytes and joint space narrowing to assign a score between 0 (no ROA) to 4 (severe ROA) (1–3). ROA is typically defined as K/L grade ≥2.

In cohort studies, standardized procedures to assign K/L scores include using a grading atlas, blinding readers to clinical features (e.g., pain), and reading radiographs paired with known sequence order (4–11). Reading single images (blinded to identity and sequence) is less sensitive to ROA progression compared to reading paired images, regardless of whether sequence is known (4,5,7). Reading paired images with known sequence has higher interrater reliability and sensitivity to ROA progression (4–11). However, blinding to sequence reduces bias (7), so although both methods do not significantly differ (5), some cohorts blind readers to sequence (12).

Reading conditions like single image versus paired are a source of error that can lead to misclassifying individuals regarding ROA prevalence (4–11). Other factors also influence scores, notably the somewhat arbitrary and subjective distinction between K/L grades 1 and 2 (13). Image-related factors include image acquisition plane, radioanatomic positioning, and image quality (14,15). Reader-related

The Cohort Hip and Cohort Knee study was initiated and funded by the Dutch Arthritis Association. Dr. Macri's work was supported by a Canadian Institutes of Health Research Banting Postdoctoral Fellowship. Dr. Bierma-Zeinstra's work was supported by The Netherlands Organization for Health Research and Development, Stichting Het Centraal Ziekenfonds, European Union, Foreum, Dutch Arthritis Association, and the Osteoarthritis Research Society International.

Erin M. Macri, PhD, Jos Runhaar, PhD, Jurgen Damen, MD, GP, PhD, Edwin H. G. Oei, MD, PhD, Sita M. A. Bierma-Zeinstra, PhD: Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands.

Dr. Bierma-Zeinstra has received consulting fees from Pfizer (less than $10,000). No other disclosures relevant to this article were reported.

Address correspondence to Erin M. Macri, PhD, Doctor Molewaterplein 40, 3015 GD Rotterdam, The Netherlands. Email: e.macri@erasmusmc.nl.

Submitted for publication June 20, 2020; accepted in revised form January 12, 2021.

SIGNIFICANCE & INNOVATIONS

- The prevalence of established hip or knee radiographic osteoarthritis (OA; Kellgren/Lawrence grade ≥2) was 2.4 to 2.9 times higher when assigning scores based on paired readings with known sequence as read by expert or trained readers compared to expert readers reading a single image.
- The highest hazard ratio for undergoing future hip or knee replacement in participants with established radiographic OA was when scores were read at a single time by expert readers (compared to paired reading in known sequence as read by expert or trained readers).
- The highest number of joints correctly classified as undergoing future hip or knee replacement occurred when images were read paired and in known sequence by expert or trained readers, and when OA was defined at a lower threshold of Kellgren/Lawrence grade ≥1.
- These findings highlight the importance of considering both radiographic scoring conditions as well as the threshold for defining OA when interpreting study results or designing new trials.

factors include training and experience (16,17). One cohort study reported "wobbles" over time whereby scores fluctuated between being classified as ROA or not (12). Further complicating the challenges of correctly classifying ROA, some researchers define ROA as K/L grade ≥1 (doubtful osteophytes), particularly in early OA research (18,19). Appreciating the extent to which reading conditions and ROA definitions influence misclassification could improve interpretation of study results and inform future study design.

The Cohort Hip and Cohort Knee (CHECK) study followed middle-aged individuals with knee or hip pain for 10 years (20). At each visit, radiographs were read and scored under different conditions as new images became available. Therefore, different CHECK publications (21–24) may have used different K/L scores, reflecting score wobble over time (12). This variation could confuse study interpretation among CHECK studies.

Our main aim was to describe and consolidate the knee and hip radiographic K/L scoring methods used in the CHECK cohort at each visit. Second, we aimed to compare the relative prevalence of baseline ROA using 2 different scoring approaches (single reading by expert readers versus paired readings of known sequence by expert and trained readers) and 2 definitions of ROA (K/L grade ≥1, K/L grade ≥2). Finally, we explored the association of baseline radiographic scores to 2 key outcomes: joint replacement and incident ROA.

PATIENTS AND METHODS

CHECK cohort. CHECK is a prospective multicenter cohort study (n = 1,002) (20). Recruitment took place at 10 hospitals throughout The Netherlands between 2002 and 2005. Individuals were ages 45–65 years with knee or hip symptoms for which they had not yet sought medical care, or who had first visited a general practitioner (GP) no more than 6 months prior to enrollment. Individuals were excluded if they had medical conditions that might otherwise explain their symptoms (e.g., rheumatic conditions, previous joint replacement); comorbidities preventing evaluations over 10 years; or malignancy in the previous 5 years. Ethics approval was provided by all participating centers, and participants provided informed written consent. Research adhered to the Helsinki Declaration.

Radiography. Radiographs of both knees and both hips were acquired at 5 time points (baseline, 2, 5, 8, and 10 years), unless a participant missed an appointment or withdrew from the study. Detailed protocols were followed at all study centers to ensure precise radioanatomic positioning, with the use of small metal balls, plexiglass frames, and foot-maps to ensure accurate, reproducible positioning across visits. We describe here only the acquired views needed for K/L scoring. For the knee, posteroanterior radiographs were taken with participants positioned in semiflexed weightbearing (23). For the hip, anteroposterior radiographs were taken with participants positioned in weightbearing.

K/L scoring procedures. Baseline images were first scored by a member of the CHECK steering committee. The steering committee consisted of senior investigators with substantial expertise in ROA research: 3 rheumatologists, 2 physical therapists, 1 rehabilitation physician, 1 physician, and 1 biologist. Prior to scoring images, the steering committee met to standardize scoring procedures, based on the original K/L scoring description (1–3), using a subset of training images. Once the steering committee was satisfied that their scoring procedures were consistent, each steering committee member scored a portion of images (no formal reliability testing was undertaken for this set of readings). All images were read blinded to symptoms, including whether pain was in the hip or knee, and which side was painful. These scores were never made available at subsequent readings.

Independently of steering committee scores, baseline and follow-up images were scored by trained readers and a GP with expertise in OA and radiograph reading (JD) (25). Extensive training was provided to the trained readers (4 readers in years 2 and 5, 5 readers in years 8 and 10) by an experienced musculoskeletal radiologist (EHGO) and the GP, described elsewhere (25). All trained readers were medical students. The GP maintained supervision over trained readers throughout the study, including answering questions or assisting with scores if needed. We previously reported training and interrater reliability using year 5 images (mean prevalence and bias adjusted $\kappa = 0.58$ [range 0.23–0.79] for knee K/L scores, and $\kappa = 0.80$ [range 0.55–0.90] for the hip)

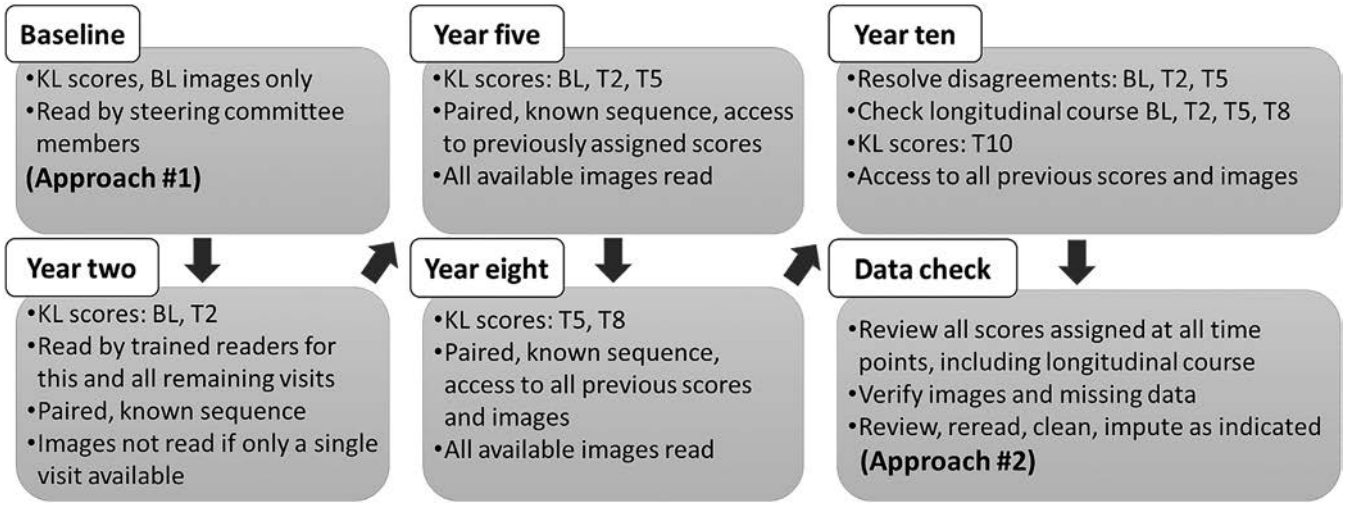


Figure 1. Flow chart of Kellgren/Lawrence (KL) scoring procedures at each visit. BL = baseline; T2, T5, T8, T10 = years 2, 5, 8, 10.

(25). In this article, we differentiate the date of image scoring from the date of image acquisition by spelling out the visit in which images were scored, and abbreviating the visit in which images were acquired: baseline (BL), year 2 (T2), year 5 (T5), year 8 (T8), and year 10 (T10). Thus at baseline, BL images were scored; at year 2, BL and T2 images were scored, and so on.

Baseline. At baseline, scores were assigned by the steering committee without access to follow-up images (Figure 1). For the present study, we defined these K/L scores as the first approach of 2 scoring approaches. The trained readers did not read or score any images at baseline, but began reading images at year 2 (see below). Scores assigned by the steering committee were never made available to trained readers at any time.

Year 2. The trained readers read and scored BL and T2 images paired with known sequence. If T2 images were missing, the BL images were not read or scored. If T2 images were available but the BL image was missing, a K/L score was assigned only to the T2 image at this visit.

Year 5. Trained readers read and scored BL, T2, and T5 radiographs paired with known sequence. Scores were assigned to all available images, even if images were only available for a single visit. Readers had access to previously assigned BL and T2 scores from year 2. Readers could assign different K/L scores for BL and T2 than had previously been assigned, if reading images across all 3 time points together justified this difference.

Year 8. Trained readers read and scored all available T5 and T8 radiographs paired with known sequence, with BL and T2 radiographs available for reference at reader discretion. Readers had knowledge of scores previously assigned at year 5 for BL images. No new scores were assigned for BL or T2.

Year 10. Trained readers first looked at all previous scores that had been read on at least 2 occasions (BL images scored at years 2 and 5; T2 images scored at years 2 and 5; T5 images scored at years 5 and 8). For any case where 2 scores differed, a third read of those images was done to resolve the disagreement, and K/L scores for subsequent time points were checked for longitudinal course. Subsequently, T10 radiographs were read and scored. Readers had access to all previously acquired radiographs and previously assigned scores, but were not explicitly instructed to use them in assigning T10 scores.

Once T10 scoring was complete, all K/L scores across all time points were reviewed with all images available, together in sequence. Further consideration was given to adjusting scores at any time point, if appropriate. For example, in cases where a K/L grade decreased from one time point to the next, all images for that participant were reviewed, and K/L scores were adjusted to better represent images across all time points. This process was done on the assumption that OA cannot regress, thus K/L grades suggesting regression were likely due to variability such as data entry error, interrater error, image quality, or radioanatomic positioning. Other reasons for image rereads included suspected data entry errors or missing scores. At year 10, there were also several cases of images that could not be found from previous time points. We could not confirm whether these images had become missing or if they had never existed, so scores were reassigned to missing. Finally, in cases of missing K/L scores: if the subsequent time point image was K/L 0, then the earlier missing data point was reclassified as K/L 0; if a previous time point had a confirmed joint arthroplasty, then subsequent visits were reclassified to arthroplasty. Remaining missing data were left as missing.

The review of all scores across all time points was done in an iterative manner, with the final review performed in August 2019, including complete score reviews (all scores assigned at all visits) and verification that radiographs existed for all assigned scores (EMM and JR), a team meeting (all coauthors), additional radiograph readings to resolve remaining uncertainties (JD), and

approval of the final data set (EMM, JR, and JD). For this study, we defined these final K/L scores as the second approach of 2 scoring approaches.

Joint replacement. Joint replacements were confirmed radiographically. For knees, we defined joint replacement as partial or total arthroplasty. Participants reported the year in which the surgery had occurred, and we recorded this value in years from baseline. If the surgery date was missing, we recorded the date as the visit in which the radiograph of the joint replacement was acquired.

Statistical analysis. All statistical analyses were done using Stata/SE software, version 15.1. We described the proportion of knees or hips with each K/L score (0–4) at baseline using both approaches: the steering committee's single time point reading (first approach), and the trained readers' year-10 final assignment of BL scores with access to images and scores across all time points (second approach). We then reported BL ROA prevalence using both scoring approaches and also using 2 ROA definitions: any ROA (K/L grade ≥1) and established ROA (K/L grade ≥2). We then compared how the 2 scoring approaches affected BL ROA prevalence (any versus established) using mixed-effects Poisson regression with robust estimates of variance.

We next compared the associations of the 4 different BL scores (2 scoring approaches, 2 ROA definitions) with undergoing joint replacement by the end of the study using Cox proportional hazards models (Stata's stcox syntax) (26). To account for correlation between both knees (or hips) within each participant, we clustered models at the participant level using the vce (cluster clustervar) option (26). We defined survival as the year in which a joint replacement occurred, or the year in which participants without joint replacement withdrew, were lost to follow-up, or completed the study.

Finally, we evaluated the associations of the 2 scoring approaches with developing incident established ROA for BL scores of K/L 1 compared to K/L 0 using Cox proportional hazards models. We defined survival as the first visit in which a joint was scored at least K/L 2 (based on the final scores assigned in year 10), or the year in which participants without ROA withdrew, were lost to follow-up, or completed the study.

RESULTS

Of 1,002 participants, 792 (79%) were women, mean ± SD age was 55.9 ± 5.2 years, and body mass index was 26.2 ± 4.0 kg/m^2. BL K/L scores differed between the 2 approaches. Using the first approach, 439 of 1,526 K/L grade 0 knees (29%) were assigned higher scores in the second approach, while 123 K/L grade 1 and 2 scores were assigned K/L 0 in the second approach, resulting overall in 20% fewer K/L 0 scores in the second approach (Table 1 and Figure 2).

Table 1. Kellgren/Lawrence (K/L) scores at baseline in the knee and hip, using 2 scoring approaches: first approach scored by steering committee at baseline without access to follow-up images versus second approach scored by trained readers with all available images and known sequence (n = 2,004 knees)*

	First approach	Second approach	PR (95% CI)
Knee K/L score			
0	1,526 (76)	1,228 (61)	–
1	359 (18)	555 (28)	–
2	79 (4)	206 (10)	–
3	8 (<1)	2 (<1)	–
4	0 (0)	0 (0)	–
Missing	32 (2)	13 (<1)	–
Knee radiographic ROA			
Any ROA†	446 (22)	763 (38)	1.7 (1.6–1.9)‡
Established ROA§	87 (4)	208 (10)	2.4 (1.8–3.1)‡
Hip K/L score			
0	1,699 (85)	1,292 (64)	–
1	209 (10)	482 (24)	–
2	67 (3)	205 (10)	–
3	7 (<1)	13 (<1)	–
4	0 (0)	0 (0)	–
Missing	22 (1)	12 (<1)	–
Hip radiographic ROA			
Any ROA†	283 (14)	700 (35)	2.5 (2.2–2.8)‡
Established ROA§	74 (4)	218 (11)	2.9 (2.3–3.7)‡

* Values are the number (%) unless indicated otherwise. 95% CI = 95% confidence interval; PR = prevalence ratio; ROA = radiographic osteoarthritis.
† Any ROA = K/L grade ≥1.
‡ Statistically significant.
§ Established ROA = K/L grade ≥2.

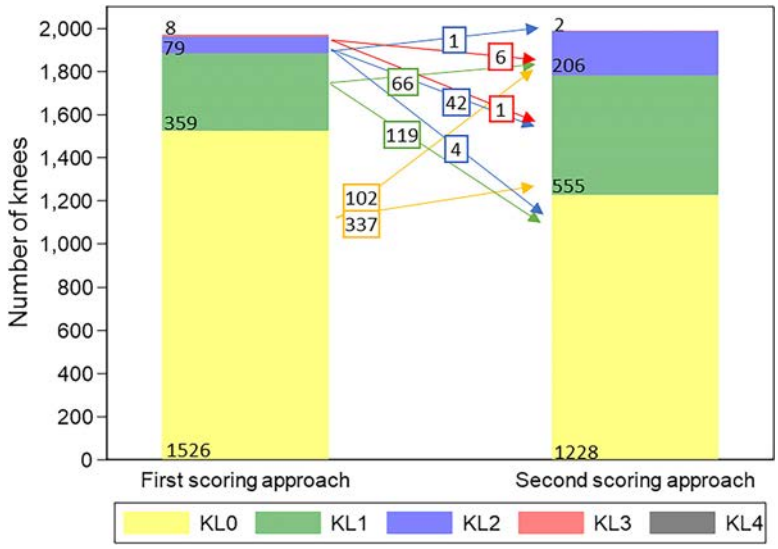


Figure 2. Knee Kellgren/Lawrence (KL) scores: differences in assigned baseline K/L scores by first scoring approach (single reading by steering committee, left column) compared to second scoring approach (paired readings with known sequence, by trained readers, right column). Numbers in columns refer to number of participants assigned each grade; numbers in small boxes refer to number of participants whose grade changed (with arrow indicating to which grade they changed) using the second scoring approach.

Similarly, 485 of 1,699 K/L 0 hips (29%) were assigned higher scores in the second approach, while 69 K/L grade 1 and 2 scores were assigned K/L 0 in the second approach, resulting overall in 24% fewer K/L 0 scores in the second approach (Table 1 and Figure 3).

Using the second approach, more participants were classified as having ROA using both ROA definitions. For knees, the prevalence ratio of the second approach compared to the first was 1.7 (95% confidence interval [95% CI] 1.6–1.9) for any ROA and 2.4 (95% CI 1.8–3.1) for established ROA (Table 1). For hips, prevalence ratios were 2.5 (95% CI 2.2–2.8) and 2.9 (95% CI 2.3–3.7), respectively (Table 1).

The hazard for undergoing knee replacement differed substantially between the 2 scoring approaches, but was only significant for established ROA (Table 2). For any ROA (compared to no ROA) at baseline, the hazard ratio (HR) for undergoing knee replacement was 9.5 (95% CI 4.8–18.6) using the first approach and 13.3 (95% CI 5.4–33.2) using the second

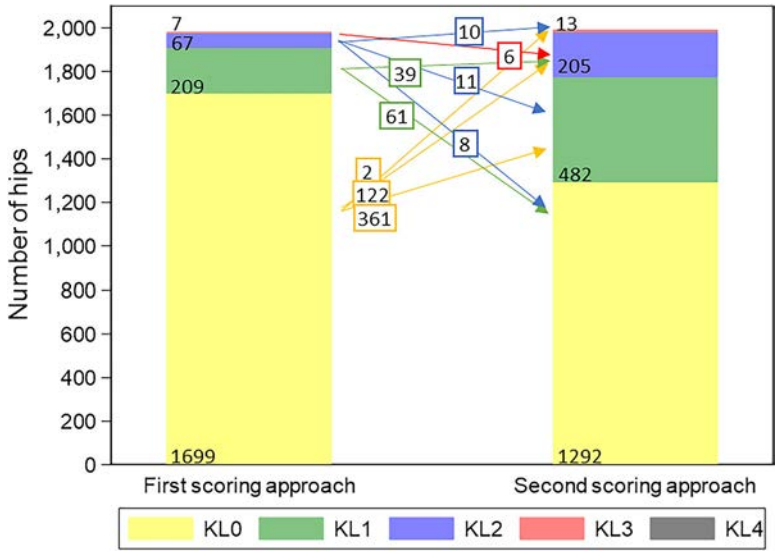


Figure 3. Hip Kellgren/Lawrence (KL) scores: differences in assigned baseline K/L scores by first scoring approach (single reading by steering committee, left column) compared to second scoring approach (paired readings with known sequence, by trained readers, right column). Numbers in columns refer to number of participants assigned each grade; numbers in small boxes refer to number of participants whose grade changed (with arrow indicating to which grade they changed) using the second scoring approach.

Table 2. Hazard ratios for undergoing joint replacement based on baseline for any (K/L grade ≥1) or established (K/L grade ≥2) OA prevalence, using 2 scoring approaches: first approach scored by steering committee at baseline without access to follow-up images versus second approach scored by trained readers with all available images and known sequence*

	First approach	HR (95% CI)	Second approach	HR (95% CI)
Knee replacement				
Any ROA†	32/446 (7)	9.5 (4.8–18.6)‡	39/763 (5)	13.3 (5.4–33.2)‡
K/L 0§	12/1,526 (0.8)	–	5/1,228 (0.4)	–
Established ROA¶	19/87 (22)‡	19.3 (10.3–36.1)‡	15/208 (7)‡	4.8 (2.4–9.6)‡
K/L 0 or 1§	25/1,885 (1)	–	29/1,783 (2)	–
Hip replacement				
Any ROA†	52/283 (18)	9.4 (6.1–14.5)‡	73/700 (10)	8.3 (4.9–14.0)‡
K/L 0§	39/1,699 (2)	–	18/1,292 (1)	–
Established ROA¶	34/74 (46)	24.4 (15.0–39.8)‡	40/218 (18)	7.7 (4.9–12.1)‡
K/L 0 or 1§	57/1,908 (3)	–	51/1,774 (3)	–

* Values are the number/total number (%) unless indicated otherwise. 95% CI = 95% confidence interval; HR = hazard ratio; K/L = Kellgren/Lawrence; OA = osteoarthritis; ROA = radiographic OA.
† Any baseline ROA K/L grade ≥1.
‡ Statistically significant.
§ Without baseline ROA.
¶ Established baseline ROA K/L grade ≥2.

approach. Moreover, 7 more knee replacements (39 of 44 compared to 32) were correctly predicted using the second approach, while at most, 310 knees were reclassified as having any ROA but did not undergo arthroplasty, though on account of right censoring (338 knees [17%]), true results may differ slightly (columns 2 and 4 in Table 2). For established ROA, the HR for undergoing knee replacement was 19.3 (95% CI 10.3–36.1) using the first approach, and decreased significantly to 4.8 (95% CI 2.4–9.6) using the second approach. Using the second approach, 4 fewer arthroplasties were correctly predicted, and at most, 125 more knees with any ROA did not undergo arthroplasty.

For hips, results were similar (Table 2). For any ROA, the HR for undergoing hip replacement was 9.4 (95% CI 6.1–14.5) using the first approach and 8.3 (95% CI 4.9–14.0) using the second approach. Despite similar HRs, the second approach correctly predicted 21 more hip replacements, while up to 396 more hips had any OA but did not undergo arthroplasty. For established ROA, the HR for undergoing hip replacement was 24.4 (95% CI 15.0–39.8) using the first approach and decreased significantly to 7.7 (95% CI 4.9–12.1) using the second approach. Despite the lower HR, the second approach correctly predicted 6 more hip replacements, while at most, 138 more knees with established ROA did not undergo arthroplasty.

The HR for developing incident established knee ROA was 2.4 (95% CI 2.0–2.8) for K/L 1 compared to K/L 0 using the first approach and 2.8 (95% CI 2.4–3.3) using the second approach (Table 3). The second approach correctly predicted 124 more knees developing established ROA, while up to 72 more knees were graded K/L 1 that did not develop ROA. For the hip, the HR was 2.1 (95% CI 1.6–2.7) using the first approach and 3.0 (95% CI 2.5–3.5) using the second approach. The second approach correctly predicted 163 more hips developing established ROA while up to 110 more knees were graded K/L 1 that did not develop ROA.

Table 3. Hazard ratios for developing incident established radiographic OA (K/L grade ≥2) for K/L grade 1 at baseline compared to K/L grade 0, using 2 scoring approaches: first approach scored by steering committee at baseline without access to follow-up images versus second approach scored by trained readers with all available images and known sequence*

	First approach	HR (95% CI)	Second approach	HR (95% CI)
Knee				
K/L 1†	269/359 (75)	2.4 (2.0–2.8)‡	393/555 (71)	2.8 (2.4–3.3)‡
K/L 0§	734/1,526 (48)	–	494/1,228 (40)	–
Hip				
K/L 1†	129/209 (62)	2.1 (1.6–2.7)‡	292/482 (61)	3.0 (2.5–3.5)‡
K/L 0§	706/1,699 (42)	–	396/1,292 (31)	–

* Values are the number/total number (%) unless indicated otherwise. 95% CI = 95% confidence interval; HR = hazard ratio; K/L = Kellgren/Lawrence; OA = osteoarthritis.
† Incident established OA with baseline K/L grade 1.
‡ Statistically significant.
§ Incident established radiographic OA with baseline K/L grade 0.

DISCUSSION

In this study, we described the methods used in the CHECK cohort to assign K/L scores to hip and knee radiographs at each visit. With these details consolidated into a single article, the reader is better equipped to compare and interpret studies published since the CHECK cohort's inception, that use K/L scores assigned at different time points. This study also illustrates how different scoring methods potentially influence cohort study results, highlighting potential implications for future trial design and interpretation.

The second scoring approach classified more hips and knees as having both any and established ROA compared to the first approach. This difference may be due to inherent challenges in determining whether a bony feature is an osteophyte, and whether it is doubtful or definite. Seeing follow-up images with progression of osteophytes may increase reader confidence in identifying and classifying baseline features as osteophytes. We acknowledge, however, that this difference could also relate to who assigned scores under the 2 approaches. Interrater reliability has previously been shown to be higher between expert radiologists than between expert radiologists and their trained readers (17,27,28). We therefore acknowledge that the differences between the 2 approaches in our study may reflect not only access to follow-up images, but also interrater reliability and relative expertise and training of the 2 groups of readers. One previous study reported that, among disagreements between an expert radiologist and trained readers, scores tended to be higher in trained readers (17). These findings are similar to ours. However, readers in that study were site investigators motivated to enroll individuals with OA features into their study, possibly introducing bias (17). Our study eligibility criteria did not include radiograph readings, removing this bias. We believe that higher scores in the second approach are more likely due to access to follow-up images and extensive data checking, though we cannot rule out reader-related factors.

Previous studies have shown that reading images paired in known sequence improves reliability and sensitivity to ROA progression, likely due to having access to more information during reading (4–11). Sensitivity to progression has been implied to suggest that, despite the bias introduced, paired reading with known sequence provides more valid scores. However, sensitivity to progression has typically been defined using the standardized response mean (SRM) (4–6). This statistic provides the equivalent of a mean effect size, so a larger SRM means more individuals are reported to have ROA progression. A gold standard has not typically been considered to confirm that larger SRMs reflect a true higher rate of progression (29). Thus while this approach may be more valid, SRM cannot confirm this increase in validity. At best, SRM provides face validity that having access to more images enables a more accurate score, but we cannot rule out that a higher SRM reflects bias introduced by a reader expecting progression to occur chronologically. Reading an image at a single time point may increase error and reduce reliability. However, such a reading also mitigates bias, may give more conservative estimates of ROA prevalence, and better reflects clinical settings where multiple images are not available.

One of the strengths of the CHECK cohort is that 2 approaches have been used to assign baseline K/L scores. This offers the unique ability to select which approach would answer specific questions best. For example, if a researcher wants to know whether baseline K/L scores are a risk factor for a future outcome, they could use scores assigned using the second approach because this method is more accurate (29). Alternatively, if researchers want to know how well radiographs in a clinical setting predict the same outcome, the first approach may provide a more conservative and clinically realistic estimate, since clinicians do not typically know the outcomes of care provided.

Our results highlight the importance of reporting absolute numbers of an outcome in addition to effect sizes: odds ratios, relative risks, or HRs reported alone may be misleading. For example, if a clinician wants to identify hip pain patients at risk for future hip replacement to be able to offer a cost-effective prevention program, the clinician could use the results of the more clinically realistic first approach (Table 2). They might be tempted to define ROA as K/L ≥2 because of the higher HR (24.4 compared to 9.4 if defining ROA as K/L ≥1). However, defining ROA as K/L ≥2 would result in not treating 18 hips that would need a future hip replacement and may benefit from treatment. In this case, treating any hip ROA despite the lower HR might be more important to the clinician. In the case of an expensive treatment, the clinician might stay with K/L 2 after all because while they would miss treating the 18 hips ultimately needing replacement, in this scenario, a clinician would theoretically avoid the need to provide costly treatment for more than 200 hips. This scenario also illustrates that the number of patients the clinician might expect to treat would be substantially overestimated had they implemented a new program based on results using the second scoring approach (700 knees with K/L 1, 218 knees with K/L 2).

The above scenario brings up the additional question of how best to define ROA. In a previous 10-year prospective population-based study of women, 62% of 90 knees with doubtful osteophytes at baseline progressed to having definite osteophytes 10 years later (18). Our findings were similar: 71–75% of knees (depending on scoring approach) and 61–62% of hips with doubtful osteophytes at baseline developed established ROA within 10 years. These results suggest that identifying middle-aged individuals with hip or knee symptoms as having OA, rather than waiting for them to develop established ROA, may offer new insights and opportunities for secondary prevention in this population.

Limitations to our study include the fact that interrater reliability was not formally assessed in the steering committee of expert readers, and trained reader reliability was assessed at years 5 and 8 but only recorded at year 5. All readers were of similar background and received similar training by the same radiologist

and GP, thus the recording of year 8 results was not felt to be necessary at the time. Also, to more accurately compare scoring methods, having the same readers assign scores using both approaches would have been advantageous. The comparisons of scoring approaches in our study represent a more pragmatic and thus generalizable comparison of approaches that capture differences due in part to having access to multiple follow-up images in known sequence, but also due to interrater reliability, differences in data-checking procedures, and reader-related factors. In addition, all participants had knee or hip symptoms, thus we had no asymptomatic reference group. However, our study design better reflects clinical reality in which patients typically seek care for existing symptoms. Finally, in the CHECK cohort, the reason for study withdrawal was not recorded. This limitation relates to use of survival analysis, and the possibility of competing risk, in particular death. While we cannot confirm this fact, the young age of our participants, combined with recollection of study investigators, suggests that death was very rare and our findings would not likely be altered.

We recommend that future studies be designed with careful consideration for radiographic scoring conditions. Evaluating a score assigned with a single reading at a single visit may give more clinically realistic predictions of future outcomes. Alternatively, evaluating scores assigned during paired readings with a known sequence may provide greater insights into the exact nature of ROA onset and progression. Both approaches are important, and thus method selection must address the specific research question. In both cases, readers must be carefully selected, with adequate experience and training to optimize score validity and reliability. We also recommend that future studies consider using an earlier definition of ROA than is typically used, particularly where researchers are interested in understanding early OA with an aim toward preventing poor clinical outcomes. Where feasible and affordable, studies incorporating magnetic resonance imaging (MRI) can also contribute meaningfully to early OA research, since MRI better visualizes soft tissues (e.g., cartilage, bone marrow lesions) and is thus more sensitive to detecting early OA features (30). In conclusion, this study of middle-aged individuals with hip or knee symptoms demonstrates that evaluation of ROA depends on radiograph scoring conditions, and the prediction of future outcomes is influenced by both scoring conditions as well as which K/L grade is used to define ROA.

ACKNOWLEDGMENTS

The CHECK cohort study was initiated by the Dutch Arthritis Association, and participating sites were: Erasmus Medical Center Rotterdam; Kennemer Gasthuis Haarlem; Leiden University Medical Center; Maastricht University Medical Center; Martini Hospital Groningen/Allied Health Care Center for Rheumatology and Rehabilitation Groningen; Medical Spectrum Twente Enschede/Ziekenhuisgroep Twente Almelo; Reade (formerly Jan van Breemen Institute/VU Medical Center Amsterdam); St. Maartenskliniek Nijmegen; and University Medical Center Utrecht and Wilhelmina Hospital Assen.

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. Macri 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. Macri, Runhaar, Damen, Oei, Bierma-Zeinstra.

Acquisition of data. Runhaar, Damen, Oei, Bierma-Zeinstra.

Analysis and interpretation of data. Macri, Runhaar, Damen, Oei, Bierma-Zeinstra.

REFERENCES

1. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957;16:494–502.
2. Schiphof D, Boers M, Bierma-Zeinstra SM. Differences in descriptions of Kellgren and Lawrence grades of knee osteoarthritis. Ann Rheum Dis 2008;67:1034–6.
3. Kellgren J, Jeffrey M, Ball J. The epidemiology of chronic rheumatism: atlas of standard radiographs of arthritis. Oxford (UK): Blackwell Scientific Publications; 1963.
4. Auleley GR, Giraudeau B, Dougados M, Ravaud P. Radiographic assessment of hip osteoarthritis progression: impact of reading procedures for longitudinal studies. Ann Rheum Dis 2000;59:422–7.
5. Gensburger D, Roux JP, Arlot M, Sornay-Rendu E, Ravaud P, Chapurlat R. Influence of blinding sequence of radiographs on the reproducibility and sensitivity to change of joint space width measurement in knee osteoarthritis. Arthritis Care Res (Hoboken) 2010;62:1699–705.
6. Botha-Scheepers S, Watt I, Breedveld FC, Kloppenburg M. Reading radiographs in pairs or in chronological order influences radiological progression in osteoarthritis. Rheumatology (Oxford) 2005;44:1452–5.
7. Van der Heijde D, Boonen A, Boers M, Kostense P, van Der Linden S. Reading radiographs in chronological order, in pairs or as single films has important implications for the discriminative power of rheumatoid arthritis clinical trials. Rheumatology (Oxford) 1999;38:1213–20.
8. Felson DT, Nevitt MC, Yang M, Clancy M, Niu J, Torner JC, et al. A new approach yields high rates of radiographic progression in knee osteoarthritis. J Rheumatol 2008;35:2047–54.
9. LaValley MP, McAlindon TE, Chaisson CE, Levy D, Felson DT. The validity of different definitions of radiographic worsening for longitudinal studies of knee osteoarthritis. J Clin Epidemiol 2001;54:30–9.
10. Kopec JA, Sayre EC, Schwartz TA, Renner JB, Helmick CG, Badley EM, et al. Occurrence of radiographic osteoarthritis of the knee and hip among African Americans and Whites: a population-based prospective cohort study. Arthritis Care Res (Hoboken) 2013;65:928–35.
11. Spector TD, Hart DJ, Byrne J, Harris PA, Dacre JE, Doyle DV. Definition of osteoarthritis of the knee for epidemiological studies. Ann Rheum Dis 1993;52:790–4.
12. The Osteoarthritis Initiative. Central reading of knee X-rays for Kellgren & Lawrence grade and individual radiographic features of tibiofemoral knee OA. URL: http://oai.epi-ucsf.org/datarelease/forms/kXR_SQ_BU_Descrip.pdf?V01XRKL.
13. Schiphof D, de Klerk BM, Kerkhof HJ, Hofman A, Koes BW, Boers M, et al. Impact of different descriptions of the Kellgren and Lawrence classification criteria on the diagnosis of knee osteoarthritis. Ann Rheum Dis 2011;70:1422–7.
14. Chaisson CE, Gale DR, Gale E, Kazis L, Skinner K, Felson DT. Detecting radiographic knee osteoarthritis: what combination of views is optimal? Rheumatology (Oxford) 2000;39:1218–21.

15. Radiography Working Group of the OARSI-OMERACT Imaging Workshop, Le Graverand MP, Mazzuca S, Lassere M, Guermazi A, Pickering E, et al. Assessment of the radioanatomic positioning of the osteoarthritic knee in serial radiographs: comparison of three acquisition techniques. Osteoarthritis Cartilage 2006;14:37–43.
16. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Third ed. Upper Saddle River (NJ): Pearson Eduction; 2009.
17. Guermazi A, Hunter DJ, Li L, Benichou O, Eckstein F, Kwoh CK, et al. Different thresholds for detecting osteophytes and joint space narrowing exist between the site investigators and the centralized reader in a multicenter knee osteoarthritis study: data from the Osteoarthritis Initiative. Skeletal Radiol 2012;41:179–86.
18. Duncan R, Peat G, Thomas E, Hay EM, Croft P. Incidence, progression and sequence of development of radiographic knee osteoarthritis in a symptomatic population. Ann Rheum Dis 2011;70:1944–8.
19. Hart DJ, Spector TD. Kellgren & Lawrence grade 1 osteophytes in the knee: doubtful or definite? Osteoarthritis Cartilage 2002;11:149–50.
20. Wesseling J, Boers M, Viergever MA, Hilberdink WK, Lafeber FP, Dekker J, et al. Cohort profile: cohort hip and cohort knee (CHECK) study. Int J Epidemiol 2014;45:36–44.
21. Schiphof D, Runhaar J, Waarsing JH, van Spil WE, van Middelkoop M, Bierma-Zeinstra SM. The clinical and radiographic course of early knee and hip osteoarthritis over 10 years in CHECK (Cohort Hip and Cohort Knee). Osteoarthritis Cartilage 2019;27:1491–500.
22. Damen J, van Rijn RM, Emans PJ, Hilberdink WK, Wesseling J, Oei EH, et al. Prevalence and development of hip and knee osteoarthritis according to American College of Rheumatology criteria in the CHECK cohort. Arthritis Res Ther 2019;21:4.
23. Lankhorst N, Damen J, Oei E, Verhaar J, Kloppenburg M, Bierma-Zeinstra S, et al. Incidence, prevalence, natural course and prognosis of patellofemoral osteoarthritis: the Cohort Hip and Cohort Knee study. Osteoarthritis Cartilage 2017;25:647–53.
24. Van Spil WE, Welsing PM, Kloppenburg M, Bierma-Zeinstra SM, Bijlsma JW, Mastbergen SC, et al. Cross-sectional and predictive associations between plasma adipokines and radiographic signs of early-stage knee osteoarthritis: data from CHECK. Osteoarthritis Cartilage 2012;20:1278–85.
25. Damen J, Schiphof D, Ten Wolde S, Cats HA, Bierma-Zeinstra SM, Oei EH. Inter-observer reliability for radiographic assessment of early osteoarthritis features: the CHECK (Cohort Hip and Cohort Knee) study. Osteoarthritis Cartilage 2014;22:969–74.
26. Stata Corporation. Stata survival analysis reference manual release 16. College Station (TX): Stata Press; 2019.
27. Bellamy N, Tesar P, Walker D, Klestov A, Muirden K, Kuhnert P, et al. Perceptual variation in grading hand, hip and knee radiographs: observations based on an Australian twin registry study of osteoarthritis. Ann Rheum Dis 1999;58:766–9.
28. Cooper C, Cushnaghan J, Kirwan J, Dieppe P, Rogers J, McAlindon T, et al. Radiographic assessment of the knee joint in osteoarthritis. Ann Rheum Dis 1992;51:80–2.
29. Felson DT, Nevitt MC. Blinding images to sequence in osteoarthritis: evidence from other diseases. Osteoarthritis Cartilage 2009;17:281–3.
30. Gudbergsen H, Lohmander L, Jones G, Christensen R, Bartels EM, Danneskiold-Samsoe B, et al. Correlations between radiographic assessments and MRI features of knee osteoarthritis: a cross-sectional study. Osteoarthritis Cartilage 2013;21:535–43.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1188–1195
DOI 10.1002/acr.24545

Lifetime Allergy Symptoms in IgG4-Related Disease: A Case–Control Study

Samantha Sanders,[1] Xiaoqing Fu,[2] Yuqing Zhang,[3] Cory A. Perugino,[3] Rachel Wallwork,[2] Emanuel Della-Torre,[4] Liam Harvey,[3] Tyler Harkness,[3] Aidan Long,[3] Hyon K. Choi,[3] John H. Stone,[3] and Zachary S. Wallace[3]

Objective. The etiology of IgG4-related disease (IgG4-RD) is unknown, and there has been controversy over the significance of allergic conditions in IgG4-RD. We examined the prevalence of lifetime allergy symptoms in IgG4-RD and the association between these and IgG4-RD.

Methods. We identified IgG4-RD patients and non-IgG4-RD controls without autoimmune conditions seen at a single center. IgG4-RD patients were classified using the American College of Rheumatology/European Alliance of Associations for Rheumatology classification criteria. Allergy symptoms were ascertained by questionnaire. We assessed the association of IgG4-RD features with allergy symptoms. We compared the proportion of cases and controls with allergy symptoms using conditional logistic regression to estimate odds ratios (ORs) and 95% confidence intervals (95% CIs) after matching cases and controls 1:1 by age and sex.

Results. Lifetime allergy symptoms were reported by 165 (71%) of 231 IgG4-RD patients. Aeroallergen symptoms were most commonly reported (n = 135, 58%), followed by skin allergy symptoms (n = 97, 42%) and food allergy symptoms (n = 47, 20%). IgG4-RD cases with a history of allergy symptoms were more likely to have head and neck involvement (OR 2.0 [95% CI 1.1–3.6]) and peripheral eosinophilia (OR 3.3 [95% CI 1.2–9.0]) than those without allergy symptoms. The prevalence of any allergy symptoms was similar between cases and controls (OR 0.7 [95% CI 0.4–1.1]); this remained consistent after stratifying by head and neck involvement.

Conclusion. Lifetime allergy symptoms are common in IgG4-RD but are not reported more often in IgG4-RD compared to non-IgG4-RD patients without autoimmune conditions. These findings suggest that allergies are not uniquely associated with the pathogenesis or presentation of IgG4-RD.

INTRODUCTION

IgG4-related disease (IgG4-RD) is an immune-mediated fibroinflammatory condition characterized by tumorous lesions, often with an elevated serum IgG4 concentration (1,2). The etiology remains poorly understood, and there has been controversy over the significance of allergic conditions and Th2 cells in the pathogenesis and presentation (3).

Th2 cells and allergies were hypothesized to be important in the pathogenesis and presentation of IgG4-RD following several clinicopathologic observations. First, allergic symptoms, especially allergic rhinitis, have been reported to be common in IgG4-RD, especially among those with manifestations in the head and neck (e.g., sialadenitis, dacryoadenitis, and orbital disease) (2,4). Second, elevated peripheral IgE concentrations, peripheral eosinophilia, and tissue infiltrating eosinophils are often observed

Dr. Perugino's work was supported by the Rheumatology Research Foundation (Scientist Development Award). Dr. Wallace's work was supported by the Rheumatology Research Foundation and the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases grants K23-AR073334 and L30-AR070520). Drs. Perugino, Stone, and Wallace's work was supported by the NIH (Autoimmune Disease Center of Excellence grant UM1-AI-144295 from the National Institute of Allergy and Infectious Diseases).

[1]Samantha Sanders, MD, MBA: Harvard Medical School, Boston, Massachusetts; [2]Xiaoqing Fu, MS, Rachel Wallwork, MD: Massachusetts General Hospital, Boston; [3]Yuqing Zhang, DSc, Cory A. Perugino, DO, Liam Harvey, BS, Tyler Harkness, BS, Aidan Long, MD, Hyon K. Choi, MD, DrPH, John H. Stone, MD, MPH, Zachary S. Wallace, MD, MSc: Harvard Medical School and Massachusetts General Hospital, Boston; [4]Emanuel Della-Torre, MD, PhD: IRCCS-San Raffaele Scientific Institute, Milan, Italy.

Dr. Perugino has received research support from UCB. Dr. Wallace has received consulting fees from Viela Bio (less than $10,000) and research support from Bristol Myers Squibb. No other disclosures relevant to this article were reported.

Address correspondence to Zachary S. Wallace, MD, MSc, Clinical Epidemiology Program, Division of Rheumatology, Allergy, and Immunology, Massachusetts General Hospital, 100 Cambridge Street, 16th Floor, Boston, MA 02114. Email: zswallace@mgh.harvard.edu.

Submitted for publication July 12, 2020; accepted in revised form December 15, 2020.

SIGNIFICANCE & INNOVATIONS

- Lifetime allergy symptoms are common in IgG4-related disease (IgG4-RD), particularly among patients with head and/or neck involvement.
- In contrast to hypothesized associations, allergy symptoms are not reported more frequently in patients with IgG4-RD than in patients without IgG4-RD or other autoimmune diseases, regardless of whether they have head and neck involvement or not.
- These observations suggest that allergies are unlikely to play a unique role in the pathogenesis or presentation of IgG4-RD.

in IgG4-RD patients, as they are in many patients with allergic conditions (4,5). Third, cytokines typically associated with Th2 cells have been reported to be present at high concentrations in tissues affected by IgG4-RD (6–8).

Despite these observations, however, mounting evidence suggests that Th2 cells are unlikely to play a pathogenic role IgG4-RD (3). Indeed, a previous study found that circulating Th2 memory cells appear to be restricted to a subset of patients with atopy (9,10). The same Th2-associated cytokines previously used to infer Th2 cell tissue infiltration, such as interleukin-4, are also produced by a specific subset of follicular helper T cells, which have been shown to accumulate in tissues affected by IgG4-RD (11). Moreover, upon rigorous quantification of CD4+ T cells infiltrating tissue, Th2 cells were found to account for only 5–10% of all CD4+ T cells on average, including salivary gland tissues and tissues from patients with IgG4-RD and concurrent atopy (12). Although the pathogenic role of Th2 cells in IgG4-RD has been called into question, there remains a lack of clarity regarding the burden and potential significance of allergic symptoms in IgG4-RD patients.

Previous studies of allergy symptoms in IgG4-RD have been limited to Asian populations, did not rely on standardized assessments of allergy symptoms, and/or did not include a reference population for comparison (2,4,13–17). Here, we sought to overcome these limitations by examining the characteristics and distributions of lifetime allergy symptoms in a US-based IgG4-RD cohort with diverse manifestations using a standardized allergy questionnaire and by measuring the association between allergy symptoms and the odds of having IgG4-RD using a case–control design.

SUBJECTS AND METHODS

IgG4-RD cohort. The Massachusetts General Hospital Center for IgG4-RD, a part of the Rheumatology Unit, maintains a database of all patients referred for evaluation in the center. The inclusion of patients with IgG4-RD was based on the classification criteria approved by the American College of Rheumatology (ACR) and European Alliance of Associations for Rheumatology (EULAR) (18). We included patients who fell into 1 of 3 categories: 1) definite IgG4-RD; 2) probable IgG4-RD; and 3) atypical IgG4-RD. Patients in the definite category fulfilled the published classification criteria. Patients who were considered probable fulfilled 2 parts of the ACR/EULAR classification criteria (i.e., had clinical involvement of a typical organ and were rigorously evaluated to ensure that they did not meet any exclusion criteria) but did not reach the threshold of 20 inclusion points according to the criteria (frequently retroperitoneal fibrosis in a typical pattern because no biopsy could be safely obtained or the biopsy was not informative). Of the 34 patients with probable IgG4-RD, 31 had ACR/EULAR classification criteria scores ≥8, which correspond to a specificity of ≥85% for IgG4-RD. The remaining 3 patients with probable IgG4-RD had typical pachymeningeal (2 patients) and bile duct (1 patient) involvement. Patients who were considered atypical met the previously established pathologic and immunostaining criteria for diagnosing IgG4-RD but presented with involvement of an atypical organ (e.g., breast, prostate) that was not considered in the ACR/EULAR classification criteria (19).

We included all IgG4-RD patients who were seen between January 19, 2012, and September 12, 2019, and who completed an allergy questionnaire. Some of the clinical and laboratory features of cases included in this study have been reported previously (9,20–24). However, this cohort's allergic histories have not been investigated using a standard questionnaire and compared to a reference population before, and the analyses pertaining to allergy symptoms reported herein are novel.

Data pertaining to demographics and IgG4-RD manifestations were collected from the Center's database. Laboratory results were extracted from the electronic health record. Age at IgG4-RD onset (index date) refers to the age at which the patient first developed symptoms ultimately attributed to IgG4-RD or the time at which the disease was first diagnosed (whichever was earlier) (20).

Control subjects. We identified controls from a sample of patients without IgG4-RD seen in the Massachusetts General Hospital rheumatology clinic between June 1, 2016, and March 30, 2020. We chose to include patients without autoimmune diseases as controls because some autoimmune rheumatic diseases, including eosinophilic granulomatosis with polyangiitis and rheumatoid arthritis, may be associated with a history of allergy symptoms (25). Accordingly, we limited our control group to patients seen in our clinic because of noninflammatory joint disease (e.g., osteoarthritis), crystalline arthritis, fibromyalgia, or osteopenia/osteoporosis. One control was matched to each case by sex and the age (± 5 years) at which the controls completed the survey relative to the age of cases at the index date. Between June 1, 2016, and July 31, 2018, potential

controls were invited to participate at the time of a routine clinic visit. Beginning August 1, 2018, potential controls were invited to participate electronically. The change in recruitment from an in-person to electronic methodology was made after our study group established electronic recruitment as a viable option to facilitate recruitment and limit the in-person resources needed for recruitment. To recruit electronically, we identified eligible patients who had been seen in clinic, obtained permission from their provider, and invited those for whom we had permission by a standardized method through our electronic patient messaging system. The proportion of patients with allergies was similar among patients recruited via in-person versus electronic methods, suggesting that this change did not impact our results.

Allergy ascertainment. We administered an allergy questionnaire to all patients following an initial visit and asked 34 questions about a lifetime history of allergy symptoms, including history of aeroallergen symptoms (e.g., hay fever-type allergy symptoms), food allergy symptoms, skin allergy symptoms (atopic dermatitis, skin reactions, and urticaria), and anaphylaxis (see Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24545/abstract). This questionnaire had similar questions to those administered in the 2005–2006 National Health and Nutrition Examination Survey (26).

Statistical analysis. Categorical variables are reported as the number (%). Continuous variables are reported as the mean ± SD or median (interquartile range) depending on their distribution. First, we examined the distribution of allergy symptoms among IgG4-RD patients. Among IgG4-RD patients, we assessed the association between reported allergy symptoms and select IgG4-RD features and manifestations using unadjusted and age- and/or sex-adjusted logistic regression. Second, we assessed for the association between allergy symptoms and the odds of having IgG4-RD. For these analyses, we used conditional logistic regression to first evaluate the association between any allergy symptom and the odds of having IgG4-RD. We then used conditional logistic regression to evaluate the independent association between each type of allergy symptom and the odds of having IgG4-RD. Associations were reported using odds ratios (ORs) and 95% confidence intervals (95% CI). We evaluated the association between allergy symptoms and IgG4-RD overall and after stratifying according to whether or not cases had head and neck involvement given previous reports of an association between head and neck manifestations and allergy symptoms among patients with IgG4-RD (2,4). In a sensitivity analysis, we evaluated whether our findings persisted when we restricted the definition of allergies to those reported to be diagnosed by a physician. For all analyses, 2-sided *P* values less than 0.05 were considered significant. This study was approved by the Partners HealthCare Institutional Review Board prior to the enrollment of any patients.

RESULTS

IgG4-RD cohort description. There were 231 patients in the IgG4-RD cohort on the date of data accession (Table 1). The mean ± SD age was 60 ± 14 years, and the majority were male (150, 65%) and White (173, 75%). Of the 231 patients, 179 (78%) had definite IgG4-RD, 34 (15%) had probable IgG4-RD,

Table 1. Demographic characteristics and features of the IgG4-related disease cohort*

Characteristic	Overall (n = 231)
Age at diagnosis, mean ± SD years	59.5 ± 13.7
Male	150 (65)
Race	
White	173 (75)
Asian	32 (14)
Black	10 (4)
Native American	1 (<1)
Unknown/Other	15 (6)
Ethnicity	
Non-Hispanic	185 (80)
Hispanic	29 (13)
Unknown/Other	17 (7)
Selected organ involvement	
Head and neck	137 (59)
Dacryoadenitis or sialadenitis	113 (49)
Lacrimal glands	50 (22)
Salivary glands	98 (42)
Orbital (non-lacrimal)	34 (15)
Other head and neck	109 (47)
Lymph nodes	63 (27)
Pulmonary	44 (19)
Aorta	21 (9)
Retroperitoneum	41 (18)
Pancreato-hepatobiliary	78 (34)
Renal	47 (20)
Laboratory results, median (IQR)	
IgG4 concentration (n = 228)	142.4 (53.2–390.8)
% ever elevated†	168 (73)
Eosinophil concentration (n = 193)	0.20 (0.10–0.40)
% elevated†	40 (21)
IgE concentration (n = 192)	104.0 (25.0–284.5)
% elevated†	98 (51)
ESR (n = 159)	24.0 (10.0–45.0)
% elevated†	66 (42)
CRP (n = 161)	3.8 (1.3–9.1)
% elevated†	53 (33)
C3 (n = 191)	115.0 (85.0–145.0)
% low C3	23 (12)
C4 (n = 193)	23.0 (12.0–31.0)
% low C4	32 (17)

* Values are the number (%) unless indicated otherwise. CRP = C-reactive protein; ESR = erythrocyte sedimentation rate; IQR = interquartile range.
† Elevated refers to serum IgG4 concentration greater than the assay's upper limit of normal, serum IgE concentration ≥100 IU/ml, eosinophils ≥0.5 10^9/liter, ESR >30 mm/hour, and CRP level >7 mg/liter.

and 18 (8%) had IgG4-RD with atypical organ involvement. The most common IgG4-RD manifestations included the head or neck (137, 59%), particularly sialadenitis and/or dacryoadenitis (113, 49%). A minority (n = 11, 5%) of patients with head and neck involvement had IgG4-RD involvement of the nasal cavities or sinuses. Other commonly affected organs included the pancreato-hepatobiliary system (n = 78, 34%), lymph nodes (n = 63, 27%), and kidneys (n = 47, 20%). The serum IgG4 concentration was elevated at any point in a patient's available medical history in 168 (73%) patients.

Features of IgG4-RD patients according to allergy symptoms. Lifetime allergy symptoms were reported by 165 (71%) IgG4-RD patients, the details of which are reported in Table 2. The proportion of patients reporting allergy symptoms was similar across those with definite (73%), probable (66%), and atypical (83%) IgG4-RD. Of the lifetime allergy symptoms reported, aeroallergen symptoms were most common (n = 135, 58%) followed by skin allergy symptoms (n = 97, 42%) and food allergy symptoms (n = 47, 20%). A history of anaphylaxis was reported in 20 (9%) subjects. Aeroallergen symptoms, food allergies, and skin allergies predated the onset of IgG4-RD in 99%, 94%, and 98% of patients, respectively, typically by many years (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24545/abstract). The majority of IgG4-RD patients (61%) with a history of allergy reported no change in their allergy symptoms following a diagnosis of IgG4-RD (Table 2). Only 6 patients reported increasing their use of antihistamines since being diagnosed with IgG4-RD.

Among patients with IgG4-RD, there were associations (Table 3) between those with allergy symptoms having head and neck involvement (adjusted OR [OR_{adj}] 2.02 [95% CI 1.12–3.62]) and having peripheral eosinophilia (OR_{adj} 3.27 [95% CI 1.19–9.02]), compared to those without head and neck involvement and without peripheral eosinophilia, respectively. The association between head and neck disease with allergy symptoms was strongly driven by the subgroup of patients with sialadenitis and/or dacryoadenitis (OR_{adj} 1.92 [95% CI 1.06–3.48]) when compared to those without these manifestations. The association between allergy symptoms and head and neck involvement by IgG4-RD persisted when we specifically examined the association between aeroallergen symptoms and these manifestations (OR_{adj} 2.24 [95% CI 1.30–3.86]) compared to those without head and neck involvement. We did not observe associations between allergy symptoms and having elevated IgG4 or IgE concentrations, having elevated inflammatory markers, or being hypocomplementemic. When we stratified patients with IgG4-RD according to prior glucocorticoid exposure, there was no difference in the proportion reporting a history of any allergy among those who had received glucocorticoids versus those who had not (72% versus 69%; $P = 0.7$).

Table 2. Characteristics of lifetime allergy symptoms in IgG4-related disease*

Features of allergy symptoms	Frequency
Any allergy symptoms	165 (71)
Aeroallergen symptoms	
History of aeroallergen symptoms	135 (58)
Self-reported allergy symptoms	129 (56)
Occurred in the last 12 months†	86 (41)
Physician-diagnosed allergies	87 (38)
Self-reported and physician-diagnosed allergies	81 (35)
Underwent aeroallergen sensitization testing	101 (44)
Reported allergen	
Seasonal allergens (e.g., grass, pollen)	32 (14)
Pet dander (e.g., cats, dogs)	18 (8)
Mold	10 (4)
Dust mites	17 (7)
Other	28 (12)
Food allergy symptoms and hypersensitivities	
History of food allergy symptoms/hypersensitivities	47 (20)
Self-reported allergy symptoms	43 (19)
Physician-diagnosed allergies	32 (14)
Self-reported and physician-diagnosed allergies	28 (12)
Underwent food allergen testing	36 (16)
Reported allergen/hypersensitivity	
Dairy/lactose	1 (<1)
Nuts	3 (1)
Shellfish	3 (1)
Other	12 (5)
Skin allergy symptoms	
History of skin allergy symptoms	97 (42)
Self-reported contact dermatitis symptoms	47 (20)
Self-reported eczema symptoms	39 (17)
Any physician-diagnosed allergy	59 (26)
Self-reported and physician-diagnosed allergies	32 (14)
Attributed causes of contact dermatitis‡	
Latex	12 (5)
Chemicals/perfumes	8 (3)
Plants/trees	4 (2)
Nickel/other metals	1 (<1)
Other	17 (7)
Anaphylaxis	20 (9)
Allergy symptoms following IgG4-RD onset	
Improved	31 (19)
No change	101 (61)
Worsened	16 (10)
Other or not reported	17 (10)

* Values are the number (%).
† Data missing in 20 subjects.
‡ % of self-reported contact dermatitis.

Case–control analysis. Of 788 potential controls invited to complete the allergy questionnaire, 208 (26%) completed the questionnaire. We matched 201 IgG4-RD cases to 201 controls with rheumatic diseases that are not associated with autoimmunity (Table 4). The cases and controls were well matched with regard to mean ± SD age (60.2 ± 12.5 years versus 60.7 ± 13.3 years, respectively) and sex (n = 121, 60% versus n = 121, 60% male, respectively). The control population included patients with gout/pseudogout (n = 67, 33%), osteoarthritis (n = 65, 32%), fibromyalgia (n = 20, 10%), other mechanical/degenerative disease (n = 19, 9%), osteoporosis/osteopenia (n = 12, 6%), and other conditions (n = 18, 9%).

Table 3. The association of select IgG4-related disease manifestations with any allergy symptoms*

	Overall, no. (n = 231)	Any allergy symptoms (n = 165)	No allergy symptoms (n = 66)	Unadjusted OR (95% CI)	Adjusted OR (95% CI)†
Head/neck disease					
Yes	137	106 (64)	31 (47)	2.03 (1.14–3.62)‡	2.02 (1.12–3.62)‡
No	94	59 (36)	35 (53)	Ref.	Ref.
Dacryoadenitis or sialadenitis					
Yes	113	88 (53)	25 (38)	1.87 (1.05–3.36)‡	1.92 (1.06–3.48)‡
No	118	77 (47)	41 (62)	Ref.	Ref.
IgG4 concentration elevated					
Yes	168	121 (73)	47 (71)	1.11 (0.59–2.10)	1.31 (0.67–2.54)
No	63	44 (27)	19 (29)	Ref.	Ref.
IgE concentration elevated					
Yes	96	76 (54)	22 (43)	1.54 (0.81–2.94)	1.58 (0.81–3.08)
No	94	65 (46)	29 (57)	Ref.	Ref.
Peripheral eosinophilia					
Yes	40	35 (25)	5 (10)	3.10 (1.14–8.42)‡	3.27 (1.19–9.02)‡
No	153	106 (75)	47 (90)	Ref.	Ref.
ESR elevated					
Yes	66	45 (39)	21 (47)	0.75 (0.37–1.49)	0.89 (0.42–1.87)
No	93	69 (61)	24 (53)	Ref.	Ref.
CRP elevated					
Yes	53	37 (32)	16 (34)	0.89 (0.43–1.83)	0.90 (0.43–1.88)
No	108	78 (68)	30 (65)	Ref.	Ref.
C3 hypocomplementemia					
Yes	23	17 (12)	6 (12)	0.97 (0.36–2.63)	1.01 (0.37–2.78)
No	168	125 (88)	43 (88)	Ref.	Ref.
C4 hypocomplementemia					
Yes	32	21 (15)	11 (22)	0.63 (0.28–1.42)	0.67 (0.29–1.54)
No	161	121 (85)	40 (78)	Ref.	Ref.

* Values are the number (%) unless indicated otherwise. 95% CI = 95% confidence interval; CRP = C-reactive protein; ESR = erythrocyte sedimentation rate; OR = odds ratio; Ref. = reference.
† Age- and sex-adjusted.
‡ Significant.

A similar proportion of cases and controls (Table 5 and Supplementary Table 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24545/abstract) reported any lifetime allergy symptoms. Any allergies were reported by 142 (71%) cases versus 158 (79%) controls (OR 0.7 [95% CI 0.4–1.1]). The prevalence of aeroallergen symptoms (OR 0.7 [95% CI 0.5–1.1]) and skin allergy symptoms (OR 1.0 [95% CI 0.6–1.5]) was similar in cases and controls. These observations remained consistent after stratifying cases according to the presence or absence of head and neck manifestations of IgG4-RD. Our findings remained unchanged after matching cases and controls on race (data not shown) and when restricting the definition of allergy symptoms (see Supplementary Table 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24545/abstract) as those reported to have been diagnosed by a physician.

Significantly fewer cases than controls reported food allergy symptoms (39 [19%] of cases versus 68 [34%] of controls; OR 0.5 [95% CI 0.3–0.9]), although this association was significantly attenuated after stratifying cases according to head and neck IgG4-RD involvement and did not persist when restricting the definition of food allergy to those diagnosed by a physician (26 [13%] of cases versus 23 [11%] of controls; OR 1.1 [95% CI 0.6–2.1]).

Table 4. Demographic features of IgG4-RD cases and matched controls*

	IgG4-related disease cases (n = 201)	Age- and sex-matched controls (n = 201)
Age, mean ± SD years	60.2 ± 12.5	60.7 ± 13.3
Male	121 (60)	121 (60)
Race		
White	155 (77)	189 (94)
Asian	24 (12)	0 (0)
Black	9 (4)	6 (3)
Unknown/Other	13 (7)	6 (3)

* Values are the number (%) unless indicated otherwise.

DISCUSSION

In this case–control study, the first of its kind in IgG4-RD, we found that lifetime allergy symptoms are common in IgG4-RD. The development of allergies did not seem to be temporally related to the presence of IgG4-RD, as allergic symptoms were typically present for many years prior to the

Table 5. The association between lifetime allergy symptoms and IgG4-related disease*

	IgG4-related disease cases, OR (95% CI)	Age- and sex-matched controls
All cases included		
Model 1		
Any allergy symptoms	0.67 (0.43–1.05)	1.0 (Ref.)
Model 2†		
Aeroallergen symptoms	0.72 (0.46–1.11)	1.0 (Ref.)
Food allergy symptoms	0.52 (0.30–0.88)	1.0 (Ref.)
Skin allergy symptoms	0.96 (0.63–1.48)	1.0 (Ref.)
Head and neck cases		
Model 1		
Any allergy symptoms	0.72 (0.39–1.32)	1.0 (Ref.)
Model 2†		
Aeroallergen symptoms	0.90 (0.49–1.65)	1.0 (Ref.)
Food allergy symptoms	0.63 (0.34–1.17)	1.0 (Ref.)
Skin allergy symptoms	0.97 (0.53–1.78)	1.0 (Ref.)
Non–head and neck cases		
Model 1		
Any allergy symptoms	0.63 (0.33–1.19)	1.0 (Ref.)
Model 2†		
Aeroallergen symptoms	0.57 (0.28–1.13)	1.0 (Ref.)
Food allergy symptoms	0.81 (0.33–1.98)	1.0 (Ref.)
Skin allergy symptoms	0.93 (0.45–1.93)	1.0 (Ref.)

* 95% CI = 95% confidence interval; OR = odds ratio; Ref. = reference.
† In model 2, each allergy symptom was included in the model to evaluate the independent association between each symptom and the odds of having IgG4-related disease.

development of IgG4-RD. Among patients with IgG4-RD, those with allergy symptoms, especially aeroallergen symptoms, were more likely to have sialadenitis and/or dacryoadenitis than those without allergy symptoms. While the prevalence of allergy symptoms in IgG4-RD patients was high, allergy symptoms were reported by IgG4-RD patients at a similar frequency to non-IgG4-RD controls, even among those with head and neck involvement by IgG4-RD. Our epidemiologic observations of allergy symptoms in IgG4-RD patients and controls complement those made in previous laboratory-based studies suggesting that allergic, Th2-mediated responses are unlikely to be pathogenic drivers of IgG4-RD or unique features of IgG4-RD presentations (3).

There are limited data on the lifetime prevalence of allergy symptoms in the general population, in part because symptoms may be managed by patients using over the counter medications without a physician diagnosis or prescription. In our study, the prevalence of lifetime allergy symptoms reported by both cases and controls are similar to a prevalence of 58% reported in a recent study (25). Compared to that study, we found a higher proportion of patients reporting any history of allergy, but these differences may have to do with differences in geography distribution of survey respondents, demographic differences of participants, and survey design.

Our study overcomes many limitations of prior studies that have evaluated allergy symptoms in IgG4-RD. These studies enrolled only Asian patients, did not systematically evaluate allergy symptoms, and/or did not use a reference population to compare the frequencies in IgG4-RD versus a control population (2,4,13–17). Our findings confirmed previous observations that allergy symptoms are commonly reported in IgG4-RD, with previous studies reporting prevalence rates for aeroallergen symptoms between 25% and 63% (2,4,13–17,27). We also found that allergy symptoms are more frequently reported in patients with head and neck involvement, particularly dacryoadenitis and/or sialadenitis (2,14,16). To our knowledge, our study is the first to report an association between self-reported allergy symptoms and peripheral eosinophilia in IgG4-RD, which may be related to our more rigorous study design, including standardized assessments of allergy symptoms and a larger sample size than in some prior studies (2,28). While this observation is not necessarily surprising, it raises the question of whether IgG4-RD patients with allergy symptoms are more likely to relapse given previous studies describing an association between peripheral eosinophilia and higher risk of IgG4-RD relapse (23,28), as well as a previous study demonstrating an association between allergy symptoms and a higher risk of relapse (15).

The difference in organ involvement among those with and without allergy symptoms might suggest etiologic heterogeneity among patients with IgG4-RD such that the onset in some patients is related to immune system dysfunction that is also contributing to allergy symptoms (29). There are several possible explanations as to why IgG4-RD patients with head and neck manifestations more often had self-reported allergy symptoms. First, it is possible that patients with head and neck disease are more likely to report allergy symptoms due to recall bias given that allergies often affect the head and neck. However, this is less likely to explain our findings given that allergy symptoms (e.g., itchy eyes, rhinitis) are likely distinguishable from IgG4-RD symptoms in the head and neck (e.g., proptosis, sialadenitis), and few patients had IgG4-RD affecting the sinonasal cavities. Second, allergen exposure could lead to a generalized activation of the immune system in the head and neck, manifesting in predisposed individuals as IgG4-RD involving the head and neck. If this were the case, we would expect those with head and neck disease to be more likely to have allergy symptoms than control patients, but our ability to detect this may have been obscured by a smaller sample size.

Future studies might further investigate the association between allergy symptoms and IgG4-RD manifestations in the head and neck, which has now been replicated across cohorts of diverse ethnic makeups, by confirming allergic diagnoses, evaluating specific allergens, considering the role of local mucosal immunity and the oral microbiome, and comparing the eosinophilic infiltrate across organs (2,14). Although a mild-to-moderate eosinophil infiltrate is frequently commented on in association with IgG4-RD, this component of the immune response has not been well investigated, especially in the head and neck (19).

In our study, we found that food allergy symptoms were more commonly reported to have ever occurred in controls than in IgG4-RD. However, this trend did not persist when we restricted our analysis to allergies reported to have been diagnosed by a physician. A potential negative association between these allergy subtypes and IgG4-RD would be somewhat surprising given our understanding of the pathogenesis of IgG4-RD. Accordingly, these observations should be interpreted cautiously but confirm our hypothesis that specific allergy symptoms are not more commonly reported by patients with IgG4-RD. Further research is needed to investigate why these allergy types may be less frequently reported by IgG4-RD patients.

Our study has several strengths, which include its sample size, use of a standardized questionnaire, and case–control design. Moreover, this is among the first studies to apply the recently defined ACR/EULAR classification criteria in an epidemiologic study. While classification criteria are not meant for diagnostic purposes, they can have an important role in observational studies such as this one for identifying patients for inclusion. Our identification of 3 groups (definite, probable, and atypical) using the entry, exclusion, and inclusion criteria of the ACR/EULAR classification criteria may be of use for future observational studies in IgG4-RD and reflects the realities of clinical practice where some patients will not fulfill definite criteria because of the organ distribution (e.g., breast, pancreas, retroperitoneum) but are accepted to have IgG4-RD based on histopathology and/or clinical assessment by expert providers.

Our study has certain limitations. First, as with any survey study, recall bias is possible. In this instance, patients may have been more likely to recall allergy symptoms given that they were asked to complete the survey in the context of medical care. Additionally, patients may have been more likely to recall certain allergy symptoms (e.g., aeroallergen symptoms) if they were administered the survey during allergy season for aeroallergens. However, both IgG4-RD patients and controls were asked to complete the survey under similar circumstances, and our survey asked about lifetime allergy symptoms. Therefore, any recall bias would be expected to be similar between cases and controls. Second, allergy symptoms were based on patient-reported symptoms and medical history. However, similar methods have been used to estimate the burden of allergic conditions in the US population through national health surveys (26). Future studies might define allergic conditions more stringently using an evaluation by an allergist with or without formal allergy testing. Third, our study did not account for the severity of the allergy disease in cases and controls, which future studies could measure. Fourth, it is possible that sinus or nasal cavity IgG4-RD involvement may have been reported as allergies; nonetheless, this most likely did not affect our outcomes given the qualitatively different nature of allergy symptoms and IgG4-RD manifestations and the low proportion of patients with sinus or nasal cavity involvement (5%). Fifth, subgroup analyses (e.g., by manifestations, laboratory results) were limited by smaller sample sizes, and we cannot rule out the possibility that associations might exist if studied in larger IgG4-RD cohorts. Sixth, our control recruitment methodology switched from in-person to electronic recruitment during the study period. However, the proportion of controls reporting allergy symptoms was similar regardless of the recruitment method, suggesting that this did not impact our findings. Finally, our study was performed at a tertiary referral center and in a cohort composed of patients who self-identified as White. Therefore, the generalizability of our findings may be limited, but we note the wide range of manifestations represented in our cohort as well as its size despite the rarity of this condition.

In conclusion, lifetime allergy symptoms are frequently reported in IgG4-RD, especially among those with head and neck involvement, but not at a higher rate than in controls without autoimmune conditions. We found a similar or lower prevalence of lifetime allergy symptoms among IgG4-RD patients when compared with age- and sex-matched controls without autoimmune conditions; these observations persisted after stratifying cases by head and neck disease involvement. This supports the hypothesis that allergies are unlikely to play a unique role in the pathogenesis or presentation of IgG4-RD.

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. Wallace 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. Sanders, Perugino, Della-Torre, Long, Choi, Stone, Wallace.

Acquisition of data. Sanders, Wallwork, Harvey, Harkness, Wallace.

Analysis and interpretation of data. Sanders, Fu, Zhang, Perugino, Della-Torre, Long, Choi, Stone, Wallace.

REFERENCES

1. Kamisawa T, Zen Y, Pillai S, Stone JH. IgG4-related disease. Lancet 2015;385:1460–71.
2. Saeki T, Kobayashi D, Ito T, Tamura M, Yoshikawa S, Yamazaki H. Comparison of clinical and laboratory features of patients with and without allergic conditions in IgG4-related disease: a single-center experience in Japan. Mod Rheumatol 2018;28:845–8.
3. Perugino CA, Mattoo H, Mahajan VS, Maehara T, Wallace ZS, Pillai S, et al. IgG4-related disease: insights into human immunology and targeted therapies. Arthritis Rheumatol 2017;69:1722–32.
4. Culver EL, Sadler R, Bateman AC, Makuch M, Cargill T, Ferry B, et al. Increases in IgE, eosinophils, and mast cells can be used in diagnosis and to predict relapse of IgG4-related disease. Clin Gastroenterol Hepatol 2017;15:1444–52.e6.
5. Saeki T, Nishi S, Imai N, Ito T, Yamazaki H, Kawano M, et al. Clinicopathological characteristics of patients with IgG4-related tubulointerstitial nephritis. Kidney Int 2010;78:1016–23.
6. Zen Y, Fujii T, Harada K, Kawano M, Yamada K, Takahira M, et al. Th2 and regulatory immune reactions are increased in immunoglobin G4-related sclerosing pancreatitis and cholangitis. Hepatology 2007; 45:1538–46.

7. Tanaka A, Moriyama M, Nakashima H, Miyake K, Hayashida JN, Maehara T, et al. Th2 and regulatory immune reactions contribute to IgG4 production and the initiation of Mikulicz disease. Arthritis Rheum 2012;64:254–63.
8. Ishiguro N, Moriyama M, Furusho K, Furukawa S, Shibata T, Murakami Y, et al. Activated M2 macrophages contribute to the pathogenesis of IgG4-related disease via toll-like receptor 7/interleukin-33 signaling. Arthritis Rheumatol 2020;72:166–78.
9. Della Torre E, Mattoo H, Mahajan VS, Carruthers M, Pillai S, Stone JH. Prevalence of atopy, eosinophilia, and IgE elevation in IgG4-related disease. Allergy 2014;69:269–72.
10. Mattoo H, Della-Torre E, Mahajan VS, Stone JH, Pillai S. Circulating Th2 memory cells in IgG4-related disease are restricted to a defined subset of subjects with atopy. Allergy 2014;69:399–402.
11. Maehara T, Mattoo H, Mahajan A, Murphy S, Yuen GJ, Ishiguro N, et al. The expansion in lymphoid organs of IL-4 + BATF + T follicular helper cells is linked to IgG4 class switching in vivo. Life Sci Alliance 2018;1.
12. Mattoo H, Mahajan VS, Maehara T, Deshpande V, Della-Torre E, Wallace ZS, et al. Clonal expansion of CD4(+) cytotoxic T lymphocytes in patients with IgG4-related disease. J Allergy Clin Immunol 2016; 138:825–38.
13. Yamamoto M, Takano KI, Kamekura R, Aochi S, Suzuki C, Ichimiya S, et al. Analysis of allergic reaction in IgG4-related disease. Mod Rheumatol 2019;29:1063–5.
14. Wang M, Zhang P, Lin W, Fei Y, Chen H, Li J, et al. Differences and similarities between IgG4-related disease with and without dacryoadenitis and sialoadenitis: clinical manifestations and treatment efficacy. Arthritis Res Ther 2019;21:44.
15. Liu Y, Zeng Q, Zhu L, Gao J, Wang Z, Wang Z, et al. Relapse predictors and serologically unstable condition of IgG4-related disease: a large Chinese cohort. Rheumatology (Oxford) 2020;59:2115–23.
16. Liu Y, Xue M, Wang Z, Zeng Q, Ren L, Zhang Y, et al. Salivary gland involvement disparities in clinical characteristics of IgG4-related disease. Rheumatology (Oxford) 2020;59:634–40.
17. Della Torre E, Germano T, Ramirez GA, Dagna L, Yacoub MR. IgG4-related disease and allergen-specific immunotherapy. Ann Allergy Asthma Immunol 2020;124:631–3.
18. Wallace ZS, Naden RP, Chari S, Choi H, Della-Torre E, Dicaire JF, et al. The 2019 American College of Rheumatology/European League Against Rheumatism classification criteria for IgG4-related disease. Arthritis Rheumatol 2020;72:7–19.
19. Deshpande V, Zen Y, Chan JK, Yi EE, Sato Y, Yoshino T, et al. Consensus statement on the pathology of IgG4-related disease. Mod Pathol 2012;25:1181–92.
20. Wallace ZS, Deshpande V, Mattoo H, Mahajan VS, Kulikova M, Pillai S, et al. IgG4-related disease: clinical and laboratory features in one hundred twenty-five patients. Arthritis Rheumatol 2015;67: 2466–75.
21. Perugino CA, Wallace ZS, Meyersohn N, Oliveira G, Stone JR, Stone JH. Large vessel involvement by IgG4-related disease. Medicine (Baltimore) 2016;95:e3344.
22. Wallace ZS, Wallace CJ, Lu N, Choi HK, Stone JH. Association of IgG4-related disease with history of malignancy. Arthritis Rheumatol 2016;68:2283–9.
23. Wallace ZS, Mattoo H, Mahajan VS, Kulikova M, Lu L, Deshpande V, et al. Predictors of disease relapse in IgG4-related disease following rituximab. Rheumatology (Oxford) 2016;55:1000–8.
24. Wallwork R, Wallace Z, Perugino CA, Sharma A, Stone JH. Rituximab for idiopathic and IgG4-related retroperitoneal fibrosis. Medicine (Baltimore) 2018;97:e12631.
25. Kronzer VL, Crowson CS, Sparks JA, Vassallo R, Davis JM III. Investigating asthma, allergic disease, passive smoke exposure, and risk of rheumatoid arthritis. Arthritis Rheumatol 2019;71:1217–24.
26. Centers for Disease Control. National Health and Nutrition Examination Survey 2005–2006 data documentation, codebook, and frequencies allergy (AGQ_D). June 2008. URL: https://wwwn.cdc.gov/Nchs/Nhanes/2005-2006/AGQ_D.htm#Interview_Setting_and_Mode_of_Administration.
27. Zhang X, Zhang P, Zhang W. Clinical significance of allergy in IgG4-related disease [abstract]. Arthritis Rheumatol 2018; 70 Suppl 10. URL: https://acrabstracts.org/abstract/clinical-significance-of-allergy-in-igg4-related-disease/.
28. Zhang X, Zhang P, Li J, He Y, Fei Y, Peng L, et al. Different clinical patterns of IgG4-RD patients with and without eosinophilia. Sci Rep 2019;9:16483.
29. Wallace ZS, Zhang Y, Perugino CA, Naden RL, Choi HK, Stone JH, et al. Clinical phenotypes of IgG4-related disease: an analysis of two international cross-sectional cohorts. Ann Rheum Dis 2019;78: 406–12.

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1196–1204
DOI 10.1002/acr.24556

American College of Rheumatology
Empowering Rheumatology Professionals

Supervised Intensive Exercise for Strengthening Exercise Health Beliefs in Patients With Axial Spondyloarthritis: A Multicenter Randomized Controlled Trial

Annelie Bilberg,[1] Hanne Dagfinrud,[2] and Silje H. Sveaas[2]

Objective. To evaluate the effect of a 3-month supervised high-intensity exercise program on exercise health beliefs in patients with axial spondyloarthritis.

Methods. This was secondary analysis of a randomized controlled trial. Participants (ages 23–69 years) were randomized to an exercise group (n = 50) or a control group (n = 50). The intervention was an individually guided cardiorespiratory and strength exercise program performed 2 times per week, plus an additional individual exercise session of personal choice. The control group received standard care and instructions to maintain their physical activity level. Exercise health beliefs using the Exercise Health Beliefs questionnaire (range 20–100, 100 = best), i.e., barriers, benefits, self-efficacy and exercise impact on arthritis, and physical activity, were assessed with self-reported questionnaires at baseline, 3 months, and 12 months after inclusion.

Results. The majority of the participants in the exercise group (76%) followed ≥80% of the prescribed exercise protocol. There was a significant effect of the intervention on exercise health beliefs at 3 months (estimated mean group differences 4.0 [95% confidence interval (95% CI) 1.4, 6.6]; $P = 0.003$) and the effect persisted at 12 months follow-up (estimated mean group differences 3.8 [95% CI 1.0, 6.6]; $P = 0.008$). Participants with higher exercise health beliefs had a higher odds ratio (1.1 [95% CI 1.0, 1.20]; $P = 0.003$) for being physically active at 12 months follow-up.

Conclusion. A supervised high-intensity exercise program had beneficial short- and long-term effects on participants' exercise health beliefs. Stronger exercise health beliefs were positively associated with a higher chance to be physically active on a health-enhancing level at 12 months follow-up.

INTRODUCTION

Axial spondyloarthritis (SpA) is a chronic inflammatory rheumatic disease, mainly affecting the spine and/or sacroiliac joints (1,2). Typical clinical features are inflammatory back pain, stiffness, and physical limitations (3), which may result in decreased health-related quality of life (3,4). Compared to the general population, aerobic capacity (5,6) and muscle strength are reduced in axial SpA (6). Additionally, patients with axial SpA have an increased risk of various comorbidities (7), including a substantially heightened risk of cardiovascular disease (8–10). Combinations of pharmacologic and nonpharmacologic treatment modalities are recommended for optimal management of patients with axial SpA (11). Exercise interventions along with appropriate medication are considered crucial elements for the management (12). Patients with inflammatory arthritic diseases are recommended to adhere to the general recommendations for physical activity with regular cardiorespiratory and muscular strength exercises (12,13). Nevertheless, a significant number of patients do not reach these recommendations (6,14,15). Several factors have been identified to account for the low adherence to exercise in axial SpA, including general and disease-related barriers (14,16). Also, a fear that exercises, especially at a higher intensity, might exacerbate the disease may hinder patients from participating in health-enhancing physical activity (17,18).

Knowledge of exercise benefits (19) and beliefs in one's own ability to master exercise are factors associated with participation in physical activity (20). Self-efficacy has especially been highlighted as an important factor influencing health behavior

ClinicalTrials.gov identifier: NCT02356874.
Supported by the Health and Medical Care Executive Board of Västra Götalands Region, the Rune and Ulla Amlöv's Foundation for Rheumatology Research, and the Renée Eanders Foundation.
[1]Annelie Bilberg, PhD: Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden; [2]Hanne Dagfinrud, PhD, Silje H. Sveaas, PhD: Diakonhjemmet Hospital, Oslo, Norway.

No potential conflicts of interest relevant to this article were reported.
Address correspondence to Annelie Bilberg, PhD, Institute of Neuroscience and Physiology, Section of Health and Rehabilitation, Physiotherapy, Sahlgrenska Academy, University of Gothenburg, Box 455, 405 30 Göteborg, Sweden. Email: annelie.bilberg@vgregion.se.
Submitted for publication May 19, 2020; accepted in revised form January 5, 2021.

SIGNIFICANCE & INNOVATIONS

- A supervised high-intensity program of cardiorespiratory and strength exercise had beneficial effects on exercise health beliefs in patients with axial spondyloarthritis (SpA).
- Stronger exercise health beliefs were positively associated with a higher chance to be physically active on a health-enhancing level over time.
- Patients with axial SpA should be given support and guidance from a physical therapist, while practicing exercise at a vigorous intensity, as this support will enhance patients exercise health beliefs and improve their confidence in mastering and adhering to exercise.

and adherence to health recommendations (20–22), including physical activity in arthritis (20,23). The exercise health belief concept is built on the "health belief mode" (24), which explains changes in health behavior based on the individual's perception and understanding. In patients with osteoarthritis and rheumatoid arthritis, exercise health belief has been found to be positively associated with participation in exercise (25). However, knowledge about exercise health belief and its influence on exercise and physical activity in axial SpA is scarce. Therefore, the aim of this study was to evaluate the effect of a supervised high-intensity exercise intervention on exercise health beliefs in patients with axial SpA, as well as to examine the relationship between exercise health beliefs and physical activity level at 12 months follow-up.

SUBJECTS AND METHODS

Design, participants, and procedure. This is a secondary analysis of a multicenter randomized controlled trial evaluating the effects of a 3-month supervised high-intensity exercise program in patients with axial SpA (26). Eligible participants were patients from 4 different outpatient rheumatology clinics, 3 in Norway (Diakonhjemmet Hospital, Martina Hansen Hospital, and the University Hospital of North Norway), and 1 in Sweden (Sahlgrenska University Hospital). Participants were also recruited through advertisement in various social media platforms. The recruitment, intervention, and data collection were performed between August 2015 and December 2017. The study was approved by the Regional Committee for Medical and Health Research Ethics (REK South East 2015/86) in Norway and the Regional Ethical Review Board Gothenburg in Sweden (032-16). All procedures followed the Declaration of Helsinki, and all participants gave written and oral informed consent before entering the study.

Inclusion criteria were fulfillment of the Assessment of SpondyloArthritis international Society criteria for axial SpA, age 18–70 years, no change in tumor necrosis inhibitor use over the past 3 months, moderate-to-high disease activity according to the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) >3.5, and no involvement in any regular cardiorespiratory and/or strength exercise (>1 time/week in the last 6 months). Exclusion criteria were established symptoms of cardiovascular disease, other comorbidity involving reduced exercise capacity, inability to participate in supervised sessions, and pregnancy. The participants were randomized in blocks of 4 subjects using a computer-generated sequence concealed in sequentially numbered, sealed, and opaque envelopes. In total, 100 patients with axial SpA were included in the study, 50 participants were allocated to the exercise group and 50 participants to the control group (Figure 1). The sample size of 100 participants is based on a power calculation for the Ankylosing Spondylitis Disease Activity Score (ASDAS), which represents disease activity in the main study (26). No power calculation was performed on the outcome measures in the present study.

Exercise intervention. The exercise program was supervised by physical therapists who were trained in the exercise protocol through participation in workshops. All physical therapists were experienced clinicians in rheumatology. The program included 40 minutes of high-intensity cardiorespiratory exercise on ergometric bicycles or treadmills (≥90–95% of maximal heart rate during intervals), followed by 20 minutes of strength exercise for large muscle groups (8–10 repetitions maximum, 2–3 sets, 2 times per week). In addition, participants also performed 1 cardiorespiratory session per week of their own choice (intensity controlled by pulse-watch, >70% of maximal heart rate). The exercise program is described in detail in Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24556.

To increase participants' self-efficacy, the supervised exercise sessions included personal guidance, positive feedback, and performance technique for high-intensity cardiorespiratory and strength exercise. To strengthen participants' autonomy for exercise, individual consultations with the physical therapist were scheduled to identify potential barriers to exercise and to find alternatives for the individual sessions of personal choice. Exercise adherence was recorded by the physical therapist as attendance at the supervised sessions and as accomplishment of the individual session of personal choice by inspection of the pulse-watch. Exercise adherence was also self-reported by the participants in a personal exercise diary, to enhance motivation. Participants in the control group received standard care and were instructed to maintain their usual physical activity level.

Assessments. All outcomes were secondary outcomes in the randomized controlled trial (26) and were assessed at baseline, immediately after completion of the intervention at 3 months and at 12 months follow-up. Information about demographic data and self-reported disease variables were obtained from a questionnaire. Clinical examinations were carried out at baseline

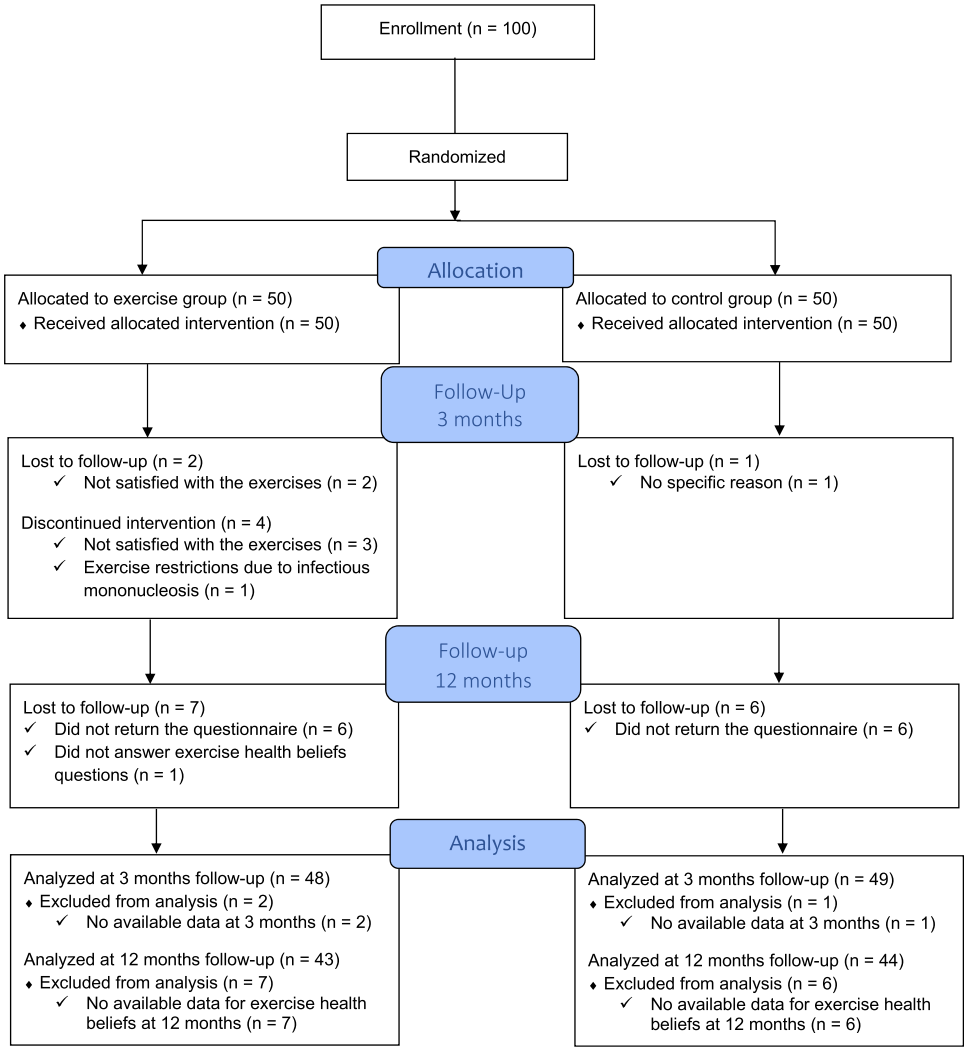


Figure 1. Flow of participants through the randomized controlled trial.

and 3 months by an independent physical therapist, blinded for group allocation for all 4 study sites. Blood samples were analyzed for C-reactive protein (CRP) level and erythrocyte sedimentation rate.

Exercise health beliefs. Exercise health beliefs were the main outcome of interest and were assessed with the Exercise Health Belief questionnaire (25), a self-administered outcome instrument identified to explain health behaviors related to exercise and to detect attitudes to exercise and changes in exercise beliefs. The questionnaire is based on a theoretical model, "the health belief model," with its origins in psychological and behavioral theories (24).The model explains changes in health behavior based on the individual's perception and understanding, as well as self-efficacy, a psychological mediator, to initiate and maintain changes in behavior (22,24,27,28). The Exercise Health Belief questionnaire consist of 20 statements divided into 4 scales; 3 items reflect barriers to exercise, 5 reflect benefits of exercise, 4 reflect self-efficacy for exercise, and 8 reflect the impact of exercise on arthritis, each scored on a 5-point Likert scale from strongly disagree to strongly agree. For each subscale, the individual item response can be summed to create a single score. The total Exercise Health Belief score ranges from 20 to 100, where a higher score represents stronger exercise health beliefs. The questionnaire is reliable for use in patients with arthritic diseases (25). A cross-culture adaption of the questionnaire was applied (29,30). The questionnaire was first translated to Swedish (by AB). The forward translated version was discussed by 3 experienced physical therapists (including AB) until consensus was reached as to its cultural appropriateness. A back translation was performed by a registered bilingual physical therapist translator. Thereafter, review of the 2 versions of the questionnaire was conducted by the first author (AB) and the translator until agreement was reached on discrepancies and cross-cultural equivalence. The same procedure was used for the Norwegian version of the questionnaire. The questionnaire was pretested in 10 patients with axial SpA who were attending the outpatient clinic at Sahlgrenska University Hospital. Patients

Table 1. Baseline description of the total study population, including the exercise group and the control group*

	All (n = 100)	Exercise group (n = 50)	Control group (n = 50)
Age, median (minimum–maximum) years	46.2 (23–69)	45.1 (23–68)	47.2 (24–69)
Male	47 (47)	25 (50)	22 (44)
Married/cohabitant	76 (76)	39 (78)	37 (74)
Working	81 (81)	42 (78)	39 (78)
Current smoking	12 (12)	5 (10)	7 (14)
Exercise frequency (calculated frequency)			
Never (0)	9 (9)	5 (10)	4 (8)
<1 time per week (0.5)	17 (17)	8 (16)	9 (18)
1 time per week (1)	29 (29)	13 (26)	16 (32)
2–3 times per week (2.5)	32 (32)	16 (32)	16 (32)
Almost every day (5)	13 (13)	8 (16)	5 (10)
Physical function			
BASFI, median (minimum–maximum)	3.2 (0.2–9.1)	2.6 (0.2–6.7)	3.0 (0.4–9.1)
BASMI, mean ± SD	2.8 ± 1.3	2.9 ± 1.3	2.6 ± 1.3
Disease characteristic			
ASDAS-CRP, mean ± SD	2.6 ± 0.7	2.6 ± 0.8	2.7 ± 0.6
BASDAI, mean ± SD	5.1 ± 1.6	4.9 ± 1.6	5.3 ± 1.5
CRP, median (minimum–maximum) mg/liter	2 (2–28)	2 (2–28)	2 (2–13)
ESR, median (minimum–maximum) mm/hour	8 (1–67)	8 (2–67)	8 (1–28)

* Values are the number (%) unless indicated otherwise. ASDAS-CRP = Ankylosing Spondylitis Disease Activity Score with C-reactive protein (inactive: <1.3, low: 1.3 to <2.1, high: 2.1–3.5, very high: >3.5); BASDAI = Bath Ankylosing Spondylitis Disease Activity Index; BASFI = Bath Ankylosing Spondylitis Functional Index; BASMI = Bath Ankylosing Spondylitis Metrology Index (all BAS instruments range 0–10, 10 = worst); ESR = erythrocyte sedimentation rate.

were asked to complete the questionnaire and to give their opinion of the relevance of the items included in the questionnaire. No further changes to the questionnaire were suggested.

Physical activity level. Physical activity was assessed with the questions from the Nord-Trøndelag Health Study (31). Participants were asked "How often do you exercise? (never, less than once a week, once a week, 2–3 times a week, and almost every day: give an average). Exercise means going for walks, skiing, swimming, and training/sports." If they exercised ≥1 time/week, they were asked about the intensity (no sweating/not out of breath [light], sweating/out of breath [moderate], or almost exhausted [vigorous]) and average duration (<15 minutes, 16–30 minutes, 30 minutes to 1 hour, or >1 hour). To categorize individuals into physical activity levels (physically active or inactive), total minutes per week were calculated by multiplying frequency and duration (frequency: never = 0, less than once a week = 0, once a week = 1, 2–3 times per week = 2.5, almost every day = 7; duration: <15 minutes = 0, 16–30 minutes = 23 minutes, 30 minutes to 1 hour = 45 minutes, >1 hour = 60 minutes) (20). Thereafter, total minutes per week and intensity were used to categorize individuals as either physically inactive (0–420 minutes with low physical activity or 0–59 minutes with moderate-to-vigorous physical activity per week) and physically active (defined as ≥60 minutes per week with moderate-to-vigorous physical activity) (20).

Disease activity. Disease activity was assessed using the ASDAS score (32), based on a calculation of the CRP level and 4 self-reported disease activity variables (neck/back/hip pain, peripheral pain, duration of morning stiffness, and patient global assessment of disease activity). The ASDAS score ranges from 0 to 10, where a higher score represents a higher disease activity. Self-reported disease activity was assessed with the BASDAI (33), a patient-reported questionnaire of 5 major symptoms: fatigue, neck-back-hip pain, peripheral joint pain, tenderness, and degree/length of morning stiffness. The BASDAI score ranges from 0 to 10, where a higher score represents a higher disease activity.

Physical function. Physical function was assessed with the Bath Ankylosing Spondylitis Functional Index, range 0–10, where a higher score indicates more functional limitations (34). Spinal mobility was assessed with the Bath Ankylosing Spondylitis Metrology Index, range 0–10, where a higher score represents more severe limitations of spinal mobility (35).

Statistical analysis. Background variables are shown as mean ± SDs, medians with minimum and maximum values or frequencies with percentages, as appropriate. To analyze the most important obstacles to physical activity at baseline, the frequency of scores from 1 to 3 (low scores) were summarized for all items on the Exercise Health Beliefs scale; thereafter, all items were ranged from high to low frequency. The effect of the intervention was assessed on the intent-to-treat population based on available data using analysis of covariance (ANCOVA) for calculation of the mean difference in change between the groups, with 95% confidence interval (95% CI). In the ANCOVA, 3- and 12-month values were used as the dependent variable, and for both timepoints, baseline values and study center were included as covariates. Normality assumptions of the ANCOVA models were

Table 2. The effect of the exercise intervention on exercise health beliefs at the end of the 3-month intervention and at 12 months follow-up*

	Exercise group			Control group			Between-group difference			
	Baseline (n = 50)	3 mo. (n = 48)	12 mo. (n = 43)	Baseline (n = 50)	3 mo. (n = 49)	12 mo. (n = 44)	3 mo. (95% CI)†	*P*	12 mo. (95% CI)†	*P*
Exercise beliefs (20–100)	84.4 ± 9.4	86.7 ± 8.7	86.7 ± 9.0	84.3 ± 8.5	82.5 ± 8.5	83.3 ± 8.8	4.0 (1.4, 6.6)	0.003	3.8 (1.0, 6.6)	0.008
Self-efficacy (4–20)	15.5 ± 2.9	17.2 ± 2.3	16.5 ± 3.0	15.7 ± 3.1	15.2 ± 3.1	15.8 ± 2.9	2.0 (1.1, 3.0)	<0.001	0.8 (–0.3, 1.9)	0.14
Barriers (3–15)	12.5 ± 2.1	12.6 ± 1.9	12.3 ± 2.3	12.0 ± 2.3	12.1 ± 2.1	11.9 ± 2.4	0.2 (–0.4, 0.8)	0.46	0.392‡	–
Benefits (5–25)	22.0 ± 2.8	22.4 ± 2.7	22.8 ± 2.7	21.6 ± 3.1	21.3 ± 3.7	21.7 ± 2.9	0.722‡	–	0.332‡	–
Impact on arthritis (8–40)	33.8 ± 4.6	34.4 ± 4.5	35.0 ± 4.3	34.6 ± 3.8	33.8 ± 4.0	33.8 ± 4.0	1.0 (–0.4, 2.4)	0.16	1.7 (0.3, 3.2)	0.01

* Values are the mean ± SD unless indicated otherwise. 95% CI = 95% confidence interval; mo. = months.
† Estimated mean group difference.
‡ Mann-Whitney U test of difference in change between the groups due to non-normally distributed residuals.

assessed by pp-plots of the residuals. For variables that were not normally distributed, group differences in change from baseline were analyzed using the Mann-Whitney U test. Differences between physically active (moderate-to-vigorous intensity) and physically inactive (low intensity or sedentary) patients in exercise health beliefs (with subscores) were analyzed with independent sample *t*-test or Mann-Whitney U test, as appropriate. The impact of exercise beliefs on physical activity level at the 12-month follow-up were analyzed using logistic regression analysis, with adjustments for age, sex, disease activity at baseline, and whether patients had received the exercise intervention. All statistical analyses were conducted in SPSS software. The level of statistical significance was set at a *P* value less than 0.05.

RESULTS

Participants. Demographic and clinical characteristics of the participants are shown in Table 1. The large majority of the participants (70%) had radiographic-verified sacroiliitis (ankylosing spondylitis), and the mean ± SD ASDAS-CRP level at baseline was 2.6 ± 0.7, which indicated a high disease level. No significant differences between clinical characteristics or disease-related variables between the groups were found at baseline. Thirty-eight participants in the exercise group (76%) followed ≥80% of the prescribed exercise protocol. Four participants left the exercise program right after the start, due to personal reasons. In the control group, 5 participants reported that they had performed cardiorespiratory or strength exercises ≥1 hour a week during the intervention period (Figure 1). One patient experienced chest pain and nausea during the exercises and completed the intervention at moderate intensity after advice from a cardiologist. Two patients reported persistent pain during exercise.

Obstacles to exercise. The most frequently reported obstacles to physical activity for the total study population at baseline were: "I'm not sure I could exercise regularly, even if I wanted to" (44%), "exercise is boring" (44%), "exercise causes disease flare-ups" (36%), "exercise takes too much time" (35%), and "exercise causes too much pain to be helpful" (23%).

Effect of the intervention on exercise health beliefs. There was a significant effect of the intervention on exercise health beliefs at 3 months of 4.0 (95% CI 1.4, 6.6; $P = 0.003$), and the effect persisted at 12 months follow-up, with 3.8 (95% CI 1.0, 6.6; $P = 0.008$) (Table 2). At 3 months, the effect was mainly seen on the subscore for exercise self-efficacy ($P < 0.001$). At 12 months follow-up, the effect on self-efficacy was reduced

Table 3. Difference in exercise health beliefs between physically inactive patients and physically active patients at 12 months follow-up*

	Inactive (n = 45)†	Active (n = 42)‡	*P*
Exercise health beliefs (20–100)	80.9 ± 8.3	87.5 ± 8.4	<0.001§
Self-efficacy (4–20)	15.5 ± 2.9	16.9 ± 2.6	0.02¶
Barriers (3–15)	11.6 ± 2.2	12.8 ± 2.3	0.002¶
Benefits (5–25)	20.6 ± 2.8	22.7 ± 2.6	<0.001¶
Impact on arthritis (8–40)	33.2 ± 3.8	35.0 ± 4.3	0.01¶

* Values are the mean ± SD.
† Physically active at a low intensity or <1 hour per week on a moderate-to-vigorous intensity. For Exercise Health Beliefs questionnaire total score and subscores, a higher score indicates stronger beliefs.
‡ ≥1 hour per week with physical activity at moderate-to-vigorous intensity level.
§ Independent sample *t*-test.
¶ Mann-Whitney U test.

and the effect was mainly seen on beliefs regarding the impact of exercise on arthritis ($P = 0.01$) (Table 2).

The impact of exercise health beliefs on physical activity level at 12 months follow-up. At 12 months follow-up, 42 of 87 participants (48%) were physically active, defined as ≥1 hour per week at a moderate-to-vigorous intensity level. Participants who were physically active had higher exercise beliefs ($P < 0.001$), including all subscores, compared to participants who were physically inactive (Table 3).

The odds for being physically active at 12 months follow-up increased, with 1.1 (95% CI 1.0, 1.20; $P = 0.003$), for every increment in exercise beliefs (range 20–100, 100 = best), adjusted for age, sex, disease activity at baseline, and whether or not the individual had received the exercise intervention (Table 4). Unadjusted exercise beliefs explained 18% of the variance in physical activity level, while the totally adjusted model explained 33%.

DISCUSSION

The main findings of the study show that the supervised intensive exercise program had a beneficial impact on participants' exercise health beliefs, and in turn, participants with a higher exercise health belief had a higher odds ratio for being physically active on a moderate-to-vigorous level 12 months after inclusion. This finding indicates that the exercise program, combining individualized motivating strategies with supervised sessions and individual sessions of personal choice, strengthened participants' exercise self-efficacy and exercise behavior.

Compared to pharmacologic treatment, exercise is a low-cost treatment with no harmful effects if it is individually adapted to the patients' health status and capacity, although exercise can produce sprains and strains of muscles, and modification of the exercise program is sometimes necessary. In the present study, some pain (≤5 on a scale from 0 to 10) was tolerated during the exercises. If the pain persisted during the following day, modification of the exercise program was done by the physical therapist. The standardized effect sizes at 3 and 12 months (0.5 and 0.4, respectively, data not shown) must be considered clinically meaningful.

Strong perceived benefits of exercise have previously been found to predict participation in exercise among patients with arthritis (16,25,36). Also, strong beliefs in the potential benefits of exercise have been suggested to be important for patients to overcome fear that exercise, especially on a vigorous intensity level, might exacerbate the arthritic disease (17,25). In the present study, the participants reported positive beliefs in the benefits of exercise at baseline, which might have facilitated the introduction of the exercise program at a vigorous intensity level.

Besides general barriers like "exercise is boring and time consuming," participants also reported disease-related obstacles for being physically active, such as pain and a risk for flare-ups. Our results are consistent with other reports suggesting that the disease itself plays an important role related to being physically active in axial SpA (14,16,19,37). However, in contrast to previous findings where perceived barriers to exercise were reported to decrease after a period of exercise (38), our intervention had no such effect, in spite of a high adherence to the prescribed exercise protocol among the participants. It has been suggested, however, that patients who exercise develop methods to overcome challenges, even if the barriers still exist (39).

Self-efficacy has been found to influence how people think, act, and motivate themselves (21,40). In other patient groups, low self-efficacy has been identified as being associated with poor exercise adherence and explaining low confidence in patients' ability to overcome obstacles to initiate and maintain exercise adherence (41). In the present study, participants' exercise self-efficacy improved significantly during the exercise period compared to the control group. Plausibly, the physical therapists' guidance and positive feedback throughout the exercise period, along with the supervised exercise sessions (including performance techniques), gave participants in the exercise group the confidence to exercise at a vigorous intensity. Also, the individual consultations with the physical therapist, addressing obstacles to exercise along with a constructive plan to overcome perceived

Table 4. The impact of exercise health beliefs on the chance of being physically active at 12 months follow-up*

	Crude estimates	*P*	Adjusted estimates†	*P*
Age, continuous	1.00 (0.96, 1.04)	0.88	1.00 (1.0, 1.1)	0.81
Sex				
Male	Ref.	–	Ref.	–
Female	1.38 (0.59, 3.20)	0.46	1.9 (0.7, 5.3)	0.24
Disease activity, continuous‡	1.1 (0.61, 2.1)	0.71	1.6 (0.8, 3.3)	0.23
Intervention				
Control	Ref.	–	Ref.	–
Exercise group	4.94 (1.99, 12.26)	0.001	4.9 (1.8, 13.4)	0.002
Exercise beliefs, continuous (20–100)	1.1 (1.0,1.2)	0.001	1.1 (1.0, 1.2)	0.003

* Values are the odds ratio (95% confidence interval) unless indicated otherwise. Being physically active was defined as ≥1 hour per week with moderate-to-high intensity physical activity. Ref. = reference.
† The totally adjusted model explained 33% of the variance in physical activity level.
‡ Ankylosing Spondylitis Disease Activity Score baseline.

barriers, might have facilitated participation and adherence to exercise. These results are in line with those of previous studies that emphasize the role of the physical therapist in patients' exercise self-efficacy (38,42,43). Although self-efficacy has been found to be easy to influence, it is temporary and situation- and task-oriented (40). Implementation over a longer period of time is suggested to be superior to enhance self-efficacy compared to a shorter period of time (20). In our study, the support from the physical therapist stopped when the exercise period ended. Possibly, a continuous support from the physical therapist, even after the end of intervention, would have resulted in a more sustained self-efficacy to exercise for the participants. However, this possibility calls for a longitudinal intervention and cannot be answered within this study.

In addition to self-efficacy, outcome expectations have been hypothesized to influence initiation and maintenance of a specific behavior (22). In the present study, participants' beliefs about the impact of exercise on the rheumatic disease enhanced over time, and a significant difference was found between the groups at 12 months follow-up. These results correspond to previous findings reporting associations between outcome expectations regarding exercise and exercise participation in patients with different arthritic diseases (25). Moreover, experiencing a successful intervention has been suggested to strengthen patients' beliefs in the treatment and illness management (20,44). Possibly, the experience of successfully managing exercise at a strenuous intensity, together with the positive effects of the intervention on disease activity and disease-related symptoms (26), increased the participants' exercise health beliefs in the current study.

While there is growing evidence on the positive effects of exercise at a high intensity in arthritis (45), doubts about the exercise mode still exist (17,18). Providing the correct information on how exercise impacts inflammatory arthritic disease can help patients overcome barriers, especially regarding physical activity at a vigorous level. Moreover, to uncover barriers, attitudes, and beliefs about exercise that can affect the outcome of an exercise intervention, we suggest identification of patients' exercise beliefs before initiation of exercise. This information can help physical therapists who are engaged in exercise to guide the start of interventions aimed at changing behavior patterns among patients. However, verbal information about the benefits of exercise and discussion of possible barriers to exercise might not always be enough to change health behaviors (16). We believe that initial support and guidance from a physical therapist, along with the practice of exercise at a vigorous intensity, will enhance patients' exercise beliefs and improve their confidence in mastering exercise at a vigorous intensity. Hopefully, this change will increase their chance of adhering to long-term physical activity at a health-enhancing level.

The major strength of the study is the multicenter, randomized controlled design and the large study population with axial SpA, including both patients with nonradiologic- and radiologic-verified axial SpA. Also, the exercise intervention following the American College of Sports Medicine recommendation for exercise (46) is a considered strength. A limitation is that the clinically meaningful change of the Exercise Health Belief questionnaire is not established, and a ceiling effect of this questionnaire cannot be excluded, as the mean score was relatively high at baseline. In addition, the physical activity level in the study population was assessed with a self-reported questionnaire, and a risk of recall bias cannot be excluded (47). People with rheumatic diseases tend to overestimate their activity level (48,49). However, objective assessment of physical activity was not applicable in this study.

This randomized controlled study of axial SpA demonstrated that a supervised high-intensity exercise program including motivation strategies, in combination with individual sessions of personal choice, strengthened participants' exercise health beliefs and exercise self-efficacy. Moreover, stronger exercise health beliefs were positively associated with a higher physical activity level at a moderate-to-vigorous intensity 12 months after inclusion. This knowledge can assist health care professionals in their work of promoting exercise at a health-enhancing level for patients with axial SpA.

ACKNOWLEDGMENTS

We thank the physical therapists and rheumatologists involved in the study, Emma Klittmar, Inger Jorid Berg, Sella Arrestad Provan, Melissa Woll Johansen, Elisabeth Pedersen, Linn Haukland, Mary Deighan Hanssen, and Kim Nielsen Martinsen.

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. Bilberg 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. Bilberg, Dagfinrud, Sveaas.
Acquisition of data. Bilberg, Dagfinrud, Sveaas.
Analysis and interpretation of data. Bilberg, Dagfinrud, Sveaas.

REFERENCES

1. Sieper J, Poddubnyy D. Axial spondyloarthritis. Lancet 2017;390:73–84.
2. Rudwaleit M, Landewé R, van der Heijde D, Listing J, Brandt J, Braun J, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part I): classification of paper patients by expert opinion including uncertainty appraisal. Ann Rheum Dis 2009;68:770–6.
3. Heikkilä S, Viitanen JV, Kautiainen H, Kauppi M. Functional long-term changes in patients with spondylarthropathy. Clin Rheumatol 2002;21:119–22.
4. Singh JA, Strand V. Spondyloarthritis is associated with poor function and physical health-related quality of life. J Rheumatol 2009;36:1012–20.
5. Halvorsen S, Vøllestad NK, Provan SA, Semb AG, van der Heijde D, Hagen KB, et al. Cardiorespiratory fitness and cardiovascular risk in

patients with ankylosing spondylitis: a cross-sectional comparative study. Arthritis Care Res (Hoboken) 2013;65:969–76.

6. O'Dwyer T, O'Shea F, Wilson F. Decreased physical activity and cardiorespiratory fitness in adults with ankylosing spondylitis: a cross-sectional controlled study. Rheumatol Int 2015;35:1863–72.
7. Baillet A, Gossec L, Carmona L, Wit M, van Eijk-Hustings Y, Bertheussen H, et al. Points to consider for reporting, screening for and preventing selected comorbidities in chronic inflammatory rheumatic diseases in daily practice: a EULAR initiative. Ann Rheum Dis 2016;75:965–73.
8. Exarchou S, Lie E, Lindstrom U, Askling J, Forsblad-d'Elia H, Turesson C, et al. Mortality in ankylosing spondylitis: results from a nationwide population-based study. Ann Rheum Dis 2016;75: 1466–72.
9. Bengtsson K, Forsblad-d'Elia H, Lie E, Klingberg E, Dehlin M, Exarchou S, et al. Are ankylosing spondylitis, psoriatic arthritis and undifferentiated spondyloarthritis associated with an increased risk of cardiovascular events? A prospective nationwide population-based cohort study. Arthritis Res Ther 2017;19:102.
10. Mathieu S, Soubrier M. Cardiovascular events in ankylosing spondylitis: a 2018 meta-analysis. Ann Rheum Dis 2019;78:e57.
11. Van der Heijde D, Ramiro S, Landewe R, Baraliakos X, Van den Bosch F, Sepriano A, et al. 2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis. Ann Rheum Dis 2017;76:978–91.
12. Rausch Osthoff AK, Niedermann K, Braun J, Adams J, Brodin N, Dagfinrud H, et al. 2018 EULAR recommendations for physical activity in people with inflammatory arthritis and osteoarthritis. Ann Rheum Dis 2018;77:1251–60.
13. Rausch Osthoff AK, Juhl CB, Knittle K, Dagfinrud H, Hurkmans E, Braun J, et al. Effects of exercise and physical activity promotion: meta-analysis informing the 2018 EULAR recommendations for physical activity in people with rheumatoid arthritis, spondyloarthritis and hip/knee osteoarthritis. RMD Open 2018;4:e000713.
14. Fongen C, Sveaas SH, Dagfinrud H. Barriers and facilitators for being physically active in patients with ankylosing spondylitis: a cross-sectional comparative study. Musculoskeletal Care 2015;13: 76–83.
15. Van Genderen S, Boonen A, van der Heijde D, Heuft L, Luime J, Spoorenberg A, et al. Accelerometer quantification of physical activity and activity patterns in patients with ankylosing spondylitis and population controls. J Rheumatol 2015;42:2369–75.
16. O'Dwyer T, McGowan E, O'Shea F, Wilson F. Physical activity and exercise: perspectives of adults with ankylosing spondylitis. J Phys Act Health 2016;13:504–13.
17. Iversen MD, Scanlon L, Frits M, Shadick NA, Sharby N. Perceptions of physical activity engagement among adults with rheumatoid arthritis and rheumatologists. Int J Clin Rheumtol 2015;10:67–77.
18. Munneke M, de Jong Z, Zwinderman AH, Ronday HK, van den Ende CH, Vliet Vlieland TP, et al. High intensity exercise or conventional exercise for patients with rheumatoid arthritis? Outcome expectations of patients, rheumatologists, and physiotherapists. Ann Rheum Dis 2004;63:804–8.
19. Rouse PC, Standage M, Sengupta R. Living with ankylosing spondylitis: an open response survey exploring physical activity experiences. Rheumatol Adv Pract 2019;3:rkz016.
20. Marks R. Self-efficacy and arthritis disability: an updated synthesis of the evidence base and its relevance to optimal patient care. Health Psychol Open 2014;1:2055102914564582.
21. Bandura A. Self-efficacy: the exercise of control New York: W. H. Freeman; 1977.
22. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977;84:191–215.
23. Sherwood NE, Jeffery RW. The behavioral determinants of exercise: implications for physical activity interventions. Annu Rev Nutr 2000; 20:21–44.
24. Glanz K, Rimer BK, Viswanath K. Health behavior and health education: theory, research, and practice. Hoboken (NJ): Wiley; 2008.
25. Gecht MR, Connell KJ, Sinacore JM, Prohaska TR. A survey of exercise beliefs and exercise habits among people with arthritis. Arthritis Care Res (Hoboken) 1996;9:82–8.
26. Sveaas SH, Bilberg A, Berg IJ, Provan SA, Rollefstad S, Semb AG, et al. High intensity exercise for 3 months reduces disease activity in axial spondyloarthritis (axSpA): a multicentre randomised trial of 100 patients. Br J Sports Med 2020;54:292–7.
27. Bandura A. The anatomy of stages of change. Am J Health Promot 1997;12:8–10.
28. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010;31:399–418.
29. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000;25:3186–91.
30. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health 2005;8:94–104.
31. Kurtze N, Rangul V, Hustvedt BE, Flanders WD. Reliability and validity of self-reported physical activity in the Nord-Trondelag Health Study: HUNT 1. Scand J Public Health 2008;36:52–61.
32. Van der Heijde D, Lie E, Kvien TK, Sieper J, Van den Bosch F, Listing J, et al. ASDAS, a highly discriminatory ASAS-endorsed disease activity score in patients with ankylosing spondylitis. Ann Rheum Dis 2009;68:1811–8.
33. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J. Rheumatol 1994;21:2286–91.
34. Calin A, Garrett S, Whitelock H, Kennedy LG, O'Hea J, Mallorie P, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J.Rheumatol 1994;21:2281–5.
35. Van der Heijde D, Landewe R, Feldtkeller E. Proposal of a linear definition of the Bath Ankylosing Spondylitis Metrology Index (BASMI) and comparison with the 2-step and 10-step definitions. Ann Rheum Dis 2008;67:489–93.
36. Ehrlich-Jones L, Lee J, Semanik P, Cox C, Dunlop D, Chang RW. Relationship between beliefs, motivation, and worries about physical activity and physical activity participation in persons with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2011;63:1700–5.
37. Niedermann K, Nast I, Ciurea A, Vliet Vlieland T, van Bodegom-Vos L. Barriers and facilitators of vigorous cardiorespiratory training in axial spondyloarthritis: surveys among patients, physiotherapists, and rheumatologists. Arthritis Care Res (Hoboken) 2019;71:839–51.
38. Mattukat K, Rennert D, Brandes I, Ehlebracht-König I, Kluge K, Mau W. Short- and long-term effects of intensive training and motivational programme for continued physical activity in patients with inflammatory rheumatic diseases. Eur J Phys Rehabil Med 2014;50: 395–409.
39. Veldhuijzen van Zanten JJ, Rouse PC, Hale ED, Ntoumanis N, Metsios GS, Duda JL, et al. Perceived barriers, facilitators and benefits for regular physical activity and exercise in patients with rheumatoid arthritis: a review of the literature. Sports Med 2015;45:1401–12.
40. Zulkosky K. Self-Efficacy: A concept analysis. Nursing Forum 2009; 44:93–102.

41. Jack K, McLean SM, Moffett JK, Gardiner E. Barriers to treatment adherence in physiotherapy outpatient clinics: a systematic review. Man Ther 2010;15:220–8.
42. Schoster B, Callahan LF, Meier A, Mielenz T, DiMartino L. The People with Arthritis Can Exercise (PACE) program: a qualitative evaluation of participant satisfaction. Prev Chronic Dis 2005;2:A11.
43. McGrane N, Galvin R, Cusack T, Stokes E. Addition of motivational interventions to exercise and traditional physiotherapy: a review and meta-analysis. Physiotherapy 2015;101:1–12.
44. Dager TN, Kjeken I, Berdal G, Sand-Svartrud AL, Bø I, Dingsør A, et al. Rehabilitation for patients with rheumatic diseases: patient experiences of a structured goal planning and tailored follow-up programme. SAGE Open Med 2017;5:2050312117739786.
45. Benatti FB, Pedersen BK. Exercise as an anti-inflammatory therapy for rheumatic diseases-myokine regulation. Nat Rev Rheumatol 2015;11:86–97.
46. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee IM, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc 2011;43:1334–59.
47. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 2008;336:601–5.
48. Yu CA, Rouse PC, Veldhuijzen Van Zanten JJ, Ntoumanis N, Kitas GD, Duda JL, et al. Subjective and objective levels of physical activity and their association with cardiorespiratory fitness in rheumatoid arthritis patients. Arthritis Res Ther 2015;17:59.
49. Van Genderen S, van den Borne C, Geusens P, van der Linden S, Boonen A, Plasqui G. Physical functioning in patients with ankylosing spondylitis: comparing approaches of experienced ability with self-reported and objectively measured physical activity. J Clin Rheumatol 2014;20:133–7.

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1205–1218
DOI 10.1002/acr.24560

American College of Rheumatology
Empowering Rheumatology Professionals

Real-World Six- and Twelve-Month Drug Retention, Remission, and Response Rates of Secukinumab in 2,017 Patients With Psoriatic Arthritis in Thirteen European Countries

Brigitte Michelsen,[1] Stylianos Georgiadis,[2] Daniela Di Giuseppe,[3] Anne G. Loft,[4] Michael J. Nissen,[5] Florenzo Iannone,[6] Manuel Pombo-Suarez,[7] Herman Mann,[8] Ziga Rotar,[9] Kari K. Eklund,[10] Tore K. Kvien,[11] Maria J. Santos,[12] Bjorn Gudbjornsson,[13] Catalin Codreanu,[14] Sema Yilmaz,[15] Johan K. Wallman,[16] Cecilie H. Brahe,[17] Burkhard Möller,[18] Ennio G. Favalli,[19] Carlos Sánchez-Piedra,[20] Lucie Nekvindova,[21] Matija Tomsic,[9] Nina Trokovic,[10] Eirik K. Kristianslund,[11] Helena Santos,[22] Thorvardur J. Löve,[23] Ruxandra Ionescu,[14] Yavuz Pehlivan,[24] Gareth T. Jones,[25] Irene van der Horst-Bruinsma,[26] Lykke M. Ørnbjerg,[17] Mikkel Østergaard,[27] and Merete L. Hetland[28]

Objective. There is a lack of real-life studies on interleukin-17 (IL-17) inhibition in psoriatic arthritis (PsA). We assessed real-life 6- and 12-month effectiveness (i.e., retention, remission, low disease activity [LDA], and response rates) of the IL-17 inhibitor secukinumab in PsA patients overall and across 1) number of prior biologic/targeted synthetic disease-modifying antirheumatic drugs (b/tsDMARDs), 2) years since diagnosis, and 3) European registries.

Methods. Thirteen quality registries in rheumatology participating in the European Spondyloarthritis Research Collaboration Network provided longitudinal, observational data collected as part of routine care for secondary use. Data were pooled and analyzed with Kaplan-Meier plots, log rank tests, Cox regression, and multiple linear and logistic regression analyses.

Results. A total of 2,017 PsA patients started treatment with secukinumab between 2015 and 2018. Overall secukinumab retention rates were 86% and 76% after 6 and 12 months, respectively. Crude (LUNDEX adjusted) 6-month remission/LDA (LDA including remission) rates for the 28-joint Disease Activity Index for Psoriatic Arthritis, the Disease Activity Score in 28 joints using the C-reactive protein level, and the Simplified Disease Activity Index (SDAI) were 13%/46% (11%/39%), 36%/55% (30%/46%), and 13%/56% (11%/47%), and 12-month rates were 11%/46% (7%/31%), 39%/56% (26%/38%), and 16%/62% (10%/41%), respectively. Clinical Disease Activity Index remission/LDA rates were similar to the SDAI rates. Six-month American College of Rheumatology 20%/50%/70% improvement criteria responses were 34%/19%/11% (29%/16%/9%); 12-month rates were 37%/21%/11% (24%/14%/7%). Secukinumab effectiveness was significantly better for b/tsDMARD-naive patients, similar across time since diagnosis (<2/2–4/>4 years), and varied significantly across the European registries.

Conclusion. In this large real-world study on secukinumab treatment in PsA, 6- and 12-month effectiveness was comparable to that in previous observational studies of tumor necrosis factor inhibitors. Retention, remission, LDA, and response rates were significantly better for b/tsDMARD-naive patients, were independent of time since diagnosis, and varied significantly across the European countries.

INTRODUCTION

Psoriatic arthritis (PsA) is a heterogeneous inflammatory rheumatic disease affecting, e.g., peripheral joints, axial spine, skin, and entheses, with significant impact on health-related quality of life (1–3). The treatment options for PsA have improved during the last few decades with the introduction of biologic disease-modifying antirheumatic drugs (bDMARDs) and targeted synthetic DMARDs (tsDMARDs) (4). Nevertheless, a recent real-world study of >14,000 patients with PsA, who started treatment with a tumor

The EuroSpA collaboration was financially supported by Novartis. Novartis had no influence on the data collection, statistical analyses, manuscript preparation, or decision to submit.

[1]Brigitte Michelsen, MD, Cand. Mag., PhD: Rigshospitalet, Glostrup, Denmark, Hospital of Southern Norway Trust, Kristiansand, Norway, and Diakonhjemmet Hospital, Oslo, Norway; [2]Stylianos Georgiadis, PhD: Rigshospitalet, Glostrup, Denmark; [3]Daniela Di Giuseppe, PhD: Karolinska Institutet, Stockholm, Sweden; [4]Anne G. Loft, MD, PhD: DANBIO Registry and Rigshospitalet, Glostrup, Denmark, and Aarhus University Hospital, Aarhus, Denmark; [5]Michael J. Nissen, MD, PhD: Geneva University Hospital, Geneva,

SIGNIFICANCE & INNOVATIONS

- Secukinumab retention, remission, low disease activity (LDA), and response rates were significantly better for biologics-naive patients after 6 as well as 12 months of treatment.
- Overall 6- and 12-month secukinumab retention rates were high; remission, LDA, and response rates were good; and overall effectiveness was comparable to that in previous observational studies of tumor necrosis factor inhibitors.
- This study is to date the largest real-world study on secukinumab effectiveness in patients with psoriatic arthritis, including 2,017 patients from 13 European national registries.
- The study documents the effectiveness of secukinumab for treatment of psoriatic arthritis in clinical practice and shows significantly better outcomes for biologics-naive patients. This may be taken into consideration in treatment decisions in routine clinical care.

necrosis factor inhibitor (TNFi), showed that less than one-half of the patients had achieved clinical remission after 6 months (5). Thus, there is an unmet need for other treatment options in patients with PsA (2,6).

The fully human IgG monoclonal interleukin-17A (IL-17A) inhibitor secukinumab was approved for use in PsA patients in the European Union in 2015 (7). Secukinumab has demonstrated good efficacy and safety in randomized controlled trials (RCTs) (8–10), whereas large observational studies on its effectiveness in patients with PsA are lacking.

Hence, the main objective of this study was to assess the overall real-life 12-month retention rate of secukinumab in PsA patients in Europe. Secondary objectives were to assess the overall 6-month secukinumab retention rate and 6- and 12-month remission, low disease activity (LDA), and response rates. These aims were assessed overall, as well as compared across number of previous b/tsDMARD treatments, time since diagnosis, and the European registries.

Switzerland; [6]Florenzo Iannone, MD, PhD: GISEA Registry and University of Bari, Bari, Italy; [7]Manuel Pombo-Suarez, MD, PhD: Hospital Clinico Universitario, Santiago de Compostela, Spain; [8]Herman Mann, MD, PhD: Charles University, Prague, Czech Republic; [9]Ziga Rotar, MD, PhD, Matija Tomsic, MD, PhD: University Medical Centre Ljubljana, Ljubljana, Slovenia; [10]Kari K. Eklund, MD, PhD, Nina Trokovic, MD, PhD: Helsinki University Hospital and ORTON Orthopaedic Hospital of the Orton Foundation, Helsinki, Finland; [11]Tore K. Kvien, MD, PhD, Eirik K. Kristianslund, MD, PhD: Diakonhjemmet Hospital, Oslo, Norway; [12]Maria J. Santos, MD, PhD: Reuma.pt Registry and Universidade de Lisboa, Lisbon, Portugal; [13]Bjorn Gudbjornsson, MD, PhD: Centre for Rheumatology Research (ICEBIO), University Hospital and University of Iceland, Reykjavik, Iceland; [14]Catalin Codreanu, MD, PhD, Ruxandra Ionescu, MD, PhD: University of Medicine and Pharmacy Carol Davila, Bucharest, Romania; [15]Sema Yilmaz, MD, PhD: Selcuk University School of Medicine, Selcuklu, Turkey; [16]Johan K. Wallman, MD, PhD: Lund University, Skåne University Hospital, Lund, Sweden; [17]Cecilie H. Brahe, MD, PhD, Lykke M. Ørnbjerg, MD, PhD: Rigshospitalet and DANBIO Registry, Glostrup, Denmark; [18]Burkhard Möller, MD, PhD: Universitätsklinik für Rheumatologie, Immunologie und Allergologie, Inselspital, Bern, Switzerland; [19]Ennio G. Favalli, MD, PhD: ASST Gaetano Pini-CTO Institute, Milan, Italy; [20]Carlos Sánchez-Piedra, MD, PhD: Spanish Society of Rheumatology, Madrid, Spain; [21]Lucie Nekvindova, MD, PhD: Charles University, Prague, Czech Republic, and Institute of Biostatistics and Analyses, Masaryk University, Brno, Czech Republic; [22]Helena Santos, MD, PhD: Reuma.pt Registry and Portuguese Institute of Rheumatology, Lisbon, Portugal; [23]Thorvardur J. Löve, MD, PhD: University of Iceland and Landspitali University Hospital, Reykjavik, Iceland; [24]Yavuz Pehlivan, MD, PhD: Uludağ University, Bursa, Turkey; [25]Gareth T. Jones, PhD: University of Aberdeen, Aberdeen, UK; [26]Irene van der Horst-Bruinsma, MD, PhD: Amsterdam University Medical Centres, VU University Medical Centre, Amsterdam, The Netherlands; [27]Mikkel Østergaard, MD, PhD, DMSc: Rigshospitalet, Glostrup, Denmark, and University of Copenhagen, Copenhagen, Denmark; [28]Merete L. Hetland, MD, PhD, DMSc: Rigshospitalet and DANBIO Registry, Glostrup, Denmark, and University of Copenhagen, Copenhagen, Denmark.

Drs. Østergaard and Hetland contributed equally to this work.

Dr. Michelsen has received consulting fees from Novartis (less than $10,000) and research support from Novartis. Dr. Georgiadis has received research support from Novartis. Dr. Loft has received consulting fees and/or speaking fees from AbbVie, Eli Lilly and Company, Janssen, MSD, Novartis, Pfizer, and UCB (less than $10,000 each). Dr. Nissen has received consulting fees and/or speaking fees from AbbVie, Celgene, Eli Lilly and Company, Novartis, and Pfizer (less than $10,000 each). Dr. Iannone has received consulting fees and/or speaking fees from AbbVie, Eli Lilly and Company, MSD, Novartis, Pfizer, Roche, Janssen, Bristol Myers Squibb, and Sanofi (less than $10,000 each). Dr. Rotar has received consulting fees and/or speaking fees from AbbVie, Amgen, Biogen, Eli Lilly and Company, Medis, MSD, Novartis, Pfizer, and Sanofi (less than $10,000 each). Dr. Kvien has received consulting fees and/or speaking fees from AbbVie, Amgen, Biogen, Celltrion, Egis, Eli Lilly and Company, Eva Pharma, Ewopharma, Gilead, Hikma, MSD, Mylan, Novartis/Sandoz, Oktal, Orion Pharma, Hospira/Pfizer, Roche, Sanofi, and UCB (less than $10,000 each) and research support to Diakonhjemmet Hospital from AbbVie, Bristol Myers Squibb, MSD, Pfizer, Roche, and UCB. Dr. Santos has received speaking fees from AbbVie, Novartis, Pfizer, and Roche (less than $10,000 each). Dr. Gudbjornsson has received speaking fees from Novartis (less than $10,000). Dr. Codreanu has received consulting fees and/or speaking fees from AbbVie, Accord Healthcare, Alfasigma, Egis, Eli Lilly and Company, Ewopharma, Genesis, Mylan, Novartis, Pfizer, Roche, Sandoz, and UCB (less than $10,000 each). Dr. Wallman has received consulting fees from Celgene, Eli Lilly and Company, and Novartis (less than $10,000 each). Dr. Brahe has received research support from Novartis. Dr. Favalli has received consulting fees and/or speaking fees from AbbVie, Bristol Myers Squibb, Eli Lilly and Company, UCB Novartis, Pfizer, and Sanofi-Genzyme (less than $10,000 each). Dr. Ionescu has received consulting fees and/or speaking fees from AbbVie, Amgen, Eli Lilly and Company, Ewopharma, Novartis/Sandoz, Pfizer, Roche, and UCB (less than $10,000 each). Dr. Jones has received research support from AbbVie, Pfizer, UCB, Amgen, and GlaxoSmithKline. Dr. van der Horst-Bruinsma has received consulting fees and/or speaking fees from AbbVie, Eli Lilly and Company, Novartis, UCB, Bristol Myers Squibb, MSD, UCB, and Pfizer (less than $10,000 each). Dr. Ørnbjerg has received research support from Novartis. Dr. Østergaard has received consulting fees and/or speaking fees from AbbVie, Bristol Myers Squibb, Boehringer-Ingelheim, Celgene, Eli Lilly and Company, Hospira, Janssen, Merck, Novartis, Novo, Orion, Pfizer, Regeneron, Roche, Sandoz, Sanofi, and UCB (less than $10,000 each) and research support from AbbVie, Bristol Myers Squibb, Celgene, Merck, and Novartis. Dr. Hetland has received consulting fees and/or speaking fees from Eli Lilly and Company, Orion Pharma, Biogen, Pfizer, CellTrion, Merck, and Samsung Bioepis (less than $10,000 each) and research support from Bristol Myers Squibb, MSD, AbbVie, Roche, Novartis, Biogen, and Pfizer. No other disclosures relevant to this article were reported.

Address correspondence to Brigitte Michelsen, MD, Cand. Mag., PhD, Copenhagen Center for Arthritis Research, Center for Rheumatology and Spine Diseases, Centre for Head and Orthopaedics, Rigshospitalet, Valdemar Hansens Vej 17, 2600 Glostrup, Denmark. Email: brigitte.michelsen@regionh.dk.

Submitted for publication August 28, 2020; accepted in revised form January 12, 2021.

PATIENTS AND METHODS

The European Spondyloarthritis Research Collaboration Network (EuroSpA RCN). The EuroSpA RCN currently includes 15 European quality registries of spondyloarthritis patients (5,11,12). The collaboration was initiated in 2016, but data collection had started as early as 1999 in some of the registries. The main aim of the collaboration is to investigate clinically relevant research questions by secondary use of prospectively collected real-life data (5,11,12). All data are anonymized in the different registries before upload to a secured central server. The data are quality checked and pooled prior to statistical analyses.

Patients. The studies in the EuroSpA collaboration are based on secondary use of real-world data already collected in the different registries, i.e., independently of the current study. In this study, we included data from PsA patients starting secukinumab for the first time between May 2015 and December 2018 in 13 countries in the EuroSpA RCN (ranked by number of patients): ARTIS (Sweden), DANBIO (Denmark), SCQM (Switzerland), GISEA (Italy), BIOBADASER (Spain), ATTRA (Czech Republic), biorx.si (Slovenia), Reuma.pt (Portugal), NOR-DMARD (Norway), ROB-FIN (Finland), ICEBIO (Iceland), RRBR (Romania), and TURKBIO (Turkey). Inclusion criteria for the current analyses were age ≥18 years at treatment initiation, a diagnosis of PsA as judged by the treating rheumatologist, and a registered start and, if relevant, stop date of secukinumab. The exclusion criterion was patients with no available clinical data.

Assessments. We included data on age, sex, time since diagnosis, current smoking status (yes/no), body mass index (kg/m^2), start and stop dates of secukinumab, previous b/tsDMARD treatment, evaluator's global assessment, patient's global assessment, pain and fatigue, C-reactive protein (CRP) level (mg/liter), erythrocyte sedimentation rate (ESR, mm/hour), 28-joint Disease Activity Index for Psoriatic Arthritis (DAPSA28) score (13), Disease Activity Score in 28 joints using the CRP level (DAS28-CRP) score (14), Clinical Disease Activity Index (CDAI) score (15), and Simplified Disease Activity Index (SDAI) score (15). The following remission/LDA and response measures were calculated at 6 and 12 months treatment: DAPSA28 remission (≤4) (13), DAPSA28 LDA (≤14) (13), DAS28-CRP remission (<2.6) (16), DAS28-CRP LDA (≤3.2) (17), CDAI remission (≤2.8) (15), CDAI LDA (≤10) (15), SDAI remission (≤3.3) (15), SDAI LDA (≤11) (15), American College of Rheumatology (ACR)/European Alliance of Associations for Rheumatology (EULAR) Boolean remission (18), change in DAPSA28, DAS28-CRP, CDAI, and SDAI, ACR 20%/50%/70% improvement criteria (ACR20/50/70) response (19), and EULAR response (moderate/good) (17).

Primary and secondary outcomes. Primary outcome was the overall 12-month secukinumab retention rate. Secondary outcomes were the overall 6-month secukinumab retention rate and 6- and 12-month remission, LDA, and response rates.

Statistical analyses. All statistical analyses were performed according to a predefined statistical analysis plan developed by the researchers in the EuroSpA collaboration. Descriptive statistics were performed for demographic data and baseline disease activity measures. All effectiveness analyses were compared across 1) the number of previous b/tsDMARDs (0/1/≥2), 2) years since diagnosis (<2/2–4/>4), and 3) the individual registries. Drug retention was explored by Kaplan-Meier analyses with log rank test and by Cox regression analyses adjusted for age, sex, and time since diagnosis (comparisons 1 and 3 above), or age and sex (comparison 2 above).

Remission, LDA, response rates, and change measures were compared by chi-square test, Fisher's exact test, and Kruskal-Wallis test, as appropriate, as well as by multiple linear and logistic regression analyses adjusted for age, sex, and time since diagnosis (comparisons 1 and 3 above), or age and sex (comparison 2 above), as appropriate. Multiple comparisons for the number of previous b/tsDMARDs (0/1/≥2) were performed by log rank test, chi-square test, Fisher's exact test, and Kruskal-Wallis with post hoc Dunn test, as appropriate, where *P* values were adjusted by applying the Holm's correction.

Significance of relevant groups was tested through likelihood ratio test or Wald test, as appropriate, by comparing 2 nested models. A significance level of 0.05 was used for all statistical tests. In adjusted analyses, multivariate imputation by chained equations (including 50 imputed data sets) was used for 463 patients with missing data for time since diagnosis (no missing data for age and sex). The variables used for imputing time since diagnosis were age, sex, country, and b/tsDMARD treatment series number. None of the other variables including outcome was imputed. To avoid inflating remission and response rates, these were provided both as crude values and with LUNDEX (20) adjustment, i.e., integrating clinical response and adherence to therapy in a composite value. In the Kaplan-Meier and Cox regression analyses, observations were censored by first occurrence of 1 of the following: end of registry follow-up or date of data extraction. Patients who stopped treatment due to remission or other reasons (e.g., pregnancy) were censored at the stop date to reflect that their withdrawal was not due to lack of effectiveness or adverse events. The baseline date was defined as the secukinumab treatment start date. To assess the robustness regarding the main outcomes, sensitivity analyses for patients 1) having ≥1 swollen joints (of 28) at baseline and 2) having date of data extraction at least 12 months after secukinumab treatment start were performed. Competing risk analysis was performed for a cumulative incidence curve showing withdrawal due to adverse events and lack of effectiveness. Numbers available for each of the analyses are shown in Supplementary Tables 1–7,

available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24560. Statistical analyses were performed with R, version 3.6.1.

Ethics. Approval of the study was obtained from the respective national data protection agencies and research ethical committees according to the individual legal regulatory requirements in the different registries/countries. The study was performed in accordance with the Declaration of Helsinki and followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines (21).

RESULTS

We included a total of 2,017 PsA patients who started secukinumab for the first time (Table 1). The number of patients included from the different European registries varied from 30 (TURKBIO) to 657 (ARTIS). Significant heterogeneity in demographic data and baseline disease activity across the European registries was found (see Supplementary Table 1, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24560). Information on doses was not registered systematically. Of 745 patients in whom doses were registered, 42% of the patients initiated secukinumab 150 mg, and 58% initiated secukinumab 300 mg.

Secukinumab retention rates. The crude 95% confidence interval secukinumab retention rates were overall 76% (74–78%) after 12 months and 86% (85–88%) after 6 months of treatment (Table 2). Secukinumab retention rates after 6 as well as 12 months of treatment were significantly higher in biologics-naive patients compared with patients previously treated with ≥2 b/tsDMARDs (Table 2 and Figure 1A). The findings were similar in 6- and 12-month adjusted Cox regression analyses (see Supplementary Table 8, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24560).

Secukinumab retention was not significantly associated with time since diagnosis, either in unadjusted or in adjusted analyses

Table 1. Demographic characteristics and baseline disease activity measures*

	All patients (n = 2,017)	b/tsDMARD naive (n = 441)	1 prior b/tsDMARD (n = 461)	≥2 prior b/tsDMARDs (n = 1,115)	*P*†
Age, years	52 (44–60)	50 (41–58)	51 (44–59)	53 (45–60)	<0.001
Men, %	43	51	46	39	<0.001
Years since diagnosis	7 (3–13)	4 (1–10)	6 (2–12)	8 (5–14)	<0.001
Current smokers, %	19	18	22	18	0.356
BMI, kg/m^2	27.5 (24.3–31.2)	28.1 (24.1–31.8)	27.3 (24.1–30.1)	27.3 (24.5–31.6)	0.309
B/tsDMARD treatment, % first (% last previous)					<0.001 (<0.001)
Adalimumab	29 (21)	–	30 (30)	28 (18)	
Certolizumab	5 (8)	–	5 (5)	5 (10)	
Etanercept	28 (22)	–	25 (25)	29 (20)	
Golimumab	10 (12)	–	9 (9)	10 (13)	
Infliximab	22 (13)	–	15 (15)	25 (12)	
Other‡	7 (24)	–	15 (15)	3 (27)	
CRP, mg/liter	5 (2–12)	7 (2–19)	4 (2–9)	5 (2–12)	<0.001
ESR, mm/hour	16 (7–31)	20 (8–36)	13 (6–27)	16 (7–30)	0.002
TJC28	4 (1–9)	5 (1–10)	3 (1–8)	4 (1–9)	<0.001
SJC28	1 (0–4)	2 (0–6)	1 (0–3)	2 (0–4)	<0.001
Patient global score	70 (50–83)	70 (51–84)	67 (42–80)	70 (50–85)	<0.001
Pain score	66 (46–80)	65 (45–78)	62 (40–78)	68 (48–81)	<0.001
Fatigue score	70 (50–85)	65 (50–80)	65 (41–80)	73 (55–87)	<0.001
Evaluator global score	40 (20–60)	57 (30–75)	35 (20–50)	35 (20–50)	<0.001
HAQ score	1.1 (0.6–1.6)	1.0 (0.5–1.5)	1.0 (0.5–1.4)	1.2 (0.8–1.8)	<0.001
DAPSA28 score	25.9 (17.4–37.6)	29.1 (19.1–41.9)	22.3 (13.5–32.4)	26.2 (18.0–37.6)	<0.001
DAS28-CRP score	4.2 (3.2–5.0)	4.5 (3.6–5.4)	3.8 (2.7–4.6)	4.2 (3.3–5.0)	<0.001
SDAI score	19.5 (12.9–28.9)	24.4 (15.3–35.4)	16.9 (10.0–24.3)	18.9 (13.0–27.5)	<0.001
CDAI score	18.0 (12.0–26.7)	22.6 (14.3–33.9)	16.0 (8.9–23.6)	17.5 (12.0–25.4)	<0.001

* Values are the median (interquartile range) unless indicated otherwise. Numbers available for each of the analyses are shown in Supplementary Table 5, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24560. b/tsDMARD = biologic/targeted synthetic disease-modifying antirheumatic drug; BMI = body mass index; CDAI = Clinical Disease Activity Index; CRP = C-reactive protein; DAPSA28 = 28-joint Disease Activity Index for Psoriatic Arthritis; DAS28-CRP = Disease Activity Score in 28 joints using the CRP level; ESR = erythrocyte sedimentation rate; HAQ = Health Assessment Questionnaire; SDAI = Simplified Disease Activity Index; SJC28 = Swollen joint count in 28 joints; TJC28 = Tender joint count in 28 joints.
† Comparisons between b/tsDMARD-naive patients and 1 prior and ≥2 prior b/tsDMARD-treated patients were performed with Kruskal-Wallis or chi-square test, as appropriate.
‡ Other previous b/tsDMARDs include ustekinumab, rituximab, abatacept, tocilizumab, apremilast, anakinra, and additionally (never as first b/tsDMARD) baricitinib and tofacitinib. Patients were included between May 2015 and December 2018.

Table 2. Treatment effectiveness after 6 and 12 months of secukinumab treatment (unadjusted analyses)*

	All patients (n = 2,017)	b/tsDMARD naive (n = 441)	1 prior b/tsDMARD (n = 461)	≥2 prior b/tsDMARDs (n = 1,115)	P†
Secukinumab drug retention rate, % (95% CI)					
6 months	86 (85–88)	90 (87–93)	86 (83–90)	85 (83–87)	0.045§
12 months	76 (74–78)	82 (78–86)	78 (74–82)	72 (70–75)	0.001§
Time in weeks to secukinumab withdrawal before 12 months due to the following‡					
Primary and secondary lack of effectiveness	24 (17, 33)	24 (17, 35)	24 (17, 30)	24 (17, 34)	0.691
Adverse events	14 (6, 28)	22 (13, 28)	15 (7, 25)	12 (5, 29)	0.395
Remission	21 (20, 43)	20 (19, 20)	–	43 (32, 43)	0.236
Other reasons	21 (12, 32)	27 (15, 40)	10 (4, 36)	21 (15, 27)	0.161
DAPSA28 score					
6 months	15.1 (8.2, 25.0)	10.1 (5.2, 17.5)	15.7 (9.0, 22.0)	16.9 (9.6, 27.1)	<0.001¶
12 months	14.9 (8.1, 24.8)	10.2 (4.1, 16.3)	15.2 (8.4, 23.6)	16.3 (10.0, 26.0)	<0.001¶
DAS28-CRP score					
6 months	3.0 (2.2, 4.0)	2.5 (1.9, 3.3)	3.1 (2.2, 3.9)	3.2 (2.4, 4.2)	<0.001#
12 months	3.0 (2.2, 4.0)	2.5 (1.7, 3.3)	3.0 (2.1, 3.9)	3.2 (2.4, 4.2)	<0.001¶
SDAI score					
6 months	10.2 (5.4, 16.7)	6.9 (3.5, 11.0)	10.4 (6.3, 15.3)	11.4 (6.6, 18.5)	<0.001¶
12 months	9.2 (5.2, 15.2)	5.7 (2.5, 9.5)	9.3 (5.8, 16.2)	10.5 (6.8, 16)	<0.001¶
CDAI score					
6 months	9.3 (4.9, 15.9)	6.2 (3.4, 10.5)	9.4 (5.5, 14.4)	10.9 (6.0, 17.8)	<0.001#
12 months	8.5 (4.4, 14.2)	5.1 (2.1, 9.3)	8.7 (5.2, 14.6)	9.8 (5.8, 14.9)	<0.001¶
Change in DAPSA28 score from baseline					
6 months	−9.5 (−20.7, −0.2)	−17.2 (−27.5, −8.3)	−8.5 (−17.6, −0.1)	−6.6 (−18.3,0.3)	<0.001¶
12 months	−10.3 (−21.9, −1.3)	−16.2 (−28.0, −8.3)	−5.0 (−10.6,1.0)	−10.3 (−21.9, −0.2)	<0.001#
Change in DAS28-CRP score from baseline					
6 months	−0.9 (−1.9, −0.1)	−2.0 (−3.0, −1.1)	−0.8 (−1.7, 0.1)	−0.6 (−1.6, 0.01)	<0.001¶
12 months	−1.1 (−2.0, −0.1)	−1.9 (−3.1, −1.0)	−0.5 (−1.3, 0.03)	−1.0 (−1.9, −0.02)	<0.001#
Change in SDAI score from baseline					
6 months	−8.9 (−17.4, −2.0)	−16.9 (−26.1, −9.3)	−7.5 (−13.5, −1.1)	−6.0 (−13.4, −0.2)	<0.001¶
12 months	−9.7 (−18.6, −2.4)	−15.0 (−24.2, −7.5)	−4.9 (−10.4, 1.3)	−9.6 (−17.9, −2.2)	<0.001#
Change in CDAI score from baseline					
6 months	−8.0 (−16.1, −1.6)	−15.1 (−24.6, −8.0)	−6.0 (−13.1, −1.4)	−5.3 (−12.2, −0.1)	<0.001¶
12 months	−8.8 (−16.0, −2.0)	−13.9 (−21.5, −7.3)	−5.0 (−10.4, 0.8)	−8.1 (−15.9, −1.5)	<0.001#
DAPSA28 score ≤4, %					
6 months					
Crude	13	23	13	10	<0.001§
LUNDEX adjusted‡	11	20	11	8	<0.001¶
12 months					
Crude	11	22	11	8	<0.001¶
LUNDEX adjusted‡	7	17	7	5	<0.001§
DAPSA28 score ≤14, %					
6 months					
Crude	46	64	45	41	<0.001¶
LUNDEX adjusted‡	39	57	37	34	<0.001¶
12 months					
Crude	46	70	46	40	<0.001¶
LUNDEX adjusted‡	31	52	30	26	<0.001¶
DAS28-CRP score <2.6, %					
6 months					
Crude	36	53	35	30	<0.001¶
LUNDEX adjusted‡	30	47	29	25	<0.001¶
12 months					
Crude	39	55	41	34	<0.001¶
LUNDEX adjusted‡	26	41	27	21	<0.001¶
DAS28-CRP score ≤3.2, %					
6 months					
Crude	55	71	57	49	<0.001¶
LUNDEX adjusted‡	46	63	47	40	<0.001¶
12 months					
Crude	56	72	55	51	<0.001¶
LUNDEX adjusted‡	38	54	37	33	<0.001¶

(Continued)

Table 2. (*Cont'd*)

	All patients (n = 2,017)	b/tsDMARD naive (n = 441)	1 prior b/tsDMARD (n = 461)	≥2 prior b/tsDMARDs (n = 1,115)	*P*†
SDAI score ≤3.3, %					
6 months					
Crude	13	24	13	9	<0.001¶
LUNDEX adjusted‡	11	21	11	8	<0.001¶
12 months					
Crude	16	32	11	11	<0.001¶
LUNDEX adjusted‡	10	24	8	7	<0.001¶
SDAI score ≤11, %					
6 months					
Crude	56	75	56	48	<0.001¶
LUNDEX adjusted‡	47	66	47	39	<0.001¶
12 months					
Crude	62	81	58	56	<0.001¶
LUNDEX adjusted‡	41	61	39	36	<0.001¶
CDAI score ≤2.8, %					
6 months					
Crude	13	19	12	10	0.004§
LUNDEX adjusted‡	10	17	10	8	0.002§
12 months					
Crude	16	32	14	11	<0.001¶
LUNDEX adjusted‡	11	24	10	7	<0.001¶
CDAI score ≤10, %					
6 months					
Crude	55	74	58	46	<0.001¶
LUNDEX adjusted‡	46	66	48	38	<0.001¶
12 months					
Crude	59	79	58	53	<0.001¶
LUNDEX adjusted‡	40	59	39	34	<0.001¶
ACR/EULAR Boolean remission, %					
6 months					
Crude	9	20	8	6	<0.001¶
LUNDEX adjusted‡	8	18	6	5	<0.001¶
12 months					
Crude	9	17	9	6	<0.001§
LUNDEX adjusted‡	6	12	6	4	<0.001§
ACR20 response, %					
6 months					
Crude	34	59	26	27	<0.001¶
LUNDEX adjusted‡	29	52	22	22	<0.001¶
12 months					
Crude	37	63	16	33	<0.001¶
LUNDEX adjusted‡	24	47	10	21	<0.001¶
ACR50 response, %					
6 months					
Crude	19	41	11	13	<0.001¶
LUNDEX adjusted‡	16	36	9	11	<0.001¶
12 months					
Crude	21	45	4	16	<0.001¶
LUNDEX adjusted‡	14	34	3	10	<0.001¶
ACR70 response, %					
6 months					
Crude	11	26	7	6	<0.001¶
LUNDEX adjusted‡	9	23	6	5	<0.001¶
12 months					
Crude	11	28	4	6	<0.001¶
LUNDEX adjusted‡	7	21	3	4	<0.001¶
EULAR good/moderate response, %					
6 months					
Crude	59	83	57	50	<0.001¶
LUNDEX adjusted‡	49	74	48	41	<0.001¶

(Continued)

Table 2. (*Cont'd*)

	All patients (n = 2,017)	b/tsDMARD naive (n = 441)	1 prior b/tsDMARD (n = 461)	≥2 prior b/tsDMARDs (n = 1,115)	*P*†
12 months					
Crude	60	79	44	59	<0.001¶
LUNDEX adjusted‡	40	59	30	38	<0.001¶

* Values are the median (interquartile range) unless indicated otherwise. Numbers available for each of the analyses are shown in Supplementary Table 6, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24560. 95% CI = 95% confidence interval; ACR = American College of Rheumatology; ACR20/50/70 = ACR 20%/50%/70% improvement criteria; b/tsDMARD = biologic/targeted synthetic disease-modifying antirheumatic drug; CDAI = Clinical Disease Activity Index; CRP = C-reactive protein; DAPSA28 = 28-joint Disease Activity Index for Psoriatic Arthritis; DAS28-CRP = Disease Activity Score in 28 joints using the CRP level; EULAR = European Alliance of Associations for Rheumatology; SDAI = Simplified Disease Activity Index.
† Drug retention rates were compared across the 3 groups with Kaplan-Meier with log rank test, continuous measures by Kruskal-Wallis test, and proportions by chi-square test or Fisher's exact test, as appropriate. Multiple comparisons between groups were conducted by log rank test, Kruskal-Wallis with post hoc Dunn test, chi-square test, or Fisher's exact test, as appropriate, with *P* values to be adjusted by applying the Holm's correction.
‡ Patients with at least 12 months from secukinumab start to date of data extraction. Patients who stopped treatment due to remission or other reasons (e.g., pregnancy) were censored at the stop date to reflect that their withdrawal was not due to lack of effectiveness or adverse events.
§ Statistically significant difference between b/tsDMARD-naive patients and patients treated with ≥2 prior b/tsDMARDs.
¶ Statistically significant difference between b/tsDMARD-naive patients and patients treated with 1 prior b/tsDMARD. Statistically significant difference between b/tsDMARD-naive patients and patients treated with ≥2 prior b/tsDMARDs.
# Statistically significant difference between b/tsDMARD-naive patients and patients treated with 1 prior b/tsDMARD. Statistically significant difference between b/tsDMARD-naive patients and patients treated with ≥2 prior b/tsDMARDs. Statistically significant difference between patients treated with 1 prior b/tsDMARD and ≥2 prior b/tsDMARDs. Significance level for all tests is 0.05.

(see Supplementary Tables 2 and 8). The number of included patients varied considerably across the European registries (from 30 to 657 patients). Significant differences in retention rates across the registries were observed, with 6-month retention rates varying between 80% (DANBIO) and 97% (TURKBIO), and 12-month retention rates varying from 51% (ROB-FIN) to 92% (RRBR and ATTRA) (Table 3 and Figure 2). Similar differences were found in adjusted analyses (see Supplementary Table 8).

Remission. Crude and LUNDEX-adjusted proportions of patients achieving DAPSA28, DAS28-CRP, SDAI, and CDAI remission after 6 and 12 months are presented in Table 2. DAPSA28, SDAI, and CDAI remission rates were similar (~10–15%), whereas approximately one-third of the patients achieved DAS28-CRP remission.

The proportion of patients achieving remission was significantly higher in biologics-naive patients than in patients previously treated with 1 and ≥2 b/tsDMARDs (Table 2, Figure 3, and Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24560). Adjusted analyses gave similar results (see Supplementary Table 9, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24560).

Crude and adjusted remission rates at 6 and 12 months of treatment were independent of time since diagnosis (see Supplementary Tables 2 and 9). Overall, heterogeneity in crude and adjusted remission rates across the European registries was found (Table 3 and Supplementary Table 7).

LDA (including remission). Crude and LUNDEX-adjusted proportions of patients achieving DAPSA28, DAS28-CRP, SDAI, and CDAI LDA after 6 and 12 months of treatment are presented in Table 2, Figure 3, and Supplementary Figure 1, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24560/abstract. Overall, crude and LUNDEX-adjusted LDA rates were significantly higher in biologics-naive patients, also in adjusted analyses (see Supplementary Table 9).

For all outcomes, achievement of LDA was independent of time since diagnosis (see Supplementary Table 2), also after adjustment (see Supplementary Table 9). Significant heterogeneities in crude (Table 3) and adjusted (see Supplementary Table 9) LDA rates were seen between the registries.

Response rates. ACR20/50/70 responses were achieved by 34%/19%/11% of the patients, and EULAR moderate/good response by 59% of the patients after 6 months. After 12 months, numbers were largely the same (Table 2). Changes in outcome measures from baseline to 6 months (and 12 months, respectively) were as follows: DAPSA28 –9.5 (–10.3), DAS28-CRP –0.9 (–1.1), SDAI –8.9 (–9.7), and CDAI –8.0 (–8.8).

Significantly better outcomes for ACR20/50/70 and EULAR moderate/good responses were observed for biologics-naive patients (Table 2, Figure 3, and Supplementary Figure 1, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24560/abstract), also after adjustment (see Supplementary Table 9).

Response rates were independent of time since diagnosis (see Supplementary Table 2), also in adjusted analyses (see Supplementary Table 9). Significant heterogeneity in response rates between the European registries was found in crude as well as adjusted analyses (Table 3 and Supplementary Table 9).

Safety. Of the 2,017 patients starting secukinumab, 1,543 patients started treatment at least 12 months before date of data

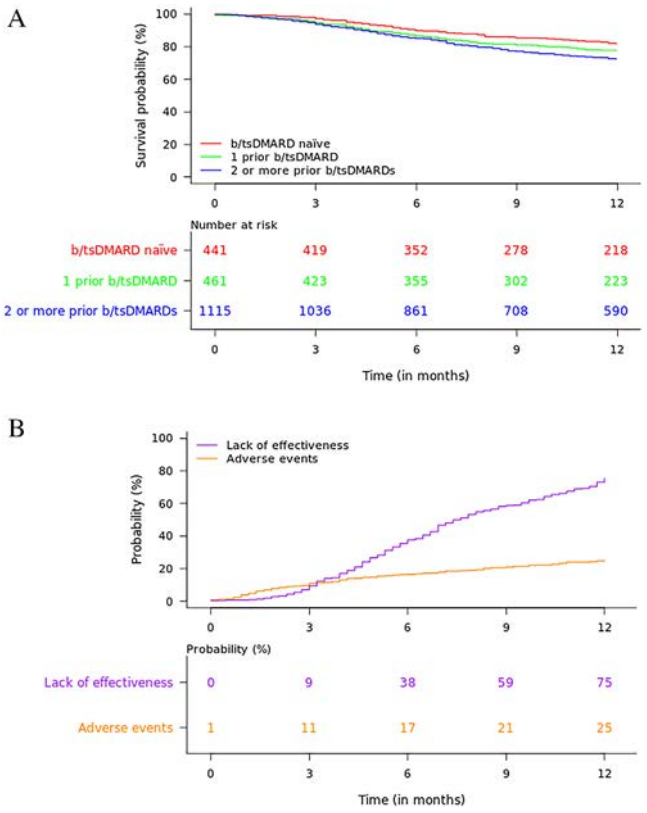


Figure 1. **A**, Pooled 12-month secukinumab retention rates stratified by number of previous biologic/targeted synthetic disease-modifying antirheumatic drugs (b/tsDMARDs) (Kaplan-Meier curve with log rank test; $P = 0.001$). **B**, Cumulative incidence curve for withdrawal of secukinumab due to adverse events and lack of effectiveness.

extraction. Of these 1,543 patients, 602 patients withdrew from secukinumab before 12 months, of whom 107 patients withdrew due to adverse events. Time in weeks to secukinumab withdrawal for these 107 patients was similar across number of previous b/tsDMARDs (0/1/≥2) (Table 2). More patients withdrew from secukinumab due to lack of effectiveness than due to adverse events (Table 2). The cumulative incidence curve, which estimates the cumulative probabilities of treatment withdrawal over time, shows that the cumulative probability of withdrawal due to lack of effectiveness is higher than adverse events after ~4 months of treatment (Figure 1B).

Sensitivity analyses. Sensitivity analyses of 976 patients with ≥1 swollen joint (of 28) at the start of secukinumab treatment showed largely similar results to the analyses in Table 2 (see Supplementary Table 3, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24560). Sensitivity analyses of patients with secukinumab initiation at least 12 months before date of data extraction also showed largely similar results but did not reach significance for the 6-month comparison of retention rates across number of previous b/tsDMARDs (b/tsDMARD naive: 89% [86–93%]; 1 prior b/tsDMARD: 85% [81–89%]; ≥2 prior b/tsDMARDs: 85% [82–87%]; $P = 0.107$ [see Supplementary Table 4]).

DISCUSSION

This large real-life study of secukinumab effectiveness (i.e., drug retention, remission, LDA, and response rates) included 2,017 patients with PsA treated as part of routine care in 13 countries across Europe. Overall, high 6-month (86%) and 12-month (76%) secukinumab retention rates were found. Secukinumab effectiveness was significantly better for biologics-naive patients after 6 as well as 12 months of treatment, was independent of time since diagnosis, and differed significantly across the European countries. Remission, LDA, and response rates were overall comparable to previous real-life observations in patients treated with a TNFi (5). Hence, this large observational study documents the effectiveness of secukinumab in the treatment of PsA patients.

Secukinumab effectiveness has previously been reported in one observational study of 76 Spanish PsA patients, in which 12-month retention rates were somewhat higher than in our study; for biologics-naive patients, it was 91%, and for non-naive patients, it was 82% (22). Good 1-year secukinumab effectiveness has also been reported in an Italian observational study of 130 PsA patients (23). In the FUTURE 1 RCT, 89% of the patients in the 150-mg secukinumab group reached 52 weeks, and ACR20/50 responses at week 24 and 52 were achieved by 50%/35% and 60%/43% of the patients, respectively (24). In our observational study, ACR20/50 responses at week 26 and 52 were lower than in the FUTURE 1 study (34%/19% and 37%/21%), probably reflecting that the study designs differed substantially (longitudinal observational study with 22% biologics-naive patients versus RCT with 71% biologics-naive patients). In the FUTURE 5 RCT, 91% of the patients treated with 150 mg of secukinumab completed 52 weeks of treatment, with ACR20/50/70 responses of 64%/41%/26%, thus substantially higher than in our study (10).

Interestingly, the overall secukinumab retention rates in this real-life study were similar to the retention rates of TNFi in a recently published observational study of 14,261 European biologics-naive PsA patients (86% versus 86% at 6 months; 76% versus 77% at 12 months, respectively) and numerically slightly higher for biologics-naive secukinumab than TNFi starters (90% versus 86% at 6 months, and 82% versus 77% at 12 months, respectively) (5). Overall, remission and response rates for patients treated with secukinumab were fairly similar to what was reported for TNFi (5) as well as to the effectiveness of TNFi reported in other, smaller observational studies (25–28).

Similar to findings in observational studies on TNFi, and in the FUTURE 2 and 5 trials, the current study demonstrated that effectiveness of secukinumab declines with increasing previous use of

Table 3. Retention, remission, low disease activity (including remission), and response rates after 6 and 12 months of secukinumab treatment across European observational registries (unadjusted analyses)*

	ARTIS (n = 657)	ATTRA (n = 151)	BIO-BADASER (n = 154)	Biorx.si (n = 79)	DANBIO (n = 313)	GISEA (n = 180)	ICEBIO (n = 38)	NOR-DMARD (n = 60)	Reuma.pt (n = 68)	ROB-FIN (n = 47)	RRBR (n = 37)	SCQM (n = 203)	TURKBIO (n = 30)	P†
Drug retention rate, % (95% CI)														
6 months	82 (79–85)	94 (90–98)	93 (89–97)	92 (87–98)	80 (75–84)	96 (93–99)	87 (77–98)	83 (74–94)	91 (84–98)	83 (73–94)	92 (83–100)	90 (85–94)	97 (90–100)	<0.001
12 months	66 (62–70)	92 (88–97)	84 (78–91)	89 (82–96)	70 (65–76)	88 (82–93)	77 (64–93)	72 (61–86)	86 (78–96)	51 (39–68)	92 (83–100)	82 (77–88)	–	<0.001
DAPSA28 score ≤4														
6 months														
Crude	8	22	–	11	12	–	0	14	14	19	–	35	21	<0.001
LUNDEX	6	21	–	10	9	–	0	12	13	16	–	31	–	<0.001
12 months														
Crude	6	23	–	10	14	–	NC	16	16	NC	–	15	NC	0.004
LUNDEX	3	20	–	9	9	–	NC	11	14	NC	–	12	–	0.002
DAPSA28 score ≤14														
6 months														
Crude	37	61	–	53	44	–	42	61	54	54	–	58	74	<0.001
LUNDEX	30	58	–	49	33	–	35	50	49	45	–	51	–	<0.001
12 months														
Crude	35	79	–	48	46	–	NC	68	63	NC	–	59	NC	<0.001
LUNDEX	19	67	–	42	29	–	NC	49	55	NC	–	47	–	<0.001
DAS28-CRP score <2.6														
6 months														
Crude	27	46	50	40	33	–	29	54	45	42	60	49	63	<0.001
LUNDEX	21	44	44	37	25	–	24	44	41	35	52	43	–	<0.001
12 months														
Crude	25	62	49	46	41	–	50	63	50	NC	NC	41	NC	<0.001
LUNDEX	14	53	36	41	26	–	37	45	44	NC	NC	33	–	<0.001
DAS28-CRP score ≤3.2														
6 months														
Crude	45	62	69	64	53	–	64	79	64	73	73	60	74	<0.001
LUNDEX	36	58	60	59	40	–	54	64	58	61	64	53	–	<0.001
12 months														
Crude	43	80	73	60	54	–	75	68	80	NC	NC	76	NC	<0.001
LUNDEX	23	69	54	52	34	–	55	49	70	NC	NC	61	–	<0.001
SDAI score ≤3.3														
6 months														
Crude	6	21	–	16	12	–	0	10	18	21	24	21	21	0.003
LUNDEX	5	20	–	15	9	–	0	8	16	17	21	18	–	0.003
12 months														
Crude	8	32	–	23	14	–	NC	25	5	NC	NC	15	NC	0.002
LUNDEX	4	27	–	20	8	–	NC	18	4	NC	NC	12	–	<0.001

(Continued)

Table 3. (*Cont'd*)

	ARTIS (n = 657)	ATTRA (n = 151)	BIO-BADASER (n = 154)	Biorx.si (n = 79)	DANBIO (n = 313)	GISEA (n = 180)	ICEBIO (n = 38)	NOR-DMARD (n = 60)	Reuma.pt (n = 68)	ROB-FIN (n = 47)	RRBR (n = 37)	SCQM (n = 203)	TURKBIO (n = 30)	*P*†
SDAI score ≤11														
6 months														
Crude	42	68	–	58	54	–	46	67	64	67	76	65	74	<0.001
LUNDEX	33	64	–	53	41	–	39	54	58	55	66	57	–	<0.001
12 months														
Crude	50	88	–	56	55	–	NC	83	85	NC	NC	67	NC	<0.001
LUNDEX	27	75	–	49	35	–	NC	59	74	NC	NC	53	–	<0.001
CDAI score ≤2.8														
6 months														
Crude	6	18	–	12	14	–	0	9	11	12	20	23	21	0.007
LUNDEX	5	17	–	11	10	–	0	7	10	10	17	21	–	0.008
12 months														
Crude	8	31	–	19	14	-	14	25	10	NC	NC	23	NC	0.003
LUNDEX	4	27	–	17	9	–	10	18	9	NC	NC	18	–	<0.001
CDAI score ≤10														
6 months														
Crude	41	68	–	60	53	–	41	59	64	62	76	64	74	<0.001
LUNDEX	33	64	–	55	40	–	35	48	58	52	66	56	–	<0.001
12 months														
Crude	44	88	–	58	56	–	57	83	80	NC	NC	63	NC	<0.001
LUNDEX	24	75	–	51	35	–	42	59	70	NC	NC	50	–	<0.001
ACR/EULAR Boolean remission														
6 months														
Crude	5	22	18	9	9	0	0	8	6	12	23	15	16	<0.001
LUNDEX	4	21	15	8	7	0	0	7	6	10	20	13	–	<0.001
12 months														
Crude	5	24	15	7	9	0	5	10	8	15	NC	7	NC	<0.001
LUNDEX	3	21	11	7	6	0	4	7	7	8	NC	5	–	<0.001
ACR20 response														
6 months														
Crude	24	55	–	59	25	–	NC	NC	56	–	–	NC	22	<0.001
LUNDEX	20	51	–	54	19	–	NC	NC	51	–	–	18	–	<0.001
12 months														
Crude	27	67	–	50	24	–	NC	NC	NC	–	–	NC	NC	<0.001
LUNDEX	14	58	–	44	15	–	NC	NC	NC	–	–	24	–	<0.001
ACR50 response														
6 months														
Crude	11	36	–	38	12	–	NC	NC	31	–	–	NC	11	<0.001
LUNDEX	9	34	–	35	9	–	NC	NC	28	–	–	NC	–	<0.001
12 months														
Crude	15	45	–	35	11	–	NC	NC	NC	–	–	NC	NC	<0.001
LUNDEX	8	39	–	30	7	–	NC	NC	NC	–	–	NC	–	<0.001

(Continued)

Table 3. (*Cont'd*)

	ARTIS (n = 657)	ATTRA (n = 151)	BIO-BADASER (n = 154)	Biorx.si (n = 79)	DANBIO (n = 313)	GISEA (n = 180)	ICEBIO (n = 38)	NOR-DMARD (n = 60)	Reuma.pt (n = 68)	ROB-FIN (n = 47)	RRBR (n = 37)	SCQM (n = 203)	TURKBIO (n = 30)	*P*†
ACR70 response														
6 months														
Crude	6	21	–	21	7	–	NC	NC	19	–	–	NC	11	0.010
LUNDEX	4	20	–	19	6	–	NC	NC	17	–	–	NC	–	0.001
12 months														
Crude	5	24	–	31	5	–	NC	NC	NC	–	–	NC	NC	0.002
LUNDEX	3	20	–	27	3	–	NC	NC	NC	–	–	NC	–	0.001
EULAR good/moderate response														
6 months														
Crude	50	88	69	83	50	–	NC	55	62	47	93	NC	39	<0.001
LUNDEX	40	82	61	76	38	–	NC	45	56	39	81	NC	–	<0.001
12 months														
Crude	48	93	63	77	60	–	NC	43	83	NC	NC	64	NC	<0.001
LUNDEX	26	79	47	68	37	–	NC	30	73	NC	NC	50	–	<0.001

* Values are the percentage unless indicated otherwise. Numbers available for each of the analyses are shown in Supplementary Table 7. Registries and countries are as follows: ARTIS (Sweden), DANBIO (Denmark), SCQM (Switzerland), GISEA (Italy), BIOBADASER (Spain), ATTRA (Czech Republic), biorx.si (Slovenia), Reuma.pt (Portugal), NOR-DMARD (Norway), ROB-FIN (Finland), ICEBIO (Iceland), RRBR (Romania), and TURKBIO (Turkey). 95% CI = 95% confidence interval; ACR = American College of Rheumatology; ACR20/50/70 = ACR 20%/50%/70% improvement criteria; CDAI = Clinical Disease Activity Index; DAPSA28 = 28-joint Disease Activity Index for Psoriatic Arthritis; DAS28-CRP = Disease Activity Score in 28 joints using the C-reactive protein level; EULAR = European Alliance of Associations for Rheumatology; NC = not calculated (because data from <10 patients available); SDAI = Simplified Disease Activity Index.

† Comparisons between the registries were performed with Kaplan-Meier with log rank test for retention rates and chi-square test or Fisher's exact test for remission and response rates, as appropriate.

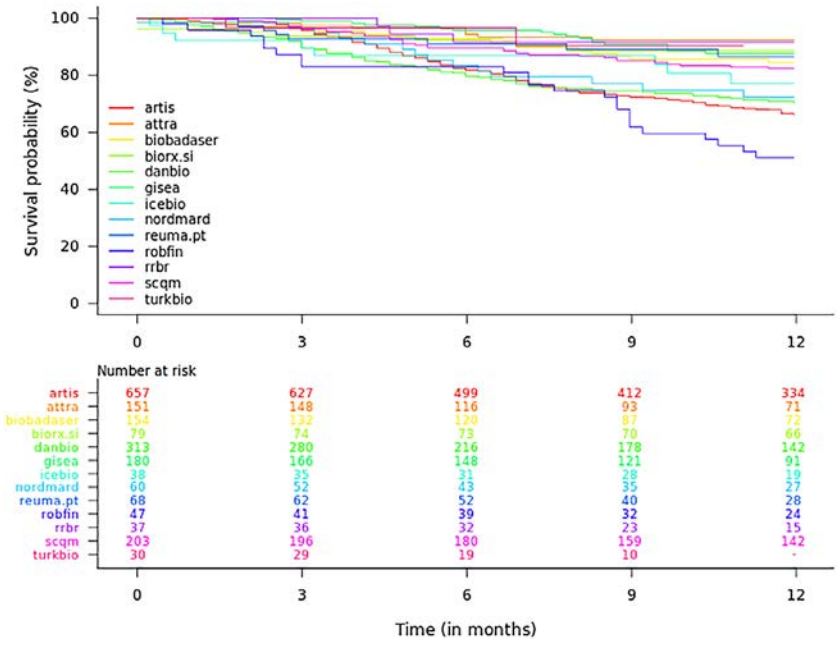


Figure 2. Twelve-month secukinumab retention rates compared across the European registries (Kaplan-Meier curve with log rank test; $P < 0.001$). Registries and countries are as follows: ARTIS (Sweden), DANBIO (Denmark), SCQM (Switzerland), GISEA (Italy), BIOBADASER (Spain), ATTRA (Czech Republic), biorx.si (Slovenia), Reuma.pt (Portugal), NOR-DMARD (Norway), ROB-FIN (Finland), ICEBIO (Iceland), RRBR (Romania), and TURKBIO (Turkey).

b/tsDMARDs, possibly reflecting confounding by indication (9,27,29,30). The similar secukinumab effectiveness for patients with different disease durations found in this study is also in accordance with previous findings for TNFi in patients with PsA (31–33).

In the 2017 updated treat-to-target recommendations for PsA, the DAPSA and minimal disease activity (MDA) are the preferred measures to define treatment target in PsA patients (34). In our study, the DAPSA (including a 66 swollen/68 tender joint count) (35) was only available in a minority of patients. Instead, we used the DAPSA28, which only requires a 28-joint count (13). The DAPSA28 has shown good criterion, correlational, and construct validity, as well as sensitivity to change, although the original DAPSA should be preferred when available (13). MDA was not assessed in the study due to lack of data on enthesitis and psoriasis in the majority of registries.

We chose the DAS28-CRP over the DAS28-ESR due to less missing data for the DAS28-CRP. Overall, the DAS28-CRP was a more liberal remission criterion than the SDAI, the CDAI, and the DAPSA28 in our study, which is consistent with previous reports (5,12,36,37). In the DAPSA28, SDAI, and CDAI LDA measures, we chose to include remission in accordance with the DAS28 LDA, as we believe that rheumatologists will be mainly interested in knowing how many patients at least were in LDA (i.e., in LDA or remission).

The major strength of this study is the 12-month longitudinal, observational study design with inclusion of a high number of PsA patients from 13 different countries. Furthermore, the data included in the study were collected independently of commercial interests as part of standard care. Hence, although Novartis supports the EuroSpA collaboration, Novartis had no influence on data collection, statistical analyses, manuscript preparation, or the decision to submit. Major limitations of the study include lack of data on extraarticular inflammatory involvement and the fact that data on the optimal number of joints (66/68) were generally not available, which may have led to underestimation of disease activity. Furthermore, the DAS28, the CDAI, and the SDAI are composite scores originally developed for RA and not PsA.

Heterogeneity in baseline characteristics and secukinumab effectiveness across the registries was found. Importantly, the number of included patients (from 30 to 657) and proportions of biologics-naive patients (from 5% to 97%) varied considerably across the registries and may explain some of the heterogeneity in effectiveness measures, e.g., a higher proportion of biologics-naive patients may positively impact upon treatment outcomes. Moreover, low patient numbers in some registries will lead to more uncertain estimates, i.e., single patients will have a higher influence on outcomes. Also, the influence of different treatment guidelines and access to treatment in the different European countries were not accounted for in this study. Hence, interpretation of the pooled analyses should be done with caution. Of note, however, consistent results in prespecified unadjusted and adjusted analyses were found.

Furthermore, as is often the case in observational studies, some missing data on disease states and response rates were observed, challenging the generalizability of the findings.

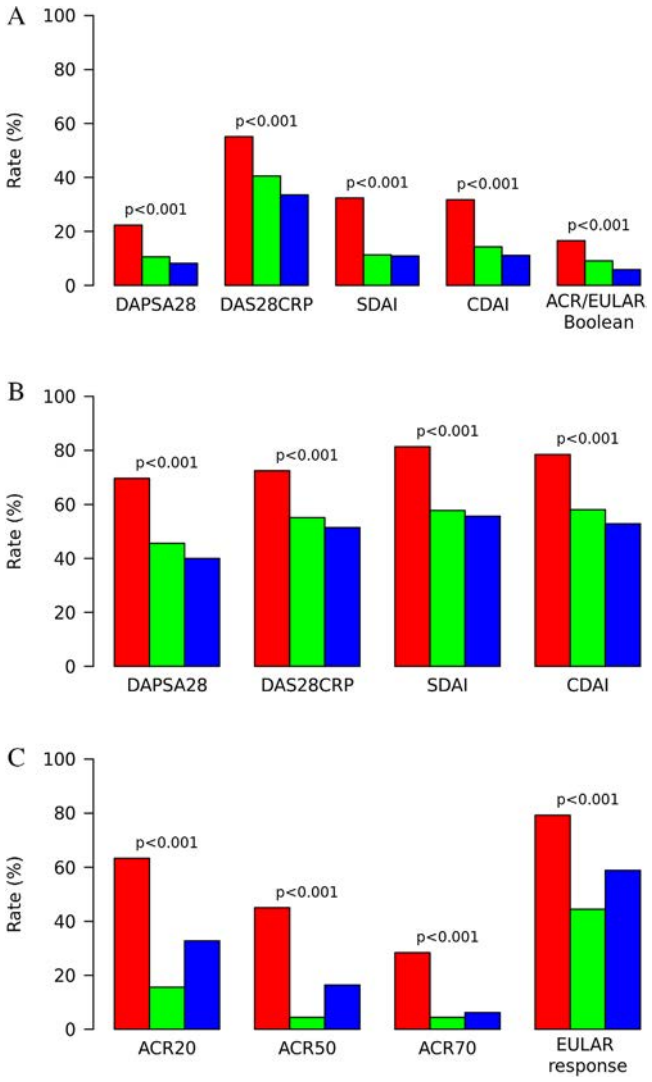


Figure 3. Bar charts of crude proportions of patients achieving remission (**A**), LDA (including remission) (**B**), and response rates (**C**) after 12 months of secukinumab treatment compared across number of previous biologic/targeted synthetic disease-modifying antirheumatic drugs (b/tsDMARDs) for b/tsDMARDs naive (red), 1 prior b/tsDMARD (green), and ≥2 prior b/tsDMARDs (blue). ACR = American College of Rheumatology; CDAI = Clinical Disease Activity Index; DAPSA28 = 28-joint Disease Activity Index for Psoriatic Arthritis; DAS28-CRP = Disease Activity Score in 28 joints using the CRP level; EULAR = European Alliance of Associations for Rheumatology; SDAI = Simplified Disease Activity Index.

However, the study is by far the largest real-life study to date on secukinumab effectiveness in patients with PsA.

In conclusion, in this longitudinal observational study of >2,000 patients with PsA treated with secukinumab, we found high retention rates after 6 and 12 months of treatment and good remission, LDA, and response rates. Secukinumab effectiveness was significantly better for biologics-naive patients, was independent of time since diagnosis, and varied across European registries.

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. Michelsen 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. Michelsen, Georgiadis, Di Giuseppe, Loft, Nissen, Iannone, Pombo-Suarez, Mann, Rotar, Eklund, Kvien, Santos, Gudbjornsson, Codreanu, Yilmaz, Wallman, Brahe, Möller, Favalli, Sánchez-Piedra, Nekvindova, Tomsic, Trokovic, Kristianslund, Santos, Löve, Ionescu, Pehlivan, Jones, van der Horst-Bruinsma, Ørnbjerg, Østergaard, Hetland.

Acquisition of data. Michelsen, Georgiadis, Di Giuseppe, Loft, Nissen, Iannone, Pombo-Suarez, Mann, Rotar, Eklund, Kvien, Santos, Gudbjornsson, Codreanu, Yilmaz, Wallman, Brahe, Möller, Favalli, Sánchez-Piedra, Nekvindova, Tomsic, Trokovic, Kristianslund, Santos, Löve, Ionescu, Pehlivan, Jones, van der Horst-Bruinsma, Ørnbjerg, Østergaard, Hetland.

Analysis and interpretation of data. Michelsen, Georgiadis, Di Giuseppe, Loft, Nissen, Iannone, Pombo-Suarez, Mann, Rotar, Eklund, Kvien, Santos, Gudbjornsson, Codreanu, Yilmaz, Wallman, Brahe, Möller, Favalli, Sánchez-Piedra, Nekvindova, Tomsic, Trokovic, Kristianslund, Santos, Löve, Ionescu, Pehlivan, Jones, van der Horst-Bruinsma, Ørnbjerg, Østergaard, Hetland.

ROLE OF THE STUDY SPONSOR

Novartis 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 Novartis.

REFERENCES

1. Gladman DD, Antoni C, Mease P, Clegg DO, Nash P. Psoriatic arthritis: epidemiology, clinical features, course, and outcome. Ann Rheum Dis 2005;64 Suppl 2:ii14–7.
2. Michelsen B, Diamantopoulos AP, Hoiberg HK, Soldal DM, Kavanaugh A, Haugeberg G. Need for improvement in current treatment of psoriatic arthritis: study of an outpatient clinic population. J Rheumatol 2017;44:431–6.
3. Michelsen B, Uhlig T, Sexton J, van der Heijde D, Hammer HB, Kristianslund EK, et al. Health-related quality of life in patients with psoriatic and rheumatoid arthritis: data from the prospective multicentre nor-DMARD study compared with Norwegian general population controls. Ann Rheum Dis 2018;77:1290–4.
4. Gossec L, Baraliakos X, Kerschbaumer A, de Wit M, McInnes I, Dougados M, et al. EULAR recommendations for the management of psoriatic arthritis with pharmacological therapies: 2019 update. Ann Rheum Dis 2020;79:700–12.
5. Brahe CH, Ornbjerg LM, Jacobsson L, Nissen MJ, Kristianslund EK, Mann H, et al. Retention and response rates in 14 261 PsA patients starting TNF inhibitor treatment-results from 12 countries in EuroSpA. Rheumatology (Oxford) 2020;59:1640–50.
6. Hetland ML. Psoriatic arthritis: still room for improvement. Lancet 2020;395:1463–5.
7. McInnes IB, Mease PJ, Kirkham B, Kavanaugh A, Ritchlin CT, Rahman P, et al. Secukinumab, a human anti-interleukin-17a monoclonal antibody, in patients with psoriatic arthritis (future 2): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet 2015;386:1137–46.
8. McInnes IB, Mease PJ, Ritchlin CT, Rahman P, Gottlieb AB, Kirkham B, et al. Secukinumab sustains improvement in signs and symptoms

of psoriatic arthritis: 2 year results from the phase 3 future 2 study. Rheumatology (Oxford) 2017;56:1993–2003.

9. Mease P, van der Heijde D, Landewe R, Mpofu S, Rahman P, Tahir H, et al. Secukinumab improves active psoriatic arthritis symptoms and inhibits radiographic progression: primary results from the randomised, double-blind, phase III future 5 study. Ann Rheum Dis 2018;77: 890–7.
10. Van der Heijde D, Mease PJ, Landewe RB, Rahman P, Tahir H, Singhal A, et al. Secukinumab provides sustained low rates of radiographic progression in psoriatic arthritis: 52-week results from a phase 3 study, future 5. Rheumatology (Oxford) 2020;59: 1325–34.
11. Ornbjerg LM, Brahe CH, Askling J, Ciurea A, Mann H, Onen F, et al. Treatment response and drug retention rates in 24 195 biologic-naive patients with axial spondyloarthritis initiating TNFi treatment: routine care data from 12 registries in the EuroSpA collaboration. Ann Rheum Dis 2019;78:1536–44.
12. Michelsen B, Ornbjerg LM, Kvien TK, Pavelka K, Nissen MJ, Nordstrom D, et al. Impact of discordance between patient's and evaluator's global assessment on treatment outcomes in 14 868 patients with spondyloarthritis. Rheumatology (Oxford) 2020;59: 2455–61.
13. Michelsen B, Sexton J, Smolen JS, Aletaha D, Krogh NS, van der Heijde D, et al. Can disease activity in patients with psoriatic arthritis be adequately assessed by a modified disease activity index for psoriatic arthritis (DAPSA) based on 28 joints? Ann Rheum Dis 2018;77: 1736–41.
14. Wells G, Becker JC, Teng J, Dougados M, Schiff M, Smolen J, et al. Validation of the 28-joint Disease Activity Score (DAS28) and European League Against Rheumatism response criteria based on C-reactive protein against disease progression in patients with rheumatoid arthritis, and comparison with the DAS28 based on erythrocyte sedimentation rate. Ann Rheum Dis 2009;68:954–60.
15. Aletaha D, Smolen JS. The Simplified Disease Activity Index (SDAI) and Clinical Disease Activity Index (CDAI) to monitor patients in standard clinical care. Best Pract Res Clin Rheumatol 2007;21:663–75.
16. Smolen JS, Braun J, Dougados M, Emery P, Fitzgerald O, Helliwell P, et al. Treating spondyloarthritis, including ankylosing spondylitis and psoriatic arthritis, to target: recommendations of an international task force. Ann Rheum Dis 2014;73:6–16.
17. Fransen J, Antoni C, Mease PJ, Uter W, Kavanaugh A, Kalden JR, et al. Performance of response criteria for assessing peripheral arthritis in patients with psoriatic arthritis: analysis of data from randomised controlled trials of two tumour necrosis factor inhibitors. Ann Rheum Dis 2006;65:1373–8.
18. Felson DT, Smolen JS, Wells G, Zhang B, van Tuyl LH, Funovits J, et al. American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. Arthritis Rheum 2011;63:573–86.
19. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38: 727–35.
20. Kristensen LE, Saxne T, Geborek P. The LUNDEX, a new index of drug efficacy in clinical practice: results of a five-year observational study of treatment with infliximab and etanercept among rheumatoid arthritis patients in southern Sweden. Arthritis Rheum 2006;54: 600–6.
21. Dixon WG, Carmona L, Finckh A, Hetland ML, Kvien TK, Landewe R, et al. EULAR points to consider when establishing, analysing and reporting safety data of biologics registers in rheumatology. Ann Rheum Dis 2010;69:1596–602.
22. Pinto Tasende JA, Maceiras Pan FJ, Mosquera Martinez JA, Fernandez Dominguez L, Correa Rey B, Garcia Porrua C. Secukinumab as biological treatment for psoriatic arthritis in real clinical practice. Reumatol Clin (Engl Ed) 2021;17:203–6.
23. Chimenti MS, Fonti GL, Conigliaro P, Sunzini F, Scrivo R, Navarini L, et al. One-year effectiveness, retention rate, and safety of secukinumab in ankylosing spondylitis and psoriatic arthritis: a real-life multicenter study. Expert Opin Biol Ther 2020;20:813–21.
24. Mease PJ, McInnes IB, Kirkham B, Kavanaugh A, Rahman P, van der Heijde D, et al. Secukinumab inhibition of interleukin-17a in patients with psoriatic arthritis. N Engl J Med 2015;373:1329–39.
25. Glintborg B, Østergaard M, Dreyer L, Krogh NS, Tarp U, Hansen MS, et al. Treatment response, drug survival, and predictors thereof in 764 patients with psoriatic arthritis treated with anti–tumor necrosis factor α therapy: results from the nationwide Danish DANBIO registry. Arthritis Rheum 2011;63:382–90.
26. Stober C, Ye W, Guruparan T, Htut E, Clunie G, Jadon D. Prevalence and predictors of tumour necrosis factor inhibitor persistence in psoriatic arthritis. Rheumatology (Oxford) 2018;57:158–63.
27. Fagerli KM, Lie E, van der Heijde D, Heiberg MS, Kalstad S, Rodevand E, et al. Switching between TNF inhibitors in psoriatic arthritis: data from the nor-DMARD study. Ann Rheum Dis 2013;72:1840–4.
28. Carmona L, Gomez-Reino JJ. Survival of TNF antagonists in spondylarthritis is better than in rheumatoid arthritis: data from the Spanish registry BIOBADASER. Arthritis Res Ther 2006;8:R72.
29. Costa L, Perricone C, Chimenti MS, Del Puente A, Caso P, Peluso R, et al. Switching between biological treatments in psoriatic arthritis: a review of the evidence. Drugs R D 2017;17:509–22.
30. Kavanaugh A, McInnes IB, Mease PJ, Hall S, Chinoy H, Kivitz AJ, et al. Efficacy of subcutaneous secukinumab in patients with active psoriatic arthritis stratified by prior tumor necrosis factor inhibitor use: results from the randomized placebo-controlled future 2 study. J Rheumatol 2016;43:1713–7.
31. Carvalho PD, Duarte C, Vieira-Sousa E, Cunha-Miranda L, Avila-Ribeiro P, Santos H, et al. Predictors of response to TNF blockers in patients with polyarticular psoriatic arthritis. Acta Reumatol Port 2017;42:55–65.
32. Fagerli KM, Lie E, van der Heijde D, Heiberg MS, Lexberg AS, Rodevand E, et al. The role of methotrexate co-medication in TNF-inhibitor treatment in patients with psoriatic arthritis: results from 440 patients included in the nor-DMARD study. Ann Rheum Dis 2014;73:132–7.
33. Perrotta FM, Marchesoni A, Lubrano E. Minimal disease activity and remission in psoriatic arthritis patients treated with anti-TNF-α drugs. J Rheumatol 2016;43:350–5.
34. Smolen JS, Schols M, Braun J, Dougados M, FitzGerald O, Gladman DD, et al. Treating axial spondyloarthritis and peripheral spondyloarthritis, especially psoriatic arthritis, to target: 2017 update of recommendations by an international task force. Ann Rheum Dis 2018;77: 3–17.
35. Schoels M, Aletaha D, Funovits J, Kavanaugh A, Baker D, Smolen JS. Application of the DAREA/DAPSA score for assessment of disease activity in psoriatic arthritis. Ann Rheum Dis 2010;69:1441–7.
36. Fleischmann R, van der Heijde D, Koenig AS, Pedersen R, Szumski A, Marshall L, et al. How much does disease activity score in 28 joints ESR and CRP calculations underestimate disease activity compared with the simplified disease activity index? Ann Rheum Dis 2015;74: 1132–7.
37. Michelsen B, Kristianslund EK, Sexton J, Hammer HB, Fagerli KM, Lie E, et al. Do depression and anxiety reduce the likelihood of remission in rheumatoid arthritis and psoriatic arthritis? Data from the prospective multicentre nor-DMARD study. Ann Rheum Dis 2017;76:1906–10.

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1219–1226
DOI 10.1002/acr.24724

American College of Rheumatology
Empowering Rheumatology Professionals

Prognostic Value of Cardiac Axis Deviation in Systemic Sclerosis–Related Pulmonary Hypertension

Justin K. Lui,[1] Ruchika A. Sangani,[1] Clara A. Chen,[1] Andreea M. Bujor,[1] Marcin A. Trojanowski,[1] Deepa M. Gopal,[1] Michael P. LaValley,[1] Renda Soylemez Wiener,[2] and Elizabeth S. Klings[1]

Objective. Systemic sclerosis–related pulmonary hypertension (SSc-PH) is a common complication of SSc associated with accelerated mortality. The present study was undertaken to investigate whether cardiac axis deviation indicates abnormalities in cardiac function allowing for prognostication of disease severity and mortality.

Methods. This was a retrospective study in which electrocardiograms (ECGs) were reviewed for cardiac axis deviation and their association with echocardiography and cardiopulmonary hemodynamics on right-sided heart catheterization. The primary outcome observed was all-cause mortality from the time of PH diagnosis.

Results. ECG results were reviewed from 169 patients with SSc-PH. Right axis deviation (RAD) and left axis deviation (LAD) occurred in 28.4% and 30.8% of patients with SSc-PH, respectively. Compared to those without RAD, patients with RAD exhibited predominantly right-sided cardiac disease on echocardiography and increased PH severity by cardiopulmonary hemodynamics including a greater mean ± SD pulmonary artery pressure (42.0 ± 12.5 mm Hg versus 29.8 ± 7.0 mm Hg) and mean ± SD pulmonary vascular resistance (645.6 ± 443.2 dynes · seconds/cm^5 versus 286.3 ± 167.7 dynes · seconds/cm^5). LAD was associated with predominantly left-sided cardiac disease on echocardiography but was not associated with PH severity on cardiopulmonary hemodynamics. Both RAD (hazard ratio 10.36 [95% confidence interval 4.90–21.93], $P < 0.001$) and LAD (hazard ratio 2.94 [95% confidence interval 1.53–5.68], $P = 0.001$) were associated with an increased hazard for all-cause mortality.

Conclusion. RAD and LAD reflect structural cardiac abnormalities and are associated with poor prognosis in patients with SSc-PH. These findings support the importance of electrocardiography, an inexpensive, widely available noninvasive test, in risk stratification.

INTRODUCTION

Systemic sclerosis (SSc) is an autoimmune disease characterized by dysregulated connective tissue repair leading to fibrosis of the skin and internal organs and vascular endothelial dysfunction (1). One of the leading causes of morbidity and mortality in patients with SSc is pulmonary hypertension (SSc-PH), conventionally classified as group 1 pulmonary arterial hypertension (PAH), occurring in 8–12% of SSc patients (2,3). However, due to its multiorgan involvement, SSc-PH can also display features of left-sided cardiac disease (group 2 PH), interstitial lung disease (ILD) (group 3 PH), and potentially, thromboembolic disease (group 4 PH) (4,5). Left-sided cardiac disease, in particular, occurs in <10–30% of patients with SSc, although subclinical disease may be evident in >70% (6–8). When present, it can involve all structures within the heart in the form of myocardial (7,9) or pericardial disease (10,11) that can impact the cardiac conduction system (12–19) and lead to the development of atrial (20,21) and ventricular arrhythmias (22–24). These electrocardiographic abnormalities occur in 25–75% of patients with SSc and are typically associated with poor outcomes (12,15,17,19,23).

The views presented herein do not necessarily reflect the views of the US Department of Veterans Affairs or the US government.

Dr. Lui's work was supported by the NIH (National Heart, Lung, and Blood Institute Institutional Training grant T32-HL-007035). Dr. Bujor's work was supported by the Rheumatology Research Foundation (Tobé and Stephen E. Malawista, MD, Endowment in Academic Rheumatology). Dr. Gopal's work was supported by the American Heart Association (grant FTF17FTF33670369). Dr. LaValley's work was supported by the NIH (Core Centers for Clinical Research grant P30-AR-072571 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases). Dr. Soylemez Wiener's work was supported in part by resources from the VA Boston Healthcare system. Dr. Klings' work was supported by the NIH (National Heart, Lung, and Blood Institute grant 1UG3-HL-143192-01A1).

[1]Justin K. Lui, MD, Ruchika A. Sangani, MD, Clara A. Chen, MHS, Andreea M. Bujor, MD, PhD, Marcin A. Trojanowski, MD, Deepa M. Gopal, MD, MS, Michael P. LaValley, PhD, Elizabeth S. Klings, MD: Boston University School of Medicine, Boston, Massachusetts; [2]Renda Soylemez Wiener, MD, MPH: Boston University School of Medicine and VA Boston Healthcare System, Boston, Massachusetts.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.24724&file=acr24724-sup-0001-Disclosureform.pdf.

Address correspondence to Justin K. Lui, MD, The Pulmonary Center, Boston University School of Medicine, 72 East Concord Street, R-304, Boston, MA 02118. Email: justin.lui@bmc.org.

Submitted for publication February 9, 2021; accepted in revised form June 1, 2021.

SIGNIFICANCE & INNOVATIONS

- Right axis deviation (RAD) was associated with increased pulmonary hypertension (PH) severity by mean pulmonary artery pressure and pulmonary vascular resistance on cardiopulmonary hemodynamics from right-sided heart catheterization.
- Both left axis deviation and RAD are associated with increased hazard for all-cause mortality in patients with systemic sclerosis (SSc)–related PH.
- Inexpensive and widely available, electrocardiography is a feasible noninvasive tool that can help guide clinical care in patients with SSc-PH by stratifying those at increased risk for death.

Central to the cardiac conduction system is the electrical cardiac axis, defined as the general direction of the conduction depolarization vector determined by the ventricular axis via the QRS complex normally lying between –30° and +90° within the frontal plane. Left axis deviation (LAD) is most commonly a normal age variant (25) but can also be associated with conduction defects such as from a left anterior fascicular block (26,27). By itself, LAD has not been associated with increased morbidity and mortality in patients with underlying cardiac disease (25). Right axis deviation (RAD) can similarly exist as a normal variant but can also indicate underlying pathology such as PH (28,29). In one study, RAD was determined to have a positive predictive value of 72.7% for severe PH based on an estimated pulmonary artery systolic pressure of ≥60 mm Hg on echocardiography (28).

To date, among patients with PH, it is unclear whether deviations in the electrical cardiac axis may also indicate poorer prognosis compared to those without. Here, we aimed to examine a cohort of patients with SSc-PH to determine if the presence of cardiac axis deviation is associated with structural disease on echocardiography and hemodynamic abnormalities on right-sided heart catheterization (RHC). Specifically, we hypothesized that both RAD and LAD may be a surrogate for more severe PH by cardiopulmonary hemodynamics and hence would be independently associated with increased all-cause mortality compared to those without these findings.

PATIENTS AND METHODS

Patient population. This was a retrospective single-center study utilizing a clinical registry from the Scleroderma Center of Research Translation (CORT) database at Boston Medical Center/Boston University School of Medicine. The research database was created in 2012, enrolling subjects with a diagnosis of SSc receiving care at Boston Medical Center. At the time of enrollment, each subject consented to inclusion in this longitudinal clinical registry. The study was approved by the Institutional Review Board at Boston University School of Medicine. Patients were included in the present study if they had at least 1 electrocardiogram (ECG) from the time of SSc diagnosis until an end date of July 31, 2020. Patients with a history of heart and/or lung transplantation were excluded from this study. Additionally, patients diagnosed with SSc-PH who did not have RHC data were also excluded. SSc-PH was confirmed by a resting mean pulmonary artery pressure (mPAP) of >20 mm Hg on RHC at any time within the patient's clinical course (3). Finally, to ensure that there was at least a 1-year follow-up of patients with SSc-PH, we excluded those who were not diagnosed with PH between July 1, 1999, and July 1, 2019, which comprised a 2-decade window for our study cohort.

Clinical data and outcome. At the time of or following diagnosis of SSc-PH on RHC, results of ECGs were reviewed to identify the earliest incidence of RAD (defined by QRS axis between +90° and +180°) and/or LAD (defined by QRS axis between –30° and –90°) using the computer-generated QRS axis. Cardiopulmonary hemodynamics including right atrial pressure, mPAP, pulmonary arterial wedge pressure (PAWP), pulmonary vascular resistance (PVR), and cardiac output were also obtained from RHC at the time of PH diagnosis (or the next RHC closest to the time of PH diagnosis, if a diagnostic RHC was unavailable). Other data that were also collected included: 1) demographic information (i.e., age at PH diagnosis, sex, SSc type, and smoking history), 2) comorbid conditions, 3) presence of autoantibodies (i.e., antinuclear antibody [ANA], anti–topoisomerase I antibody [anti–Scl-70]), 4) results from pulmonary function testing, 5) modified Rodnan skin thickness scores (MRSS), 6) New York Heart Association (NYHA) functional classification, and 7) initial PAH-specific therapy. For all of these data, we utilized the time point that was closest to the time of RHC or PH diagnosis.

Results from echocardiography were reviewed for each patient over the entire clinical course from the time of SSc diagnosis to the July 31, 2020 end date to determine whether each patient had at least 1 incidence of the following: 1) valvular abnormalities (i.e., aortic regurgitation, mitral regurgitation, tricuspid regurgitation, pulmonic regurgitation, aortic stenosis, mitral stenosis); 2) left-sided cardiac abnormalities (i.e., left ventricular [LV] systolic dysfunction, LV dilation, LV diastolic dysfunction, LV hypertrophy, left atrial dilation); 3) right-sided cardiac abnormalities (i.e., right atrial [RA] dilation, right ventricular [RV] dilation, RV systolic dysfunction); 4) pericardial effusion; and/or 5) regional wall motion abnormalities.

Additionally, the LV ejection fraction was also obtained in all patients with SSc-PH with the lowest value recorded, when available. The primary clinical outcome of this study was all-cause mortality from the time of PH diagnosis.

Statistical analysis. Continuous variables were described by the mean ± SD, and categorical variables were described by

their frequencies and percentages for each group. Student's *t*-test was used for analysis of continuous variables, and chi-square test (or Fisher's exact test) was used for analysis of categorical variables. Both univariable and multivariable Cox proportional hazards regression models were applied to determine associations with all-cause mortality. Both RAD and LAD were introduced as time-varying covariates and were adjusted for age at PH diagnosis, sex, SSc subtype classification, NYHA functional classification, and forced vital capacity (FVC) in the multivariable model. In addition, receipt of PAH-specific therapy was also incorporated into the model as a time-varying covariate. These results were presented as a hazard ratio (HR) with 95% confidence intervals (95% CIs) and *P* values. In the final multivariable Cox proportional hazards regression model, interaction between RAD and LAD was assessed, and all covariates were evaluated for the proportional hazards assumption by Schoenfeld residuals. In the event that either RAD or LAD did not meet the proportional hazards assumption, the time was divided into subintervals: ≤1 year after PH diagnosis and >1 year after PH diagnosis, with an HR determined for each subinterval. Statistical significance was determined by a 2-sided *P* value less than 0.05. All statistical analyses were conducted in RStudio using the survival package.

RESULTS

Study population. Within the CORT database, 587 SSc patients were identified after removal of duplicate patients; 387 had at least 1 ECG from the time of SSc diagnosis to August 1, 2020. Ten of the 387 (2.6%) had a prior heart and/or lung transplant and were excluded from the study. Of the remaining 377 patients, 173 were diagnosed as having PH; 2 did not have RHC identified, and 2 were outside the study timeframe (July 1, 1999, to July 1, 2019) and therefore were excluded from the study (Figure 1). The mean ± SD time for pulmonary function test assessment was 14.6 ± 23.7 months from the time of PH diagnosis, and the timing for MRSS measurements was 20.6 ± 34.6 months from the time of PH diagnosis. The demographic and patient characteristics are summarized in Table 1. In those with SSc-PH, LAD (30.8%) was more common than RAD (28.4%). There were 9 (5.3%) patients who exhibited both LAD and RAD at different time points. The age and sex distributions were similar across groups exhibiting RAD, LAD, and no RAD/LAD. Patients with RAD had a greater frequency of diffuse cutaneous SSc subtype with a lower autoantibody positivity rate and a reduced degree of skin disease compared to those with LAD and those with no RAD/LAD.

Cardiopulmonary hemodynamics and echocardiography. Among 169 patients with SSc-PH, 161 had cardiopulmonary hemodynamics from RHC at the time of PH diagnosis (95.3%), whereas the remaining 8 patients had hemodynamics from the next available RHC after diagnosis. Cardiopulmonary hemodynamics on RHC are outlined in Table 2. Those with RAD on ECG predominantly exhibited precapillary PH with a greater mPAP and PVR compared to those with LAD and no RAD/LAD. On echocardiography, RAD was associated with greater frequency of right-sided cardiac abnormalities, including RA dilation, RV dilation, and systolic dysfunction. Those with LAD on ECG demonstrated a slightly greater frequency of isolated postcapillary PH compared to those with RAD and no RAD/LAD, as well as a marginally greater PAWP consistent with LV diastolic and systolic dysfunction, including LV dilation observed on echocardiography. Compared to those without, patients with RAD and LAD had an increased mPAP and PVR by RHC and increased frequencies of tricuspid or pulmonic regurgitation, RA dilation, RV dilation, or systolic dysfunction on echocardiography.

NYHA functional classification. Within the entire cohort of patients with SSc-PH, most patients were symptomatically

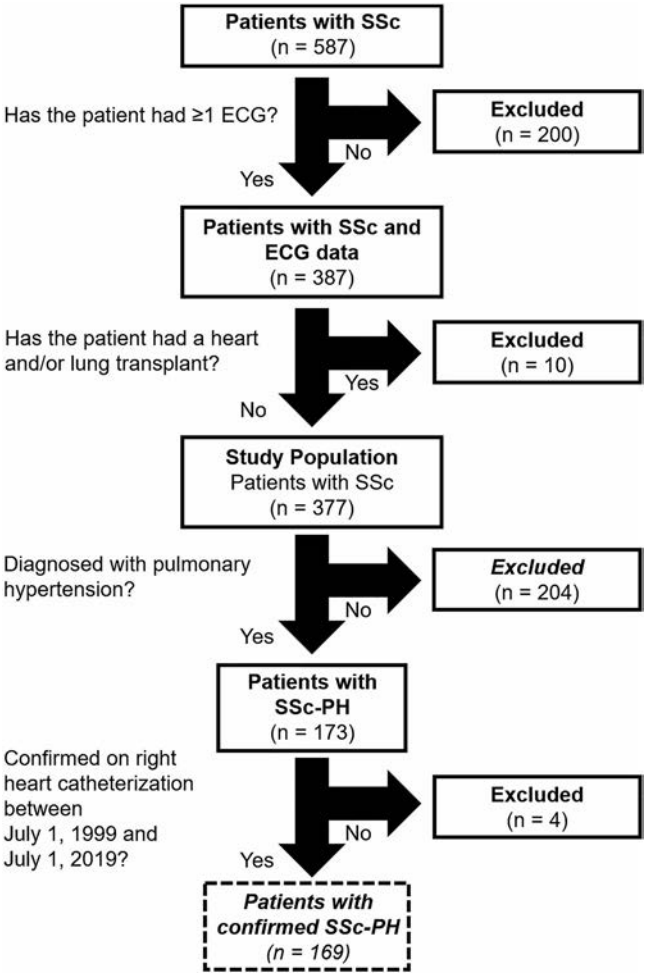


Figure 1. Schematic of study design. In total, 587 patients were identified from the Scleroderma Center of Research Translation database as diagnosed with systemic sclerosis (SSc), of which 169 were diagnosed as having pulmonary hypertension (PH). ECG = electrocardiogram.

Table 1. Patient demographic characteristics and clinical features*

Characteristic	No RAD/LAD (n = 78)	RAD (n = 48)	LAD (n = 52)
Age at PH diagnosis, mean ± SD years	58.4 ± 10.9	56.2 ± 11.5	61.7 ± 10.6
Sex			
Male	11 (14.1)	10 (20.8)	14 (26.9)
Female	67 (85.9)	38 (79.2)	38 (73.1)
SSc type			
Limited cutaneous SSc	52 (66.7)	38 (79.2)	33 (63.5)
Diffuse cutaneous SSc	23 (29.5)	10 (20.8)	18 (34.6)
Undifferentiated SSc	3 (3.8)	0	1 (1.9)
Smoking status			
Never smoker	37 (47.4)	27 (56.3)	28 (53.8)
Former/current smoker	41 (52.6)	21 (43.8)	24 (46.2)
Comorbid conditions			
Interstitial lung disease	45 (57.7)	21 (43.8)	33 (63.5)
Coronary artery disease	8 (10.3)	4 (8.3)	3 (5.8)
Atrial fibrillation/atrial flutter	15 (19.2)	10 (20.8)	16 (30.8)
Venous thromboembolism	10 (12.8)	7 (14.6)	9 (17.3)
Diabetes mellitus	2 (2.6)	2 (4.2)	5 (9.6)
Hypertension	33 (42.3)	24 (50.0)	33 (63.5)
Hyperlipidemia	30 (38.5)	15 (31.3)	18 (34.6)
Pulmonary function testing, no.	76	43	45
FVC, mean ± SD % predicted	76.9 ± 16.9	71.8 ± 18.5	68.6 ± 16.8
FEV_1, mean ± SD % predicted	75.8 ± 16.5	70.9 ± 19.9	71.0 ± 16.5
FEV_1/FVC ratio, mean ± SD	77.4 ± 10.2	75.9 ± 10.9	78.9 ± 8.4
DLco, mean ± SD % predicted	45.1 ± 16.9	40.8 ± 15.7	43.0 ± 16.4
ANA, no.	62	33	32
ANA	59 (95.2)	29 (87.9)	31 (96.9)
Anti–topoisomerase I antibody, no.	48	20	20
Anti–Scl-70 antibody	12 (25.0)	3 (15.0)	8 (32.0)
Degree of skin disease, no.	56	28	36
MRSS, mean ± SD	11.5 ± 10.0	8.3 ± 9.8	10.3 ± 9.8

* Values are the number (%) unless indicated otherwise. ANA = antinuclear antibodies; DLco = diffusing capacity for carbon monoxide; FEV_1 = forced expiratory volume in 1 second; FVC = forced vital capacity; LAD = left axis deviation; MRSS = modified Rodnan skin thickness score; PH = pulmonary hypertension; RAD = right axis deviation; SSc = systemic sclerosis.

NYHA functional class III (58.6%) at the time of PH diagnosis. More patients were NYHA functional class III–IV (61.7%) than NYHA functional class I–II (38.3%), indicating significant functional impairment in our SSc-PH cohort. Furthermore, those with RAD had a greater number of NYHA functional class III–IV symptoms compared to those with LAD and no RAD/LAD (Table 2).

Initiation of PAH-specific therapy. There were a total of 107 (63.3%) patients with SSc-PH who were initiated on PAH-specific therapy. PDE5 inhibitors were the most commonly prescribed (45.8%), followed by endothelin receptor antagonists (38.3%) and prostacyclin analogs (PCA) (25.2%), with most patients initially receiving monotherapy (56.8%). RAD was associated with a greater frequency of PAH-specific therapy, specifically, with PCA compared to LAD and no RAD/LAD. Table 3 summarizes the clinical characteristics of patients in whom PAH-specific therapy was initiated. Among those who did not receive PAH-specific therapy, interestingly, there was a greater frequency of diffuse cutaneous SSc compared to those who did. Not surprisingly, patients who were started on PAH-specific therapy predominantly exhibited precapillary PH with greater PH severity by cardiopulmonary hemodynamics, specifically, mPAP and PVR.

Associations with all-cause mortality. After excluding the 25 (14.8%) patients who were lost to follow-up and the 22 (13.0%) patients without 5 years of follow-up, the 5-year all-cause mortality in patients with SSc-PH was 41.8%. Using a univariable Cox proportional hazards regression, male sex, FVC % predicted, NYHA functional class III–IV, and treatment with PAH-specific medications were significantly associated with all-cause mortality, with HRs of 2.34 (95% CI 1.36–4.04, $P = 0.002$), 0.98 (95% CI 0.96–0.99, $P = 0.001$), 1.87 (95% CI 1.08–3.24, $P = 0.026$), and 4.08 (95% CI 2.17–7.66, $P < 0.001$), respectively (Table 4). Age at PH diagnosis or diffuse cutaneous SSc subtype classification were not associated with mortality. The presence of RAD within 1 year after PH diagnosis was associated with an HR of 10.06 (95% CI 5.35–18.95, $P < 0.001$) for all-cause mortality. The presence of RAD >1 year after PH diagnosis was associated with an HR of 1.37 for all-cause mortality but was not statistically significant (95% CI

Table 2. Characteristics of pulmonary hypertension (PH) related to systemic sclerosis*

Characteristic	No RAD/LAD (n = 78)	RAD (n = 48)	LAD (n = 52)
Hemodynamic definition of PH			
Precapillary PH	36 (46.2)	36 (75.0)	25 (48.1)
Isolated postcapillary PH	9 (11.5)	4 (8.3)	7 (13.5)
Combined pre- and postcapillary PH	5 (6.4)	2 (4.2)	3 (5.8)
Cardiopulmonary hemodynamics, no.	78	48	52
RAP, mean ± SD mm Hg	6.4 ± 4.3	8.4 ± 6.2	8.5 ± 6.0
mPAP, mean ± SD mm Hg	29.8 ± 7.0	42.0 ± 12.5	34.4 ± 12.0
PAWP, mean ± SD mm Hg	11.8 ± 5.0	10.5 ± 6.5	12.2 ± 5.3
CO, mean ± SD liters/minute	5.4 ± 1.4	4.7 ± 1.7	5.1 ± 1.5
PVR, mean ± SD dynes · seconds/cm^5	286.3 ± 167.7	645.6 ± 443.2	388.1 ± 299.3
Echocardiography, no.	78	47	49
Aortic regurgitation	15 (19.2)	9 (19.1)	9 (18.4)
Mitral regurgitation	46 (59.0)	18 (38.3)	27 (55.1)
Tricuspid regurgitation	55 (70.5)	39 (83.0)	40 (81.6)
Pulmonic regurgitation	20 (25.6)	20 (42.6)	24 (49.0)
Aortic stenosis	6 (7.7)	2 (4.3)	8 (16.3)
Mitral stenosis	6 (7.7)	1 (2.1)	3 (6.1)
LV systolic dysfunction (LVEF ≤40%)	5 (6.4)	3 (6.4)	10 (20.4)
LVEF, mean ± SD %	58.3 ± 7.7	57.7 ± 7.6	53.0 ± 12.8
LV dilation	2 (2.6)	2 (4.3)	6 (12.2)
LV diastolic dysfunction	42 (53.8)	22 (46.8)	28 (57.1)
LV hypertrophy	34 (43.6)	20 (42.6)	26 (53.1)
LA dilation	48 (61.5)	19 (40.4)	29 (59.2)
Regional wall motion abnormalities	8 (10.3)	11 (23.4)	17 (34.7)
RA dilation	24 (30.8)	38 (80.9)	28 (57.1)
RV dilation	26 (33.3)	38 (80.9)	29 (59.2)
RV systolic dysfunction	17 (21.8)	31 (66.0)	25 (51.0)
Pericardial effusion	33 (42.3)	32 (68.1)	29 (59.2)
NYHA functional class, no.	72	47	51
NYHA class I	4 (5.6)	2 (4.3)	3 (5.9)
NYHA class II	26 (36.1)	12 (25.5)	17 (33.3)
NYHA class III	41 (56.9)	30 (63.8)	29 (56.9)
NYHA class IV	1 (1.4)	3 (6.4)	2 (3.9)
Initial PAH-specific therapy, no.	78	50	54
Endothelin receptor antagonist	16 (20.5)	16 (33.3)	11 (21.2)
Phosphodiesterase 5 inhibitor	22 (28.2)	11 (22.9)	17 (32.7)
Prostacyclin analog	3 (3.8)	20 (41.7)	8 (15.4)
Monotherapy	31 (39.7)	41 (85.4)	31 (59.6)
Dual therapy	5 (6.4)	4 (8.3)	3 (5.8)

* Values are the number (%) unless indicated otherwise. CO = cardiac output; LA = left atrial; LAD = left axis deviation; LV = left ventricular; LVEF = left ventricular ejection fraction; mPAP = mean pulmonary artery pressure; NYHA = New York Heart Association; PAH = pulmonary arterial hypertension; PAWP = pulmonary artery wedge pressure; PH = pulmonary hypertension; PVR = pulmonary vascular resistance; RA = right atrial; RAD = right axis deviation; RAP = right atrial pressure; RV = right ventricular.

0.57–3.27, $P = 0.478$). Similarly, LAD was associated with increased hazards for all-cause mortality, with an HR of 3.35 (95% CI 2.03–5.53, $P < 0.001$). From the multivariable Cox proportional hazards regression, after adjusting for age at PH diagnosis, male sex, diffuse cutaneous SSc subtype classification, FVC % predicted, NYHA functional classification, and receipt of PAH-specific medications, RAD within 1 year after PH diagnosis was associated with an HR of 10.36 (95% CI 4.90–21.93, $P < 0.001$), and RAD >1 year after PH diagnosis was associated with an HR of 3.69 (95% CI 1.24–10.98, $P = 0.019$) for all-cause mortality. Likewise, in the adjusted model, LAD was also associated with an HR of 2.94 (95% CI 1.53–5.68, $P = 0.001$) for all-cause mortality. Of note, 1 patient (0.6%) who died from a pulmonary artery rupture during the RHC was excluded from the Cox proportional hazards regression.

DISCUSSION

Despite its classification as group 1 PAH, SSc-PH is a heterogeneous disease exhibiting a spectrum of clinical phenotypes including contributions from left-sided cardiac disease (group 2), ILD (group 3), and thromboembolic disease (group 4) among individual patients (4,5). Left-sided cardiac disease, in particular, may be underdiagnosed among SSc patients, as only 10–30% develop clinical disease in which subclinical features of cardiac dysfunction may develop in many others (6–8). Previously thought

Table 3. Clinical characteristics of patients in whom initiated PAH-specific therapy was initiated*

Demographic characteristic and feature	Patients not initiated on PAH-specific therapy (n = 62)	Patients initiated on PAH-specific therapy (n = 107)	*P*
Age at PH diagnosis, mean ± SD years	59.7 ± 11.1	58.5 ± 11.3	0.480
Male sex	10 (16.1)	25 (23.4)	0.263
Diffuse cutaneous SSc	26 (41.9)	24 (22.4)	0.007†
Interstitial lung disease	38 (61.3)	56 (52.3)	0.259
Precapillary PH	12 (19.4)	79 (73.8)	<0.001†
Cardiac axis deviation, no.	62	107	
RAD	4 (6.5)	44 (41.1)	<0.001†
LAD	17 (27.4)	35 (32.7)	0.473
Pulmonary function testing, no.	59	98	
FVC, mean ± SD % predicted	75.3 ± 16.7	72.0 ± 17.9	0.252
FEV_1, mean ± SD % predicted	75.3 ± 16.5	71.7 ± 17.9	0.202
FEV_1/FVC ratio, mean ± SD	78.2 ± 8.8	77.0 ± 10.7	0.445
DLco, mean ± SD % predicted	50.7 ± 17.4	38.8 ± 14.2	<0.001†
Cardiopulmonary hemodynamics, no.	62	107	
RAP, mean ± SD mm Hg	7.4 ± 5.5	7.4 ± 5.3	0.983
mPAP, mean ± SD mm Hg	27.3 ± 7.1	38.2 ± 11.3	<0.001†
PAWP, mean ± SD mm Hg	14.0 ± 5.9	10.2 ± 5.0	<0.001†
CO, mean ± SD liters/minute	5.5 ± 1.3	4.9 ± 1.6	0.006†
PVR, mean ± SD dynes · seconds/cm^5	203.4 ± 91.9	529.9 ± 368.4	<0.001†
NYHA functional class, no.	61	101	
NYHA class III–IV	37 (60.7)	63 (62.4)	0.827

* Values are the number (%) unless indicated otherwise. CO = cardiac output; DLco = diffusing capacity for carbon monoxide; FEV_1 = forced expiratory volume in 1 second; FVC = forced vital capacity; LAD = left axis deviation; mPAP = mean pulmonary artery pressure; NYHA = New York Heart Association; PAH = pulmonary arterial hypertension; PAWP = pulmonary artery wedge pressure; PH = pulmonary hypertension; PVR = pulmonary vascular resistance; RAD = right axis deviation; RAP = right atrial pressure; SSc = systemic sclerosis.
† Statistically significant.

to occur in 19–32% of patients with SSc, cardiac conduction abnormalities remain poorly understood, particularly in those with SSc-PH, a major cause of mortality in this patient population. Furthermore, the association between cardiac conduction abnormalities and clinical outcomes has remained unclear (17,30). This present study is the largest cohort to date investigating cardiac conduction system abnormalities, specifically cardiac axis deviation in patients with SSc-PH. We found that cardiac axis deviation is common in SSc-PH, with RAD observed in 28.4%, and LAD observed in 30.8% of patients.

Not surprisingly, RAD was associated with right-sided structural abnormalities on echocardiography, specifically, in the RA and RV, likely as a consequence of PH. This was reflected by an increased mPAP and PVR on RHC, suggesting that RAD may indicate a greater degree of PH severity. In contrast, LAD was associated with left-sided structural abnormalities, specifically LV diastolic and systolic dysfunction as well as LV dilation, a known complication of SSc. While RAD has long been recognized as an indicator of PH (28,29), these data suggest that RAD may also be a marker of disease severity and mortality in patients with

Table 4. Factors impacting all-cause mortality in systemic sclerosis–related pulmonary hypertension (SSc-PH)*

Covariates	Univariable Cox proportional hazards regression			Multivariable Cox proportional hazards regression		
	HR	95% CI	*P*	HR	95% CI	*P*
Age at PH diagnosis	1.01	0.98–1.03	0.491	1.02	0.99–1.05	0.166
Male sex	2.34	1.36–4.04	0.002†	1.95	1.00–3.82	0.051
Diffuse cutaneous SSc	1.21	0.70–2.10	0.495	1.35	0.68–2.68	0.392
FVC, % predicted	0.98	0.96–0.99	0.001†	0.98	0.96–1.00	0.045†
NYHA functional class III–IV	1.87	1.08–3.24	0.026†	1.80	0.97–3.34	0.062
Receipt of PAH-specific medications	4.08	2.17–7.66	<0.001†	1.90	0.94–3.83	0.074
RAD ≤1 year after PH diagnosis	10.06	5.35–18.95	<0.001†	10.36	4.90–21.93	<0.001†
RAD >1 year after PH diagnosis	1.37	0.57–3.27	0.478	3.69	1.24–10.98	0.019†
LAD	3.35	2.03–5.53	<0.001†	2.94	1.53–5.68	0.001†

* 95% CI = 95% confidence interval; FVC = forced vital capacity; HR = hazard ratio; LAD = left axis deviation; NYHA = New York Heart Association; PAH = pulmonary arterial hypertension; PH = pulmonary hypertension; RAD = right axis deviation; SSc = systemic sclerosis.
† Statistically significant.

SSc-PH. Interestingly, the effect of RAD on mortality is more pronounced when occurring within 1 year following the time of PH diagnosis, with a near 3-fold mortality increase compared to LAD, suggesting that it may be an important prognosticator in patients with SSc-PH. The association of LAD with increased hazards for all-cause mortality emphasizes the importance of left-sided cardiac disease in contributing to poorer outcomes in SSc-PH. Taken together, these findings underline the importance of ECG in risk stratifying patients with SSc-PH. Given that they are widely available, inexpensive, and noninvasive, ECGs are a very feasible tool to help guide clinical care and decision-making in the care of these patients.

There were several limitations to this study. Most importantly, this study is retrospective, in which collection of data is dictated by clinical practice, leading to inconsistencies in data availability and timing. As a result, while we were able to delineate the onset of cardiac axis deviation in our patient cohort, these data were based on the availability of ECGs. Most ECGs were obtained at the time of RHC. The patients who were excluded were typically those with limited or localized SSc without any cardiopulmonary manifestations to warrant further investigation with RHC. We limited our study cohort to investigate only patients with SSc-PH so as to not create selection bias in all patients with SSc. While most patients with SSc and SSc-PH had RHCs performed at our institution, there were a few who had RHCs performed elsewhere. In those instances, the ECGs were not consistently available. One future direction of this work will be to determine whether cardiac axis deviation can predict PH among patients with SSc to better risk stratify patients for RHC. This will require a prospective study in which ECGs are routinely and systematically collected in all SSc patients as opposed to limiting data to those with high clinical suspicion of PH.

There were frequent missing data to contend with in the current study. Of the 169 patients diagnosed with SSc-PH, 22 (13.0%) were lost to follow-up, which may have impacted the observed mortality. This limited some of the variable selection for our multivariable Cox proportional hazards model. For instance, missing data for the MRSS did not allow us to utilize this variable. Similarly, we did not have data on disease duration since onset of non–Raynaud's phenomenon symptoms. We opted to adjust for the diffuse cutaneous SSc subtype classification given that these patients are more likely to have cardiopulmonary involvement and poorer survival compared to those with limited cutaneous SSc (31,32). Furthermore, given that ILD and PH constitute the 2 leading causes of mortality among patients, we opted to adjust for ILD severity using FVC based on prior studies (33–35). While the diffusing capacity for carbon monoxide (DLco) is also capable of quantifying the degree of ILD severity, it is reduced with increasing PH severity. The inclusion of DLco with FVC may lead to overfitting of our multivariable Cox proportional hazards model.

Finally, this data set spans >2 decades in which changes in clinical practice occurred reflective of increased understanding of the disease and its management. One of these critical paradigm shifts was the updated hemodynamic classification of PH, which changed the hemodynamic threshold from a mPAP of >25 mm Hg to >20 mm Hg in 2019 (3). This may explain the lower rates of initiation of PAH-specific therapy. Overall, the rationale for not starting PAH therapy was likely multifactorial. Analysis of these data demonstrates that those in which PAH therapy was not commenced had hemodynamics slanted toward postcapillary PH with a normal PVR and borderline PAWP. It will be important to confirm our findings within a large prospective study.

In conclusion, abnormalities in cardiac conduction, particularly cardiac axis deviation, are common in SSc-PH. RAD was associated with increase PH severity by cardiopulmonary hemodynamics. Furthermore, both RAD and LAD were associated with increased hazard of mortality, suggesting a role for ECGs, an inexpensive, widely available noninvasive test, in the prognostication of patients with SSc-PH.

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. Lui 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. Lui, Sangani, Gopal, Klings.
Acquisition of data. Lui, Sangani, Chen, Bujor, Trojanowski.
Analysis and interpretation of data. Lui, LaValley, Soylemez Wiener, Klings.

REFERENCES

1. Denton CP, Khanna D. Systemic sclerosis. Lancet 2017;390:1685–99.
2. Ioannidis JP, Vlachoyiannopoulos PG, Haidich AB, Medsger TA Jr, Lucas M, Michet CJ, et al. Mortality in systemic sclerosis: an international meta-analysis of individual patient data. Am J Med 2005;118:2–10.
3. Simonneau G, Montani D, Celermajer DS, Denton CP, Gatzoulis MA, Krowka M, et al. Haemodynamic definitions and updated clinical classification of pulmonary hypertension. Eur Respir J 2019;53:1801913.
4. Launay D, Montani D, Hassoun PM, Cottin V, Le Pavec J, Clerson P, et al. Clinical phenotypes and survival of pre-capillary pulmonary hypertension in systemic sclerosis. PLoS One 2018;13:e0197112.
5. Launay D, Sobanski V, Hachulla E, Humbert M. Pulmonary hypertension in systemic sclerosis: different phenotypes. Eur Respir Rev 2017;26:170056.
6. Hung G, Mercurio V, Hsu S, Mathai SC, Shah AA, Mukherjee M. Progress in understanding, diagnosing, and managing cardiac complications of systemic sclerosis. Curr Rheumatol Rep 2019;21:68.
7. Kahan A, Allanore Y. Primary myocardial involvement in systemic sclerosis. Rheumatology (Oxford) 2006;45 Suppl 4:iv14.
8. Tyndall AJ, Bannert B, Vonk M, Airò P, Cozzi F, Carreira PE, et al. Causes and risk factors for death in systemic sclerosis: a study from the EULAR Scleroderma Trials and Research (EUSTAR) database. Ann Rheum Dis 2010;69:1809.
9. Bissell LA, Md Yusof MY, Buch MH. Primary myocardial disease in scleroderma: a comprehensive review of the literature to inform the UK Systemic Sclerosis Study Group cardiac working group. Rheumatology (Oxford) 2017;56:882–95.

10. Byers RJ, Marshall DA, Freemont AJ. Pericardial involvement in systemic sclerosis. Ann Rheum Dis 1997;56:393–4.
11. Hosoya H, Derk CT. Clinically symptomatic pericardial effusions in hospitalized systemic sclerosis patients: demographics and management. Biomed Res Int 2018;2018:6812082.
12. Draeger HT, Assassi S, Sharif R, Gonzalez EB, Harper BE, Arnett FC, et al. Right bundle branch block: a predictor of mortality in early systemic sclerosis. PLoS One 2013;8:e78808.
13. Eisen A, Arnson Y, Dovrish Z, Hadary R, Amital H. Arrhythmias and conduction defects in rheumatological diseases: a comprehensive review. Semin Arthritis Rheum 2009;39:145–56.
14. Follansbee WP, Curtiss EI, Rahko PS, Medsger TA Jr, Lavine SJ, Owens GR, et al. The electrocardiogram in systemic sclerosis (scleroderma). Study of 102 consecutive cases with functional correlations and review of the literature. Am J Med 1985;79: 183–92.
15. Muresan L, Petcu A, Pamfil C, Muresan C, Rinzis M, Mada RO, et al. Cardiovascular profiles of scleroderma patients with arrhythmias and conduction disorders. Acta Reumatol Port 2016;41:26–39.
16. Ridolfi RL, Bulkley BH, Hutchins GM. The cardiac conduction system in progressive systemic sclerosis. Clinical and pathologic features of 35 patients. Am J Med 1976;61:361–6.
17. Roberts NK, Cabeen WR Jr, Moss J, Clements PJ, Furst DE. The prevalence of conduction defects and cardiac arrhythmias in progressive systemic sclerosis. Ann Intern Med 1981;94:38–40.
18. Seferović PM, Ristić AD, Maksimović R, Simeunović DS, Ristić GG, Radovanović G, et al. Cardiac arrhythmias and conduction disturbances in autoimmune rheumatic diseases. Rheumatology (Oxford) 2006;45 Suppl 4:iv39–42.
19. Vacca A, Meune C, Gordon J, Chung L, Proudman S, Assassi S, et al. Cardiac arrhythmias and conduction defects in systemic sclerosis. Rheumatology (Oxford) 2014;53:1172–7.
20. Mercurio V, Peloquin G, Bourji KI, Diab N, Sato T, Enobun B, et al. Pulmonary arterial hypertension and atrial arrhythmias: incidence, risk factors, and clinical impact. Pulm Circ 2018;8:2045894018769874.
21. Mizuno R, Fujimoto S, Nakano H, Nakajima T, Kimura A, Nakagawa Y, et al. Atrial conduction abnormalities in patients with systemic progressive sclerosis. Eur Heart J 1997;18:1995–2001.
22. Bienias P, Ciurzyński M, Kisiel B, Chrzanowska A, Ciesielska K, Siwicka M, et al. Comparison of non-invasive assessment of arrhythmias, conduction disturbances and cardiac autonomic tone in systemic sclerosis and systemic lupus erythematosus. Rheumatol Int 2019;39:301–10.
23. Kostis JB, Seibold JR, Turkevich D, Masi AT, Grau RG, Medsger TA Jr, et al. Prognostic importance of cardiac arrhythmias in systemic sclerosis. Am J Med 1988;84:1007–15.
24. Sebestyén V, Szűcs G, Páll D, Ujvárosy D, Ötvös T, Csige I, et al. Electrocardiographic markers for the prediction of ventricular arrhythmias in patients with systemic sclerosis. Rheumatology (Oxford) 2020;59: 478–86.
25. Ostrander LD Jr. Left axis deviation: prevalence, associated conditions, and prognosis: an epidemiologic study. Ann Intern Med 1971; 75:23–8.
26. Grayzel J, Neyshaboori M. Left-axis deviation: etiologic factors in one-hundred patients. Am Heart J 1975;89:419–27.
27. Das G. Left axis deviation: a spectrum of intraventricular conduction block. Circulation 1976;53:917–9.
28. Al-Naamani K, Hijal T, Nguyen V, Andrew S, Nguyen T, Huynh T. Predictive values of the electrocardiogram in diagnosing pulmonary hypertension. Int J Cardiol 2008;127:214–8.
29. Kovacs G, Avian A, Foris V, Tscherner M, Kqiku X, Douschan P, et al. Use of ECG and other simple non-invasive tools to assess pulmonary hypertension. PLoS One 2016;11:e0168706.
30. Ferri C, Bernini L, Bongiorni MG, Levorato D, Viegi G, Bravi P, et al. Noninvasive evaluation of cardiac dysrhythmias, and their relationship with multisystemic symptoms, in progressive systemic sclerosis patients. Arthritis Rheum 1985;28:1259–66.
31. Ferri C, Valentini G, Cozzi F, Sebastiani M, Michelassi C, La Montagna G, et al. Systemic sclerosis: demographic, clinical, and serologic features and survival in 1,012 Italian patients. Medicine (Baltimore) 2002;81:139.
32. Scussel-Lonzetti L, Joyal F, Raynauld JP, Roussin A, Rich E, Goulet JR, et al. Predicting mortality in systemic sclerosis: analysis of a cohort of 309 French Canadian patients with emphasis on features at diagnosis as predictive factors for survival. Medicine (Baltimore) 2002;81:154.
33. Hoa S, Bernatsky S, Steele RJ, Baron M, Hudson M, Canadian Scleroderma Research Group. Association between immunosuppressive therapy and course of mild interstitial lung disease in systemic sclerosis. Rheumatology (Oxford) 2020;59:1108–17.
34. Tashkin DP, Elashoff R, Clements PJ, Goldin J, Roth MD, Furst DE, et al. Cyclophosphamide versus placebo in scleroderma lung disease. N Engl J Med 2006;354:2655–66.
35. Tashkin DP, Roth MD, Clements PJ, Furst DE, Khanna D, Kleerup EC, et al. Mycophenolate mofetil versus oral cyclophosphamide in scleroderma-related interstitial lung disease (SLS II): a randomised controlled, double-blind, parallel group trial. Lancet Respir Med 2016;4:708–19.