Arthritis Care & Research

Aims and Scope

Arthritis Care & Research is an official journal of the American College of Rheumatology and the Association of Rheumatology Professionals, a division of the College. *Arthritis Care & Research* is a peer-reviewed journal that publishes both original research and review articles that promote excellence in the clinical practice of rheumatology. Relevant to the care of individuals with arthritis and related disorders, major topics are evidence-based practice studies, clinical problems, practice guidelines, health care economics, health care policy, educational, social, and public health issues, and future trends in rheumatology practice.

Volume 75 No. 9 September 2023

Arthritis Care & Research

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

Cover design: Sandra Pulmano

♾ This journal is printed on acid-free paper.

Arthritis Care & Research

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

VOLUME 75 • SEPTEMBER 2023 • NO. 9

Cover image: The image on the cover (from Chen et al, page 1949), shows an example of change in tophus composition using dual energy computed tomography during two years of urate-lowering therapy.

Arthritis Care & Research
Vol. 75, No. 9, September 2023, pp 1859–1870
DOI 10.1002/acr.25090

American College of Rheumatology
Empowering Rheumatology Professionals

Assessing the Costs of Neuropsychiatric Disease in the Systemic Lupus International Collaborating Clinics Cohort Using Multistate Modeling

Ann E. Clarke,[1] John G. Hanly,[2] Murray B. Urowitz,[3] Yvan St. Pierre,[4] Caroline Gordon,[5] Sang-Cheol Bae,[6] Juanita Romero-Diaz,[7] Jorge Sanchez-Guerrero,[3] Sasha Bernatsky,[4] Daniel J. Wallace,[8] David A. Isenberg,[9] Anisur Rahman,[9] Joan T. Merrill,[10] Paul R. Fortin,[11] Dafna D. Gladman,[3] Ian N. Bruce,[12] Michelle Petri,[13] Ellen M. Ginzler,[14] Mary Anne Dooley,[15] Rosalind Ramsey-Goldman,[16] Susan Manzi,[17] Andreas Jönsen,[18] Graciela S. Alarcón,[19] Ronald F. Van Vollenhoven,[20] Cynthia Aranow,[21] Meggan Mackay,[21] Guillermo Ruiz-Irastorza,[22] S. Sam Lim,[23] Murat Inanc,[24] Kenneth C. Kalunian,[25] Soren Jacobsen,[26] Christine A. Peschken,[27] Diane L. Kamen,[28] Anca Askanase,[29] and Vernon Farewell[30]

Objective. To estimate direct and indirect costs associated with neuropsychiatric (NP) events in the Systemic Lupus International Collaborating Clinics inception cohort.

Methods. NP events were documented annually using American College of Rheumatology definitions for NP events and attributed to systemic lupus erythematosus (SLE) or non-SLE causes. Patients were stratified into 1 of 3 NP states (no, resolved, or new/ongoing NP event). Change in NP status was characterized by interstate transition rates using multistate modeling. Annual direct costs and indirect costs were based on health care use and impaired productivity over the preceding year. Annual costs associated with NP states and NP events were calculated by averaging all observations in each state and adjusted through random-effects regressions. Five- and 10-year costs for NP states were predicted by multiplying adjusted annual costs per state by expected state duration, forecasted using multistate modeling.

Results. A total of 1,697 patients (49% White race/ethnicity) were followed for a mean of 9.6 years. NP events (n = 1,971) occurred in 956 patients, 32% attributed to SLE. For SLE and non-SLE NP events, predicted annual, 5-, and 10-year direct costs and indirect costs were higher in new/ongoing versus no events. Direct costs were 1.5-fold higher and indirect costs 1.3-fold higher in new/ongoing versus no events. Indirect costs exceeded direct costs 3.0 to 5.2 fold. Among frequent SLE NP events, new/ongoing seizure disorder and cerebrovascular disease accounted for the largest increases in annual direct costs. For non-SLE NP events, new/ongoing polyneuropathy accounted for the largest increase in annual direct costs, and new/ongoing headache and mood disorder for the largest increases in indirect costs.

Conclusion. Patients with new/ongoing SLE or non-SLE NP events incurred higher direct and indirect costs.

INTRODUCTION

Approximately 50% of patients with systemic lupus erythematosus (SLE) experience neurologic and/or psychiatric (NP) events (1,2) ranging from common syndromes such as mild cognitive dysfunction, anxiety, and headache to infrequent manifestations such as psychosis and neuropathy (3). Approximately 30% of these NP events are reported to be directly attributable to SLE (4). NP events

The Systemic Lupus International Collaborating Clinics research network received partial funding for this study from UCB Pharmaceuticals. The Montreal General Hospital Lupus Clinic is partially supported by the Singer Family Fund for Lupus Research. The Hopkins Lupus Cohort is supported by the NIH (grants AR-43727 and AR-69572). Dr. Clarke holds the Arthritis Society Chair in Rheumatic Diseases at the University of Calgary. Dr. Hanly's work was supported by the Canadian Institutes of Health Research (grant MOP-88526). Dr. Gordon's work was supported by Lupus UK, Sandwell and West Birmingham Hospitals NHS Trust, and the Birmingham NIHR/Wellcome Trust Clinical Research Facility. Dr. Bae's work was supported in part by the National Research Foundation of Korea, funded by the Ministry of Education (grant NRF-2021R1A6A1A03038899). Dr. Bernatsky holds a James McGill Research Chair. Drs. Isenberg and Rahman's work was supported by the NIHR University College London Hospitals Biomedical Research Center. Dr. Fortin holds a tier 1 Canada Research Chair on Systemic Autoimmune Rheumatic Diseases at Université Laval. Dr. Bruce's work was supported by Versus Arthritis UK, the NIHR Manchester Biomedical Research Centre, and the NIHR Manchester Clinical Research Facility. Dr. Dooley's work was supported by the NIH (grant RR-00046). Dr. Ramsey-Goldman's work was supported by the NIH (grants 5UL1-TR-001422-02 [formerly 8UL1-TR-000150 and UL1-RR-025741], K24-AR-02318, and P30-AR-072579 [formerly P60-AR-064464 and P60-AR-48098]). Dr. Ruiz-Irastorza's work was supported by the Department of Education,

SIGNIFICANCE & INNOVATIONS

- This is the first study to assess the long-term economic burden of neuropsychiatric (NP) lupus in an international, multiethnic inception cohort using multistate modeling to characterize transition between onset, remission, and relapse of NP events.
- For SLE and non-SLE NP events, annual, 5-, and 10-year direct costs were higher in those with new/ongoing versus no events and resolved versus no events. For SLE and non-SLE NP events, annual, 5-, and 10-year indirect costs were higher in those with new/ongoing versus no events, and 5- and 10-year indirect costs were higher in new/ongoing versus resolved events.
- Among frequent SLE NP events, new/ongoing seizure disorder and cerebrovascular disease accounted for the largest increases in annual direct costs. For non-SLE NP events, new/ongoing polyneuropathy accounted for the largest increase in annual direct costs, and new/ongoing headache and mood disorder for the largest increases in indirect costs.
- The high economic burden associated with NP events in SLE, in addition to the previously documented negative impact on health-related quality of life and mortality, underline the importance of improving care for this component of SLE.

in SLE patients negatively impact health-related quality of life (5,6) and increase mortality (1,7), but little is known about their economic impact.

A few studies have reported the direct and indirect costs associated with NPSLE (8–12), but most were limited as they relied on administrative data (8–10), provided only direct (8–10) or short-term (8,10–12) cost estimates, or involved a single center (11,12). The long-term economic burden has never been assessed in an international, multiethnic cohort such as the Systemic Lupus International Collaborating Clinics (SLICC) inception cohort using multistate modeling. Multistate modeling can characterize the transition of SLE patients between onset, remissions, and relapses of different disease states, capturing both the likelihood of moving between states and state durations. We have previously used multistate modeling developed in the SLICC cohort to estimate long-term costs associated with renal involvement (13) and damage accrual (14). Hanly et al recently have described dynamic changes in NP events, both attributable (SLE NP events) and not attributable to SLE (non-SLE NP events) using reversible multistate modeling (6). In the current study, we calculated annual direct and indirect costs for each SLE and non-SLE NP state and used the interstate transition probabilities predicted in the models to estimate the expected duration in each state. Five- and 10-year cumulative costs were then estimated by multiplying the annual costs associated with each NP state with the expected duration in that state, providing predictions of long-term costs for states with limited observations. We also provide cost estimates for individual SLE and non-SLE NP events.

PATIENTS AND METHODS

Inception cohort. Between 1999 and 2011, patients from 31 centers in 11 countries fulfilling the American College of Rheumatology (ACR) revised classification criteria for SLE (15) were enrolled in the SLICC inception cohort within 15 months of diagnosis and assessed longitudinally. For this study, data collection continued until December 2019. Each patient provided informed

Universities, and Research of the Basque Government. Dr. Jacobsen's work was supported by the Danish Rheumatism Association (grant A3865) and the Independent Research Fund Denmark (grant 0134-00473B).

[1]Ann E. Clarke, MD, MSc: University of Calgary, Calgary, Alberta, Canada; [2]John G. Hanly, MD: Queen Elizabeth II Health Sciences Centre and Dalhousie University, Halifax, Nova Scotia, Canada; [3]Murray B. Urowitz, MD, Jorge Sanchez-Guerrero, MD, MS, Dafna D. Gladman, MD: Toronto Western Hospital and University of Toronto, Toronto, Ontario, Canada; [4]Yvan St. Pierre, MSc, Sasha Bernatsky, MD, PhD: McGill University Health Center, Montreal, Quebec, Canada; [5]Caroline Gordon, MD: University of Birmingham, Birmingham, UK; [6]Sang-Cheol Bae, MD, PhD, MPH: Hanyang University Hospital for Rheumatic Diseases and Hanyang University, Seoul, Republic of Korea; [7]Juanita Romero-Diaz, MD, MS: Instituto Nacional de Ciencias Médicas y Nutricion, Mexico City, Mexico; [8]Daniel J. Wallace, MD: Cedars-Sinai and University of California, Los Angeles; [9]David A. Isenberg, MD, Anisur Rahman, PhD: University College London, London, UK; [10]Joan T. Merrill, MD: Oklahoma Medical Research Foundation, Oklahoma City; [11]Paul R. Fortin, MD, MPH: CHU de Québec-Université Laval, Quebec City, Quebec, Canada; [12]Ian N. Bruce, MD: The University of Manchester and Central Manchester University Hospitals National Health Service Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK; [13]Michelle Petri, MD, MPH: Johns Hopkins University School of Medicine, Baltimore, Maryland; [14]Ellen M. Ginzler, MD, MPH: State University of New York Downstate Health Sciences University, Brooklyn; [15]Mary A. Dooley, MD, MPH: University of North Carolina, Chapel Hill; [16]Rosalind Ramsey-Goldman, MD, DrPh: Northwestern University, Chicago, Illinois; [17]Susan Manzi, MD, MPH: Alleghany Health Network, Pittsburgh, Pennsylvania; [18]Andreas Jönsen, MD, PhD: Lund University, Lund, Sweden; [19]Graciela S. Alarcón, MD, MPH: University of Alabama at Birmingham; [20]Ronald F. van Vollenhoven, MD, PhD: Amsterdam University Medical Centers, Amsterdam, The Netherlands; [21]Cynthia Aranow, MD, Meggan Mackay, MD, MS: Feinstein Institute for Medical Research, Manhasset, New York; [22]Guillermo Ruiz-Irastorza, MD, PhD: University of the Basque Country, Barakaldo, Spain; [23]S. Sam Lim, MD, MPH: Emory University School of Medicine, Atlanta, Georgia; [24]Murat Inanc, MD: Istanbul University, Istanbul, Turkey; [25]Kenneth C. Kalunian, MD: University of California San Diego, La Jolla; [26]Soren Jacobsen, MD, DMSc: Rigshospitalet and Copenhagen University Hospital, Copenhagen, Denmark; [27]Christine A. Peschken, MD, MSc: University of Manitoba, Winnipeg, Manitoba, Canada; [28]Diane L. Kamen, MD, MSCR: Medical University of South Carolina, Charleston; [29]Anca Askanase, MD, MPH: Hospital for Joint Diseases and New York University Seligman Center for Advanced Therapeutics, New York, New York; [30]Vernon Farewell, PhD: University of Cambridge, Cambridge, UK.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.25090&file=acr25090-sup-0001-Disclosureform.pdf.

Address correspondence via email to Ann E. Clarke, MD, MSc, at aeclarke@ucalgary.ca.

Submitted for publication July 5, 2022; accepted in revised form January 19, 2023.

consent, and research ethics boards at each site approved the study.

At enrollment, data were collected on age, sex, and self-reported race and ethnicity, and at enrollment and annually, on disease activity (16), damage (17), NP events (using the ACR case definitions) (18), postsecondary education, smoking, and alcohol consumption (19). At enrollment and annually, data were also collected on hospitalizations and medications (regardless of attribution to SLE) in the year preceding each visit. The cohort was originally created to assess cardiovascular, NP, and renal outcomes, and therefore, data on diagnostic/therapeutic procedures were limited.

Beginning in 2015, 18 sites collected supplemental economic data annually on patients still followed in the cohort (see Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25090); the supplemental data included: 1) additional health resource utilization that was not captured in the preexisting data collection (i.e., physicians, nonphysician health care professionals, emergency room visits, laboratory tests, radiologic and other diagnostic procedures, outpatient surgeries, and help obtaining medical care) and 2) lost productivity in labor force and nonlabor force activity over the year (20,21) preceding the assessment. All health care use and all health-related lost productivity were included regardless of attribution to SLE.

Statistical analysis. *Multistate modeling.* At enrollment and annually, patients were assessed for NP events attributed to SLE (SLE NP events) or non-SLE causes (non-SLE NP events). NP events were attributed to SLE based on published attribution decision rules (6) and were attributed if they: 1) had their onset within 10 years of SLE diagnosis and were still present within the enrollment window or occurred subsequently; 2) had no concurrent non-SLE causes; and 3) were not one of the common NP events in the normal population, as described by Ainiala et al (22). Separate patient-level models were developed for SLE and non-SLE NP events, including the following 3 states: 1) no NP event ever; 2) resolved NP event, i.e., no current NP event but ≥1 in the past (state entry was time of resolution of the NP event); and 3) new/ongoing NP event with state entry at onset of the event (see Supplementary Figure 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25090).

At each assessment, patients were assigned to 1 of 3 NP states. The SLE NP and non-SLE NP models were estimated independently, and all patients were included in both. When fitting the model for SLE NP events, non-SLE NP events were ignored, and vice versa. Therefore, we did not estimate costs for SLE NP events with or without concurrent non-SLE NP events. As costs were only collected at assessments prior to death and not over the interval between the last follow-up visit and death, death was not included in the economic models, although it was allowed for in the multistate modeling. Transition rates were estimated through maximum likelihood estimation using the R (23) package "msm" (24).

Calculating annual direct costs. At each assessment, annual direct costs were based on health resource utilization over the preceding year and annual indirect costs on lost time in labor force and nonlabor force activity over the preceding year (depending on the cost data set available; refer to cost data set description below). Annual costs associated with each SLE and non-SLE NP state were calculated by averaging costs for all patients contributing an observation to that state.

Health care costs were calculated by multiplying each health resource by its corresponding 2021 Canadian unit cost (sources of unit cost for health care components are provided in Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25090). As the objective of this research was to compare health care costs between SLE patients with new/ongoing versus resolved versus no NP events rather than to provide country-specific estimates of costs, health care prices and wages essentially served as a set of weights to aggregate resources and lost productivity into a single cost measure. Canadian prices were chosen because the largest proportion of patient observations was from Canada, and prices are set in a single-payer universal public system covering the entire Canadian population and therefore better reflect the direct cost of resources.

Calculating annual indirect costs. Total indirect costs consisted of the sum of the following components: 1) absenteeism, 2) presenteeism, and 3) opportunity costs. Absenteeism referred to self-report of time lost from paid labor because of poor health; presenteeism referred to self-report of how productivity, while engaged in labor and nonlabor force activities, was affected by health based on a visual analog scale anchored at 0% for health having no impact and 100% for complete inability to work; opportunity costs referred to additional time that patients would be working in labor force and nonlabor force activities if not ill. Opportunity costs were calculated as the difference between the time patients reported working versus the time worked by an age, sex, and geographic-matched general population (25–28). Indirect costs from labor force activities were valued using age-and-sex-specific wages from Statistics Canada (29). Indirect costs from nonlabor force activities were valued using opportunity costs (i.e., age-and-sex specific wages rather than expected earnings of service workers).

Cost data sets. Based on our method of collecting data on health resource utilization and lost productivity, we have 2 types of cost data: 1) partial direct costs based on the data provided by the full cohort. These partial direct costs included hospitalizations, medications, selected procedures, and dialysis; and 2) complete direct and indirect costs for the cohort subset who completed the annual supplemental economic questionnaire introduced in 2015.

To take full advantage of both cost data sets, we used a multiple imputation strategy to predict all missing values for the patients in the full cohort who did not provide complete direct and indirect costs for all observations. All models for imputing complete direct and indirect costs included partial direct costs and NP state (time-varying) as well as education and geographic location as final covariates, with the direct cost model also including age at diagnosis, and the indirect cost model also including race and ethnicity. Ten sets of imputations were derived from these models, and all subsequent analyses in this setting involved pooling and averaging all estimates across imputed sets, while their variances were computed by applying standard combination rules.

Adjusting annual costs and predicting 5- and 10-year cumulative costs. Within each of the 3 data settings (i.e., partial direct costs for the full cohort, unimputed complete costs for the cohort subset, and imputed complete costs for the full cohort), multivariate random-effects linear regression modeling was used to adjust for possible confounding of demographic variables on the association of annual direct and indirect costs and NP state. Potential covariates included age at diagnosis, sex, race and ethnicity, education, and geographic regions as well as the following time-varying covariates: age, disease duration, smoking, and high-risk alcohol use. Using the average values of significant covariates, predictions were obtained for adjusted annual costs; 95% confidence intervals (95% CIs) were calculated using bootstrapping except in the multiple imputation setting, where bootstrapping does not appear to provide realistic variance estimates (30). All statistical computations were done using Stata, version 17.

For each NP state, cumulative adjusted costs over the following 5 and 10 years were predicted by multiplying adjusted annual costs by the expected duration in each state for each of the following years. Annual change in NP state was determined using transition probabilities derived from the multistate model. Future costs were discounted at an annual rate of 3%.

Assessing costs associated with individual SLE and non-SLE NP events. The increase in annual costs associated with the 4 most frequent SLE and non-SLE NP events was also estimated. Random-effects linear regression models were developed using the imputed complete costs for the full cohort with annual direct and indirect costs as the outcomes for SLE and non-SLE NP events. In each model, predictors included indicator variables for whether any of the 4 most frequent events or any other NP events (SLE or non-SLE, depending on the model) had been ongoing at any time over each observed patient-year, as well as other statistically significant covariates, i.e., race and ethnicity and disease duration for direct costs; disease duration, region, and education for indirect costs. This allows cost increases associated with specified new/ongoing NP events to be estimated independently of any co-occurring NP event and compared to no and resolved NP events.

RESULTS

Patients. A total of 1,827 patients were recruited in the SLICC inception cohort, and 1,697 provided utilization data on hospitalizations, medications, and selected procedures. Of these 1,697 patients, 672 patients were still being followed in 2015 when the annual questionnaire on additional health resource utilization and lost productivity was introduced. In the full cohort of 1,697 patients, 88.7% were female subjects, 48.8% were of White race and ethnicity, and their mean ± SD age and mean disease duration at cohort enrollment were 35.1 ± 13.3 years and 0.5 years (range 0–1.3 years), respectively (Table 1). In total, 1,971 unique NP events occurred in 956 patients, 32% attributed to SLE. Mood disorder (121 of 624 SLE NP events, 19.4%), seizure disorder (19.2%), cerebrovascular disease (19.1%), and mononeuropathy (7.7%) were the most frequent SLE NP events, and headache (940 of 1,347 non-SLE NP events, 69.8%), mood disorder (14.8%), anxiety (7.0%), and polyneuropathy (2.8%) were the most frequent non-SLE NP events (see Supplementary

Table 1. Demographic and clinical characteristics at the time of cohort entry for the full sample, providing partial direct costs, and for the cohort subset, providing complete direct and indirect costs*

Characteristic	Full sample (n = 1,697)	Subset (n = 672)
Age, mean ± SD years	35.1 ± 13.3	33.2 ± 12.0
Sex, female	88.7	89.3
Education, any postsecondary	61.8	61.1
Race/ethnicity		
White	48.8	40.9
African	16.7	11.5
Hispanic	15.8	18.0
Asian	15.0	26.5
Geographic region		
US	27.9	14.6
Europe	26.8	8.2
Canada	23.2	41.4
Mexico	12.6	16.1
Republic of Korea	9.5	19.8
Disease duration, mean (range) years	0.5 (0.0–1.3)	0.4 (0.0–1.3)
SLEDAI–2K score, mean ± SD	5.4 ± 5.4	6.1 ± 5.6
SDI score at first annual follow-up, mean ± SD	0.44 ± 0.87	0.36 ± 0.78
Medications		
Glucocorticoids	70.9	72.3
Antimalarials	67.7	68.4
Immunosuppressants	40.9	42.9
Smoking, ever	35.0	30.2
High-risk alcohol consumption†	1.3	0.6
Employed‡	–	59.8

* Values are the percentage unless indicated otherwise. SDI = Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index; SLEDAI-2K = Systemic Lupus Erythematosus Disease Activity Index 2000.
† Refers to >15 drinks per week for men and >10 drinks per week for women (19).
‡ At the time of the completion of the first economic questionnaire; data were only available for the subcohort completing the economic questionnaire.

Tables 2 and 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25090).

In the subset of 672 patients providing complete economic data, 89.3% were female subjects, 40.9% were of White race and ethnicity, and their mean ± SD age and mean disease duration at time of enrollment in the inception cohort were 33.2 ± 12.0 years and 0.4 years (range 0–1.3 years), respectively. Their mean disease duration at the time of introduction of the economic questionnaire was 10.8 years (range 3.9–19.1 years). The cohort subset had a larger proportion of Asian patients than the full cohort, and Canada, Mexico, and Korea contributed a higher proportion of patients to this subset than the full cohort. Transition probabilities are shown in Supplementary Table 4, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25090.

Partial direct costs on full cohort. *Annual costs and predictors.* For the 1,697 patients, there was a mean follow-up of 9.6 years, yielding 13,987 observations (Table 2; see Supplementary Table 5, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25090, for the distribution of observations per country). In the regression model that examined the association between annual partial direct costs and SLE NP states, older age at diagnosis and White race and ethnicity were associated with lower costs, whereas longer disease duration was associated with higher costs (see Supplementary Table 6, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25090, panel A, model 1). A similar relationship was observed in the model for annual partial direct costs and non-SLE NP states (see Supplementary Table 6, panel B, model 1). Adjusted annual partial direct costs were higher in those with new/ongoing SLE NP events ($7,028 [2021 Canadian]) versus those with no SLE NP events ($4,212; difference $2,816 [95% CI $1,139, $4,493]) (see Supplementary Table 7, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25090).

Five- and 10-year cumulative costs. For SLE NP events, patients with new/ongoing versus no events at the beginning of the 5-year period incurred higher predicted 5-year partial direct costs (i.e., new/ongoing [$34,580] versus no events [$23,149; difference $11,431 (95% CI $5,293, $17,570)]) (see Supplementary Table 7, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25090). Similarly, patients with new/ongoing versus no events at the beginning of the 10-year period incurred higher predicted 10-year partial direct costs (i.e., new/ongoing [$67,407] versus no events [$48,416; difference $18,992 (95% CI $8,774, $29,210)]). For the non-SLE NP events, 5- and 10-year partial direct costs were also higher in those with new/ongoing versus no events.

Complete direct and indirect costs on cohort subset. *Annual costs and predictors.* For the 672 patients in the cohort subset completing the economic questionnaire starting in 2015, there was a mean follow-up of 2.7 years, yielding 1,594 observations (Table 3). Across all SLE and non-SLE NP states, indirect costs exceeded direct costs by an average of 4.4 fold; within indirect costs, unpaid labor costs exceeded paid labor costs by an average of 1.6 fold.

In the regression model that examined the association between annual complete direct costs and SLE NP states (see Supplementary Table 6, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25090, panel A, model 2), no additional variables were associated with costs, whereas in the model examining the association between annual complete direct costs and non-SLE NP states (see Supplementary Table 6, panel B, model 2), longer disease duration was associated with higher costs. In the model examining the association between annual indirect costs and SLE NP states (see Supplementary Table 6, panel A, model 3) and non-SLE NP states (see Supplementary Table 6, panel B, model 3), longer disease duration was associated with higher costs, whereas postsecondary education and residing outside of North America were associated with lower costs.

Adjusted annual complete direct costs were higher in those with new/ongoing SLE NP events ($13,825) versus those with no SLE NP events ($7,505; difference $6,320 [95% CI $1,399, $11,241]) (Table 4). Adjusted annual indirect costs were also higher in those with new/ongoing ($42,695) versus no SLE NP events ($33,347; difference $9,348 [95% CI $1,004, $17,692]). Similarly, adjusted annual direct and indirect costs were higher in those with new/ongoing non-SLE NP events versus no non-SLE NP events.

Five- and 10-year cumulative costs. For the SLE NP events, predicted 5-year complete direct costs were higher in those with new/ongoing ($62,071) versus those with no events ($36,948; difference $25,123 [95% CI $6,566, $43,680]) (Table 4). Similarly, 10-year complete direct costs were higher in those with new/ongoing ($110,682) versus no events ($69,870; difference $40,812 [95% CI $7,186, $74,438]). Five-year cumulative indirect costs were higher in the new/ongoing ($209,893) versus no SLE NP event ($177,634; difference $32,259 [95% CI $2,380, $62,138]). For the non-SLE NP events, 5- and 10-year complete direct and indirect costs were higher in the new/ongoing versus no event.

Imputed complete direct and indirect costs on full cohort. *Annual costs and predictors.* Unadjusted imputed annual direct and indirect costs for the full cohort are shown in Supplementary Table 8, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25090. In the regression model that examined the association between imputed annual complete direct costs and SLE NP states (see Supplementary Table 6, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25090, panel A, model 4) and non-SLE NP states (see Supplementary

Table 2. Unadjusted annual partial direct costs (in 2021 Canadian dollars) for the full cohort (n = 1,697) stratified by systemic lupus erythematosus (SLE) and non-SLE neuropsychiatric (NP) events*

Direct cost components	SLE NP events			Non-SLE NP events		
	No NP event	Resolved NP event	New/ongoing NP event	No NP event	Resolved NP event	New/ongoing NP event
Patients, no.	1,487	235	273	1,280	504	620
Observations, no. (%)	11,155 (80)	1,562 (11)	1,270 (9)	8,169 (58)	2,922 (21)	2,896 (21)
Direct costs	3,948 (3,609, 4,287)	5,054 (4,163, 5,945)	6,916 (5,864, 7,967)	3,559 (3,253, 3,866)	5,405 (4,637, 6,173)	5,472 (4,579, 6,365)
Hospital	1,148 (1,032, 1,263)	1,682 (1,211, 2,154)	2,246 (1,797, 2,695)	1,179 (1,040, 1,318)	1,597 (1,285, 1,910)	1,376 (1,156, 1,597)
Medications	1,833 (1,560, 2,107)	2,219 (1,620, 2,819)	2,459 (1,934, 2,984)	1,629 (1,412, 1,846)	2,237 (1,672, 2,802)	2,485 (1,732, 3,238)
Tests	72 (66, 78)	85 (68, 102)	162 (138, 186)	76 (69, 83)	64 (54, 73)	117 (102, 132)
Dialysis	895 (743, 1,047)	1,067 (636, 1,498)	2,049 (1,352, 2,746)	676 (527, 826)	1,507 (1,116, 1,898)	1,494 (1,097, 1,890)

* Values are the mean (95% confidence interval) unless indicated otherwise.

Table 3. Unadjusted annual complete direct, indirect, and total costs (in 2021 Canadian dollars) for the cohort subset (n = 672), providing complete cost data, stratified by systemic lupus erythematosus (SLE) and non-SLE neuropsychiatric (NP) events*

	SLE NP events			Non-SLE NP events		
Direct and indirect cost components	No NP event	Resolved NP event	New/ongoing NP event	No NP event	Resolved NP event	New/ongoing NP event
Patients, no.	537	93	60	361	191	167
Observations, no. (%)	1,250 (78)	214 (13)	130 (8)	822 (52)	403 (25)	369 (23)
Direct costs	7,240 (6,150, 8,331)	9,245 (6,565, 11,926)	10,938 (7,383, 14,493)	6,125 (5,172, 7,078)	9,009 (6,186, 11,832)	10,259 (8,373, 12,145)
Hospital	1,156 (824, 1,487)	1,644 (738, 2,550)	1,683 (898, 2,468)	1,153 (709, 1,597)	1,185 (730, 1,641)	1,598 (975, 2,221)
Medications	2,254 (1,443, 3,065)	2,040 (1,073, 3,008)	3,518 (1,524, 5,512)	1,807 (1,356, 2,259)	2,948 (573, 5,323)	2,813 (2,017, 3,608)
Physicians	1,010 (928, 1,092)	1,241 (905, 1,577)	1,376 (1,106, 1,645)	898 (813, 983)	1,114 (955, 1,274)	1,409 (1,170, 1,648)
Tests	732 (641, 823)	698 (559, 837)	1,108 (822, 1,394)	639 (554, 725)	819 (688, 951)	956 (721, 1,191)
Dialysis	1,203 (657, 1,749)	2,397 (575, 4,220)	2,030 (–208, 4,268)	850 (293, 1,407)	1,910 (706, 3,113)	2,201 (832, 3,571)
Other†	886 (759, 1,012)	1,225 (804, 1,645)	1,224 (872, 1,575)	777 (643, 911)	1,033 (747, 1,319)	1,282 (1,017, 1,547)
Indirect costs	33,108 (30,767, 35,450)	34,444 (28,959, 39,928)	44,571 (38,140, 51,001)	29,820 (27,167, 32,473)	36,095 (32,189, 40,001)	41,984 (37,042, 46,926)
Paid labor costs	12,449 (10,674, 14,224)	13,998 (10,437, 17,559)	22,127 (18,150, 26,104)	10,873 (8,961, 12,786)	13,437 (10,741, 16,133)	19,188 (15,259, 23,117)
Absenteeism	1,621 (722, 2,519)	993 (622, 1,364)	514 (150, 878)	922 (700, 1,144)	1,523 (1,046, 2,000)	2,529 (–447, 5,505)
Presenteeism	4,646 (4,101, 5,191)	3,850 (2,615, 5,086)	3,001 (1,672, 4,331)	4,254 (3,631, 4,877)	5,474 (4,369, 6,580)	3,575 (2,711, 4,438)
Opportunity	6,182 (4,441, 7,923)	9,155 (5,107, 13,203)	18,612 (13,881, 23,343)	5,697 (3,627, 7,767)	6,440 (3,284, 9,596)	13,084 (9,911, 16,258)
Unpaid labor costs	20,659 (19,075, 22,244)	20,446 (15964, 24,927)	22,444 (17,381, 27,507)	18,947 (16,959, 20,935)	22,658 (19,681, 25,634)	22,796 (19,899, 25,693)
Presenteeism	7,979 (7,161, 8,798)	7,364 (5,628, 9,100)	8,627 (5,816, 11,439)	6,641 (5,755, 7,527)	9,291 (7,748, 10,834)	9,399 (7,705, 11,093)
Opportunity	12,680 (10,845, 14,515)	13,082 (8,061, 18,102)	13,817 (7,932, 19,702)	12,306 (10,081, 14,531)	13,366 (9,933, 16,800)	13,397 (9,826, 16,968)
Total costs	40,349 (37,691, 43,007)	43,689 (37,443, 49,935)	55,509 (47,659, 63,359)	35,945 (33,132, 38,759)	45,104 (39,994, 50,214)	52,243 (46,747, 57,740)

* Values are the mean (95% confidence interval) unless indicated otherwise.
† Other includes non-physician health care professional, emergency room visits, outpatient surgeries, and help obtaining medical care.

Table 4. Predicted annual and 5- and 10-year direct and indirect costs (in 2021 Canadian dollars) for the cohort subset (n = 672), providing complete cost data, stratified by systemic lupus erythematosus (SLE) and non-SLE neuropsychiatric (NP) events*

	No NP event	Resolved NP event	New/ongoing NP event	Difference between resolved and no NP event	Difference between ongoing/new and no NP event	Difference between new/ongoing and resolved NP event
SLE NP events						
Direct costs						
1 year	7,505	10,704	13,825	3,199 (–4,140, 10,538)	6,320 (1,399, 11,241)†	3,121 (–3,260, 9,501)
5 years	36,948	51,118	62,071	14,170 (–18,588, 46,928)	25,123 (6,566, 43,680)†	10,953 (–20,059, 41,965)
10 years	69,870	96,196	110,682	26,326 (–31,628, 84,280)	40,812 (7,186, 74,438)†	14,486 (–26,529, 55,500)
Indirect costs‡						
1 year	33,347	32,941	42,695	–406 (–7,332, 6,520)	9,348 (1,004, 17,692)†	9,754 (1,006, 18,502)†
5 years	177,634	175,660	209,893	–1,974 (–32,889, 28,940)	32,259 (2,380, 62,138)†	34,234 (–3,825, 72,292)
10 years	366,003	364,372	409,647	–1,630 (–56,335, 53,706)	43,645 (–2,387, 89,677)	45,275 (–5,059, 95,608)
Non-SLE NP events						
Direct costs§						
1 year	6,606	9,893	11,181	3,287 (–435, 7,010)	4,575 (1,145, 8,006)†	1,288 (–2,156, 4,733)
5 years	37,984	51,051	55,364	13,067 (–2,510, 28,645)	17,380 (5,974, 28,787)†	4,313 (–12,635, 21,261)
10 years	80,277	103,034	108,532	22,757 (–3,104, 48,617)	28,255 (10,588, 45,922)†	5,498 (–16,108, 27,104)
Indirect costs‡						
1 year	30,391	35,545	40,665	5,154 (–451, 10,759)	10,274 (3,493, 17,056)†	5,120 (–1,687, 11,927)
5 years	166,771	186,798	203,940	20,027 (–3,438, 43,492)	37,169 (14,976, 59,361)†	17,142 (–12,313, 46,597)
10 years	347,053	382,680	404,533	35,627 (–3,311, 74,565)	57,480 (24,978, 89,982)†	21,853 (–15,697, 59,402)

* Values are the mean (95% confidence interval) unless indicated otherwise.
† Significant difference (as the 95% CI does not include 0).
‡ Adjusted for disease duration, education, and residing outside North America.
§ Adjusted for disease duration.

Table 6, panel B, model 4), longer disease duration was associated with higher costs, whereas White race and ethnicity was associated with lower costs. In the model examining the association between imputed annual indirect costs and SLE NP states (see Supplementary Table 6, panel A, model 5) and non-SLE NP states (see Supplementary Table 6, panel B, model 5), White race and ethnicity was associated with higher costs, whereas residing outside of North America was associated with lower costs.

Adjusted imputed annual complete direct costs were higher in those with new/ongoing SLE NP events ($10,471) versus those with no SLE NP events ($6,668; difference $3,803 [95% CI $2,136, $5,471]) (Table 5; expressed as US dollars using 2021 purchasing power parity [31] in Supplementary Table 9, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25090). Adjusted imputed annual complete direct costs were also higher in the resolved ($9,089) versus no SLE NP event ($6,668; difference $2,421 [95% CI $859, $3,983]). Adjusted imputed annual indirect costs were higher in those with new/ongoing ($37,197) versus no SLE NP events ($26,248; difference $10,950 [95% CI $376, $21,523]). For the non-SLE NP events, adjusted imputed annual complete direct costs were higher in the new/ongoing versus no event and the resolved versus no event. Adjusted imputed annual indirect costs were higher in the new/ongoing versus no event and new/ongoing versus resolved event.

Five- and 10-year cumulative costs. For the SLE NP events, imputed 5- and 10-year complete direct costs were higher in the new/ongoing versus no event and in the resolved versus no event (Table 5). Imputed 5- and 10-year indirect costs were higher in the new/ongoing versus no event and new/ongoing versus resolved event. For the non-SLE NP events, imputed 5- and 10-year complete direct costs were higher in the new/ongoing versus no event and in the resolved versus no event. Imputed 5-year indirect costs were higher in the new/ongoing versus no event and new/ongoing versus resolved event, and imputed 10-year indirect costs were higher in the new/ongoing versus no event, resolved versus no event, and new/ongoing versus resolved event.

Costs of individual SLE and non-SLE NP events. For SLE NP events, new/ongoing seizure disorder, cerebrovascular disease, and NP event(s) other than the 4 most frequent (i.e., mood disorder, seizure disorder, cerebrovascular disease, and mononeuropathy), respectively, accounted for increases in annual direct costs of $10,179 (95% CI $7,114, $13,245), $3,907 (95% CI $920, $6,893), and $4,383 (95% CI $2,272, $6,494) (Table 6). Only new/ongoing SLE NP events other than the 4 most frequent were associated with an increase in annual indirect costs ($8,065 [95% CI $22, $16,108]).

For non-SLE NP events, new/ongoing headache, polyneuropathy, and NP event(s) other than the 4 most frequent (i.e., headache, mood disorder, anxiety disorder, and polyneuropathy), respectively, accounted for increases in annual direct costs of $1,216 (95% CI $202, $2,229), $9,168 (95% CI $5,392, $12,943), and $8,939 (95% CI $5,564, $12,314) (Table 6). New/ongoing headache and mood disorder,

Table 5. Predicted imputed annual and 5- and 10-year complete direct and indirect costs (in 2021 Canadian dollars) for the full cohort (n = 1,697) stratified by systemic lupus erythematosus (SLE) and non-SLE neuropsychiatric (NP) events*

	No NP event	Resolved NP event	New/ ongoing NP event	Difference between resolved and no NP event	Difference between new/ongoing and no NP event	Difference between new/ongoing and resolved NP event
SLE NP events						
Direct costs†						
1 year	6,668	9,089	10,471	2,421 (859, 3,983)‡	3,803 (2,136, 5,471)‡	1,382 (–602, 3,366)
5 years	35,324	46,066	50,916	10,742 (3,781, 17,704)‡	15,592 (9,601, 21,584)‡	4,850 (–3,149, 12,849)
10 years	71,906	91,667	98,081	19,761 (7,443, 32,079)‡	26,176 (16,707, 35,644)‡	6,415 (–4,164, 16,993)
Indirect costs§						
1 year	26,248	27,103	37,197	855 (–2,759, 4,469)	10,950 (376, 21,523)‡	10,094 (–2,136, 22,505)
5 years	139,617	143,244	178,672	3,627 (–11,907, 19,161)	39,055 (6,181, 71,930)‡	35,428 (363, 70,493)‡
10 years	286,295	294,893	341,747	8,599 (–18,925, 36,122)	55,453 (9,002, 101,904)‡	46,855 (480, 93,229)‡
Non-SLE NP events						
Direct costs†						
1 year	6,264	8,045	8,931	1,781 (438, 3,124)‡	2,667 (1,471, 3,864)‡	886 (–514, 2,287)
5 years	34,086	41,139	44,106	7,052 (1,491, 12,614)‡	10,019 (6,045, 13,994)‡	2,967 (–3,012, 8,947)
10 years	70,090	82,417	86,200	12,327 (3,094, 21,560)‡	16,110 (9,913, 22,306)‡	3,783 (–3,840, 11,405)
Indirect costs§						
1 year	24,286	29,059	35,732	4,772 (–214, 9,759)	11,446 (7,532, 15,360)‡	6,673 (1,179, 12,168)‡
5 years	134,332	152,589	174,931	18,256 (–748, 37,260)	40,598 (28,356, 52,841)‡	22,342 (2,858, 41,826)‡
10 years	279,010	311,965	340,447	32,955 (1,403, 64,507)‡	61,437 (41,757, 81,117)‡	28,482 (3,644, 53,320)‡

* Values are the mean (95% confidence interval) unless indicated otherwise.
† Adjusted for disease duration and White race and ethnicity.
‡ Significant difference (as the 95% CI does not include 0).
§ Adjusted for White race and ethnicity and residing outside North America.

Table 6. Regression models for direct and indirect costs stratified by individual systemic lupus erythematosus (SLE) and non-SLE neuropsychiatric (NP) events*

	Direct	Indirect
SLE NP events		
Mood disorder (new/ongoing)†	−1,147 (−3,374, 1,081)	6,495 (−3,927, 16,916)
Seizure disorder (new/ongoing)	10,179 (7,114, 13,245)‡	9,365 (−1,469, 20,200)
Cerebrovascular disease (new/ongoing)	3,907 (920, 6,893)‡	4,222 (−3,460, 11,904)
Mononeuropathy (new/ongoing)	1,899 (−1,699, 5,498)	4,205 (−5,625, 14,035)
Other NP event (new/ongoing)	4,383 (2,272, 6,494)‡	8,065 (22, 16,108)‡
White race/ethnicity	−2,380 (−3,452, −1,309)‡	–
Disease duration	278 (203, 354)‡	1,213 (367, 2,059)‡
Residing outside of North America§	–	−12,907 (−18,658, −7,157)‡
Postsecondary education	–	−5,866 (−10,074, −1,657)‡
Non-SLE NP events		
Headache (new/ongoing)	1,216 (202, 2,229)‡	6,824 (3,441, 10,208)‡
Mood disorder (new/ongoing)	−580 (−2,424, 1,263)	4,660 (229, 9,091)‡
Anxiety disorder (new/ongoing)	2,299 (−218, 4,816)	6,901 (−194, 13,996)
Polyneuropathy (new/ongoing)	9,168 (5,392, 12,943)‡	7,448 (−3,375, 18,270)
Other NP event (new/ongoing)	8,939 (5,564, 12,314)‡	3,597 (−3,290, 10,485)
White race/ethnicity	−2,502 (−3,626, −1,377)‡	–
Disease duration	282 (202, 363)‡	1,276 (796, 1,756)‡
Residing outside of North America§	–	−11,472 (−18,352, −4,592)‡
Postsecondary education	–	−5,821 (−10,333, −1,310)‡

* Values are the regression coefficient (95% confidence interval). Empty cells refer to variables that were included as potential covariates but were not retained in the final model, as they were not significant.
† Reference group is no event or resolved event.
‡ Significant difference (as the 95% CI does not include 0).
§ North America includes Canada, the US, and Mexico.

respectively, were associated with increases in annual indirect costs of $6,824 (95% CI $3,441, $10,208) and $4,660 (95% CI $229, $9,091).

DISCUSSION

We have provided the first estimates of annual and long-term costs stratified by patients with NP events attributed to both SLE and non-SLE causes and in varying stages of evolution (new/ongoing versus resolved). For SLE and non-SLE NP events, predicted annual, 5-, and 10-year direct costs were higher in the new/ongoing versus no events and resolved versus no events. For SLE and non-SLE NP events, annual, 5-, and 10-year indirect costs were higher in the new/ongoing versus no events, and 5- and 10-year indirect costs were higher in new/ongoing versus resolved events. Direct costs were 1.2- to 1.8-fold higher, and indirect costs 1.1- to 1.5-fold higher in patients with new/ongoing versus no NP events, and indirect exceeded direct costs between 3.0 and 5.2 fold. The higher direct and indirect costs in those with new/ongoing versus no event is to be expected based on the significantly poorer health-related quality of life experienced by those with NP lupus, as previously documented in this cohort (6). The relationship between costs and health-related quality of life is likely complex and bidirectional. Although Hanly et al have reported that patients in this cohort with NP events attributed to SLE generally have a more favorable outcome than patients with NP events attributed to non-SLE causes (32), we did not consistently observe lower costs in those with SLE NP events.

While a few studies have assessed costs associated with NPSLE (8–12), only 2 studies defined NPSLE based on ACR NP cases definitions (10,12), and 1 of these relied on claims data (10) to identify NP events. Both only included NP events attributable to SLE. Mean annual direct and indirect costs for a clinical cohort in Hong Kong with NPSLE (n = 83) were estimated at $16,590 and $9,240 (2021 US dollars) (12,31,33), respectively, whereas mean annual direct costs in NPSLE patients identified from a US claims database were $38,408 (10). The other NP cost studies (8,9) examined costs associated with damage accrual in the NP domain of the SLICC/ACR Damage Index (SDI), which includes only a subset of the items in our much broader definition of SLE and non-SLE NP events. For patients with damage in the NP domain of the SDI identified in a US claims database (8), mean annual direct costs were $28,191; for patients identified in Taiwanese National Health Insurance database (9), mean annual direct costs ranged between $2,558 for cranial or peripheral neuropathy and $19,949 for recurrent cerebrovascular accidents.

Annual direct and indirect costs in our patients with new/ongoing SLE NP events were $8,136 and $28,902 (2021 US dollars) and new/ongoing non-SLE NP events, $6,939 and $27,764. While our indirect cost estimates ($28,902 and $27,764) exceeded those in the Hong Kong cohort ($9,240) (12), our direct costs estimates ($8,136 and $6,939) were substantially lower than those from US administrative databases ($38,408 and $28,191) (8,10). Costs are expected to vary widely across studies due to a variety of factors. Direct costs are influenced by both the method of ascertainment (i.e., patient self-

report, medical chart review, or insurance claims databases) and source of valuation of health care resources (i.e., single-payer national health insurance or private medical insurer). Similarly, indirect costs depend on the method of measuring relevant time inputs (i.e., human capital or friction cost approach), whether presenteeism is accounted for, and valuation of lost productivity. Our estimates of indirect costs exceeded direct costs across all NP states, which is consistent with other SLE cost-of-illness studies (which do not provide cost estimates specifically for NPSLE) (34).

Annual direct cost increases associated with specified new/ongoing NP events in our cohort ranged from $1,216 for non-SLE headaches to $10,179 for SLE-associated seizure disorder, and indirect cost increases ranged from $4,660 for non-SLE mood disorder to $8,065 for SLE NP events other than the 4 most frequent. It is noteworthy that some ongoing non-SLE NP events such as headaches and mood disorder, despite appearing to require none or relatively modest additional health care resources, accounted for significant annual productivity losses (respectively, $6,824 and $4,660).

Our study is limited, as we were unable to collect data on direct and indirect costs in the interval between the last annual follow-up visit and death, and therefore our cost estimates do not represent costs incurred in the year prior to death, and our predictions are only applicable to individuals who would survive the entire predicted period. Further, we did not collect complete direct and indirect costs on the full cohort for the entire observation period. However, as we had collected data on the major sources of direct costs on the full cohort for the entire study and complete direct and indirect costs on a cohort subset, we believed that multiple imputation would allow us to accurately predict complete direct and indirect costs for the full cohort. Costs in the cohort subset were measured later in the disease course when patients were more likely to have accumulated more damage and experienced more NP events. Consistent with this, our estimates based on imputed data were more conservative than when using only unimputed data. Adjusted total costs observed in the cohort subset ranged from 16% to 24% higher than imputed total costs for the full cohort. By combining these imputed costs with interstate transition probabilities predicted in multistate models, we provide the first comprehensive long-term cost estimates for patients with no, active, and resolved NP events.

Additionally, we are not providing country-specific cost estimates for NPSLE. Rather, our purpose was to compare the costs of new/ongoing versus resolved versus no NP event, and we used Canadian prices and wages to aggregate resources and lost productivity into a single measure of direct or indirect costs. The use of Canadian prices results in an underestimation (or overestimation) of NP costs in countries where the prices of health care services are higher (or lower). Finally, although we assessed costs associated with varying states of NPSLE, all costs incurred by a patient were included in our estimates. Therefore, it was not possible to determine if cost differentials between NP states were directly attributable to an NP event or other SLE manifestations or comorbidities that may be correlated with NP events. Although dialysis, for example, may be a cost item that could be correlated without being causally linked to NP events, it should be noted that unadjusted partial direct costs for the full cohort excluding the portion due to dialysis remained higher in those with new/ongoing and resolved NP events versus no SLE or non-SLE NP events.

Both SLE and non-SLE NP events are important components of the economic costs associated with SLE. It is important to consider non-SLE NP events, as patients with SLE may be affected differently or experience different sequelae than individuals unaffected by SLE experiencing the event. Accordingly, current models of SLE care should consider allocating more health care resources to the detection and treatment of NP events, particularly the costliest, i.e., SLE-associated seizure disorder and cerebrovascular disease and non-SLE polyneuropathy, headache, and mood disorder. Further, the incorporation of economic outcomes in observational studies and clinical trials of NPSLE could help determine if the benefits of interventions are commensurate with their costs.

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. Clarke 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. Clarke, Hanly, St. Pierre, Farewell.

Acquisition of data. Clarke, Hanly, Urowitz, Gordon, Bae, Romero-Diaz, Sanchez-Guerrero, Bernatsky, Wallace, Isenberg, Rahman, Merrill, Fortin, Gladman, Bruce, Petri, Ginzler, Dooley, Ramsey-Goldman, Manzi, Jönsen, Alarcón, Van Vollenhoven, Aranow, Mackay, Ruiz-Irastorza, Lim, Inanc, Kalunian, Jacobsen, Peschken, Kamen, Askanase.

Analysis and interpretation of data. Clarke, Hanly, Urowitz, St. Pierre, Farewell.

ROLE OF THE STUDY SPONSOR

UCB Pharmaceuticals 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 UCB Pharmaceuticals.

REFERENCES

1. Unterman A, Nolte JE, Boaz M, et al. Neuropsychiatric syndromes in systemic lupus erythematosus: a meta-analysis. Semin Arthritis Rheum 2011;41:1–11.
2. Govoni M, Bortoluzzi A, Padovan M, et al. The diagnosis and clinical management of the neuropsychiatric manifestations of lupus. J Autoimmun 2016;74:41–72.

3. Schwartz N, Stock AD, Putterman C. Neuropsychiatric lupus: new mechanistic insights and future treatment directions. Nat Rev Rheumatol 2019;15:137–52.
4. Govoni M, Hanly JG. The management of neuropsychiatric lupus in the 21st century: still so many unmet needs? Rheumatology (Oxford) 2020;59 Suppl 5:v52–62.
5. Hanly JG, Urowitz MB, Su L, et al. Prospective analysis of neuropsychiatric events in an international disease inception cohort of patients with systemic lupus erythematosus. Ann Rheum Dis 2010;69: 529–35.
6. Hanly JG, Urowitz MB, Gordon C, et al. Neuropsychiatric events in systemic lupus erythematosus: a longitudinal analysis of outcomes in an international inception cohort using a multistate model approach. Ann Rheum Dis 2020;79:356–62.
7. Zirkzee EJ, Huizinga TW, Bollen EL, et al. Mortality in neuropsychiatric systemic lupus erythematosus (NPSLE). Lupus 2014;23:31–8.
8. Bell CF, Ajmera MR, Meyers J. An evaluation of costs associated with overall organ damage in patients with systemic lupus erythematosus in the United States. Lupus 2022;31:202–11.
9. Chiu YM, Chuang MT, Lang HC. Medical costs incurred by organ damage caused by active disease, comorbidities, and side effect of treatments in systemic lupus erythematosus patients: a Taiwan nationwide population-based study. Rheumatol Int 2016;36: 1507–14.
10. Furst DE, Clarke A, Fernandes AW, et al. Medical costs and healthcare resource use in patients with lupus nephritis and neuropsychiatric lupus in an insured population. J Med Econ 2013;16:500–9.
11. Zhu TY, Tam LS, Lee VW, et al. The impact of flare on disease costs of patients with systemic lupus erythematosus. Arthritis Rheum 2009; 61:1159–67.
12. Zhu TY, Tam LS, Lee VW, et al. Systemic lupus erythematosus with neuropsychiatric manifestation incurs high disease costs: a cost-of-illness study in Hong Kong. Rheumatology (Oxford) 2009;48:564–8.
13. Barber MR, Hanly JG, Su L, et al. Economic evaluation of lupus nephritis in the Systemic Lupus International Collaborating Clinics Inception Cohort using a multistate model approach. Arthritis Care Res (Hoboken) 2018;70:1294–302.
14. Barber MR, Hanly JG, Su L, et al. Economic evaluation of damage accrual in an international systemic lupus erythematosus inception cohort using a multistate model approach. Arthritis Care Res (Hoboken) 2020;72:1800–8.
15. 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.
16. Bombardier C, Gladman DD, Urowitz MB, et al. Derivation of the SLEDAI: a disease activity index for lupus patients. Arthritis Rheum 1992; 35:630–40.
17. Gladman D, Ginzler E, Goldsmith C, 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.
18. American College of Rheumatology Ad Hoc Committee on Neuropsychiatric Lupus Syndromes. The American College of Rheumatology nomenclature and case definitions for neuropsychiatric lupus syndromes. Arthritis Rheum 1999;42:599–608.
19. Butt P, Beirness D, Gliksman L, et al. Alcohol and health in Canada: a summary of evidence and guidelines for low-risk drinking. 2011. URL: https://www.uvic.ca/research/centres/cisur/assets/docs/report-alcohol-and-health-in-canada.pdf.
20. Zhang W, Bansback N, Boonen A, et al. Development of a composite questionnaire, the valuation of lost productivity, to value productivity losses: application in rheumatoid arthritis. Value Health 2012;15: 46–54.
21. Zhang W, Bansback N, Kopec J, et al. Measuring time input loss among patients with rheumatoid arthritis: validity and reliability of the Valuation of Lost Productivity questionnaire. J Occup Environ Med 2011;53:530–6.
22. Ainiala H, Hietaharju A, Loukkola J, et al. Validity of the new American College of Rheumatology criteria for neuropsychiatric lupus syndromes: a population-based evaluation. Arthritis Rheum 2001;45: 419–23.
23. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2018.
24. Jackson C. Multi-state models for panel data: the msm Package for R. J Stat Softw 2011;38:1–28.
25. Turcotte M, Wendt M, Houle P. General Social Survey, 2015 Cycle 29: Time Use - Public Use Microdata File Documentation and User's Guide. 2017.
26. US Bureau of Labor Statistics. American Time Use Survey (ATUS), table 3. 2017. URL: https://www.bls.gov/webapps/legacy/tustab3.htm.
27. OECD. OECD family database. URL: https://www.oecd.org/els/family/database.htm.
28. Statistics Korea. Time use survey - average time spent on activities by age group. 2014. URL: https://kostat.go.kr/portal/eng/pressReleases/11/6/index.board?bmode=read&aSeq=347182&pageNo=&rowNum=10&amSeq=&sTarget=&sTxt=.
29. Statistics Canada. Employee wages by industry, annual. Ontario; 2022. Table 14–10–0064–01. URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410006401.
30. Brand J, van Buuren S, le Cessie S, et al. Combining multiple imputation and bootstrap in the analysis of cost-effectiveness trial data. Stat Med 2019;30;38:210–20.
31. OECD. Purchasing power parities (PPP). 2022. URL: https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm.
32. Hanly JG, Gordon C, Bae SC, et al. Neuropsychiatric events in systemic lupus erythematosus: predictors of occurrence and resolution in a longitudinal analysis of an international inception cohort. Arthritis Rheumatol 2021;73:2293–302.
33. US Bureau of Labor Statistics. Consumer price index (R-CPI-U-RS). 1977–2021. URL: https://www.bls.gov/cpi/research-series/r-cpi-u-rs-home.htm.
34. Meacock R, Dale N, Harrison MJ. The humanistic and economic burden of systemic lupus erythematosus: a systematic review. Pharmacoeconomics 2013;31:49–61.

Arthritis Care & Research
Vol. 75, No. 9, September 2023, pp 1871–1877
DOI 10.1002/acr.25091

American College of Rheumatology
Empowering Rheumatology Professionals

Risk of End-Stage Renal Disease in Patients With Systemic Lupus Erythematosus and Diabetes Mellitus: A Danish Nationwide Cohort Study

Renata Baronaite Hansen,[1] Titilola Falasinnu,[2] Mikkel Faurschou,[3] Søren Jacobsen,[4] and Julia F. Simard[2]

Objective. The risk of end-stage renal disease (ESRD) is increased in patients with systemic lupus erythematosus (SLE). This study was undertaken to determine whether diabetes mellitus (DM) increases ESRD risk in a large inception cohort of SLE patients.

Methods. By means of the Danish National Patient Registry, we identified 3,178 adult patients diagnosed as having SLE between January 1, 1996, and July 31, 2018. DM was defined as the date of first hospital contact for DM or date of a first prescription of an antidiabetic drug. ESRD was defined as first registration of dialysis, renal transplant, or terminal renal insufficiency in the Danish National Patient Registry. ESRD incidence was compared between SLE patients with DM (SLE–DM) and those without DM (SLE–non-DM). Hazard ratios (HRs), adjusted for sex, age, educational level, and occupational status at baseline were calculated for sex, age, educational level, and hypertension (at baseline or during follow-up) strata. The overall hazard ratio (HR) was also adjusted for hypertension.

Results. The SLE–DM group included 290 patients, of whom 77% were female, compared with 85% of the 2,859 patients in the SLE–non-DM group. SLE–DM patients had a 3 times higher risk of ESRD compared with SLE–non-DM patients (multivariable-adjusted HR 3.3 [95% confidence interval 1.8–6.1]). In stratified multivariable-adjusted analyses, DM increased the rate of ESRD in women and men, patients ≥50 years old at baseline, those with low educational level at baseline, and those with concomitant hypertension.

Conclusion. Our findings indicate that SLE patients with DM have a markedly higher risk of developing ESRD compared with SLE patients without DM.

INTRODUCTION

End-stage renal disease (ESRD) is one of the most severe manifestations of systemic lupus erythematosus (SLE). Up to 40% of SLE patients develop chronic kidney disease (CKD) (lupus nephritis) over the course of the disease (1,2), and about 10% of those patients progress to ESRD (3,4). ESRD greatly reduces quality of life and increases both morbidity and mortality in SLE patients (5) as well as in the general population (6). The global prevalence of CKD is ∼10% in the general population (6–8), and in most countries the leading cause of ESRD is diabetes mellitus (DM), accounting for up to 50% of ERSD cases in the general population (9). Improved screening, glycemic control, and new therapies for DM have reduced the risk of diabetic kidney disease (diabetic nephropathy). However, DM remains a global problem with its incidence increasing and many people living undiagnosed (7).

SLE is associated with the development of DM (10–12). During a 5-year follow up, ∼10% of patients with incident SLE developed type 2 DM in a recent population-based study (13). Concurrent autoimmune diseases (including type 1 DM), metabolic syndrome, inflammation, and medications such as glucocorticoids may contribute to a higher prevalence of DM in SLE patients compared with the general population (14,15).

The pathogenic mechanisms of ESRD development in SLE and DM are different (16), and whether DM further increases the risk of ESRD in SLE has not yet been established. Another significant cause of ESRD in the general population is hypertension (8). Besides frequently coexisting with DM and synergistically

Supported by Bristol Myers Squibb. Dr. Jacobsen's work was supported by Bristol Myers Squibb.

[1]Renata Baronaite Hansen, MD: Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark, and Stanford University School of Medicine, Stanford, California; [2]Titilola Falasinnu, PhD, Julia F. Simard, ScD: Stanford University School of Medicine, Stanford, California; [3]Mikkel Faurschou, DMSci: Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark; [4]Søren Jacobsen, DMSci: Copenhagen University Hospital, Rigshospitalet and University of Copenhagen, Copenhagen, Denmark.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.25091&file=acr25091-sup-0001-Disclosureform.pdf.

Address correspondence via email to Renata Baronaite Hansen, MD, at renatabhansen@gmail.com.

Submitted for publication August 18, 2022; accepted in revised form January 24, 2023.

SIGNIFICANCE & INNOVATIONS

- Both diabetes mellitus (DM) and systemic lupus erythematosus (SLE) are associated with developing end-stage renal disease (ESRD).
- The risk of developing ESRD is 3 times higher in patients with SLE and DM compared with SLE patients without DM.
- DM increases the rate of ESRD in female and male SLE patients ≥50 years old at baseline, those with low educational level at baseline, and those with concomitant hypertension.
- Future studies are warranted to further explore how the risk of developing ESRD is influenced by lupus nephritis, DM, and hypertension in SLE patients.

worsening outcomes due to vascular disease, hypertension is also one of the most frequent comorbidities in SLE patients (10,11,17,18). In this study in a nationwide cohort of Danish SLE patients, we investigated whether DM is associated with an increased risk of ESRD and whether hypertension modifies this association.

PATIENTS AND METHODS

Data sources. In this population-based study, we used prospectively collected national registry data. The Danish Civil Registration System was established in 1968; it assigns a central person registration number, issued by law, at birth, or upon immigration to all persons residing in Denmark, and contains continuously updated information regarding vital status (19). The following registries were linked using central person registration numbers.

The Danish National Patient Registry contains International Classification of Diseases (ICD) codes of diagnoses and procedures recorded by physicians during discharges from inpatient wards (since 1977), emergency wards (since 1995), and outpatient hospital clinics (since 1995) (20,21). The data are organized as patient "contacts," wherein inpatient and emergency care contacts always contain only 1 visit and include a start and end date. However, individual outpatient contacts are initially assigned a start date and may contain 1 or several follow-up visits under the same primary diagnosis. For each patient care contact, ≤20 diagnoses can be registered, including a primary diagnosis and any secondary, optional diagnoses. ICD, Eighth Revision (ICD-8) codes were used from 1977 until 1993, followed by ICD, Tenth Revision (ICD-10) codes since 1994.

The Danish National Prescription Registry was established in 1995 and contains all prescription dispensing data from Danish community pharmacies (21). The Income Statistics Registry and the Population's Education Registry contain data concerning the occupational status (affiliation to the labor market) and educational status, respectively, of Danish residents (21).

Study population. We established a national SLE inception cohort comprising patients ≥18 years old with a first-time Danish National Patient Registry registration of SLE between January 1, 1996, and July 31, 2018. Registration codes used to define SLE included 73419 (ICD-8) as well as M32.1, M32.8, and M32.9 (ICD-10). We required ≥1 SLE ICD code (primary or secondary) registered at a rheumatology, nephrology, or dermatology department (inpatient or outpatient) or a primary code registered at any inpatient department. Baseline was the date when this definition was fulfilled. Using a previously reported data set for validation of SLE registration in the Danish National Patient Register (22), we identified 194 individuals fulfilling our case definition. Among these, 145 (75%) fulfilled the American College of Rheumatology classification criteria for SLE (23,24) and 160 (83%) had a physician-based clinical diagnosis of SLE.

Patients with prevalent ESRD at baseline were excluded (n = 29). The positive predictive value of registration of moderate-to-severe renal disease in the Danish National Patient Register is 100% (25).

Study variables and outcomes. *Outcome and follow-up.* We defined incident ESRD as the first registration of terminal (stage 5) kidney insufficiency, chronic dialysis, or renal transplant according to ICD diagnosis and procedure codes (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25091). Individuals contributed person-time until ESRD, death, emigration, or July 31, 2018, whichever occurred first.

Clinical factors: DM and hypertension. To define DM (both types 1 and 2) and hypertension, we used the first relevant ICD codes in the Danish National Patient Register or the date of the first filled prescription (available from 1995–2018) for these conditions, whichever came first (see Supplementary Table 1, at http://onlinelibrary.wiley.com/doi/10.1002/acr.25091). The primary exposure, DM, was a time-dependent factor. When categorized, DM was presented as prevalent at baseline, incident during follow-up, and no DM. The exposure variable (DM) was considered a risk factor for developing ESRD, and hypertension was regarded as a clinical factor that could be an effect modifier. In patients with incident DM, hypertension was considered "present" if it was prevalent at baseline or diagnosed prior to incident DM. In the time-dependent analyses, both DM and hypertension are covariates considered absent until they were diagnosed.

Parameters of socioeconomic status. The highest level of education achieved and occupational status at baseline were used as socioeconomic factors. Education was divided into 5 categories based on International Standard Classification of Education (ISCED) 2011 classification: 1) early childhood education, primary, and lower secondary education (ISCED levels 0–2);

2) general upper secondary education, high school (ISCED level 3); 3) vocational upper secondary education (ISCED level 4); 4) short- or medium-length tertiary education, bachelor, or equivalent (ISCED levels 5 and 6); and 5) long-length higher education or master's degree, doctorate, or equivalent (ISCED levels 7 and 8) (26).

For adjustment purposes, all 5 educational level categories were used. When stratifying, educational-level categories were dichotomized as primary and lower secondary education (1) compared with upper secondary and higher education (2–5). In the educational-level stratified analyses, patients with missing data regarding educational level were excluded. In the adjusted analyses, "missing" was considered to be a separate educational level category.

Occupational status was categorized as being 1) affiliated to the labor market, 2) under education, 3) retired, and 4) unemployed or on welfare. Information regarding occupational status was available through 2016; thus, for patients with baseline visit dates in 2017 or 2018, we carried forward their 2016 data.

Data and statistical analysis. *Incidence of ESRD.* We estimated the crude incidence rate (IR) of ESRD per 1,000 person-years as the number of ESRD cases divided by DM-exposed and DM-unexposed person-years at risk separately. In those with incident DM during follow-up, person-time between baseline and date of incident DM was classified as non-DM person-time. We calculated crude IR ratios (IRRs) and corresponding 95% confidence intervals (95% CIs), stratified by sex, age, educational level (at baseline), and hypertension, and we used tests of homogeneity to assess differences between strata.

Multivariable-adjusted analyses. Multivariable-adjusted Cox proportional hazards models were used to estimate hazard ratios (HRs) for ESRD relative to the time-dependent exposure of DM, adjusting for confounders that were determined based on a priori knowledge of the DM–ESRD association. The models were then stratified by sex, age at baseline (<50 years versus ≥50 years), educational level category at baseline, and hypertension (prevalent at baseline or incident prior to incident DM versus no hypertension). All HRs were adjusted for sex, age, educational level, and occupational status at baseline. Further, the overall HR was also adjusted for hypertension at baseline or prior to incident DM. Both DM and hypertension were time dependent in these analyses.

To evaluate if further development of hypertension could act as an effect modifier, multivariable-adjusted Cox modeling was extended to also include a time-dependent interaction term between DM and hypertension. To account for the possible competing risk of death (death is a competing risk for ESRD because ESRD can no longer occur if a patient dies), we used 2 different Cox proportional hazards models: cause-specific hazards and cumulative incidence (proportional subdistributional hazards) (27). Data analyses were performed using SAS software version 9.4.

Table 1. Characteristics of systemic lupus erythematosus (SLE) patients stratified by concomitant diabetes mellitus (DM)*

	DM (n = 290)	No DM (n = 2,859)
Female sex	224 (77.2)	2,441 (85.4)
Age, mean ± SD		
At SLE diagnosis	53.4 ± 14.8	46.2 ± 16.4
At DM diagnosis	54.3 ± 15.8	NA
Occupational status at baseline		
Affiliation to the labor market	107 (36.9)	1,378 (48.2)
Under education	3 (1.0)	198 (6.9)
Unemployed or on welfare	51 (17.6)	489 (17.1)
Retired	128 (44.1)	790 (27.6)
Unknown or missing data	1 (0.3)	4 (0.1)
Highest educational level at baseline		
Primary and lower secondary	114 (39.3)	844 (29.5)
General upper secondary (high school)	14 (4.8)	238 (8.3)
Vocational upper secondary	92 (31.7)	889 (31.1)
Short- or medium-length tertiary	42 (14.5)	640 (22.4)
Long-length higher (master's or doctorate)	15 (5.2)	143 (5.0)
Unknown or missing data	13 (4.5)	105 (3.7)
Hypertension		
At baseline	105 (36.2)	668 (23.4)
During follow-up†	54 (18.6)	1,215 (42.5)
At baseline or during follow-up	159 (54.8)	1,883 (65.9)
ESRD during follow-up	13	63

* Except where indicated otherwise, values are the number (%) of patients. NA = not applicable.
† Prior to incident DM (excludes hypertension following incident DM); in patients with end-stage renal disease (ESRD), incident before ESRD.

Ethics and approvals. The study was approved by the Danish Data Protection Agency (approval no. VD-2018-175). Informed participant consent was not required due to the anonymized, registry-based data retrieval provided by Statistics Denmark.

RESULTS

Baseline characteristics. In this cohort of SLE patients, 290 had DM at baseline or developed DM during follow-up, whereas 2,859 did not have DM. Those with DM were more likely to be men (23% versus 15%) and older (mean age at baseline 53 versus 46 years) compared with the group of SLE patients without DM (Table 1). Among the SLE patients with DM, 119 (41%) already had a DM diagnosis at baseline. In the remaining 171 SLE patients (59%), DM presented during follow-up a mean ± SD of 6.4 ± 5.8 years after baseline. Hypertension was more common at baseline in those with DM compared with patients who did not have DM (36% versus 23%).

Incidence of ESRD. There were 13 cases of incident ESRD during 1,672 person-years at risk in the SLE–DM group (7.8 cases per 1,000 person-years) and 63 ESRD cases during

Table 2. Measures of occurrence and association of ESRD (per 1,000 person-years) in SLE patients stratified by concomitant DM and by age, sex, educational level, and hypertension*

	DM			No DM				
	No. ESRD/ total no.	Person-years	IR (95% CI)	No. ESRD/ total no.	Person-years	IR (95% CI)	IRR (95% CI)	P†
Age at baseline								0.405
≥50 years	8/172	843	9.5 (4.7–19.0)	22/1,181	9,824	2.2 (1.5–3.4)	4.2 (1.9–9.5)	
<50 years	5/118	829	6.0 (2.5–14.5)	41/1,678	17,081	2.4 (1.8–3.3)	2.5 (1.0–6.4)	
Sex								0.612
Male	6/66	324	18.5 (8.3–41.2)	17/418	3,350	5.1 (3.1–8.2)	3.6 (1.4–9.3)	
Female	7/224	1,348	5.2 (2.5–10.9)	46/2,441	23,546	1.9 (1.5–2.6)	2.7 (1.2–5.9)	
Educational level at baseline‡								0.278
Primary and lower secondary	8/114	595	13.4 (6.7–26.9)	28/844	8,457	3.3 (2.3–4.8)	4.1 (1.8–8.9)	
Upper secondary and higher	4/163	1,043	3.8 (1.4–10.2)	34/1,910	17,619	1.9 (1.4–2.7)	2.0 (0.7–5.6)	
Hypertension at baseline or prior to incident DM								0.048
Yes	7/159	618	11.3 (5.4–23.8)	36/1,883	18,378	2.0 (1.4–2.7)	5.8 (2.6–13.0)	
No	6/131	1,066	5.7 (2.5–12.7)	27/976	8,528	3.2 (2.2–4.6)	1.8 (0.7-4.4)	

* 95% CI = 95% confidence interval; DM = diabetes mellitus; ESRD = end-stage renal disease; IR = incidence rate; SLE = systemic lupus erythematosus.
† *P* value for the test of homogeneity of incidence rate ratio (IRR) in each stratum.
‡ Patients with unknown or missing education level were excluded (n = 118).

26,905 person-years at risk in the SLE–non-DM group (2.3 cases per 1,000 person-years). The overall ESRD incidence rate among SLE patients was 2.7 (95% CI 2.1–3.3) cases per 1,000 person-years. ESRD occurred more frequently in SLE patients with DM compared with those without DM, irrespective of stratification for age, sex, or education (Table 2). At baseline or before incident DM, the unadjusted association between DM and ESRD was higher in SLE patients with hypertension (IRR 5.8 [95% confidence interval (95% CI) 2.6–13.0]) compared with those without hypertension (IRR 1.8 [95% CI 0.74–4.35]) (Table 2).

Table 3. Hazard ratios (HRs) of ESRD in SLE patients with DM, stratified according to age at baseline, sex, education category at baseline, and hypertension*

	HR	95% CI
Overall, multivariable-adjusted†	3.3	1.8–6.2
Overall, including adjustment for hypertension†	3.3	1.8–6.1
Age at baseline‡		
≥50 years	4.0	1.8–9.0
<50 years	2.6	0.99–6.7
Sex§		
Male	3.3	1.2–8.9
Female	3.2	1.4–7.2
Educational level at baseline¶		
Primary and lower secondary	3.8	1.7–8.4
Upper secondary and higher	2.2	0.8–6.5
Hypertension at baseline or prior to incident DM#		
Yes	4.9	2.1–11.3
No	2.2	0.9–5.9

* Diabetes mellitus (DM) is time dependent; hypertension is time dependent when adjusting. 95% CI = 95% confidence interval; ESRD = end-stage renal disease; SLE = systemic lupus erythematosus.
† Adjusted for sex, age, education, and occupational status at baseline.
‡ Adjusted for sex, education, and occupational status at baseline.
§ Adjusted for age, education, and occupational status at baseline.
¶ Patients with unknown or missing educational level excluded (n = 118). Adjusted for sex, age, and occupational status at baseline.
# Hypertension was not time dependent in this stratum. Adjusted for sex, age, education, and occupational status at baseline.

Multivariable-adjusted analyses. SLE patients with DM had a 3-fold higher rate of ESRD compared with SLE patients without DM (multivariable [sex, age, hypertension, education, and occupational status at baseline] adjusted HR 3.3 [95% CI 1.8–6.2]) (Table 3). In stratified, multivariable-adjusted Cox proportional hazards models, DM increased the rate of ESRD in women and men, individuals in the older age group at baseline, those with low educational level at baseline, and those with hypertension at baseline or during follow-up (incident prior to incident DM) (Table 3). Although the highest incidence rate for ESRD was found in SLE patients with both DM and a history of hypertension (Table 2), no time-dependent interaction between DM and hypertension was observed in the adjusted Cox proportional hazards models (HR 0.6 [95% CI 0.2–1.9]).

ESRD was associated with DM in SLE patients using either a model of cause-specific hazard by treating competing risks (death) as a censored observation (HR 2.9 [95% CI 1.6–5.3]) or a model based on proportional subdistribution hazard ratios (cumulative incidence; HR 2.4 [95% CI 1.3–4.3]). Thus, the observed increased incidence of ESRD in SLE–DM patients could not be explained by competing risks.

DISCUSSION

In this nationwide cohort of 3,178 patients with incident SLE, we found that the SLE patients with DM developed ESRD at a 3-fold higher rate compared with those without DM. It is well known that DM and autoimmune diseases, such as SLE, are risk

factors for developing CKD and ESRD (1,3,4,28). Few studies have shown an increased risk of ESRD in SLE patients with DM. In a cohort study of 601 adults with biopsy-proven glomerulonephritis, which included individuals with lupus nephritis (21%), DM was independently associated with a 2-fold greater risk of ESRD during a median of 39 months of follow-up (29). In a study of 1,317 patients with SLE and incident DM, matched with 1,317 patients with SLE and without DM, the crude HR for ESRD by DM was 2.7; the HR was reduced to 1.6 when adjusted for confounders such as age, sex, and comorbidities, including hypertension (30). Thus, the slightly varying estimates of ESRD risk by DM could partly be attributable to varying definitions and incidences of DM and hypertension in the reported cohorts.

In our study, the rate of ESRD was highest among SLE–DM patients in the lowest educational stratum, which in part could be related to lower educational level being reported as a determinant of nonadherence to antimalarials in SLE (31). Adherence to antimalarials in SLE has been found to be protective of type 2 DM (13) as well as renal damage (32).

The highest incidence rate of ESRD was observed in male SLE–DM patients, but the multivariable-adjusted HRs were nearly equivalent in women and men. Although male sex in some studies has been associated with poor renal outcome in SLE (33,34), this is not a general finding (35) and in diabetic nephropathy, the risk of ESRD does not appear to be sex dependent (36).

We also found that DM was associated with ESRD in the older group at baseline. Whether this is due to a longer period of DM exposure, even before SLE diagnosis, or other factors, is unknown. The durations of DM and hypertension—prior to or following SLE diagnosis—are also likely to influence the risk of developing lupus nephritis and ESRD. Our study design could not address this accurately since the start of follow-up begins with the SLE case definition and not DM/hypertension definition.

The most frequent cause of kidney damage in SLE is lupus nephritis (glomerulonephritis). In a recent review, 7–31% of patients had lupus nephritis at SLE diagnosis; 31–48% developed lupus nephritis after SLE diagnosis, most within 5 years (37). Less frequent causes of ESRD in SLE include thrombotic microangiopathy/antiphospholipid nephropathy, non–immune complex podocytopathy, tubulointerstitial nephritis, acute tubular necrosis, renovascular disease, or nephrotoxicity from medications (38). Recently, a case of coexistence of diabetic nephropathy and lupus-related renal manifestation, lupus podocytopathy, was reported (39).

Patients with lupus nephritis often receive aggressive treatment with immunosuppressive medications and high doses of glucocorticoids. It is well known that glucocorticoid use is associated with various complications, including DM (40). Thus, in some SLE patients in our study, DM might have developed as a consequence of lupus nephritis treatment and not been related to the development of ESRD. Some ESRD cases in our study could also have been caused or exacerbated by other, less frequent, lupus-related renal manifestations, as mentioned above.

In this study, we aimed to evaluate the association between DM and ESRD in SLE in general. To what extent DM is an independent risk factor or has a synergistic effect with lupus nephritis or other possible causes of ESRD, leading to a progression from CKD to ESRD is yet to be investigated. Due to the register-based nature of the study, we were not able specifically to address relevant clinical descriptors of lupus nephritis. Future studies evaluating the potential influence of renal changes as determined by histologic changes on the association between diabetes and hypertension in SLE patients are warranted.

In our study, concomitant hypertension was associated with a higher risk of ESRD in SLE–DM patients compared with SLE–non-DM patients. Hypertension is currently the second leading cause of ESRD in the general population (following DM) (6) and also a potential adverse event in glucocorticoid use (40). Some studies report an additional effect of DM (diabetic nephropathy) and hypertension (hypertensive nephrosclerosis) on the risk of developing ESRD in the general population, due to effects on the macrovasculature and microvasculature (8,18). In our study, adjustment for hypertension did not change the HR for development of ESRD, nor did we find an interaction between DM and hypertension on the risk of ESRD in SLE patients.

We did not have access to any blood pressure measurements, laboratory results, or kidney biopsy results in our registry-based study and were thus not able to assess effectiveness of DM and hypertension management or specify the histologic patterns in the kidneys. African ancestry is associated with a markedly increased risk of ESRD in SLE patients (41,42), hence evaluating race as a possible effect modifier may have been of interest. However, this ancestral group constitutes only 1% of SLE patients in Denmark (43), and therefore our study was not sufficiently powered for such an analysis.

Strengths of this study include the large inception cohort of SLE patients with prospectively collected, population-based longitudinal data, including inpatient and outpatient diagnoses, and medication data. The positive predictive value of registration of moderate/severe renal disease and DM in the Danish National Patient Register is high (100% and 96%, respectively) (25). Use of medication prescription data in addition to the ICD codes reduced misclassification in defining DM and hypertension.

In our study we found that DM is a significant risk factor for ESRD in SLE. Whether it is independent or related to specific SLE phenotypes, other autoimmunity, or treatment remains to be investigated. Our study also emphasizes the need for the assessment and treatment of DM and hypertension in the care of patients with SLE. Further studies are warranted to explore how the risk of ESRD is influenced by glomerulonephritis, DM, and hypertension in SLE patients.

ACKNOWLEDGMENTS

We would like to thank Bristol Myers Squibb and Peloton Advantage, LLC, an OPEN Health company, for providing professional editing assistance.

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. Hansen 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. Hansen, Falasinnu, Jacobsen, Simard.
Acquisition of data. Hansen.
Analysis and interpretation of data. Hansen, Falasinnu, Faurschou, Jacobsen, Simard.

ROLE OF THE STUDY SPONSOR

Bristol Myers Squibb funded this trial and reviewed and approved the manuscript prior to submission. The authors independently collected the data, interpreted the results, and had the final decision to submit the manuscript for publication. Publication of this article was not contingent upon approval by Bristol Myers Squibb.

REFERENCES

1. Hanly JG, O'Keeffe AG, Su L, et al. The frequency and outcome of lupus nephritis: results from an international inception cohort study. Rheumatology (Oxford) 2016;55:252–62.
2. Fanouriakis A, Kostopoulou M, Cheema K, et al. 2019 Update of the Joint European League Against Rheumatism and European Renal Association-European Dialysis and Transplant Association (EULAR/ERA-EDTA) recommendations for the management of lupus nephritis. Ann Rheum Dis 2020;79:713–23.
3. Tektonidou MG, Dasgupta A, Ward MM. Risk of end-stage renal disease in patients with lupus nephritis, 1971–2015: a systematic review and Bayesian meta-analysis. Arthritis Rheumatol 2016;68:1432–41.
4. Barber MR, Hanly JG, Su L, et al. Economic evaluation of lupus nephritis in the Systemic Lupus International Collaborating Clinics inception cohort using a multistate model approach. Arthritis Care Res (Hoboken) 2018;70:1294–302.
5. Gasparotto M, Gatto M, Binda V, et al. Lupus nephritis: clinical presentations and outcomes in the 21st century. Rheumatology (Oxford) 2020;59 Suppl 5:v39–51.
6. NIH/National Institute of Diabetes and Digestive and Kidney Diseases. 2020 US Renal Data System annual data report: epidemiology of kidney disease in the United States. URL: https://usrds-adr.niddk.nih.gov/2020.
7. Hallan SI, Coresh J, Astor BC, et al. International comparison of the relationship of chronic kidney disease prevalence and ESRD risk. J Am Soc Nephrol 2006;17:2275–84.
8. GBD Chronic Kidney Disease Collaboration. Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2020; 395:709–33.
9. Tuttle KR, Bakris GL, Bilous RW, et al. Diabetic kidney disease: a report from an ADA Consensus Conference. Diabetes Care 2014; 37:2864–83.
10. Hansen RB, Simard JF, Faurschou M, et al. Distinct patterns of comorbidity prior to diagnosis of incident systemic lupus erythematosus in the Danish population. J Autoimmun 2021;123:102692.
11. Kuo CF, Chou IJ, Rees F, et al. Temporal relationships between systemic lupus erythematosus and comorbidities. Rheumatology (Oxford) 2019;58:840–8.
12. Bruce IN, Urowitz MB, Gladman DD, et al. Risk factors for coronary heart disease in women with systemic lupus erythematosus: the Toronto Risk Factor Study. Arthritis Rheum 2003;48:3159–67.
13. Salmasi S, Sayre EC, Antonio Avina-Zubieta J, et al. 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;73:702–6.
14. Tsokos GC. Autoimmunity and organ damage in systemic lupus erythematosus. Nat Immunol 2020;21:605–14.
15. Nikpour M, Gladman DD, Urowitz MB. Premature coronary heart disease in systemic lupus erythematosus: what risk factors do we understand? Lupus 2013;22:1243–50.
16. Falasinnu T, O'Shaughnessy MM, Troxell ML, et al. A review of non-immune mediated kidney disease in systemic lupus erythematosus: a hypothetical model of putative risk factors. Semin Arthritis Rheum 2020;50:463–72.
17. Gandelman JS, Khan OA, Shuey MM, et al. Increased incidence of resistant hypertension in patients with systemic lupus erythematosus: a retrospective cohort study. Arthritis Care Res (Hoboken) 2020;72: 534–43.
18. Climie RE, van Sloten TT, Bruno RM, et al. Macrovasculature and microvasculature at the crossroads between type 2 diabetes mellitus and hypertension. Hypertension 2019;73:1138–49.
19. Schmidt M, Pedersen L, Sorensen HT. The Danish Civil Registration System as a tool in epidemiology. Eur J Epidemiol 2014;29:541–9.
20. Schmidt M, Schmidt SA, Sandegaard JL, et al. The Danish National Patient Registry: a review of content, data quality, and research potential. Clin Epidemiol 2015;7:449–90.
21. Schmidt M, Schmidt SA, Adelborg K, et al. The Danish health care system and epidemiological research: from health care contacts to database records. Clin Epidemiol 2019;11:563–91.
22. Hermansen ML, Lindhardsen J, Torp-Pedersen C, et al. Incidence of systemic lupus erythematosus and lupus nephritis in Denmark: a nationwide cohort study. J Rheumatol 2016;43:1335–9.
23. Tan EM, Cohen AS, Fries JF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1982;25:1271–7.
24. 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.
25. Thygesen SK, Christiansen CF, Christensen S, et al. The predictive value of ICD-10 diagnostic coding used to assess Charlson comorbidity index conditions in the population-based Danish National Registry of Patients. BMC Med Res Methodol 2011;11:83.
26. Jensen VM, Rasmussen AW. Danish Education Registers. Scand J Public Health 2011;39 Suppl:91–4.
27. Andersen PK, Geskus RB, de Witte T, et al. Competing risks in epidemiology: possibilities and pitfalls. Int J Epidemiol 2012;41:861–70.
28. Tuttle KR, Bakris GL, Bilous RW, et al. Diabetic kidney disease: a report from an ADA Consensus Conference. Am J Kidney Dis 2014; 64:510–33.
29. Lim CC, Choo JC, Tan HZ, et al. Changes in metabolic parameters and adverse kidney and cardiovascular events during glomerulonephritis and renal vasculitis treatment in patients with and without diabetes mellitus. Kidney Res Clin Pract 2021;40:250–62.
30. Jiang MY, Hwang JC, Feng IJ. Impact of diabetes mellitus on the risk of end-stage renal disease in patients with systemic lupus erythematosus. Sci Rep 2018;8:6008.
31. Mehat P, Atiquzzaman M, Esdaile JM, et al. Medication nonadherence in systemic lupus erythematosus: a systematic review. Arthritis Care Res (Hoboken) 2017;69:1706–13.
32. Pons-Estel GJ, Alarcon GS, McGwin G Jr, et al. Protective effect of hydroxychloroquine on renal damage in patients with lupus nephritis: LXV, data from a multiethnic US cohort. Arthritis Rheum 2009;61: 830–9.

33. Riveros Frutos A, Casas I, Rua-Figueroa I, et al. Systemic lupus erythematosus in Spanish males: a study of the Spanish Rheumatology Society Lupus Registry (RELESSER) cohort. Lupus 2017;26:698–706.
34. Tan TC, Fang H, Magder LS, et al. Differences between male and female systemic lupus erythematosus in a multiethnic population. J Rheumatol 2012;39:759–69.
35. Faurschou M, Dreyer L, Kamper AL, et al. Long-term mortality and renal outcome in a cohort of 100 patients with lupus nephritis. Arthritis Care Res (Hoboken) 2010;62:873–80.
36. Wang Y, Zhang J, Zhang J, et al. Sex differences in biopsy-confirmed diabetic kidney disease. Front Endocrinol (Lausanne) 2021;12: 670674.
37. Mahajan A, Amelio J, Gairy K, et al. Systemic lupus erythematosus, lupus nephritis and end-stage renal disease: a pragmatic review mapping disease severity and progression. Lupus 2020;29:1011–20.
38. Parikh SV, Almaani S, Brodsky S, et al. Update on lupus nephritis: core curriculum 2020. Am J Kidney Dis 2020;76:265–81.
39. Liu L, Murray B, Tomaszewski JE. Lupus podocytopathy superimposed on diabetic glomerulosclerosis: a case report. Medicine (Baltimore) 2021;100:e27077.
40. Rice JB, White AG, Scarpati LM, et al. Long-term systemic corticosteroid exposure: a systematic literature review. Clin Ther 2017;39: 2216–29.
41. Petri M, Barr E, Magder LS. Risk of renal failure within 10 or 20 years of systemic lupus erythematosus diagnosis. J Rheumatol 2021;48: 222–7.
42. Lewis MJ, Jawad AS. The effect of ethnicity and genetic ancestry on the epidemiology, clinical features and outcome of systemic lupus erythematosus. Rheumatology (Oxford) 2017;56 Suppl 1:i67–77.
43. Leffers HC, Troldborg A, Voss A, et al. Smoking associates with distinct clinical phenotypes in patients with systemic lupus erythematosus: a nationwide Danish cross-sectional study. Lupus Sci Med 2021;8:e000474.

Arthritis Care & Research
Vol. 75, No. 9, September 2023, pp 1878–1885
DOI 10.1002/acr.25094

American College of Rheumatology
Empowering Rheumatology Professionals

Effect of Systemic Lupus Erythematosus and Immunosuppressive Agents on COVID-19 Vaccination Antibody Response

Michelle Petri,[1] Daniel Joyce,[1] Kristin Haag,[1] Andrea Fava,[1] Daniel W. Goldman,[1] Diana Zhong,[1] Shaoming Xiao,[1] Aaron Milstone,[1] and Laurence S. Magder[2]

Objective. The risk of COVID-19 infection is increased in patients with systemic lupus erythematosus (SLE) versus those without SLE. Some immunosuppressive medications increase COVID-19 infection and decrease the efficacy of vaccination. Consensus documents have suggested management strategies for handling immunosuppressive medications to increase vaccine efficacy, but the benefit of such strategies has not been proven. The current study was undertaken to determine the effect of immunosuppressive drugs on vaccine response in SLE.

Methods. We collected information on COVID-19 infection, vaccination history, and COVID-19 antibodies in the Hopkins Lupus Cohort. A cohort of health care workers was used for comparison. Outcome measures included SARS–CoV-2 antibody IgG levels after vaccination over time in both cohorts and effect of immunosuppressive medications on postvaccination IgG levels in SLE patients.

Results. The analysis was based on 365 observations from 334 different patients in the SLE cohort, and 2,235 observations from 1,887 different health care workers. SLE patients taking immunosuppressive medications had lower vaccine IgG levels than SLE patients who were not; but both groups had lower levels than health care workers. Holding mycophenolate for 1 week after vaccination increased postvaccine IgG levels significantly without leading to clinical flares. In multiple variable models, mycophenolate mofetil, tacrolimus, and belimumab all significantly reduced antibody response to vaccination.

Conclusion. SLE patients, regardless of background immunosuppressive therapy, had lower vaccine IgG levels than health care workers. Mycophenolate, tacrolimus, and belimumab significantly reduced IgG response to vaccination. Holding mycophenolate for 1 week improved vaccine efficacy, providing clinical benefit on vaccine response without leading to clinical flares.

INTRODUCTION

Both systemic lupus erythematosus (SLE) itself and the use of some immunosuppressants may impact the risk of COVID-19 (1–3). In particular, corticosteroids and B cell–depleting biologics such as rituximab impair vaccine response (4–6). SLE increases the risk of severe COVID-19 (hospitalization, intensive care unit stay, or intubation) both before and after vaccination (7). In some studies, immunosuppressant therapies have been grouped together without sufficient data to separate out different therapies (or within therapies, the dose response) on vaccine response (8,9). The American College of Rheumatology (ACR) guidance on management of immunosuppressant therapies during COVID-19 vaccination was based on expert opinion due to limited data (10). In the Hopkins Lupus Center, we adopted a policy of holding mycophenolate mofetil and azathioprine on the day of and for 1 week following each COVID-19 vaccination and of not giving methotrexate the week after COVID-19 vaccination. Tacrolimus and belimumab were not held. The Hopkins Lupus Cohort's structured visit design allowed us to ascertain the

Supported in part by the NIH (National Institute of Allergy and Infectious Diseases grants T32-AI-007291 to Dr. Zhong and K24-AI-141580 to Dr. Milstone, and National Institute of Arthritis and Musculoskeletal and Skin Diseases grant R01-AR-069572 to Dr. Petri) and the Maren Foundation (grant to Dr. Petri).

[1]Michelle Petri, MD, MPH, Daniel Joyce, BA, Kristin Haag, BA, Andrea Fava, MD, Daniel W. Goldman, PhD, MCS, Diana Zhong, MD, Shaoming Xiao, MS, Aaron Milstone, MD, MHS: Johns Hopkins University School of Medicine, Baltimore, Maryland; [2]Laurence S. Magder, PhD, MPH: University of Baltimore School of Medicine, Baltimore, Maryland.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.25094&file=acr25094-sup-0001-Disclosureform.pdf.

Address correspondence via email to Michelle Petri, MD, MPH, at mpetri@jhmi.edu.

Submitted for publication September 5, 2022; accepted in revised form January 26, 2023.

SIGNIFICANCE & INNOVATIONS

- This is the first multiple variable analysis of the effect of immunosuppressive treatments and doses on COVID-19 vaccine response.
- Mycophenolate (>1,000 mg/day), tacrolimus, and belimumab are the immunosuppressive drugs that have the most impact on response to vaccination.

effectiveness on IgG response to COVID-19 vaccination of holding mycophenolate mofetil, azathioprine, and methotrexate, per the above protocol.

MATERIALS AND METHODS

The Hopkins Lupus Cohort has been approved by the Johns Hopkins University School of Medicine Institutional Review Board on a yearly basis. Patients with SLE in the Hopkins Lupus Cohort gave written informed consent and are seen at quarterly visits during which information on SLE activity, medications, and immunologic measures are recorded. Patients self-reported race by selection from the National Institutes of Health set of categories. For this study, additional data were collected on COVID-19 history and vaccination status prior to May 2021, and SARS–CoV-2 antibodies were assessed. A positive SARS–CoV-2 RNA test result was required to confirm COVID-19 infection. Dates of vaccination and vaccine type were recorded. All patients met the ACR (11) and/or Systemic Lupus International Collaborating Clinics criteria (12) for SLE. Lupus Activity Index scores (0–3 visual acuity scale) (13), Safety of Estrogens in Lupus Erythematosus National Assessment (SELENA) version of the Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) scores (14), physician global assessment of disease activity (PhGA) scores, and laboratory values were collected at each visit. All SLE patients received 2 doses of an mRNA COVID-19 vaccine (Pfizer or Moderna), the standard of care at that time.

Beginning in June 2020, 3,015 health care workers at 5 regional hospitals in the Johns Hopkins Health System consented to participate and were enrolled in a prospective cohort study to determine the seroprevalence of spike antibodies to SARS–CoV-2 (15). This study was approved by the Johns Hopkins University Institutional Review Board, and written informed consent was obtained from all participants. Participants provided serum samples and completed surveys (including providing demographic data and exposures) every 3–4 months after enrollment. SARS–CoV-2 polymerase chain reaction testing results and immunization data were collected from electronic health records. Health care workers who participated in a study visit between March 10 and April 8, 2021, were included in this analysis if their serum sample was collected ≥14 days after receiving dose 2 of either mRNA vaccine. This control cohort has been previously published (15,16).

Serum specimens from both cohorts were tested using an enzyme-linked immunosorbent assay (Euroimmun) that targets the S1 subunit of the SARS–CoV-2 spike protein and measures optical density ratios. We applied an internally derived IgG cutoff ratio (>1.23) for greater sensitivity and specificity with an upper threshold of 11 based on assay saturation (17).

Data analysis approach. This analysis was based only on those subjects who received 2 doses of Moderna or Pfizer vaccinations and excluded those observations that occurred after an infection with COVID-19. To estimate the mean level of antibody by the time since the second vaccination by cohort and immunosuppression history, we used loess smoothing. To assess the statistical significance of differences between the lupus and health-worker cohorts with respect to mean antibody levels, we fit a longitudinal regression model allowing for a difference between the groups at baseline (14 days post vaccination) and a difference between the groups with respect to the degree of decline in antibodies over time. To assess the relationship between SLE patient characteristics (disease activity, treatments) and the magnitude of antibody response in the SLE cohort, we used regression models, adjusting for time since vaccination and time squared. All regression models were fit using generalized estimating equations to account for the fact that some participants provided >1 antibody measure. To assess the effect of withholding immunosuppressants at the time of vaccination on

Table 1. Demographic characteristics of the Hopkins Lupus Cohort (HLC) and Hopkins Health Care Workers (HCW) cohorts*

Variable	HCW (n = 1,887)	HLC (n = 342)
No. of observations	2,235	342
Sex		
Female	1,511 (80)	308 (92)
Male	376 (20)	26 (8)
Race		
White	1,505 (90)	171 (51)
African American	98 (5)	113 (34)
Asian	240 (13)	38 (11)
Other	44 (2)	12 (4)
Age, years		
<30	288 (15)	21 (6)
30–44	865 (46)	96 (29)
45–59	534 (28)	118 (36)
60+	200 (11)	96 (29)
Vaccine		
Pfizer	1,530 (81)	196 (59)
Moderna	357 (18)	138 (41)
Days since vaccination		
14–59	948 (42)	17 (5)
60–119	864 (39)	110 (30)
120–179	345 (15)	176 (48)
180+	78 (3)	62 (17)

* Values are the number (%) unless indicated otherwise.

disease activity, only patients who were taking immunosuppressants at the time of their first vaccination, had a cohort visit within 60 days prior to vaccination, and had a cohort visit within 60 days after vaccination were included. A PhGA flare was defined as an increase of ≥1.0 points. A SELENA flare was defined as in Petri et al (14). Analyses were performed using SAS 9.4 statistical software.

RESULTS

Comparison of the SLE cohort with the Health Care Worker cohort with respect to IgG levels after vaccination. The analysis was based on 365 observations from 334 different patients in the SLE cohort and 2,235 observations from 1,887 different health care workers. Table 1 shows the

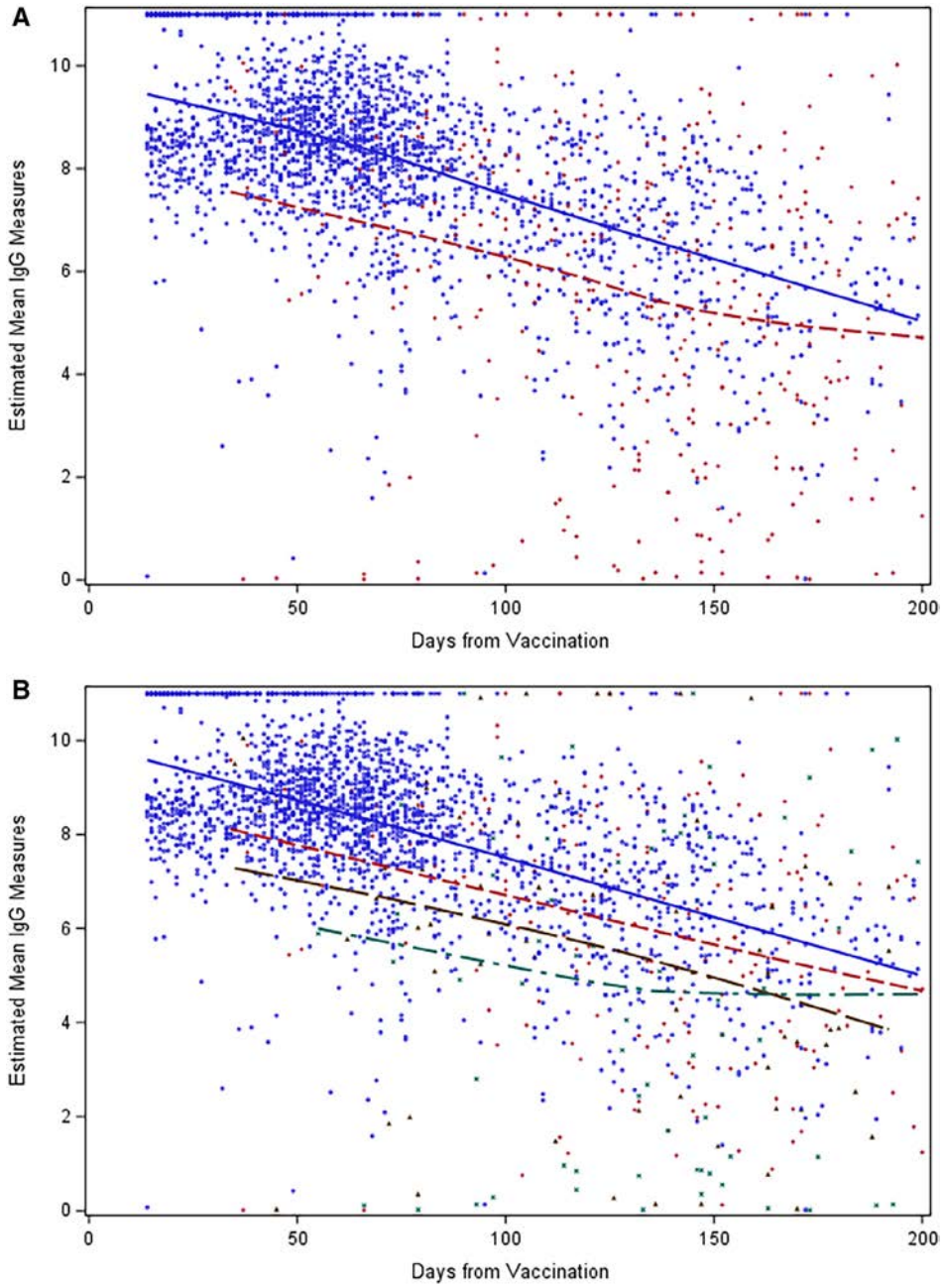

Figure 1. Mean SARS–CoV-2 spike protein IgG antibody level by days since the second vaccination for each cohort. **A**, Estimated mean SARS–CoV-2 spike protein IgG measure by days since the second vaccination, by cohort, for patients without systemic lupus erythematosus (SLE) (solid blue line) and patients with SLE (broken red line). **B**, Estimated mean SARS–CoV-2 spike protein IgG measurement over time since second vaccination, by cohort and use of immunosuppressants, for patients with SLE not taking immunosuppressants (broken red line), patients with SLE taking immunosuppressants (broken green line), patients with SLE with immunosuppressants held (broken black line), and patients without SLE (solid blue line). Each dot represents the antibody level for each subject.

Table 2. Mean SARS–CoV-2 spike protein IgG levels by time since vaccination and cohort*

Cohort	Antibody measurement by days since vaccination				*P* for difference between groups on day 14†	*P* for difference between groups with respect to change over time†
	14–59 days	60–119 days	120–179 days	180+ days		
Hopkins Lupus cohort	7.4 ± 3.2	6.5 ± 2.8	5.1 ± 3.0	4.8 ± 2.5		
Health Care Workers cohort	8.9 ± 1.4	8.1 ± 1.5	6.3 ± 1.8	5.0 ± 2.1	<0.0001	0.0033

* Values are the mean ± SD unless indicated otherwise.
† *P* values were based on a regression model allowing for a different baseline and slope in the 2 groups, adjusting for age, race, and sex, and were fit by generalized estimating equations to account for the fact that some participants contributed >1 antibody measure.

demographic comparison of the previously published Health Care Worker cohort (15) and the Hopkins Lupus Cohort. As expected, the SLE cohort had a higher frequency of female and African American patients.

Figure 1 shows the estimated mean SARS–CoV-2 IgG levels over time since the second mRNA vaccination by cohort. Patients with SLE had lower average levels shortly after vaccination, but over time, the 2 group means appeared to converge. The differences shortly after vaccination and the different slope of the lines were statistically significant (Table 2).

The relationship between SLE patient characteristics (disease activity, treatments) and the magnitude of antibody response. Table 3 shows the effects of demographic characteristics and medications at the time of vaccination on SARS–CoV-2 IgG levels. Men had a significantly lower mean antibody level. Taking prednisone at the time of vaccination (≥10 mg/day) was associated with a significant reduction in mean antibody levels. Mycophenolate mofetil, tacrolimus, and antihypertensive use were all associated with lower vaccine response (both mean IgG and positive/negative response). Holding mycophenolate mofetil for just 1 week significantly improved vaccine response. Holding azathioprine and methotrexate had no benefit. Belimumab reduced vaccine response.

To assess the effect of withholding immunosuppressants on disease activity, change from before the first vaccination to after the second vaccination was compared in those for whom immunosuppressants were held versus those for whom immunosuppressants were not held. There was no significant difference in change in SLEDAI score (mean ± SD 0.24 ± 3.56 versus 0.51 ± 2.96; $P = 0.72$), in PhGA (mean ± SD 0.12 ± 0.71 versus –0.05 ± 2.96; $P = 0.26$), in number of mild-to-moderate SELENA flares ($P = 0.84$), or in number of PhGA flares ($P = 0.66$).

We examined whether SLE activity around the time of vaccination was associated with lower antibody response. This analysis was based on 257 patients who had a cohort visit within 45 days of their vaccination (the mean number of days from vaccination was 20). A higher PhGA score around the time of vaccination (but not SLEDAI score) was associated with lower SARS–CoV-2 IgG levels and positive/negative response ($P = 0.023$ and 0.080, respectively) (Table 4). We also examined whether SLE activity at the time of the antibody measure was associated with antibody levels based on 335 antibody measures within 7 days of a cohort visit. Again, a high PhGA score was associated with lower measured antibody levels (Table 4).

Table 5 presents the multivariable model of the effect of immunosuppressive medication on postvaccination SARS–CoV-2 IgG levels, allowing for the fact that patients could be taking >1 medication. A mycophenolate mofetil dose of >1,000 mg, tacrolimus, and belimumab remained highly statistically significant in terms of reduction in SARS–CoV-2 IgG levels.

DISCUSSION

We used a previously published health care worker cohort as our control group (15). This large, well-characterized control cohort allowed us to create Figure 1 with great precision (compared to previous studies with very small control groups [9]). This clearly showed the difference between controls and SLE patients in vaccine-induced SARS–CoV-2 IgG levels as well as the decline in antibody levels over time. The Centers for Disease Control and Prevention guidance on immunocompromised patients changed to 3 primary doses of mRNA vaccines after the SLE patients in this study had been vaccinated. Thus, we could not study the effect of 3 primary mRNA vaccine doses.

Past studies did not have a large enough sample size or sufficiently detailed information to separate the effect of different immunosuppressive drugs and actual doses (9). We were able to show that taking mycophenolate mofetil, tacrolimus, and belimumab at the time of vaccination was associated with reduced SARS–CoV-2 IgG levels. Surprisingly, azathioprine and methotrexate were not. Methotrexate has been thought to have a modest effect (in non-SLE studies) (8,18). In fact, studies of holding methotrexate in rheumatoid arthritis patients showed benefit in influenza vaccination (19). Our results indicated no benefit of holding either methotrexate or azathioprine in SLE.

Our study showed a strong association of antihypertensive use with reduced vaccine response. There is no obvious

Table 3. Effect of demographic characteristics and medications taken at the time of vaccination on mean SARS–CoV-2 spike protein IgG levels*

Variable	No.	IgG, mean ± SD	*P*†	Positive for IgG, no. (%)	*P*†
Sex			0.028		0.15
Female	338	5.7 ± 2.9		306 (91)	
Male	27	4.1 ± 3.1		21 (78)	
Race			0.23		0.38
White	182	5.4 ± 2.8		167 (92)	
African American	129	6.0 ± 3.1		115 (89)	
Other	54	5.1 ± 3.2		45 (83)	
Age, years			0.66		0.44
<30	24	5.4 ± 3.1		20 (83)	
30–44	103	5.7 ± 2.8		96 (93)	
45–59	131	5.7 ± 3.1		115 (88)	
60+	107	5.3 ± 3.0		96 (90)	
Immunosuppressant use			0.047		0.0002
Not taking immunosuppressants	209	5.8 ± 2.6		200 (96)	
Taking immunosuppressants but held	71	5.5 ± 3.1		64 (90)	
Taking immunosuppressants	76	5.0 ± 3.5		56 (74)	
Prednisone dose			0.0045		0.044
None	295	5.9 ± 2.8		272 (92)	
<10 mg/day	56	4.5 ± 3.3		45 (80)	
≥10 mg/day	14	3.6 ± 2.8		10 (71)	
Mycophenolate mofetil use			<0.0001		<0.0001
None	271	6.0 ± 2.7		259 (96)	
Held	50	5.2 ± 3.0		45 (90)	
≤1,000 mg/day	17	5.2 ± 3.0		11 (65)	
>1,000 mg/day	27	2.3 ± 2.6		12 (44)	
Tacrolimus use			0.0009		0.0007
No	322	5.8 ± 2.9		299 (93)	
Yes	43	4.0 ± 3.5		28 (65)	
Azathioprine use			0.14		Too few
No	333	5.5 ± 3.0		296 (89)	
Held	8	6.3 ± 2.4		7 (88)	
Used	24	6.9 ± 3.1		24 (100)	
Belimumab use			0.018		0.18
No	353	5.7 ± 3.0		319 (90)	
Yes	12	2.9 ± 2.5		8 (67)	
Methotrexate use			0.23		Too few
No	341	5.5 ± 3.0		304 (89)	
Held	11	6.7 ± 3.4		10 (91)	
Used	13	7.0 ± 2.4		13 (100)	
Hydroxychloroquine use			0.50		0.95
No	40	6.0 ± 2.9		36 (90)	
Yes	325	5.5 ± 3.0		291 (90)	
NSAID use			0.64		0.081
No	323	5.6 ± 3.0		289 (89)	
Yes	30	5.7 ± 3.0		29 (97)	
Aspirin use			0.060		0.062
No	233	5.4 ± 3.0		204 (88)	
Yes	132	5.8 ± 2.9		123 (93)	
Clopidogrel use			0.78		0.82
No	346	5.6 ± 3.0		312 (90)	
Yes	8	5.7 ± 2.5		7 (88)	
Diuretic use			0.22		0.22
No	286	5.7 ± 2.9		260 (91)	
Yes	79	5.1 ± 3.1		67 (85)	
ACE-ARB use			0.043		0.0007
No	211	5.8 ± 2.8		200 (95)	
Yes	154	5.2 ± 3.2		127 (82)	
Calcium-channel blocker use			0.023		0.040
No	313	5.8 ± 2.9		286 (91)	
Yes	52	4.5 ± 3.4		41 (79)	
Other antihypertensive agent use			0.0055		0.0027
No	282	5.8 ± 2.8		262 (93)	
Yes	83	4.7 ± 3.3		65 (78)	

(Continued)

Table 3. (*Cont'd*)

Variable	No.	IgG, mean ± SD	*P*†	Positive for IgG, no. (%)	*P*†
Statin use			0.62		0.58
No	242	5.7 ± 2.9		219 (91)	
Yes	123	5.4 ± 3.2		108 (88)	

* ACE = angiotensin-converting enzyme; ARB = angiotensin II receptor blocker; NSAID = nonsteroidal antiinflammatory drug.
† Based on a generalized estimating equation model adjusting for time since vaccination (and time squared).

explanation for this association. The COVID-19 virus enters cells through binding of spike protein to the angiotensin-converting enzyme 2 (ACE2) receptor. The consensus is that ACE inhibitors do not inhibit this binding or affect the course of COVID-19 (20). Multiple studies have looked at the role of the renin–angiotensin system (RAS) in immune responsiveness and have suggested that inhibition of the RAS could inhibit the immune system (21), which might explain an impaired vaccine response (but no studies have actually looked at whether these inhibitors block vaccine response). β-adrenergic blockade can affect vaccine response in animal models, but the effects are either positive (22) or negative depending on the model (23).

Our data clearly showed the benefit of our policy of holding mycophenolate mofetil on the day of and for 1 week after the mRNA vaccine. We were able to do this without increasing activity or flares. We did not hold tacrolimus (as our nephritis patients taking tacrolimus were all also taking mycophenolate, we thought it too risky to hold both). Our findings are similar to a recent study that found that a 10-day methotrexate pause after the COVID vaccine booster enhanced immunity against the omicron variant

Table 4. Mean SARS–CoV-2 spike protein IgG levels by systemic lupus erythematosus–related variables measured at a cohort visit around the time of Pfizer or Moderna COVID-19 vaccination (2 doses) or antibody testing*

Variable	No.	IgG, mean ± SD	*P*†	Positive for IgG, no. (%)	*P*†
SLEDAI score at time of vaccination‡			0.68		0.39
<2	102	5.4 ± 2.7		94 (92)	
2–3	71	5.6 ± 3.0		64 (90)	
4+	83	5.3 ± 3.3		70 (84)	
PhGA score at time of vaccination‡			0.023		0.080
<0.5	90	5.4 ± 3.0		80 (89)	
0.5–1.0	127	5.8 ± 2.7		118 (93)	
>1	40	4.3 ± 3.4		31 (78)	
Anti-dsDNA at time of vaccination‡			0.43		0.55
0	150	5.5 ± 2.9		135 (90)	
>0	100	5.3 ± 3.1		87 (87)	
C3 at time of vaccination‡			0.11		0.35
Low (<79)	237	5.5 ± 3.0		212 (89)	
Not low	13	4.3 ± 3.0		10 (77)	
C4 at time of vaccination‡			0.51		0.88
Low (<10)	239	5.4 ± 3.0		212 (89)	
Not low	11	4.6 ± 2.9		10 (91)	
SLEDAI score at time of antibody test§			0.87		0.97
<2	137	5.6 ± 2.8		123 (90)	
2–3	89	5.5 ± 3.1		79 (89)	
4+	108	5.7 ± 3.0		96 (89)	
PhGA score at time of antibody test§			0.020		0.039
<0.5	131	5.4 ± 3.0		115 (88)	
0.5–1	153	6.1 ± 2.8		143 (93)	
>1	50	4.6 ± 3.2		40 (80)	
Anti-dsDNA at time of antibody test§			0.27		0.26
0	202	5.7 ± 2.8		183 (91)	
>0	132	5.4 ± 3.2		114 (86)	
C3 at time of antibody test§			0.070		0.11
Low (<79)	32	4.5 ± 3.2		25 (78)	
Not low	302	5.7 ± 2.9		272 (90)	
C4 at time of antibody test§			0.33		0.38
Low (<10)	17	5.0 ± 2.7		16 (94)	
Not low	317	5.6 ± 3.0		281 (89)	

* Anti-dsDNA = anti–double-stranded DNA; PhGA = physician global assessment of disease activity; SLEDAI = Systemic Lupus Erythematosus Disease Activity Index.
† Based on a generalized estimating equation model adjusting for time since vaccination (and time squared).
‡ Based on 257 observations with a clinic visit within 45 days of the date of first vaccination.
§ Based on 335 observations with a clinic visit within 7 days of the date of antibody assessment.

Table 5. Estimated effect of treatment at the time of vaccination on mean SARS–CoV-2 spike protein IgG levels based on a multivariable regression model*

Treatment	Estimated effect on mean IgG (95% CI)†	*P*†
Prednisone dose <10 mg/day (versus nonuse)	−0.46 (−1.44, 0.51)	0.35
Prednisone dose ≥10 mg/day (versus nonuse)	−1.25 (−3.52, 1.02)	0.28
MMF dose >1,000 (versus lower dose of MMF, nonuse, or use of MMF being held)	−2.38 (−3.57, −1.19)	<0.0001
Tacrolimus use	−1.49 (−2.60, −0.37)	0.0092
Belimumab use	−2.29 (−4.13, −0.45)	0.015
Antihypertensive medication use	−0.58 (−1.18, 0.020)	0.058

* 95% CI = 95% confidence interval; MMF = mycophenolate mofetil.
† Estimated effect and *P* value were based on a multivariable generalized estimating equation model, adjusting for sex, time since vaccination, time since vaccination squared, and all other treatments in the table.

(24), although our study did not show statistical significance. Increases in antibody level do have an impact on the risk of COVID-19 infection. In a separate study using the same lupus cohort, we observed that the risk of COVID-19 infection significantly decreased with increasing antibody titers (25). We think that these data will be helpful to clinicians now and in the future and will help revise the ACR guidelines, which had recommended holding up to 2 weeks (10). Importantly, holding immunosuppressive drugs for 1 week did not increase lupus flares.

We did not hold belimumab due to past data suggesting that it did not affect efficacy of other vaccines (26,27). Our data surprisingly showed that it did affect COVID-19 antibody response in the univariate and in the multiple variable model (this is important, as many of our patients were taking belimumab plus an oral immunosuppressive drug). We do not think holding belimumab is logical, however, as holding for 1 weekly subcutaneous dose would not be long enough to allow a rebound in B cells. Our patients were receiving subcutaneous dosing rather than intravenous dosing due to the pandemic.

This was an outpatient study. Visits followed the Hopkins Lupus Cohort protocol and were not timed to the vaccine date. Data on the rate and severity of breakthrough infections were not available for the control group and therefore could not be studied. Serologic responses were instead quantified as demonstrated to correlate with vaccine efficacy (28). Patients whose laboratory tests were not done at our hospital could not have the Euroimmun assay performed. However, this is the largest SLE study with the largest control group, the only one to include a balance of White and African American SLE patients, the only one to include immunosuppressant type and dose, and the only multivariable analysis. It is also the only study to evaluate a strategy, holding immunosuppressive drugs for 1 week, and to have found it effective for mycophenolate mofetil without leading to SLE flares.

ACKNOWLEDGMENTS

We acknowledge the collective community of donors to the Johns Hopkins University School of Medicine and the Johns Hopkins Health System for COVID-19 research. We thank Amanda Debes, PhD, JHBSPH International Health, for her contributions to the Health Care Workers study.

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. Petri 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. Petri, Magder.

Acquisition of data. Joyce, Haag, Fava, Goldman, Zhong, Xiao, Milstone.

Analysis and interpretation of data. Petri, Milstone, Magder.

REFERENCES

1. Strangfeld A, Schäfer M, Gianfrancesco MA, et al. Factors associated with COVID-19-related death in people with rheumatic diseases: results from the COVID-19 Global Rheumatology Alliance physician-reported registry. Ann Rheum Dis 2021;80:930–42.
2. Ugarte-Gil MF, Alarcón GS, Izadi Z, et al. Characteristics associated with poor COVID-19 outcomes in individuals with systemic lupus erythematosus: data from the COVID-19 Global Rheumatology Alliance. Ann Rheum Dis 2022;81:970–8.
3. FAI2R/SFR/SNFMI/SOFREMIP/CRI/IMIDIATE consortium and contributors. Severity of COVID-19 and survival in patients with rheumatic and inflammatory diseases: data from the French RMD COVID-19 cohort of 694 patients. Ann Rheum Dis 2021;80:527–38.
4. Vijenthira A, Gong I, Betschel SD, et al. Vaccine response following anti-CD20 therapy: a systematic review and meta-analysis of 905 patients. Blood Adv 2021;5:2624–43.
5. Galmiche S, Luong Nguyen LB, Tartour E, et al. Immunological and clinical efficacy of COVID-19 vaccines in immunocompromised populations: a systematic review. Clin Microbiol Infect 2022;28:163–77.
6. Jena A, Mishra S, Deepak P, et al. Response to SARS-CoV-2 vaccination in immune mediated inflammatory diseases: systematic review and meta-analysis. Autoimmun Rev 2022;21:102927.
7. Wu SS, Atefi G, Moynihan M, et al. Overall and post-vaccination prevalence of severe covid-19-related events among commercially insured patients with systemic lupus erythematosus and the general U.S. population [abstract]. Arthritis Rheumatol 2022;74 Suppl 9. URL: https://acrabstracts.org/abstract/overall-and-post-vaccination-prevalence-of-severe-covid-19-related-events-among-commercially-insured-patients-with-systemic-lupus-erythematosus-and-the-general-u-s-population/.
8. Sieiro Santos C, Calleja Antolin S, Moriano Morales C, et al. Immune responses to mRNA vaccines against SARS-CoV-2 in patients with immune-mediated inflammatory rheumatic diseases. RMD Open 2022;8:e001898.
9. Izmirly PM, Kim MY, Samanovic M, et al. Evaluation of immune response and disease status in systemic lupus erythematosus patients following SARS–CoV-2 vaccination. Arthritis Rheumatol 2022;74:284–94.

10. Curtis JR, Johnson SR, Anthony DD, et al. American College of Rheumatology guidance for COVID-19 vaccination in patients with rheumatic and musculoskeletal diseases: version 3. Arthritis Rheumatol 2021;73:e60–75.
11. Tan EM, Cohen AS, Fries JF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1982;25:1271–7.
12. Petri M, Orbai AM, Alarcón GS, et al. Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum 2012;64:2677–86.
13. Petri M, Hellmann D, Hochberg M. Validity and reliability of lupus activity measures in the routine clinic setting. J Rheumatol 1992;19:53–9.
14. Petri M, Kim MY, Kalunian KC, et al. Combined oral contraceptives in women with systemic lupus erythematosus. N Engl J Med 2005;353:2550–8.
15. Debes AK, Xiao S, Colantuoni E, et al. Association of vaccine type and prior SARS-CoV-2 infection with symptoms and antibody measurements following vaccination among health care workers. JAMA Intern Med 2021;181:1660–2.
16. Zhong D, Xiao S, Debes AK, et al. Durability of antibody levels after vaccination with mRNA SARS-CoV-2 vaccine in individuals with or without prior infection. JAMA 2021;326:2524–6.
17. Caturegli G, Materi J, Howard BM, et al. Clinical validity of serum antibodies to SARS-CoV-2 : a case-control study. Ann Intern Med 2020;173:614–22.
18. Haberman RH, Herati RS, Simon D, et al. Methotrexate hampers immunogenicity to BNT162b2 mRNA COVID-19 vaccine in immune-mediated inflammatory disease. Ann Rheum Dis 2021;80:1339–4.
19. Park JK, Lee YJ, Shin K, et al. Impact of temporary methotrexate discontinuation for 2 weeks on immunogenicity of seasonal influenza vaccination in patients with rheumatoid arthritis: a randomised clinical trial. Ann Rheum Dis 2018;77:898–904.
20. Peng M, He J, Xue Y, et al. Role of hypertension on the severity of COVID-19: a review. J Cardiovasc Pharmacol 2021;78:e648–55.
21. Ranjbar R, Shafiee M, Hesari A, et al. The potential therapeutic use of renin–angiotensin system inhibitors in the treatment of inflammatory diseases. J Cell Physiol 2019;234:2277–95.
22. Daher C, Vimeux L, Stoeva R, et al. Blockade of β-adrenergic receptors improves CD8+ T-cell priming and cancer vaccine efficacy. Cancer Immunol Res 2019;7:1849–63.
23. Bufan B, Arsenović-Ranin N, Živković I, et al. B-cell response to seasonal influenza vaccine in mice is amenable to pharmacological modulation through β-adrenoceptor. Life Sci 2022;301:120617.
24. Habermann E, Gieselmann L, Tober-Lau P, et al. Pausing methotrexate prevents impairment of omicron BA.1 and BA.2 neutralisation after COVID-19 booster vaccination. RMD Open 2022;8:e002639.
25. Magder LS, Fava A, Goldman DW. The incidence of COVID-19 infection in an SLE cohort, and its association with immunosuppressants, SLE disease activity, vaccination status, and COVID-19 antibodies [abstract]. Arthritis Rheumatol 2022;74 Suppl 9. URL: https://acrabstracts.org/abstract/the-incidence-of-covid-19-infection-in-an-sle-cohort-and-its-association-with-immunosuppressants-sle-disease-activity-vaccination-status-and-covid-19-antibodies/.
26. Chatham W, Chadha A, Fettiplace J, et al. A randomized, open-label study to investigate the effect of belimumab on pneumococcal vaccination in patients with active, autoantibody-positive systemic lupus erythematosus. Lupus 2017;26:1483–90.
27. Nagel J, Saxne T, Geborek P, et al. Treatment with belimumab in systemic lupus erythematosus does not impair antibody response to 13-valent pneumococcal conjugate vaccine. Lupus 2017;26:1072–81.
28. Gilbert PB, Montefiori DC, McDermott AB, et al. Immune correlates analysis of the mRNA-1273 COVID-19 vaccine efficacy clinical trial. Science 2022;375:43–50.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 75, No. 9, September 2023, pp 1886–1896
DOI 10.1002/acr.25097

Associations of Postdischarge Follow-Up With Acute Care and Mortality in Lupus: A Medicare Cohort Study

Maria Schletzbaum, Nadia Sweet, Brad Astor, Ang Yu, W. Ryan Powell, Andrea Gilmore-Bykovskyi, Farah Kaiksow, Ann Sheehy, Amy J. Kind, and Christie M. Bartels

Objective. Patients with systemic lupus erythematosus experience the sixth highest rate of 30-day readmissions among chronic diseases. Timely postdischarge follow-up is a marker of ambulatory care quality that can reduce readmissions in other chronic conditions. Our objective was to test the hypotheses that 1) beneficiaries from populations experiencing health disparities, including patients from disadvantaged neighborhoods, will have lower odds of completed follow-up, and that 2) follow-up will predict longer time without acute care use (readmission, observation stay, or emergency department visit) or mortality.

Methods. This observational cohort study included hospitalizations in January–November 2014 from a 20% random sample of Medicare adults. Included hospitalizations had a lupus code, discharge to home without hospice, and continuous Medicare A/B coverage for 1 year before and 1 month after hospitalization. Timely follow-up included visits with primary care or rheumatology within 30 days. Thirty-day survival outcomes were acute care use and mortality adjusted for sociodemographic information and comorbidities.

Results. Over one-third (35%) of lupus hospitalizations lacked 30-day follow-up. Younger age, living in disadvantaged neighborhoods, and rurality were associated with lower odds of follow-up. Follow-up was not associated with subsequent acute care or mortality in beneficiaries age <65 years. In contrast, follow-up was associated with a 27% higher hazard for acute care use (adjusted hazard ratio [HR] 1.27 [95% confidence interval (95% CI) 1.09–1.47]) and 65% lower mortality (adjusted HR 0.35 [95% CI 0.19–0.67]) among beneficiaries age ≥65 years.

Conclusion. One-third of lupus hospitalizations lacked follow-up, with significant disparities in rural and disadvantaged neighborhoods. Follow-up was associated with increased acute care, but 65% lower mortality in older systemic lupus erythematosus patients. Further development of lupus-specific postdischarge strategies is needed.

INTRODUCTION

One-fourth of people living with systemic lupus erythematosus (SLE or lupus) are hospitalized each year, with one-third rehospitalized within 30 days (1–3), making SLE the sixth highest chronic disease cause of readmission in the US (4). These inpatient stays contribute to high health care costs and a burden for patients with lupus (5,6) who are disproportionately female and people identifying as Black, Hispanic, Asian, or American Indian (7). With approximately half of US lupus hospitalizations and emergency department visits covered by public insurance (8,9), readmission reduction strategies may produce significant health and cost benefits in this high-risk population (10–12).

In a move toward value-based care, the Centers for Medicare and Medicaid Services (CMS) implemented the Hospital Readmissions Reduction Program focused on reducing 30-day rehospitalization rates in 2013. Timely postdischarge follow-up visits have been promoted as one method to reduce readmissions (13,14). Among patients with heart failure, timely follow-up within 14 and 30 days has been associated with lower readmission and mortality rates (15–17). However, in other populations, such as patients with chronic obstructive pulmonary disease (COPD) and age >65 years, results have been mixed on the

Supported by research awards from the National Institute on Minority Health and Health Disparities (grants F30-MD-015211 and R01-MD-010243), the University of Wisconsin Medical Scientist Training Program, and the NIH/National Institute of General Medical Sciences (grant T32-GM-140935).

Maria Schletzbaum, PhD, Nadia Sweet, MD, Brad Astor, PhD, MPH, Ang Yu, MA, MS, W. Ryan Powell, PhD, Andrea Gilmore-Bykovskyi, PhD, RN, Farah Kaiksow, MD, MPP, Ann Sheehy, MD, MS, Amy J. Kind, MD, PhD, Christie M. Bartels, MD, MS: University of Wisconsin, Madison.

Author disclosures and a graphical abstract can be found online at https://onlinelibrary.wiley.com/doi.10.1002/acr.25097.

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

Submitted for publication March 2, 2022; accepted in revised form January 31, 2023.

SIGNIFICANCE & INNOVATIONS

- In one of the first studies assessing rates of postdischarge follow-up among patients with lupus, we found that 35% lacked follow-up within 30 days, worse than reported in other chronic conditions.
- Patients with lupus residing in highly disadvantaged neighborhoods and rural areas had significantly lower odds of timely postdischarge follow-up, demonstrating disparities.
- In systemic lupus erythematosus patients ages ≥65 years, receiving timely follow-up was associated with 65% lower 30-day mortality, indicating the importance of follow-up in quality lupus care.
- Follow-up among older lupus patients was associated with increased acute care use, which may indicate the need for lupus-specific transitional care management and postdischarge support.

impact of follow-up on acute care use, including readmissions and emergency department visits (18–20). In lupus, the frequency of postdischarge follow-up and associations with subsequent acute care use and mortality have not been assessed.

The National Institute on Minority Health and Health Disparities' research framework acknowledges that differences in health care access, quality, and individual and community socioeconomic resources contribute to US health disparities (21,22). Given that postdischarge follow-up indicates ambulatory access and quality care, we hypothesized that lupus patients who are part of populations experiencing health disparities (21) are less likely to receive follow-up. Further, we hypothesized that receiving timely follow-up would be associated with lower rates of acute care use and mortality. Therefore, the first objective of this study was to assess health disparities in timely postdischarge follow-up with a primary care provider (PCP) or rheumatologist by race and ethnicity, neighborhood disadvantage, and rural-urban context. The second objective was to evaluate the association of timely follow-up with subsequent acute care use and mortality within 30 days of discharge among Medicare beneficiaries with SLE.

PATIENTS AND METHODS

Study population. The study cohort comprised hospitalizations occurring between January 1, 2014 and November 30, 2014 of patients in a 20% random sample of Medicare beneficiaries. Inclusion required the hospitalization to be associated with an SLE diagnosis code at any position (International Classification of Diseases, Ninth Revision, Clinical Modification 710.0; positive predictive value for SLE of 99.4%) (1,23,24). The hospitalized beneficiary had to be age ≥18 years, alive at discharge, and discharged to a home setting without hospice to focus on the population relevant to transitional care management and acute care use (16,24,25). The hospitalized beneficiary had to have at least a year of continuous Medicare A and B coverage prior to the index admission to allow assessment of comorbidities and for 30 days after discharge, excluding end of benefits due to death (Figure 1). To capture complete claims, hospitalizations were excluded if the beneficiary was enrolled in a Medicare Advantage plan or had railroad benefits. Consistent with CMS readmission definitions, hospitalizations at long-term acute care facilities and psychiatric, children's, cancer, and rehabilitation hospitals were also excluded (26). A given beneficiary could have multiple hospitalizations included in the sample if they had multiple qualifying hospitalizations within the study period. Each hospitalization was treated as an index hospitalization with its own 30-day postdischarge period in which follow-up, acute care, and mortality were measured.

A Health Sciences Institutional Review Board at the University of Wisconsin School of Medicine and Public Health approved this de-identified medical claims study as minimal risk with a waiver of individual informed consent. Study design and findings are reported per the Strengthening the Reporting of Observational Studies in Epidemiology guidelines for cohort studies (27).

Follow-up care. Follow-up care was defined as receipt of an ambulatory visit within 30 days of discharge with a PCP (defined as general internal medicine, family medicine, general practice, pediatrics, geriatrics, obstetrics, and gynecology) or rheumatologist. Eligible follow-up visits were identified by Physician Specialty Code or National Provider Identifier (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25097). Timing of follow-up was determined by the first eligible visit after the discharge date. Follow-up received within 14 days of discharge was also examined based on CMS transitional care management reimbursement requirements (14).

Acute care use and mortality. Our primary outcome was survival time until the first acute care use encounter within 30 days of index hospitalization discharge. Acute care included inpatient hospital readmissions, inpatient observation stays, and emergency department visits. Observation stays were identified using Medicare revenue codes 0760, 0762, and 0769; revenue code 0761 was excluded as it is generally used for planned procedural interventions in hospital settings (28). Our secondary outcome was survival time until mortality within 30 days of discharge, using death as reported in CMS data (29).

Beneficiary and hospitalization characteristics. Variables defining the beneficiary's demographic, socioeconomic, geographic, and health status were included in analyses. Patient demographic factors included age at index hospitalization and sex. Race and ethnicity were included as a proxy for lived experiences of these groups, including the effects of structural and institutional racism. For descriptive statistics, race and

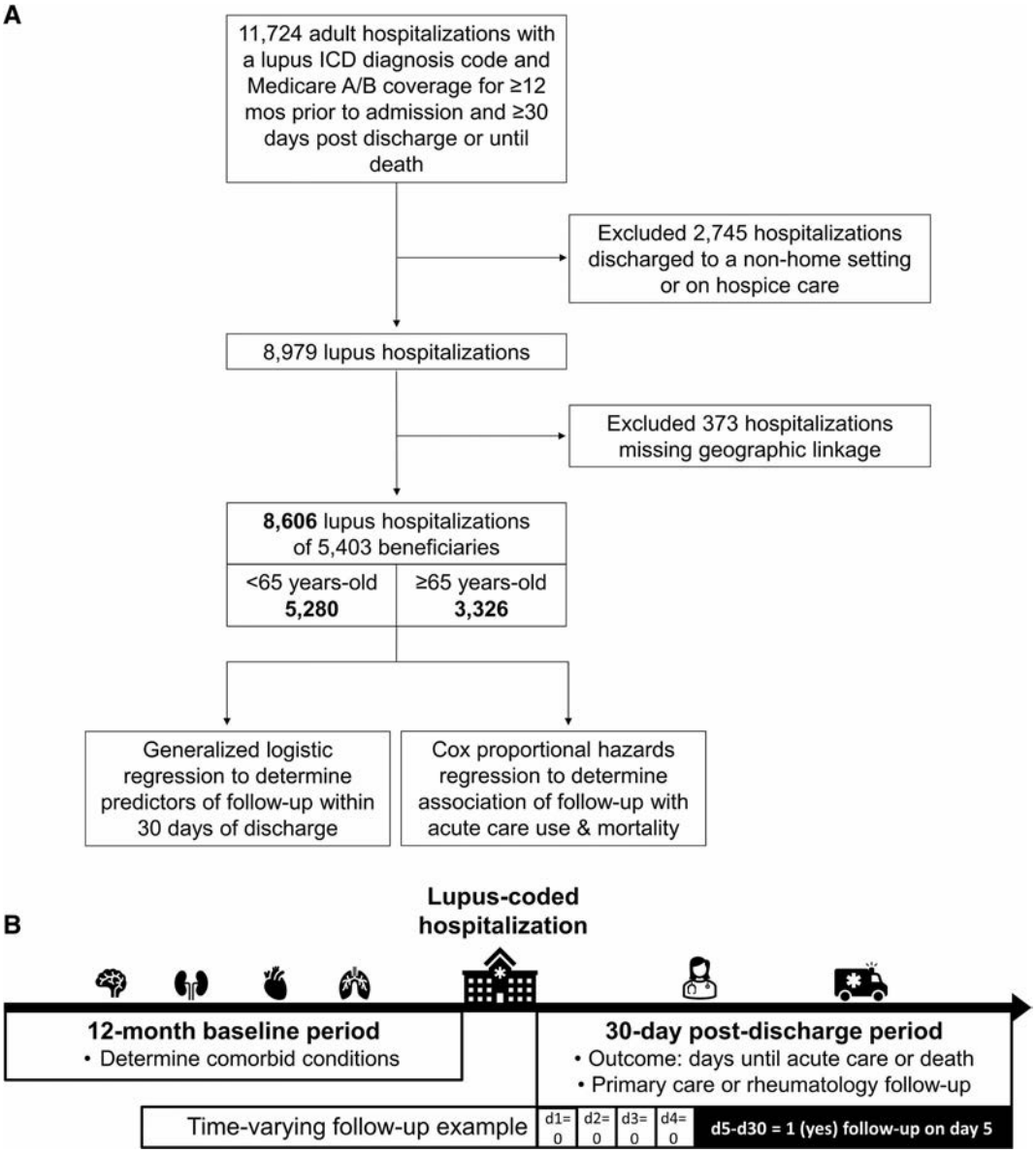

Figure 1. **A**, Flow diagram showing the creation of a cohort of lupus-coded hospitalizations from a 20% national Medicare beneficiary sample between January 1, 2014 and November 30, 2014. **B**, Graphical depiction of data structure, including the 12-month baseline period for collecting information on comorbid conditions. Following hospitalization was the 30-day postdischarge period during which follow-up visits with primary care providers and rheumatologists were captured and time until first acute care use encounter and death were recorded. The bottom row provides an example of how time-varying follow-up exposure was coded. ICD = International Classification of Diseases; mos = months.

ethnicity responses were categorized as Asian and Pacific Islander, Black, American Indian and Alaska Native, Hispanic or Latino, White, and other based on the Medicare-reported Research Triangle Institute race and ethnicity variable (30–33). To meet statistical assumptions for the regression methods used and to reduce the risk of identifiability, race and ethnicity were consolidated to Non-Hispanic White (hereafter White), Non-Hispanic Black (hereafter Black), Hispanic or Latino (hereafter Hispanic), and other for analyses. Neighborhood-level disadvantage was measured by the national decile of the Area Deprivation Index (ADI), a validated geospatial metric incorporating 17 measures of housing, education, employment, and income (34,35). The ADI was linked to each hospitalization using the ZIP+4 postal code of the beneficiary. We used rural–urban commuting area (RUCA) codes to determine rurality (36). Dual eligibility for Medicaid was also included.

Indicators of the beneficiary's health status included CMS's hierarchical condition category (HCC) community comorbidity score (37) and Elixhauser comorbidity indicators for renal failure, congestive heart failure, diabetes mellitus, depressive disorders, and alcohol use disorder (38). CMS Chronic Condition Warehouse condition indicators were included for anxiety disorders,

tobacco use disorder, and drug or opioid use disorder (39). Disability as the original reason for Medicare eligibility and length of the index hospitalization, in days, were also included. For each hospitalization, the clinical classifications software first-level category code for the primary discharge diagnosis was reported, as were billing codes for transitional care management services, as defined by CMS (14).

Statistical analysis. Characteristics of the included hospitalizations were described for the entire cohort as well as stratified by age <65 years and ≥65 years, given the different ways these groups can qualify for Medicare. Multivariable logistic regression was used to identify sociodemographic and comorbidity factors of beneficiaries associated with contributing >1 hospitalization. Cumulative incidence functions, accounting for censoring and mortality, were found for follow-up and acute care, with the functions for the components of acute care also reported. A mortality-only cumulative incidence function was calculated.

We performed generalized logistic regressions to determine predictors of postdischarge follow-up care with a PCP or rheumatologist within 14 and within 30 days of discharge. Analysis was clustered by beneficiary to account for correlation among multiple hospitalizations of the same beneficiary.

Cox proportional hazards regressions, clustered by beneficiary, were then used to evaluate the association of postdischarge follow-up with subsequent acute care use and with mortality within 30 days of discharge. Martingale residuals were assessed for covariate variable specification and Schoenfeld residuals to affirm that the proportional hazards assumption was not violated. In the acute care use survival analysis, death was treated as a competing event using Fine and Gray's proportional subdistribution hazards model (40). Receipt of follow-up care was included as a time-varying covariate, which allows it to change over the 30-day outcome period (20). A binary follow-up variable was created for each day postdischarge through day 30 for every hospitalization. The variable was coded as "no follow-up" until the day the beneficiary had a qualifying follow-up visit, when it then switched to follow-up for that and subsequent days (Figure 1B). For example, for a hospitalization where the beneficiary had follow-up on day 5 postdischarge, the follow-up variables for days 1 to 4 would indicate no follow-up and for days 5 to 30 would indicate follow-up.

For beneficiaries without follow-up, all 30-day variables would indicate no follow-up. This method allows the survival model to evaluate and compare the beneficiary's follow-up status on the day they had an outcome event. This method also helps address immortal time bias by assigning days at risk to the appropriate category, i.e., days prior to the follow-up visit as non–follow-up time and days on or after the follow-up visit as follow-up time. Additional analyses were conducted to assess for the effect of rheumatology follow-up and PCP follow-up separately and for 30-day readmissions and emergency department visits, with death as a competing risk.

Supplementary analyses were conducted by creating 2 distinct periods: a follow-up period from postdischarge days 0 through 14 and an outcome period from days 15 through 30. Follow-up within 14 days was dichotomized and associations of follow-up with subsequent acute care use and mortality in the outcome period were found using Cox proportional hazards regression. Hospitalizations with acute care within the 14-day follow-up period were excluded from the acute care outcome analysis, and those who died in the follow-up period were excluded from the mortality analysis. Analyses were performed using SAS software, version 9.4.

RESULTS

There were 8,606 hospitalizations with a lupus diagnosis included in the sample, representing 5,403 beneficiaries with 1,663 beneficiaries (30.8% of beneficiaries) contributing multiple hospitalizations (Figure 1A). After adjusting for sociodemographic factors and comorbidities, beneficiaries with multiple included hospitalizations were less likely to be older (adjusted odds ratio [OR_{adj}] per decade 0.84 [95% confidence interval (95% CI) 0.80–0.89]) (see Supplementary Table 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25097) and more commonly of Hispanic ethnicity (OR_{adj} 1.30 [95% CI 1.04–1.64]).

Beneficiaries with >1 hospitalization were also more likely to have a higher HCC comorbidity score (OR_{adj} 1.27 [95% CI 1.22–1.32]), tobacco use disorder (OR_{adj} 1.22 [95% CI 1.05–1.41]), drug or opioid use disorder (OR_{adj} 1.53 [95% CI 1.24–1.89]), anxiety (OR_{adj} 1.33 [95% CI 1.16–1.53]), depression (OR_{adj} 1.49 [95% CI 1.29–1.73]), renal failure (OR_{adj} 1.76 [95% CI 1.52–2.05]), or congestive heart failure (OR_{adj} 1.51 [95% CI 1.29–1.78]). The 3 most common primary discharge diagnosis categories were diseases of the circulatory system (20.4%), digestive system (13.6%), and respiratory system (12.2%). The cumulative incidence of follow-up care was 46.9% by day 14 and 64.5% within 30 days of discharge (Figure 2A). Among those who had follow-up within 30 days, 18.3% had initial follow-up with a rheumatologist (28.2% in those with rheumatology visits in the year before hospitalization versus 3.6% in those without). Transitional care management codes within 31 days of discharge were associated with 4.9% of hospitalizations (n = 419); only 1 hospitalization with initial follow-up with rheumatology was associated with a transitional care management billing code.

Observed follow-up within 30 days was less common among those with higher HCC comorbidity scores, longer index hospitalization length of stay, rural residence, and greater neighborhood disadvantage (Table 1). Among patients living in the most disadvantaged neighborhood decile, 48.7% received follow-up within 30 days, nearly 10% lower than in the least disadvantaged decile (57.0%). Follow-up was received by 54.4% of White patients within 30 days but only 49.6% of Black patients,

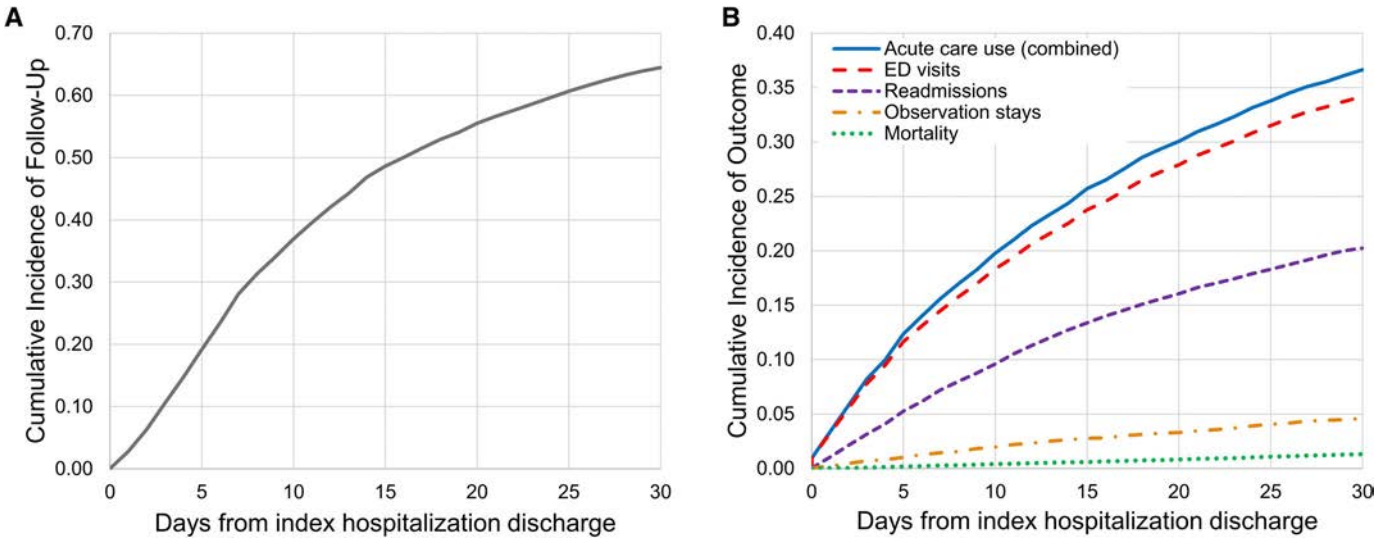

Figure 2. **A**, Cumulative incidence of follow-up; **B**, Acute care use, readmissions, observation stays, emergency department (ED) visits, and mortality outcomes within 30 days of discharge.

50.3% of Hispanic patients, and 42.1% of patients in the other race and ethnicity category (Table 2). For urban and suburban residents, follow-up was 53.0% compared to 45.1% among rural residents.

Predictors of follow-up. In multivariate generalized logistic models, older age was significantly associated with higher 30-day (OR_{adj} 1.09 [95% CI 1.04–1.14]) (Table 2) and 14-day follow-up (OR_{adj} 1.12 [95% CI 1.07–1.17]). Meanwhile, lower rates of 30-day and 14-day follow-up visits were associated with rural residence compared to a suburban residence (OR_{adj} 0.70 [95% CI 0.57–0.86] and OR_{adj} 0.76 [95% CI 0.62–0.94], respectively) and having renal failure (OR_{adj} 0.69 [95% CI 0.61–0.78] and OR_{adj} 0.73 [95% CI 0.64–0.82], respectively). For follow-up within 14 days, greater neighborhood disadvantage was associated with a lower rate of follow-up (OR_{adj} 0.98 [95% CI 0.96–0.99]). For follow-up within 30 days, other race and ethnicity was associated with a lower rate of follow-up (OR_{adj} 0.73 [95% CI 0.55–0.96]). See Supplementary Table 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25097, for predictors of follow-up within 30 and 14 days stratified by age group.

Acute care use. The cumulative incidence of acute care use within 30 days was 36.6%, with a cumulative incidence of 34.2% for emergency department visits, 4.6% for observation stays, and 20.3% for readmissions (Figure 2B). Among those age ≥65 years, follow-up was significantly associated with a higher hazard rate for acute care use (adjusted hazard ratio [HR_{adj}] 1.27 [95% CI 1.09–1.47]) (Table 3). Other significant predictors of greater acute care use included longer length of index hospitalization (HR_{adj} 1.03 [95% CI 1.02–1.05]), higher HCC comorbidity score (HR_{adj} 1.11 [95% CI 1.07–1.15]), depression (HR_{adj} 1.25 [95% CI 1.08–1.46]), and Medicaid eligibility (HR_{adj} 1.26 [95% CI 1.06–1.50]). Follow-up within 14 days of discharge had a similar effect to follow-up within 30 days (HR_{adj} 1.26 [95% CI 1.09–1.47]) (see Supplementary Table 4, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25097) and was somewhat attenuated in supplemental analysis using distinct follow-up and outcome periods (HR_{adj} 1.14 [95% CI 0.92–1.42]).

In contrast, in those patients age <65 years, 30-day follow-up was not associated with acute care use (HR_{adj} 1.06 [95% CI 0.96–1.17]) (Table 3). Longer length of index hospitalization (HR_{adj} 1.01 [1.00–1.02]; $P < 0.05$), higher HCC comorbidity score (HR_{adj} 1.09 [95% CI 1.07–1.11]), and depression (HR_{adj} 1.17 [95% CI 1.06–1.29]) were still associated with higher levels of acute care use. Additionally, greater acute care use was associated with greater neighborhood disadvantage (HR_{adj} 1.04 [95% CI 1.02–1.06]), alcohol use disorder (HR_{adj} 1.26 [1.01–1.55]), drug or opioid use disorders (HR_{adj} 1.45 [95% CI 1.30–1.62]), anxiety (HR_{adj} 1.17 [95% CI 1.06–1.29]), and congestive heart failure (HR_{adj} 1.17 [95% CI 1.06–1.30)]. Conversely, older age within this age stratum was associated with a lower hazard for acute care (HR_{adj} 0.83 [95% CI 0.79–0.87]). Follow-up within 14 days similarly had no association with acute care in those age <65 years when using time-varying follow-up (HR_{adj} 1.03 [95% CI 0.93–1.14]) (see Supplementary Table 4, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25097) or with distinct follow-up and outcome periods (HR_{adj} 1.11 [95% CI 0.95–1.29]).

In sensitivity analyses, the association of PCP follow-up with acute care was consistent with the results for combined follow-up with PCP or rheumatology among all adult ages. The HR_{adj} of PCP follow-up on acute care use in all ages was 1.13 (95% CI 1.04–1.23). However, there was not an association of

Table 1. Baseline characteristics of Medicare SLE hospitalizations by age group and receipt of follow-up within 30 days*

	All ages cohort (n = 8,606)	Age <65 years		Age ≥65 years	
		Follow-up (n = 2,593)	No follow-up (n = 2,687)	Follow-up (n = 1,888)	No follow-up (n = 1,438)
Patient variables					
Admission age, mean ± SD	57.6 ± 16.4	48.0 ± 11.1	46.3 ± 11.6	74.0 ± 6.3	74.2 ± 6.9
Age group, years					
18–35	1,065 (12.4)	457 (17.6)	608 (22.6)	–	–
36–64	4,215 (49.0)	2,136 (82.4)	2,079 (77.4)	–	–
≥65	3,326 (38.6)	–	–	1,888 (100)	1,438 (100)
Female	7,648 (88.9)	2,360 (91.0)	2,389 (88.9)	1,645 (87.1)	1,254 (87.2)
Race and ethnicity					
White	4,790 (55.7)	1,131 (43.6)	1,066 (39.7)	1,475 (78.1)	1,118 (77.8)
Black	2,619 (30.4)	1,045 (40.3)	1,110 (41.3)	255 (13.5)	209 (14.5)
Hispanic or Latino	867 (10.1)	328 (12.7)	358 (13.3)	108 (5.7)	73 (5.1)
Other: Asian and Pacific Islander†	123 (1.4)	26 (1.0)	57 (2.1)	50 (2.7)	38 (2.6)
American Indian and Alaska Native	93 (1.1)	33 (1.3)	42 (1.6)	–	–
Unknown	114 (1.3)	30 (1.2)	54 (2.0)	–	–
ADI disadvantage, mean ± SD	55.8 ± 27.3	59.1 ± 26.7	60.7 ± 26.5	48.1 ± 27.1	51.1 ± 27.4
Most disadvantaged quintile	2,158 (25.1)	738 (28.5)	831 (30.9)	312 (16.5)	277 (19.3)
Rural–urban commuting area					
Urban	6,323 (73.5)	1,983 (76.5)	1,975 (73.5)	1,371 (72.6)	994 (69.1)
Suburban	745 (8.7)	204 (7.9)	226 (8.4)	191 (10.1)	124 (8.6)
Large rural	906 (10.5)	256 (9.9)	287 (10.7)	191 (10.1)	172 (12.0)
Small town/rural	632 (7.3)	150 (5.8)	199 (7.4)	135 (7.2)	148 (10.3)
Medicaid beneficiary	4,104 (47.7)	1,688 (65.1)	1,733 (64.5)	355 (18.8)	328 (22.8)
Disability as Medicare reason	5,960 (69.3)	2,423 (93.4)	2,407 (89.6)	625 (33.1)	505 (35.1)
HCC comorbidity score, mean ± SD	3.5 ± 2.4	3.7 ± 2.5	3.9 ± 2.7	3.0 ± 2.0	3.2 ± 2.1
Index hospital stay, mean ± SD days	4.6 ± 4.3	4.7 ± 4.2	4.8 ± 4.9	4.1 ± 3.4	4.5 ± 4.4
Renal failure	3,377 (39.2)	1,029 (39.7)	1,411 (52.5)	510 (27.0)	427 (29.7)
Congestive heart failure	2,258 (26.2)	670 (25.8)	762 (28.4)	463 (24.5)	363 (25.2)
Diabetes mellitus	2,627 (30.5)	812 (31.3)	870 (32.4)	541 (28.7)	404 (28.1)
Anxiety	4,289 (49.8)	1,511 (58.3)	1,472 (54.8)	745 (39.5)	561 (39.0)
Depression	2,706 (31.4)	960 (37.0)	989 (36.8)	408 (21.6)	349 (24.3)
Tobacco use disorder	2,662 (30.9)	980 (37.8)	1,059 (39.4)	339 (18.0)	284 (19.8)
Alcohol use disorder	216 (2.5)	88 (3.4)	100 (3.7)	15 (0.8)	13 (0.9)
Drug or opioid use disorder	1,245 (14.5)	496 (19.3)	608 (22.6)	100 (5.3)	41 (2.9)
Follow-up type					
Primary care provider	3,661 (42.5)	2,028 (78.2)	–	1,633 (86.5)	–
Rheumatology	820 (9.5)	565 (21.8)	–	255 (13.5)	–
30-day outcomes					
Emergency department visits	2,942 (34.2)	972 (37.5)	1,081 (40.2)	499 (26.4)	390 (27.1)
Observation stays	419 (4.9)	155 (6.0)	157 (5.8)	60 (3.2)	47 (3.3)
Readmissions	1,928 (22.4)	590 (23.0)	736 (29.3)	315 (16.9)	287 (21.1)
Deaths	114 (1.3)	14 (0.5)	31 (1.2)	13 (0.7)	56 (3.9)

* Values are the number (%) unless indicated otherwise. ADI = Area Deprivation Index; HCC = hierarchical condition category; SLE = systemic lupus erythematosus.
† Due to small, potentially identifiable numbers, patients age ≥65 years in the Asian and Pacific Islander, American Indian and Alaska Native, and unknown categories are reported together.

rheumatology follow-up with acute care use among all ages (HR_{adj} 0.95 [95% CI 0.82–1.11]). PCP follow-up was still associated with less mortality (HR_{adj} 0.58 [95% CI 0.36–0.92]) and rheumatology follow-up trended toward less mortality but was not statistically significant (HR_{adj} 0.25 [95% CI 0.06–1.04]).

Mortality. The cumulative incidence of mortality among all hospitalizations was 1.3% (Figure 2B). In the population age ≥65 years, follow-up was significantly associated with a 65% lower hazard for mortality (HR_{adj} 0.35 [95% CI 0.19–0.67]) (Table 4). Conversely, longer index hospitalization (HR_{adj} 1.08 [95% CI 1.04–1.12]), greater HCC comorbidity score (HR_{adj} 1.22 [95% CI 1.10–1.35]), and tobacco use disorder (HR_{adj} 1.90 [1.02–3.53]) were associated with higher levels of mortality, as was older age (HR_{adj} 1.53 [95% CI 1.00–2.32]; $P < 0.05$). Among hospitalizations of patients age < 65 years, mortality was not associated with follow-up (HR_{adj} 0.91 [95% CI 0.46–1.80]). The length of the index hospitalization (HR_{adj} 1.04 [95% CI 1.01–1.07]) and HCC comorbidity score (HR_{adj} 1.20 [95% CI 1.06–1.35]) were significantly associated with mortality in this younger stratum. Follow-up within 14 days similarly had no association with mortality for those age < 65 years, but an association with lower

Table 2. Observed 30-day follow-up rate and adjusted ORs for predictors of follow-up within 30 and 14 days postdischarge (n = 8,606)*

	Follow-up 30 days		Follow-up 14 days
	Observed rate, no. (%)	Adjusted OR (95% CI)	Adjusted OR (95% CI)
Age at admission (per decade), years	–	1.09 (1.04–1.14)†	1.12 (1.07–1.17)†
<65	2,593 (49.1)	–	–
≥65	1,888 (50.7)	–	–
Sex			
Male	476 (49.7)	Ref.	Ref.
Female	4,005 (52.4)	1.07 (0.90–1.27)	1.00 (0.85–1.18)
Race and ethnicity			
White	2,606 (54.4)	Ref.	Ref.
Black	1,300 (49.6)	1.01 (0.87–1.17)	1.06 (0.92–1.22)
Hispanic	436 (50.3)	0.97 (0.79–1.19)	1.06 (0.87–1.28)
Other	139 (42.1)	0.73 (0.55–0.96)†	0.78 (0.59–1.03)
Rural–urban commuting area			
Suburban	395 (53.0)	Ref.	Ref.
Urban	3,354 (53.0)	0.96 (0.80–1.17)	0.91 (0.76–1.10)
Large rural	447 (49.3)	0.85 (0.71–1.02)	0.96 (0.80–1.14)
Small town/rural	285 (45.1)	0.70 (0.57–0.86)†	0.76 (0.62–0.94)†
ADI disadvantage (per decile)	–	0.98 (0.96–1.00)	0.98 (0.96–0.99)†
Most disadvantaged quintile	1,050 (48.7)	–	–
Medicaid beneficiary	2,043 (49.8)	1.08 (0.94–1.23)	1.04 (0.91–1.18)
Disability as Medicare reason	3,048 (51.1)	1.04 (0.90–1.19)	1.08 (0.95–1.24)
HCC comorbidity score	–	1.00 (0.97–1.03)	1.00 (0.99–1.01)
Hospital length of stay (per day)	–	0.99 (0.98–1.00)	1.00 (0.99–1.01)
Tobacco use disorder	1,319 (49.6)	0.90 (0.79–1.03)	0.94 (0.83–1.06)
Alcohol use disorder	103 (47.7)	0.90 (0.61–1.32)	0.78 (0.55–1.12)
Drug or opioid use disorder	596 (47.9)	0.93 (0.77–1.11)	1.03 (0.87–1.22)
Anxiety	2,256 (52.6)	1.11 (0.99–1.25)	1.08 (0.96–1.21)
Depression	1,368 (50.6)	0.96 (0.84–1.08)	0.91 (0.80–1.03)
Renal failure	1,539 (45.6)	0.69 (0.61–0.78)†	0.73 (0.64–0.82)†
Diabetes mellitus	1,353 (51.5)	1.02 (0.90–1.15)	1.04 (0.9–1.17)
Congestive heart failure	1,133 (50.2)	1.05 (0.91–1.22)	1.10 (0.96–1.26)

* 95% CI = 95% confidence interval; ADI = Area Deprivation Index; HCC = hierarchical condition category; OR = odds ratio; Ref. = reference.
† Statistically significant.

levels of mortality in those age ≥ 65 years in analyses using both time-varying follow-up (HR_{adj} 0.47 [95% CI 0.25–0.88]) (see Supplementary Table 4, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25097) and distinct follow-up and outcome periods (HR_{adj} 0.43 [95% CI 0.20–0.91]).

DISCUSSION

In this large study of Medicare lupus hospitalizations, we found that timely follow-up was associated with 27% higher acute care use and 65% lower mortality in those age ≥65 years. The association of timely follow-up with lower mortality fits with prior studies in heart failure, COPD, and the Medicare population age ≥65 years that consistently show lower mortality (13,17,20,41). Prior studies suggest that while follow-up visits alone decrease mortality, they may increase care utilization (13), which has been observed for follow-up after emergency department visits (13,20). Another key finding was that >1 in 3 hospitalizations of lupus patients lacked follow-up. Thirty-day follow-up in this lupus study (64.5%) was lower than rates reported among patients with heart failure, myocardial infarction, pneumonia, and COPD (66.9–75.4%) (17,18,25,41). Notable disparities included fewer beneficiaries from highly disadvantaged neighborhoods and rural areas receiving timely follow-up, even after adjusting for health status. Neighborhood disadvantage was also associated with greater acute care use. Our study reinforces the role of timely follow-up in good clinical practice and the importance of considering health outcomes, such as mortality, in addition to acute care use (13,20,42). The results indicate a need for better ambulatory access or outreach to lupus patients, particularly those residing in rural and disadvantaged neighborhoods (43).

Heart failure specialists have developed and implemented clinical treatment guidelines and best practices for 30-day postdischarge follow-up (44–46), which could inform postdischarge lupus care, recognizing similarly high rehospitalization rates (1). Among lupus patients who did receive follow-up within 30 days, 18.3% initially saw a rheumatologist, which exceeded specialist follow-up in heart failure (7.5%) (15) or COPD (13–19%) (18,41). These numbers reflect greater rheumatology specialist involvement in lupus care. Notably, rheumatology follow-up, in contrast to PCP follow-up, was not associated with higher subsequent

Table 3. Adjusted hazard ratios for predictors of acute care use within 30 days postdischarge stratified by age group*

	All ages (n = 8,606)	<65 years (n = 5,280)	≥65 years (n = 3,326)
Timely follow-up (within 30 days)	1.11 (1.02–1.21)†	1.06 (0.96–1.17)	1.27 (1.09–1.47)†
Age at admission (per decade)	0.89 (0.87–0.92)†	0.83 (0.79–0.87)†	1.07 (0.96–1.19)
Female	0.96 (0.86–1.08)	1.02 (0.88–1.19)	0.91 (0.75–1.10)
Race and ethnicity			
White	Ref.	Ref.	Ref.
Black	0.99 (0.90–1.09)	1.03 (0.92–1.15)	0.91 (0.75–1.12)
Hispanic	0.96 (0.84–1.09)	0.98 (0.85–1.14)	0.90 (0.68–1.19)
Other	0.93 (0.75–1.15)	0.95 (0.77–1.17)	0.85 (0.53–1.35)
Rural–urban commuting area			
Suburban	Ref.	Ref.	Ref.
Urban	1.08 (0.93–1.24)	1.05 (0.89–1.24)	1.07 (0.84–1.36)
Large rural	1.15 (0.97–1.37)	1.13 (0.92–1.38)	1.20 (0.91–1.60)
Small town/rural	1.04 (0.85–1.26)	1.06 (0.84–1.35)	0.98 (0.70–1.36)
ADI disadvantage (per decile)	1.03 (1.01–1.04)†	1.04 (1.02–1.06)†	1.00 (0.97–1.02)
Medicaid beneficiary	1.12 (1.02–1.22)†	1.05 (0.95–1.17)	1.26 (1.06–1.50)†
Disability as Medicare reason	0.86 (0.79–0.94)†	0.93 (0.80–1.08)	1.02 (0.88–1.18)
HCC comorbidity score	1.09 (1.07–1.11)†	1.09 (1.07–1.11)†	1.11 (1.07–1.15)†
Hospital length of stay (per day)	1.02 (1.01–1.03)†	1.01 (1.00–1.02)†	1.03 (1.02–1.05)†
Tobacco use disorder	1.05 (0.97–1.14)	1.07 (0.97–1.17)	1.10 (0.94–1.29)
Alcohol use disorder	1.25 (1.02–1.53)†	1.26 (1.02–1.56)†	1.27 (0.68–2.36)
Drug or opioid use disorder	1.41 (1.28–1.56)†	1.45 (1.30–1.62)†	1.05 (0.79–1.42)
Anxiety	1.13 (1.04–1.22)†	1.17 (1.06–1.30)†	1.06 (0.93–1.22)
Depression	1.19 (1.10–1.29)†	1.17 (1.07–1.29)†	1.25 (1.08–1.46)†
Renal failure	1.06 (0.97–1.15)	1.08 (0.97–1.20)	0.96 (0.82–1.12)
Diabetes mellitus	1.05 (0.97–1.14)	1.07 (0.97–1.18)	1.12 (0.97–1.30)
Congestive heart failure	1.18 (1.08–1.29)†	1.17 (1.06–1.30)†	1.15 (0.98–1.34)

* Values are the adjusted hazard ratio (95% confidence interval). Acute care use defined as inpatient hospital readmission, inpatient observation stay, or emergency department visit. ADI = Area Deprivation Index; HCC = hierarchical condition category; Ref. = reference.
† Statistically significant.

acute care. Therefore, developing strategies, guidelines, and best practices for postdischarge management tailored for both primary care and rheumatology providers could be important steps in improving lupus care. Likewise, both rheumatology and primary care could potentially better leverage transitional care practices, including medication reconciliation and connection to support services.

Systems-level, multidisciplinary team-based interventions, such as transitional care management, care coordination, and patient navigation, can improve postdischarge care by addressing the complex medical, socioeconomic, and self-management factors that contribute to acute care use (43,47,48). In 1 study, a team-based transitional care intervention for high-risk lupus patients in China substantially reduced 30-day readmissions from 21.3% to 4.7% (11). Likewise, a study in New York state showed a 50% reduction in 30-day readmissions after implementing team-based care coordination (43). The utility of telemedicine visits, home health, and outreach clinics could also be explored to overcome barriers to postdischarge follow-up. The impact of an intervention on lupus care quality, health outcomes, and health disparities should inform implementation (22,48).

Beyond the strengths of this study, including a large national sample of hospitalizations in patients with lupus with robust geolinked indices of neighborhood disadvantage, we also acknowledge limitations. First, given that administrative claims data only reflect completion or noncompletion of follow-up, we cannot assess scheduled versus missed or canceled visits. This gap could lead to overestimation of the effect of follow-up on mortality as patients without follow-up may have been too sick to attend clinic visits. However, there was no difference in the HCC comorbidity score or hospital length of stay between those with or without follow-up within 30 days after adjustment, suggesting that there were not residual systematic health status differences between the groups. Second, visits with nephrology providers were not included in our measure of timely follow-up, although many lupus patients have kidney disease. Outcomes of acute care and mortality were only measured within the first 30 days postdischarge, and results may differ over longer periods of time. Next, for patients age <65 years, the Medicare sample may not be completely generalizable to all US lupus patients given oversampling of patients with end-stage renal disease and disability. Nevertheless, up to a third of patients with lupus have public insurance, and over half of lupus hospitalizations are covered by Medicare, which underscores the policy relevance of our study (8,9). While we were unable to control specifically for lupus severity, duration, or treatments in this administrative sample, validated definitions for comorbidities and health care utilization using

Table 4. Adjusted hazard ratios for predictors of mortality within 30 days postdischarge stratified by age group*

	All ages (n = 8,606)	<65 years (n = 5,280)	≥65 years (n = 3,326)
Timely follow-up (within 30 days)	0.53 (0.33–0.84)†	0.91 (0.46–1.80)	0.35 (0.19–0.67)†
Age at admission (per decade)	1.36 (1.17–1.59)†	1.20 (0.87–1.65)	1.53 (1.00–2.32)†
Female	0.94 (0.54–1.62)	0.82 (0.33–2.03)	1.03 (0.49–2.15)
Race and ethnicity			
White	Ref.	Ref.	Ref.
Black	0.99 (0.59–1.67)	1.88 (0.89–3.99)	0.33 (0.10–1.07)
Hispanic	0.77 (0.34–1.76)	1.04 (0.35–3.11)	0.66 (0.19–2.03)
Other	1.12 (0.39–3.20)	1.47 (0.32–6.84)	1.28 (0.29–5.54)
Rural–urban commuting area			
Suburban	Ref.	Ref.	Ref.
Urban	1.16 (0.56–2.41)	0.57 (0.19–1.65)	1.73 (0.60–4.95)
Large rural	0.98 (0.38–2.52)	0.62 (0.15–2.64)	1.33 (0.37–4.79)
Small town/rural	1.68 (0.66–4.28)	1.15 (0.26–5.11)	2.26 (0.65–7.87)
ADI disadvantage (per decile)	1.02 (0.95–1.09)	1.05 (0.95–1.15)	1.02 (0.93–1.13)
Medicaid beneficiary	0.89 (0.54–1.47)	0.86 (0.39–1.89)	0.95 (0.46–1.98)
Disability as Medicare reason	0.72 (0.48–1.07)	1.46 (0.49–4.35)	0.70 (0.37–1.32)
HCC comorbidity score	1.20 (1.12–1.30)†	1.20 (1.06–1.35)†	1.22 (1.10–1.35)†
Hospital length of stay (per day)	1.05 (1.03–1.08)†	1.04 (1.01–1.07)†	1.08 (1.04–1.12)†
Tobacco use disorder	1.34 (0.84–2.13)	1.06 (0.54–2.10)	1.90 (1.02–3.53)†
Alcohol use disorder	2.11 (0.80–5.52)	2.50 (0.77–8.12)	2.13 (0.37–12.22)
Drug or opioid use disorder	0.57 (0.26–1.26)	0.41 (0.15–1.15)	1.09 (0.32–3.70)
Anxiety	1.27 (0.82–1.96)	1.22 (0.57–2.61)	1.38 (0.81–2.34)
Depression	0.64 (0.40–1.02)	0.81 (0.38–1.74)	0.55 (0.29–1.03)
Renal failure	1.15 (0.74–1.78)	1.48 (0.63–3.47)	0.94 (0.53–1.65)
Diabetes mellitus	0.93 (0.60–1.43)	0.73 (0.34–1.54)	1.28 (0.76–2.17)
Congestive heart failure	1.07 (0.65–1.76)	1.44 (0.66–3.15)	0.76 (0.37–1.55)

* Values are the adjusted hazard ratio (95% confidence interval). ADI = Area Deprivation Index; HCC = hierarchical condition category; Ref. = reference.
† Statistically significant.

CMS's HCC risk adjustment score were incorporated in multivariate modeling. Yet residual confounding based on health status may still be present in this observational analysis.

In addition, reported race and ethnicity in Medicare data have known issues with misclassification (30,31). While we used the more valid Research Triangle Institute measure, approximately 4% of patients may be misclassified, particularly American Indian and Alaska Native, and Asian and Pacific Islander patients (31). Though the number of deaths in the study was small, limiting power to identify factors associated with mortality, overall findings suggest that improving follow-up may decrease mortality in the 30-day postdischarge period. Future studies should examine multipayor cohorts, especially in younger patients with lupus, and evaluate associations with mortality risk in cohorts with a greater number of outcome events.

We found that postdischarge follow-up was associated with higher level of subsequent acute care use in older lupus patients but was also associated with reduced mortality, 65% lower within 30 days. This finding suggests the importance of ambulatory follow-up for positive health outcomes in lupus. Importantly, timely follow-up among lupus patients was approximately 10% lower than rates reported in heart failure and other conditions with high readmission risk (17,18,25,41). Patients residing in highly disadvantaged neighborhoods and rural areas had even lower odds of follow-up, demonstrating disparities. Future studies should identify barriers to timely postdischarge follow-up among lupus patients and test interventions to overcome these barriers in primary and rheumatology care, not only to improve care and disease outcomes but also to reduce health disparities in lupus.

ACKNOWLEDGMENT

The authors would like to thank Monica Messina for manuscript formatting.

AUTHOR CONTRIBUTIONS

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

Study conception and design. Schletzbaum, Astor, Yu, Bartels.

Acquisition of data. Schletzbaum, Powell, Gilmore-Bykovskyi, Kaiksow, Sheehy, Kind, Bartels.

Analysis and interpretation of data. Schletzbaum, Sweet, Astor, Yu, Powell, Gilmore-Bykovskyi, Kind, Bartels.

REFERENCES

1. Bartels CM, Chodara A, Chen Y, et al. One quarter of Medicare hospitalizations in patients with systemic lupus erythematosus readmitted within thirty days. Semin Arthritis Rheum 2021;51:477–85.

2. Chakravarty EF, Bush TM, Manzi S, et al. Prevalence of adult systemic lupus erythematosus in California and Pennsylvania in 2000: estimates obtained using hospitalization data. Arthritis Rheum 2007;56:2092–6.
3. Nangit A, Lin C, Ishimori ML, et al. Causes and predictors of early hospital readmission in systemic lupus erythematosus. J Rheumatol 2018;45:929–33.
4. Elixhauser A, Steiner C. Healthcare cost and utilization project, statistical brief #153: readmissions to U.S. hospitals by diagnosis, 2010. 2013. URL: https://www.hcup-us.ahrq.gov/reports/statbriefs/sb153.pdf.
5. Kan HJ, Song X, Johnson BH, et al. Healthcare utilization and costs of systemic lupus erythematosus in Medicaid. Biomed Res Int 2013;2013:808391.
6. Anandarajah AP, Luc M, Ritchlin CT. Hospitalization of patients with systemic lupus erythematosus is a major cause of direct and indirect healthcare costs. Lupus 2017;26:756–61.
7. Pons-Estel GJ, Alarcon GS, Scofield L, et al. Understanding the epidemiology and progression of systemic lupus erythematosus. Semin Arthritis Rheum 2010;39:257–68.
8. Ward MM. Hospital experience and mortality in patients with systemic lupus erythematosus: which patients benefit most from treatment at highly experienced hospitals? J Rheumatol 2002;29:1198–206.
9. Panopalis P, Gillis JZ, Yazdany J, et al. Frequent use of the emergency department among persons with systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2010;62:401–8.
10. McIlvennan CK, Eapen ZJ, Allen LA. Hospital readmissions reduction program. Circulation 2015;131:1796–803.
11. Xie X, Song Y, Yang H, et al. Effects of transitional care on self-care, readmission rates, and quality of life in adult patients with systemic lupus erythematosus: a randomized controlled trial. Arthritis Res Ther 2018;20:184.
12. Wasfy JH, Zigler CM, Choirat C, et al. Readmission rates after passage of the Hospital Readmissions Reduction Program: a pre-post analysis. Ann Intern Med 2017;166:324–31.
13. Bindman AB, Cox DF. Changes in health care costs and mortality associated with transitional care management services after a discharge among Medicare beneficiaries. JAMA Intern Med 2018;178:1165–71.
14. Centers for Medicare and Medicaid Services. Medicare Learning Network, Transitional care management services. 2022. URL: https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/Downloads/Transitional-Care-Management-Services-Fact-Sheet-ICN908628.pdf.
15. Hernandez AF, Greiner MA, Fonarow GC, et al. Relationship between early physician follow-up and 30-day readmission among Medicare beneficiaries hospitalized for heart failure. JAMA 2010;303:1716–22.
16. Lee KK, Yang J, Hernandez AF, et al. Post-discharge follow-up characteristics associated with 30-day readmission after heart failure hospitalization. Med Care 2016;54:365–72.
17. McAlister FA, Youngson E, Bakal JA, et al. Impact of physician continuity on death or urgent readmission after discharge among patients with heart failure. CMAJ 2013;185:E681–9.
18. Sharma G, Kuo YF, Freeman JL, et al. Outpatient follow-up visit and 30-day emergency department visit and readmission in patients hospitalized for chronic obstructive pulmonary disease. Arch Intern Med 2010;170:1664–70.
19. Field TS, Ogarek J, Garber L, et al. Association of early post-discharge follow-up by a primary care physician and 30-day rehospitalization among older adults. J Gen Intern Med 2015;30:565–71.
20. Lin MP, Burke RC, Orav EJ, et al. Ambulatory follow-up and outcomes among Medicare beneficiaries after emergency department discharge. JAMA Netw Open 2020;3:e2019878.
21. Alvidrez J, Castille D, Laude-Sharp M, et al. The National Institute on Minority Health and Health Disparities research framework. Am J Public Health 2019;109:S16–20.
22. Wasserman J, Palmer RC, Gomez MM, et al. Advancing health services research to eliminate health care disparities. Am J Public Health 2019;109:S64–9.
23. Hanly JG, Thompson K, Skedgel C. Identification of patients with systemic lupus erythematosus in administrative healthcare databases. Lupus 2014;23:1377–82.
24. Yazdany J, Marafino BJ, Dean ML, et al. Thirty-day hospital readmissions in systemic lupus erythematosus: predictors and hospital- and state-level variation. Arthritis Rheumatol 2014;66:2828–36.
25. DeLia D, Tong J, Gaboda D, et al. Post-discharge follow-up visits and hospital utilization by Medicare patients, 2007–2010. Medicare Medicaid Res Rev 2014;4:mmrr.004.02.a01.
26. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med 2009;360:1418–28.
27. Von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007;147:573–7.
28. Powell WR, Kaiksow FA, Kind AJ, et al. What is an observation stay? Evaluating the use of hospital observation stays in Medicare. J Am Geriatr Soc 2020;68:1568–72.
29. Centers for Medicare and Medicaid Services. Hospital-wide all-cause risk standardized mortality measure: Department of Health and Human Services, 2021. URL: https://cmit.cms.gov/CMIT_public/ViewMeasure?MeasureId=3199.
30. Jarrin OF, Nyandege AN, Grafova IB, et al. Validity of race and ethnicity codes in Medicare administrative data compared with gold-standard self-reported race collected during routine home health care visits. Med Care 2020;58:e1–8.
31. Grafova IB, Jarrin OF. Beyond black and white: mapping misclassification of Medicare beneficiaries race and ethnicity. Med Care Res Rev 2021;78:616–26.
32. Eicheldinger C, Bonito A. More accurate racial and ethnic codes for Medicare administrative data. Health Care Financ Rev 2008;29:27–42.
33. Bonito AJ, Bann, C, Eicheldinger C. Agency for Healthcare Research and Quality publication #08-0029-EF. Creation of new race-ethnicity codes and socioeconomic status (SES) indicators for Medicare beneficiaries: final report, sub-task 2. Rockville (MD): ARHQ; 2008.
34. Singh GK. Area deprivation and widening inequalities in US mortality, 1969–1998. Am J Public Health 2003;93:1137–43.
35. Kind AJ, Jencks S, Brock J, et al. Neighborhood socioeconomic disadvantage and 30-day rehospitalization: a retrospective cohort study. Ann Intern Med 2014;161:765–74.
36. WWAMI RUCA Rural Health Research Center. Map classifications, 2011. URL: http://depts.washington.edu/uwruca/ruca-maps.php.
37. Pope GC, Kautter J, Ellis RP, et al. Risk adjustment of Medicare capitation payments using the CMS-HCC model. Health Care Financ Rev 2004;25:119–41.
38. Elixhauser A, Steiner C, Harris DR, et al. Comorbidity measures for use with administrative data. Med Care 1998;36:8–27.
39. Centers for Medicare and Medicaid Services. Chronic conditions data warehouse: condition categories, 2020. URL: https://www2.ccwdata.org/web/guest/condition-categories.
40. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc 1999;94:496–509.
41. Fidahussein SS, Croghan IT, Cha SS, et al. Posthospital follow-up visits and 30-day readmission rates in chronic obstructive pulmonary disease. Risk Manag Healthc Policy 2014;7:105–12.

42. Kangovi S, Grande D. Hospital readmissions: not just a measure of quality. JAMA 2011;306:1796–7.
43. Anandarajah A. Designing an intervention to improve management of high-risk lupus patients through care coordination. Rheum Dis Clin North Am 2020;46:723–34.
44. Soucier RJ, Miller PE, Ingrassia JJ, et al. Essential elements of early post discharge care of patients with heart failure. Curr Heart Fail Rep 2018;15:181–90.
45. Metra M, Gheorghiade M, Bonow RO, et al. Postdischarge assessment after a heart failure hospitalization: the next step forward. Circulation 2010;122:1782–5.
46. Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines. J Am Coll Cardiol. 2013;62:e147–239.
47. Kangovi S, Barg FK, Carter T, et al. Challenges faced by patients with low socioeconomic status during the post-hospital transition. J Gen Intern Med 2014;29:283–9.
48. Centers for Medicare and Medicaid Services. Guide to reducing disparities in readmissions, 2018. URL: https://www.cms.gov/About-CMS/Agency-Information/OMH/Downloads/OMH_Readmissions_Guide.pdf.

Arthritis Care & Research
Vol. 75, No. 9, September 2023, pp 1897–1907
DOI 10.1002/acr.25101

American College of Rheumatology
Empowering Rheumatology Professionals

Core Recommendations for Osteoarthritis Care: A Systematic Review of Clinical Practice Guidelines

Brooke Conley,[1] Samantha Bunzli,[2] Jonathan Bullen,[3] Penny O'Brien,[1] Jennifer Persaud,[4] Tilini Gunatillake,[1] Michelle M. Dowsey,[1] Peter F. M. Choong,[1] and Ivan Lin[5]

Objective. To evaluate the quality of clinical practice guidelines (CPGs) for interventions in management of osteoarthritis (OA) and to provide a synthesis of high-quality CPG recommendations.

Methods. Five databases (OvidSP Medline, Cochrane, Cumulative Index to Nursing and Allied Health Literature [CINAHL], Embase, and the Physiotherapy Evidence Database [PEDro]) and 4 online guideline repositories were searched. CPGs for the management of OA were included if they were 1) written in English and published from January 2015 to February 2022, focused on adults age ≥18 years, and met the criteria of a CPG as defined by the Institute of Medicine; and 2) were rated as high quality on the Appraisal of Guidelines for Research and Evaluation II (AGREE II) instrument. CPGs for OA were excluded if they were available via institutional access only, only addressed recommendations for the system/organization of care and did not include interventional management recommendations, and/or included other arthritic conditions.

Results. Of 20 eligible CPGs, 11 were appraised as high quality and included in the synthesis. Of interest were the hip, knee, hand, and glenohumeral joints and/or polyarticular OA. Consistent recommendations were that care should be patient centered and include exercise, education, and weight loss (where appropriate). Nonsteroidal antiinflammatory drugs and surgical interventions were recommended for disabling OA that had not improved with nonsurgical care. Hand orthoses should be recommended for patients with hand OA.

Conclusion. This synthesis of high-quality CPGs for OA management offers health care providers with clear, simple guidance of recommended OA care to improve patient outcomes.

INTRODUCTION

Osteoarthritis (OA) is a degenerative joint disease that can affect any joint, but it most commonly occurs in the hip, knee, and hand (1,2). Symptoms often include joint pain, stiffness, and reduced range of movement (2). OA affects 303 million people worldwide, with prevalence expected to increase with aging populations and rising obesity rates globally (3,4). OA is a leading cause of pain and disability among adults worldwide and inflicts a significant burden on the individuals affected, including activity limitations and reduced quality of life (5,6). OA is associated with substantial direct health care costs due to health care visits, diagnostic procedures, medications and surgery, and indirect costs related to lost workplace productivity (4,6).

Clinical practice guidelines (CPGs) are a set of health care recommendations developed by reviewing scientific literature and consensus from an expert panel (7). The aim of CPGs is to guide health care decision-making, thereby reducing practice variability and improving patient outcomes (8,9). Several global CPGs have been published in recent years for the management of OA

Supported by the Australian Commonwealth Government (Arthritis Australia grant GA76063). Professor Dowsey's work was supported by the Australian National Health and Medical Research Council (Career Development Fellowship grant APP1122526) and the University of Melbourne (Dame Kate Campbell Fellowship). Professor Choong's work was supported by the Australian National Health and Medical Research Council (Practitioner Fellowship grant APP1154203).

[1]Brooke Conley, BExSci, MPhysio, Penny O'Brien, BHealthSci (Hons), PhD, Tilini Gunatillake, BSc (Hons), PhD, Michelle M. Dowsey, BHealthSci, MEpi, PhD, Peter F. M. Choong, MBBS, MD, FRACS, FAOrthA, FAAHMS, MAICD: The University of Melbourne, Melbourne, Victoria, Australia; [2]Samantha Bunzli, BPhty (Hons), PhD: The University of Melbourne, Melbourne, Victoria, Australia, Griffith University, Nathan, Queensland, Australia, and Royal Brisbane and Women's Hospital, Brisbane, Queensland, Australia; [3]Jonathan Bullen, Bus.Info.Sys., GradDipPsychStudies, PhD: Curtin University, Perth, Western Australia, Australia; [4]Jennifer Persaud, BSc (Hons) Physio, MSc Sports and Ex Med, GradCert Social Impact: Arthritis and Osteoporosis Western Australia and Sir Charles Gairdner Hospital, Perth, Western Australia, Australia; [5]Ivan Lin, BSc(Physio), M. Manip. Ther., PhD, FACP: University of Western Australia and Geraldton Regional Aboriginal Medical Service, Geraldton, Western Australia, Australia.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.25101&file=acr25101-sup-0001-Disclosureform.pdf.

Address correspondence via email to Brooke Conley, BExSci, MPhysio, at bccon@student.unimelb.edu.au.

Submitted for publication October 25, 2022; accepted in revised form February 7, 2023.

SIGNIFICANCE & INNOVATIONS

- Eleven clinical practice guidelines for osteoarthritis (OA) were appraised as high quality.
- Consistent recommendations were that care should be patient centered and include exercise, education, and weight loss (where appropriate). Nonsteroidal antiinflammatory drugs and surgical interventions are recommended for disabling OA that had not improved with nonsurgical care. Hand orthoses should be recommended for patients with hand OA.
- To implement recommendations in practice, future priorities include identifying core skill sets and competencies among health care workers, developing training/education resources, and creating a framework to improve quality of OA care.

(10–29). These typically include nonpharmacologic (e.g., exercise and education), pharmacologic (e.g., acetaminophen and nonsteroidal antiinflammatory drugs [NSAIDs]), and surgical options such as total joint replacement (10–29). However, uptake of recommendations from CPGs into practice is variable, especially for first-line, nonpharmacologic treatments such as exercise (30). Gaps between evidence and practice may in part be due to a lack of clarity about what is being recommended and conflicting recommendations across different CPGs, a situation exacerbated when CPGs for OA are not developed rigorously and their recommendations are less trustworthy (30–36). As CPGs are costly to develop, this is an inefficient use of resources and further adds to the confusion for clinicians (37). To encourage the uptake of evidence and delivery of appropriate OA care, clinicians require clear, consistent management recommendations (31).

The purpose of this systematic review was to evaluate the quality of the CPGs for the management of OA and to provide a synthesis of high-quality CPG recommendations. By synthesizing recommendations across high-quality CPGs, the aim was to offer health care providers with clear, simple guidance of recommended OA care to improve patient outcomes.

MATERIALS AND METHODS

Search strategy and eligibility criteria. This systematic review was registered on the Open Science Framework (DOI 10.17605/OSF.IO/UB3Y7) and followed the Preferred Reporting Items for Systematic Reviews (PRISMA) guidelines. A database search of Medline, Cochrane, CINAHL, Embase, and the Physiotherapy Evidence Database (PEDro), and a further search of 4 online guideline repositories (Guidelines International Network, National Health and Medical Research Council, Agency for Health Care Research and Quality, and the National Institute for Health and Care Excellence) was conducted to identify all relevant CPGs. The search strategy was developed in consultation with a reference librarian. Medical subject headings and key words associated with CPGs (e.g., guideline*.mp. or Practice Guideline/ or Guideline/) and OA were used (see Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25101). Other arthritis conditions were included in the search terms as this systematic review is part of a wider body of work to inform arthritis management. Eligibility criteria are presented in Table 1. The search included CPGs published between January 2015 and December 2020 and was updated to include CPGs published between December 2020 and February 14, 2022. This cutoff date was selected as CPGs >5 years old may be out of date (7).

Protocol changes. In the original protocol, CPGs addressing 1 treatment modality (e.g., medication prescribing) were excluded. To improve comprehensiveness, the scope was expanded during the study selection phase to include all OA management options. Due to time that has elapsed since the original search, the original timeline (January 2015 and December 2020) was extended to include CPGs published up until February 2022.

Study selection. After importing search results into EndNote (Clarivate), duplicates were removed and titles/abstracts were uploaded into Covidence systematic review software (Veritas Health Innovation; available at www.covidence.org). Titles and abstracts were screened by 2 independent reviewers (BC and TG or IL); any disagreements were resolved through consensus discussion with a third reviewer. Following this, full texts were uploaded into Covidence and screened through the same process.

Data appraisal (quality assessment of guidelines). The Appraisal of Guidelines for Research and Evaluation II (AGREE II) instrument was used to assess CPG quality (38). This is an internationally validated tool that has been widely applied in CPG research (38,39). It consists of 23 items grouped into 6 domains: 1) scope and purpose; 2) stakeholder involvement; 3) rigor of development; 4) clarity of presentation; 5) applicability; and 6) editorial independence. Seven reviewers (BC, SB, JB, PO, JP, TG, and IL) were

Table 1. Clinical practice guidelines (CPGs) selection criteria

Inclusion criteria
Published between January 2015 and February 14, 2022
For the interventional management of osteoarthritis
For adults (individuals age ≥18 years)
Published in the English language or has a complete English language version available
Is a CPG, as defined by inclusion of a systematic review of the literature, and developed by an expert multidisciplinary panel (33)
Represents an original body of work, i.e., not solely an adaptation or systematic review of existing guidelines
Exclusion criteria
Does not include interventional management recommendations
Includes other arthritic conditions
Only addresses recommendations for the system/organization of care
Unavailable via institutional access, i.e., requires additional payment

provided with the AGREE II user manual and undertook the online AGREE II practice exercise to participate in CPG quality appraisal (38,40). In accordance with the AGREE II manual, each item was rated independently by 2 reviewers using a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree) (38). To calculate scores for each domain, the following formula was used: obtained score – minimum possible score / maximum possible score – minimum possible score (38). There is no uniform criterion for overall quality; the AGREE II developers recommend that research teams define their own criteria based on their own study context (38). For the purposes of this review, and consistent with previous reviews in musculoskeletal pain management, the authors defined a quality cutoff score of ≥60% of the maximum possible score in 3 domains deemed the most important for validity: stakeholder involvement (domain 2); rigor of development (domain 3); and editorial independence (domain 6) (36,37,41,42). CPGs that did not meet this definition were excluded.

Interrater agreement. The domain percentages and overall quality rating (%) were independently calculated for each reviewer. We defined acceptable interrater agreement as excellent with intraclass coefficient values of ≥80 and domain percentages and an overall quality rating of ≤20% difference between reviewers (43,44). Where variation of ≥20% between scores existed, a consensus discussion took place with a third reviewer engaged when necessary to agree on a rating (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25101).

Data extraction. Data extraction was performed by the first author (BC) using a purpose-designed Excel (Microsoft) spreadsheet. Extracted data comprised CPG characteristics (e.g., title, country of publication), methodology, and guideline topic target users (see Supplementary Table 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25101). From each CPG, extracted recommendations were ranked as either "should do," "could do," "do not do," or "uncertain" (see Supplementary Table 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25101). Recommendation ratings were consistent with language used in the CPGs and definitions from a previous musculoskeletal systematic review of CPGs (36) (Table 2). Language among the CPGs varied, although recommendations were ranked according to the same criteria, either the Grading of Recommendations Assessment, Development and Evaluation (GRADE) method (10,12,13,20,22,26,29), Oxford Centre for Evidence-Based Medicine standards (13,24), or PEDro scores (17). Extracted data and recommendation rankings were checked by 2 authors (SB and IL), and any discrepancies were resolved by consensus discussion between the 3 reviewers while consulting the original citation.

Narrative summary. A narrative summary was drafted by the first author (BC) and then reviewed and refined by 2 authors (SB and IL). The summary detailed how many CPGs reported on an intervention, what the recommendations involved, and how consistent/inconsistent recommendations were across CPGs regarding OA interventions (see Supplementary Table 4, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25101).

Table 2. Recommendation classification, definition, and examples of terminology for each classification*

Recommendation classification	Definition (37)	Examples of terminology from CPGs
Should do	"Should do" recommendations were those that the authors determined should be applied in all circumstances unless there is a rationale not to. These were based on strong evidence, for example, multiple high-quality studies reporting clinically relevant positive effects, benefits that outweigh risks, or when in the opinion of CPG development group members the benefits were unequivocal.	"Should do" (18) and "strongly recommend" (11)
Could do	"Could do" recommendations were those that the authors determined could be applied depending on the circumstances of individual patients. They were usually based on consistent evidence from multiple lesser-quality studies or 1 high-quality study and where benefits outweigh harms.	"Could be used" (18), "may be beneficial" (9,18), "can" and "consider" (26)
Do not do	"Do not do" recommendations were those for which the authors determined that there was strong evidence of no benefit and/or harms outweighing benefits.	"Are not recommended" (18), "do not recommend" (11), and "do not offer" (26)
Uncertain	"Uncertain" recommendations were those for which the authors determined that there was no recommendation for or against a practice because of incomplete or inconsistent research findings. Not all CPGs provided uncertain recommendations.	"Cannot recommend for or against" (9) and "unable to recommend either for or against" (11)

* CPG = clinical practice guideline.

RESULTS

Characteristics of included CPGs. Twenty CPGs met the eligibility criteria, and 11 were CPGs appraised as high quality and included (11–13,17–19,21,24,25,27,29) (Figure 1). Most CPGs were developed by medical societies (n = 9, 82%), while the remaining were developed by an expert panel (n = 2, 18%). Five CPGs were published by medical societies or expert panels in the US (10,11,19,25,29), 4 in Europe (13,21,24,27), 1 in Canada (17), and 1 in Australia (12). Of interest were the hip (n = 6), knee (n = 7), hand (n = 4), and glenohumeral (n = 1) joints and/or polyarticular OA (n = 1). Target users included health professionals, decision/policy makers, patients, their families, the pharmaceutical industry, health insurance companies, and those responsible for commissioning care (see Supplementary Table 2, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25101).

Quality of CPGs. The AGREE II quality assessment scores for each CPG are provided in Supplementary Table 1, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25101. The quality assessment results of each guideline included in the systematic review are provided in Table 3. Those that were excluded based on not achieving a high-quality cutoff score are presented in Supplementary Table 1. The quality of the included CPGs was assessed across the following 6 domains: scope and purpose (range 75–100%); stakeholder involvement (range 58–89%); rigor

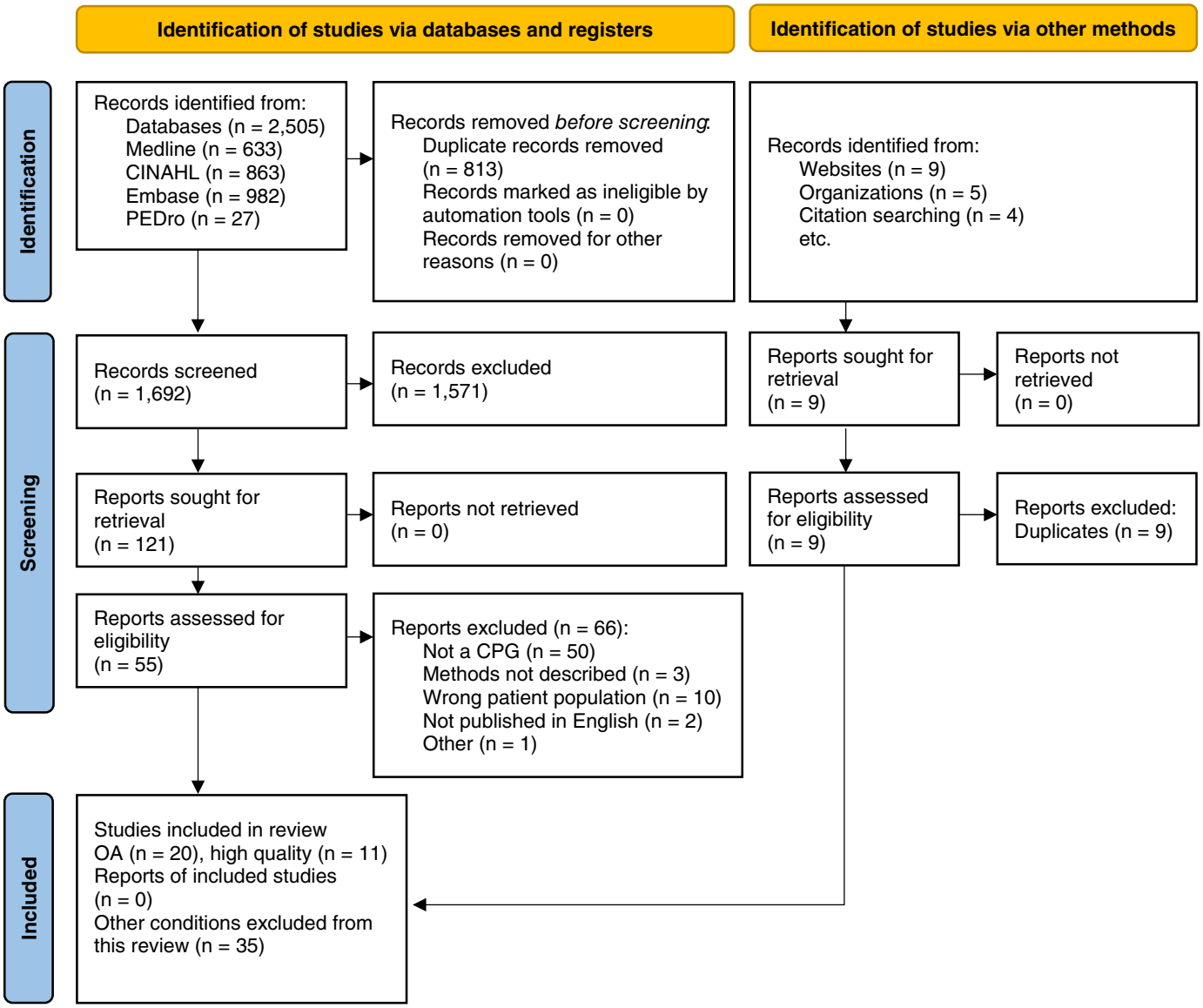

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 flow diagram for new systematic reviews that included searches of databases, registers, and other sources. CINAHL = Cumulative Index to Nursing and Allied Health Literature; CPG = clinical practice guideline; OA = osteoarthritis; PEDro = Physiotherapy Evidence Database.

of development (range 59–96%); clarity of presentation (range 53–100%); applicability (range 2–42%); and editorial independence (range 33–100%). The mean ± SD AGREE II scores for each item, domain, and overall scores across all guidelines are displayed in Supplementary Table 5, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25101. The domain with the lowest mean ± SD score was applicability (21.14% ± 15.0%), and the highest mean ± SD score was for scope and purpose (88.77% ± 9.8%).

Consensus recommendations ("should do"). After synthesis (see Supplementary Table 4, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25101), 7 "should do" recommendations were identified. The following recommendations were all found in at least 2 CPGs, where the majority strongly recommended the intervention.

Exercise. Eight CPGs strongly recommended strengthening, aerobic exercise, and tai chi exercise therapy for management of knee, hip, polyarticular, and/or hand OA (12,13,18,19,21,24,25,27). CPGs recommended several modes of exercise therapy, acknowledging that there is currently no consensus on the type of exercise that elicits the greatest benefit (12). Programs should be individualized and progressively overloaded with frequency, duration, and intensity consistent with the patient's preference and capability and the availability of local facilities (12,13).

Education. Five CPGs strongly recommended patient education for managing knee, hip, hand, and polyarticular OA (13,19,21,24,25). They recommended that education be an ongoing intervention that is patient centered and include information to enhance understanding about OA, its management options, education and training in exercise therapy, ergonomic principles, and pacing and assistive devices (13,21,24).

NSAIDs. Five CPGs strongly and 2 CPGs conditionally recommended the use of oral NSAIDs for people with knee, hip, hand, and/or polyarticular OA unless contraindicated (11–13,19,21,24,25). CPGs recommended that clinicians prescribe a low dose for a short period of time and discontinue if not effective, monitoring for side effects or adverse events (12,24).

Four CPGs strongly recommended the use of topical NSAIDS for knee, hip, and/or hand OA (13,19,24,25). Two CPGs conditionally recommended topical NSAIDS for patients with hand, knee, hip, and/or polyarticular OA and some comorbidities (21,25). One CPG was unable to recommend for or against the use of topical NSAIDS for people with knee and/or hip OA (12). However, the authors stated that it might be reasonable to trial topical NSAIDs for a short period and then discontinue use if not effective. Topical NSAIDs are seen to be safe and effective and should be recommended for older adults (>75 years) with only a few symptomatic joints (13). Clinicians should monitor for side effects or adverse events (12,21).

Weight loss. Four CPGs strongly recommended weight loss or management for people who are either overweight (body mass index [BMI] ≥25 kg/m^2) or obese (BMI ≥30 kg/m^2) with hip and/or knee OA (12,13,19,25). People with OA should be educated about the importance of maintaining a healthy body weight, while those who are overweight or obese should be encouraged to achieve a minimum weight loss target of 5.0–7.5% of body weight, with greater weight loss being linked to symptomatic benefits (12).

Hand orthosis. Three CPGs strongly recommended the use of a hand orthosis for OA of the carpometacarpal joint and conditionally recommended it for OA of other hand joints (13,24,25). Two CPGs stated that hand orthoses are suitable for both short-term and long-term use as they provide symptom relief, improve function, and prevent progression of degenerative changes (13,24).

Patient-centered care. Two CPGs strongly recommended that care be patient centered for people with OA of the knee, hip, and/or hand (13,24). This included shared decision-making between the patient and health professional and care that is individualized to the patient's circumstances.

Surgery. Two CPGs strongly recommended considering surgery for people with hip, knee, and hand OA in certain circumstances (13,24). The patient should have radiographic evidence of OA, marked disability, and reduced quality of life, and other treatment modalities should have been unsuccessful in relieving pain (13,24).

Consensus recommendations ("could do"). The following recommendations were found in at least 2 CPGs in which the majority conditionally recommended (or where there was an even number of strongly and conditionally recommended recommendations) that these could apply in a given patient's circumstances: balance exercises; yoga; weight management and exercise; cognitive behavioral therapy; assistive devices; ultrasound-guided injections; duloxetine; and glucocorticoid injections for knee and hand OA. In surgical contexts, preoperative physical therapy and postoperative physical therapy or exercise can be recommended after joint replacement surgery. General and neuraxial anesthesia and tranexamic acid could be considered during surgery (see Supplementary Table 4, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25101).

Consensus recommendations ("do not do"). After synthesis (see Supplementary Table 4), 5 "do not do" recommendations were identified (Table 4). The following recommendations were found in at least 2 CPGs, where the majority recommended against the intervention.

Therapeutic ultrasound and pharmacologic interventions (bisphosphonates, colchicine, methotrexate, diacerein). Three CPGs recommended against the use of therapeutic ultrasound for people with knee, hip, and/or polyarticular OA (12,21,27). Two CPGs recommended against the use of bisphosphonates, colchicine, hydroxychloroquine, and methotrexate in people with knee, hip, and/or hand OA (12,25). Similarly, 1 CPG recommended against the use of biologic disease-modifying

Table 3. Clinical practice guidelines included in the systematic review*

Author, year (ref.)	Title/type	Domain 1, scope and purpose	Domain 2, stakeholder involvement	Domain 3, rigor of development	Domain 4, clarity of presentation	Domain 5, applicability	Domain 6, editorial independence	Over all assessments core	Domains 2,3, and 6 combined value
RACGP, 2018 (12)	RACGP	97	58	81	89	35	79	83	73
Bannuru et al, 2019 (21)	OARSI	89	72	59	89	8	100	58	77
Kolasinski et al, 2020 (25)	ACR	97	89	85	86	21	63	75	79
AAOS, 2020 (10)	AAOS (GH joint)	94	61	90	69	15	92	75	81
AAOS, 2017 (11)	AAOS (hip OA)	97	61	92	53	42	71	75	75
Ariani et al, 2019 (13)	ISR	81	64	65	72	2	67	58	65
Kloppenburg et al, 2019 (24)	EULAR	78	81	77	81	6	79	67	79
Van Doormaal et al, 2020 (27)	–	75	67	81	75	42	33	58	60
Brosseau et al, 2018 (17)	OTTAWA	75	61	75	56	6	92	92	76
AAOS, 2015 (29)	AAOS (knee OA surgical)	100	81	96	100	21	100	92	92
AAOS, 2021 (19)	AAOS (knee OA non-arthroplasty)	94	61	96	100	35	92	92	83

* Values are the domain scores in % on the Appraisal of Guidelines for Research and Evaluation II (AGREE II) instrument. AAOS = American Academy of Orthopaedic Surgeons; ACR = American College of Rheumatology; GH = glenohumeral; ISR = Italian Society for Rheumatology; OA = osteoarthritis; OARSI = Osteoarthritis Research Society International; OTTAWA = Ottawa Panel; RACGP = Royal Australian College of General Practitioners; Ref. = reference.

antirheumatic drugs for people with hand OA (24). Two CPGs recommended against the use of diacerein for people with knee and/or hip OA (12,21).

Glucosamine and chondroitin combined. Two CPGs recommended against the use of glucosamine and chondroitin for knee, hip, and polyarticular OA (12,21). One CPG was unable to recommend for or against their combined use in people with OA of the glenohumeral joint (10). Moreover, 1 CPG conditionally recommended this intervention for people with knee OA, noting that further research is warranted to determine structural effects, patients' suitability, and cost-to-benefit ratio (13).

Postsurgical continuous passive motion (CPM) and postsurgical cryotherapy devices. Two CPGs recommended against the use of CPM after total joint replacement for patients with knee and/or hip OA, as research found no improvement in outcomes (27,29). One CPG recommended against the use of cryotherapy devices for patients after total knee arthroplasty (TKA) (29). In contrast, 1 CPG conditionally recommended the use of cryotherapy or cold packs following total shoulder replacement while acknowledging that this decision was based on the opinion of the working group and not strong/reliable evidence (10).

Table 4. Consensus recommendations*

Should do
Exercise therapies (strengthening, aerobics, and/or tai chi)
Education
Weight loss
Hand orthosis
Patient-centered care
Nonsteroidal antiinflammatory drugs (oral and topical)
Surgery
Could do
Balance exercises
Yoga
Assistive devices
Weight management and exercise
Cognitive behavioral therapy
Glucocorticoid injection (knee and hand OA)
Ultrasound-guided injections
Duloxetine
Preoperative physical therapy
Tranexamic acid
Neuraxial anesthesia
Pre- and postoperative physical therapy
Do not do
Therapeutic ultrasound
Bisphosphonates
Colchicine
Methotrexate
Diacerein
Glucosamine and chondroitin combined (hip and polyarticular OA)
Postsurgical continuous passive motion and cryotherapy devices
No consensus
Aquatic therapy
Balneotherapy
Manual therapy
Acupuncture
Massage therapy
Dry needling
Heat and cold therapy
Electrotherapy
Taping and braces
Shoe orthotics
Footwear
Topical capsaicin
Glucocorticoid injection (hip and polyarticular OA)
Intraarticular hyaluronic acid injections
Platelet-rich plasma injections
Stem cell injection
Acetaminophen
Oral opioids
Glucosamine and chondroitin, individually or combined (GH joint OA)
Nutraceuticals

* GH = glenohumeral; OA = osteoarthritis.

Recommendations with no consensus. The following were conflicting recommendations found in at least 2 CPGs: aquatic therapy; balneotherapy; massage therapy; manual therapy; acupuncture; dry needling; heat and cold therapy; electrotherapy; taping and braces; shoe orthotics; footwear; opioids; injections; topical capsaicin; glucosamine and chondroitin individually or combined for OA of the glenohumeral joint; acetaminophen; and nutraceuticals (see Supplementary Table 4).

DISCUSSION

Following quality assessment, 9 CPGs were rated as low quality, and 11 CPGs were high quality and included in the final synthesis. Overall, CPGs recorded the highest score for the AGREE II domain "scope and purpose" and the lowest score for the domain "applicability." This is consistent with the findings of similar systematic reviews (37,45). The AGREE II "applicability" domain assesses whether CPGs provide advice and/or tools for how to apply the guideline in practice, considers the facilitators, barriers, and resource implications, and includes monitoring and/or auditing criteria (38). Poor applicability has been identified as a barrier to the uptake of CPG recommendations into practice (37). Given that developing CPGs is expensive, development of fewer, higher-quality CPGs that focus on implementation (as reflected in higher scoring in the applicability domain on the AGREE II tool) is recommended.

Recommendations from 11 high-quality CPGs were that first-line care should be patient centered and include exercise therapy, patient education, and weight loss (if appropriate). These interventions can be beneficial in reducing pain and in improving function, performance, and quality of life outcomes (46–49). This should be followed by pharmacologic strategies such as NSAIDs in oral or topical form before considering surgical interventions as second- and third-line care. For people with hand OA, orthosis should be used for symptom relief and improved function and to prevent progression of degenerative changes (13,24). This synthesis of recommendations provides evidence-based guidance for clinicians on what should be delivered for best practice in OA care. These recommendations could also be used as a minimum

standard for health services to assess OA care and to provide the basis for clear consumer information about recommended OA management.

We identified a substantial number of recommendations that were inconsistent between CPGs, which may contribute to confusion among clinicians and to varied management. For example, manual therapy recommendations were inconsistent across the CPGs, with a majority recommending against, yet these are still widely used in clinical practice (50). Similarly, the majority of CPGs recommended against opioids, although 2 CPGs recommended that opioids can be considered in particular circumstances, when pain is severe or if patients do not respond, are intolerant, or contraindicated to NSAIDs, or when other alternatives have been exhausted (13,25). Despite this, opioids are often prescribed for persistent musculoskeletal pain conditions, including OA, and opioid-related harms are of increasing concern (51,52). Additional conflicting recommendations included acupuncture/dry needling, shoe orthotics, taping/braces, glucosamine and chondroitin, and injection therapies, e.g., platelet-rich plasma, stem cell, and intra-articular hyaluronic acid (for a comprehensive list of conflicting recommendations, see Supplementary Table 4, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25101).

Many of these recommendations are routinely utilized in clinical care. Further high-quality trials are needed to determine their efficacy and/or their suitability for certain subgroups of individuals with OA in order to guide clinicians' practice. CPG recommendations vary, potentially due to differences in evidence included based on the year of CPG publication, assessment of evidence quality, and involvement of expert panels or societies members. For example, 1 CPG recommended against heat therapy, while another CPG recommended that it can be considered as an adjunctive management option for people with hip and/or knee OA (12,27). Both CPGs acknowledged limited evidence supporting this intervention; however, interpretation of evidence by the respective guideline development groups led to conflicting recommendations.

While exercise, weight management, and education were supported across the CPGs and have been recommended as first-line interventions for almost 2 decades, translation into practice remains an issue (53). In Western health care settings including Australia, Europe, the UK, and the US, a majority of patients do not receive care consistent with CPGs (54). Conservative management interventions are often overlooked in favor of pharmacologic and surgical care despite being associated with higher financial costs and risks (e.g., medication side effects or surgical complications) (6,30,55,56). Globally, utilization of exercise and education is low, while pharmacologic therapy and surgical referrals are common (54,57). In Australia, joint replacement surgeries are a substantial cost to the health care system, estimated at between $19,000 and $30,000 (Australian; between $13,000 and $20,600 US dollars) per patient for total knee or hip replacement, resulting in an expenditure of $1.2 billion (Australian) annually on both public and private hospital services (58–60); similar findings have been documented in the UK and the US (61,62). Surgery is a successful and cost-effective intervention for people with end-stage hip and knee OA, although overuse of surgery in patients who could benefit from conservative care remains a challenge (63–66).

Implementation of high-value care such as exercise and weight loss is needed (67). One way is through OA management programs such as OA models of care that operationalize what and how recommended care should be delivered (68). In order to achieve better care, priorities include training/education of OA health care workers, identifying core skill sets and competencies, developing resources, and creating a framework to improve quality of care (69). Outcomes from models of care suggest that this has been an effective way to translate evidence into practice, although definitive evidence for OA management is currently lacking (70). Structured exercise therapies, with or without education and dietary interventions, are cost effective and clinically effective (71). Implementation research that operationalizes recommended care, especially for populations who experience a higher burden of OA, including low- and middle-income countries and First Nations people, is a pressing future priority (72,73). We excluded CPGs that were not published in the English language and that addressed assessment and/or diagnosis of OA without management or treatment recommendations. It is possible that we may have overlooked other CPGs containing recommendations related to OA care. To mitigate this risk, all authors checked the list of full-text CPGs to augment the search process, including authors who are expert clinician researchers in the field of OA (MMD and PC).

Strengths of this systematic review include the involvement of a multidisciplinary team and the use of the AGREE II tool. The research team defined high-quality CPGs as ≥60% in the 3 domains of interest on the AGREE II instrument. These domains are consistent with other high-quality musculoskeletal reviews (36), while 60% is supported by other arthritis and osteoporosis reviews (41,42). Grading of interventions and consensus statements (e.g., "should do," "could do," "do not do," or "unsure") were based on the language used in CPGs and required interpretation by the research team. Consensus statements were cross-checked by 2 authors (SB and IL) to mitigate the risk of misinterpretation. It is important to acknowledge that the majority of the literature regarding OA management is based on hip and knee OA, often neglecting OA in other joints. For transparency, we have listed the affected joint for each recommendation in Supplementary Tables 3 and 4, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25101.

In conclusion, 7 consistent "should do" recommendations were identified across the 11 CPGs. Exercise therapy, education, and weight loss (if relevant) should be recommended for people with OA before considering pharmacologic or surgical

interventions, with care being patient centered. Hand orthosis should be considered for those with hand OA. These core tenets of OA care can be used by health care providers to improve consistency and quality of OA care.

ACKNOWLEDGMENT

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

AUTHOR CONTRIBUTIONS

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

Acquisition of data. Conley, Gunatillake.

Analysis and interpretation of data. Conley, Bunzli, Bullen, O'Brien, Persaud, Gunatillake, Dowsey, Choong, Lin.

REFERENCES

1. Charlesworth J, Fitzpatrick J, Perera NK, et al. Osteoarthritis: a systematic review of long-term safety implications for osteoarthritis of the knee. BMC Musculoskeletal Disord 2019;20:151.
2. Fu K, Robbins SR, McDougall JJ. Osteoarthritis: the genesis of pain. Rheumatology (Oxford) 2018;57 Suppl 4:iv43–50.
3. Global Burden of Disease 2017 Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017 [correction published in Lancet 2019;393:132] [correction published in Lancet 2019; 393:e44]. Lancet 2018;392:1923–94.
4. Safiri S, Kolahi AA, Smith E, et al. Global, regional and national burden of osteoarthritis 1990–2017: a systematic analysis of the Global Burden of Disease Study 2017. Ann Rheum Dis 2020;79:819–28.
5. National Institute for Health and Care Excellence. Clinical guideline. Osteoarthritis: care and management. 2014;CG177:1–27. URL: https://www.nice.org.uk/guidance/cg177.
6. Hunter DJ, Schofield D, Callander E. The individual and socioeconomic impact of osteoarthritis. Nat Rev Rheumatol 2014;10:437–41.
7. Graham R, Mancher M, Miller Wolman D, et al, editors. Institute of Medicine Committee on standards for developing trustworthy clinical practice guidelines: clinical practice guidelines we can trust [review]. Washington (DC): National Academies Press; 2011.
8. Ng JY, Mohiuddin U, Azizudin AM. Clinical practice guidelines for the treatment and management of low back pain: a systematic review of quantity and quality. Musculoskelet Sci Pract 2021;51:102295.
9. Romão VC, Talarico R, Scirè CA, et al. Sjögren's syndrome: state of the art on clinical practice guidelines. RMD Open 2018;4 Suppl 1: e000789.
10. American Academy of Orthopaedic Surgeons. Management of glenohumeral joint osteoarthritis evidence based clinical practice guideline. 2020. URL: https://www.aaos.org/globalassets/quality-and-practice-resources/glenohumeral/gjo-cpg.pdf.
11. American Academy of Orthopaedic Surgeons. Management of osteoarthritis of the hip evidence based clinical practice guideline. 2017. URL: https://www.aaos.org/quality/quality-programs/lower-extremity-programs/osteoarthritis-of-the-hip/.
12. Royal Australian College of General Practitioners. Guideline for the management of knee and hip osteoarthritis. 2nd ed. East Melbourne (Australia): RACGP; 2018. URL: https://www.racgp.org.au/download/Documents/Guidelines/Musculoskeletal/guideline-for-the-management-of-knee-and-hip-oa-2nd-edition.pdf.
13. Ariani A, Manara M, Fioravanti A, et al. The Italian Society for Rheumatology clinical practice guidelines for the diagnosis and management of knee, hip and hand osteoarthritis. Reumatismo 2019;71:5–21.
14. Brosseau L, Taki J, Desjardins B, et al. The Ottawa Panel clinical practice guidelines for the management of knee osteoarthritis. Part one: introduction, and mind-body exercise programs. Clin Rehabil 2017; 31:582–95.
15. Brosseau L, Taki J, Desjardins B, et al. The Ottawa Panel clinical practice guidelines for the management of knee osteoarthritis. Part three: aerobic exercise programs. Clin Rehabil 2017;31:612–24.
16. Brosseau L, Taki J, Desjardins B, et al. The Ottawa Panel clinical practice guidelines for the management of knee osteoarthritis. Part two: strengthening exercise programs. Clin Rehabil 2017;31:596–611.
17. Brosseau L, Thevenot O, MacKiddie O, et al. The Ottawa Panel guidelines on programmes involving therapeutic exercise for the management of hand osteoarthritis. Clin Rehabil 2018;32:1449–71.
18. Brosseau L, Wells, GA, Pugh AG, et al. Ottawa Panel evidence-based clinical practice guidelines for therapeutic exercise in the management of hip osteoarthritis. Clin Rehabil 2016;30:935–46.
19. American Academy of Orthopaedic Surgeons. Management of osteoarthritis of the knee (nonarthroplasty) evidence-based clinical practice guideline. 2021. URL: https://www.aaos.org/oak3cpg.
20. Ayhan FF, Sunar İ, Umay E, et al. The Turkish League Against Rheumatism recommendations for the management of hand osteoarthritis under guidance of the current literature and 2018 European League Against Rheumatism recommendations. Arch Rheumatol 2020;35: 309–20.
21. Bannuru RR, Osani MC, Vaysbrot EE, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthritis Cartilage 2019;27:1578–89.
22. Bruyère O, Honvo G, Veronese N, et al. An updated algorithm recommendation for the management of knee osteoarthritis from the European Society for Clinical and Economic Aspects of Osteoporosis, Osteoarthritis and Musculoskeletal Diseases (ESCEO). Semin Arthritis Rheum 2019;49:337–50.
23. Eymard F, Ornetti P, Maillet J, et al. Intra-articular injections of platelet-rich plasma in symptomatic knee osteoarthritis: a consensus statement from French-speaking experts [correction published in Knee Surg Sports Traumatol Arthrosc 2021;29:3211–12]. Knee Surg Sports Traumatol Arthrosc 2021;29:3195–210.
24. Kloppenburg M, Kroon FP, Blanco FJ, et al. 2018 update of the EULAR recommendations for the management of hand osteoarthritis. Ann Rheum Dis 2019;78:16–24.
25. Kolasinski SL, Neogi T, Hochberg MC, et al. 2019 American College of Rheumatology/Arthritis Foundation Guideline for the Management of Osteoarthritis of the Hand, Hip, and Knee [correction published in Arthritis Rheumatol 2021;73:799]. Arthritis Rheumatol 2020;72: 220–33.
26. Rillo O, Riera H, Acosta C, et al. PANLAR consensus recommendations for the management in osteoarthritis of hand, hip, and knee. J Clin Rheumatol 2016;22:345–54.
27. Van Doormaal MC, Meerhoff GA, Vliet Vlieland TP, et al. A clinical practice guideline for physical therapy in patients with hip or knee osteoarthritis. Musculoskeletal Care 2020;18:575–95.

28. Zhang Z, Huang C, Jiang Q, et al. Guidelines for the diagnosis and treatment of osteoarthritis in China (2019 edition). Ann Transl Med 2020;8:1213.
29. American Academy of Orthopaedic Surgeons. Surgical management of osteoarthritis of the knee: clinical practice guideline. 2015. URL: https://www.aaos.org/quality/quality-programs/lower-extremity-programs/surgical-management-of-osteoarthritis-of-the-knee/.
30. Basedow M, Esterman A. Assessing appropriateness of osteoarthritis care using quality indicators: a systematic review. J Eval Clin Pract 2015;21:782–9.
31. Correa VC, Lugo-Agudelo LH, Aguirre-Acevedo DC, et al. Individual, health system, and contextual barriers and facilitators for the implementation of clinical practice guidelines: a systematic metareview. Health Res Policy Syst 2020;18:74.
32. Ferreira de Meneses S, Rannou F, Hunter DJ. Osteoarthritis guidelines: barriers to implementation and solutions. Ann Phys Rehabil Med 2016;59:170–3.
33. Swaithes L, Paskins Z, Dziedzic K, et al. Factors influencing the implementation of evidence-based guidelines for osteoarthritis in primary care: a systematic review and thematic synthesis. Musculoskeletal Care 2020;18;101–10.
34. Allen KD, Golightly YM, White DK. Gaps in appropriate use of treatment strategies in osteoarthritis. Best Pract Res Clin Rheumatol 2017;31:746–59.
35. Nelson AE, Allen KD, Golightly YM, et al. A systematic review of recommendations and guidelines for the management of osteoarthritis: the chronic osteoarthritis management initiative of the U.S. Bone and Joint Initiative. Semin Arthritis Rheum 2014;43:701–12.
36. Lin I, Wiles L, Waller R, et al. What does best practice care for musculoskeletal pain look like? Eleven consistent recommendations from high-quality clinical practice guidelines: systematic review. Br J Sports Med 2020;54:79–86.
37. Lin I, Wiles LK, Waller R, et al. Poor overall quality of clinical practice guidelines for musculoskeletal pain: a systematic review. Br J Sports Med 2018;52:337–43.
38. Brouwers MC, Kho ME, Browman GP, et al. Appraisal of Guidelines for Research Evaluation (AGREE) II. The AGREE II instrument. 2017. URL: https://www.agreetrust.org/wp-content/uploads/2017/12/AGREE-II-Users-Manual-and-23-item-Instrument-2009-Update-2017.pdf.
39. Murad MH. Clinical practice guidelines: a primer on development and dissemination. Mayo Clin Proc 2017;92:423–33.
40. AGREE Research Trust. AGREE II tutorial and practice exercise. AGREE Enterprise. URL: http://www.agreetrust.org/resource-centre/agree-ii-training-tools.
41. Smith CA, Toupin-April K, Jutai JW, et al. A systematic critical appraisal of clinical practice guidelines in juvenile idiopathic arthritis using the Appraisal of Guidelines for Research and Evaluation II (AGREE II) instrument. PLoS One 2015;10:e0137180.
42. Hayawi LM, Graham ID, Tugwell P, et al. Screening for osteoporosis: a systematic assessment of the quality and content of clinical practice guidelines, using the AGREE II instrument and the IOM standards for trustworthy guidelines [correction published in PLoS One 2019;14:e0216196]. PLoS One 2018;13:e0208251.
43. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
44. Fleiss JL, Paik MC. Statistical methods for rates and proportions. 2nd ed. New York (NY): John Wiley & Sons; 1981.
45. Armstrong JJ, Rodrigues IB, Wasiuta T, et al. Quality assessment of osteoporosis clinical practice guidelines for physical activity and safe movement: an AGREE II appraisal. Arch Osteoporos 2016;11:6.
46. Goff AJ, De Oliveira Silva D, Merolli M, et al. Patient education improves pain and function in people with knee osteoarthritis with better effects when combined with exercise therapy: a systematic review. J Physiother 2021;67:177–89.
47. Goh SL, Persson MS, Stocks 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. Sports Med 2019;49:743–61.
48. Wellsandt E, Golightly Y. Exercise in the management of knee and hip osteoarthritis. Curr Opin Rheumatol 2018;30:151–9.
49. Gay C, Chabaud A, Guilley E, et al. Educating patients about the benefits of physical activity and exercise for their hip and knee osteoarthritis: systematic literature review. Ann Phys Rehabil Med 2016;59:174–83.
50. Xu Q, Chen B, Wang Y, et al. The effectiveness of manual therapy for relieving pain, stiffness, and dysfunction in knee osteoarthritis: a systematic review and meta-analysis. Pain Physician 2017:20;229–43.
51. Ashaye T, Hounsome N, Carnes D, et al. Opioid prescribing for chronic musculoskeletal pain in UK primary care: results from a cohort analysis of the COPERS trial. BMJ Open 2017;8:e019491.
52. Van den Driest JJ, Schiphof D, de Wilde M, et al. Opioid prescriptions in patients with osteoarthritis: a population-based cohort study. Rheumatology (Oxford) 2020:59;2462–70.
53. Meneses SR, Goode AP, Nelson AE, et al. Clinical algorithms to aid osteoarthritis guideline dissemination. Osteoarthritis Cartilage 2016;24;1487–99.
54. Hagen KB, Smedslund G, Østerås N, et al. Quality of community-based osteoarthritis care: a systematic review and meta-analysis. Arthritis Care Res (Hoboken) 2016:68;1443–52.
55. Lützner J, Kasten P, Günther KP, et al. Surgical options for patients with osteoarthritis of the knee. Nat Rev Rheumatol 2009:5;309–16.
56. Ackerman IN, Bohensky MA, de Steiger R, et al. Substantial rise in the lifetime risk of primary total knee replacement surgery for osteoarthritis from 2003 to 2013: an international, population-level analysis. Osteoarthritis Cartilage 2017:25;455–61.
57. Brand CA, Harrison C, Tropea J, et al. Management of osteoarthritis in general practice in Australia. Arthritis Care Res (Hoboken) 2014:66;551–8.
58. Independent Health and Aged Care Pricing Authority. National hospital cost data collection public hospitals report - round 19 (financial year 2014–15). 2017. URL: https://www.ihacpa.gov.au/resources/national-hospital-cost-data-collection-nhcdc-public-hospitals-report-round-19-financial-year-2014-15.
59. Australian Institute of Health and Welfare. Disease expenditure in Australia 2018–2019: Australian Burden of Disease conditions. 2021. URL: https://www.aihw.gov.au/reports/health-welfare-expenditure/disease-expenditure-australia/contents/australian-burden-of-disease-conditions.
60. Ackerman IN, Bohensky MA, Zomer E, et al. The projected burden of primary total knee and hip replacement for osteoarthritis in Australia to the year 2030. BMC Musculoskelet Disord 2019:20;90.
61. Healey EL, Afolabi EK, Lewis M, et al. Uptake of the NICE osteoarthritis guidelines in primary care: a survey of older adults with joint pain. BMC Musculoskelet Disord 2018;19;295.
62. Hunter DJ, Neogi T, Hochberg MC. Quality of osteoarthritis management and the need for reform in the US. Arthritis Care Res (Hoboken) 2011:63;31–8.
63. Kamaruzaman H, Kinghorn P, Oppong R. Cost-effectiveness of surgical interventions for the management of osteoarthritis: a systematic review of the literature. BMC Musculoskelet Disord 2017:18;183.
64. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. Lancet 2019:393;1745–59.
65. Cronström A, Nero H, Lohmander LS, et al. On the waiting list for joint replacement for knee osteoarthritis: are first-line treatment recommendations implemented? Osteoarthr Cartil Open 2020:2;100056.

66. King LK, Marshall DA, Faris P, et al. Use of recommended non-surgical knee osteoarthritis management in patients prior to total knee arthroplasty: a cross-sectional study. J Rheumatol 2020:47;1253–60.
67. Briggs AM, Page CJ, Shaw BR, et al. A model of care for osteoarthritis of the hip and knee: development of a system-wide plan for the health sector in Victoria, Australia. Healthc Policy 2018:14;47–58.
68. Allen KD, Choong PF, Davis AM, et al. Osteoarthritis: models for appropriate care across the disease continuum. Best Pract Res Clin Rheumatol 2016:30;503–35.
69. Eyles JP, Hunter DJ, Bennell KL, et al. Priorities for the effective implementation of osteoarthritis management programs: an OARSI international consensus exercise. Osteoarthritis Cartilage 2019:27;1270–9.
70. Toomey CM, Kennedy N, MacFarlane A, et al. Implementation of clinical guidelines for osteoarthritis together (IMPACT): protocol for a participatory health research approach to implementing high value care. BMC Musculoskelet Disord 2022:23;643.
71. Mazzei DR, Ademola A, Abbott JH, et al. Are education, exercise and diet interventions a cost-effective treatment to manage hip and knee osteoarthritis? A systematic review. Osteoarthritis Cartilage 2021:29; 456–70.
72. Lin IB, Bunzli S, Mak DB, et al. Unmet needs of Aboriginal Australians with musculoskeletal pain: a mixed-method systematic review. Arthritis Care Res (Hoboken) 2018:70;1335–47.
73. Eyles JP, Sharma S, Telles RW, et al. Implementation of best-evidence osteoarthritis care: perspectives on challenges for, and opportunities from, low and middle-income countries. Frontiers Rehabil Sci 2022:2;826765.

Arthritis Care & Research
Vol. 75, No. 9, September 2023, pp 1908–1913
DOI 10.1002/acr.25083

American College of Rheumatology
Empowering Rheumatology Professionals

Physical Activity and Features of Knee Osteoarthritis on Magnetic Resonance Imaging in Individuals Without Osteoarthritis: A Systematic Review

Dawei Xu, Marienke Van Middelkoop, Sebastia M. A. Bierma-Zeinstra, and Jos Runhaar

Objective. To systematically review all studies that have evaluated the association between physical activity (PA) levels and features of knee osteoarthritis (OA) on magnetic resonance imaging (MRI) for subjects without OA.

Methods. The inclusion criteria for prospective studies were as follows: 1) subjects without OA; 2) average age 35–80 years; and 3) any self-reported PA or objective measurement of PA. The eligible MRI outcomes were OA-related measures of intraarticular knee joint structures. Exclusion criteria were evaluations of instant associations with transient structural changes after PA.

Results. Two randomized controlled trials and 16 observational studies were included. One of 11 studies found that PA was harmfully related to cartilage volume or thickness, but 4 studies found a significant protective association. Four of 10 studies found that PA was harmfully related to cartilage defects, while others showed no significant associations. Two of 3 studies reported a significantly increased cartilage T2 value in individuals with more PA. All 3 studies reported no significant association between PA and bone marrow lesions. Two studies assessed the association between PA and meniscus pathology, in which only occupational PA involving knee bending was associated with a greater risk of progression.

Conclusion. Within the sparse and diverse evidence available, no strong evidence was found for the presence or absence of an association between PA and the presence or progression of features of OA on MRI among subjects without OA. Therefore, more research is required before PA in general and also specific forms of PA can be deemed safe for knee joint structures.

INTRODUCTION

As a modifiable behavior, physical activity (PA) is one of the highly recommended public health and clinical management interventions for secondary and tertiary prevention of osteoarthritis (OA) (1–3). Among patients at risk for OA, previous studies have reported that PA had no effects (4,5) or protective effects against joint degeneration (6,7). However, in terms of the safety of PA for the primary prevention or early-onset of OA, there are few studies, and the findings to date are conflicting (8).

There is a concern that some weight-bearing forms of PA may increase the risk of knee OA development (9–11). However, it may take years to observe radiographic OA or symptomatic OA among individuals free of signs and symptoms. Even before the onset of symptomatic OA, structural changes are already developing, including the presence of bone marrow lesions (BMLs), cartilage loss, and changes in the meniscus. Therefore, detecting early structural changes in the knee among the population without OA could be meaningful to judge the safety of PA.

Several studies have used magnetic resonance imaging (MRI) to capture features of OA, such as cartilage defects and meniscal pathologies, in the early stage of OA. Cartilage abnormalities, such as reductions in cartilage volume and thickness, may be associated with knee pain and joint space narrowing (12–14). Knee cartilage defects play an important role in early knee OA, which could result in increased cartilage breakdown and lead to decreased cartilage volume and joint space narrowing (15). Also, cartilage T2 relaxation time mapping is used to detect early articular cartilage degeneration (16), with higher cartilage T2 values being associated with the development of radiographic

Mr. Xu's work was supported by the China Scholarship Council (grant 201806380153).

Dawei Xu, MSc, Marienke Van Middelkoop, PhD, Sebastia M. A. Bierma-Zeinstra, PhD, Jos Runhaar, PhD: Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.25083&file=acr25083-sup-0001-Disclosureform.pdf.

Address correspondence via email to Jos Runhaar, PhD, at j.runhaar@erasmusmc.nl.

Submitted for publication February 2, 2022; accepted in revised form December 20, 2022.

SIGNIFICANCE & INNOVATIONS

- Data on the effects of physical activity (PA) and the presence or progression of features of osteoarthritis (OA) on magnetic resonance imaging (MRI) among subjects without OA are sparse and highly diverse.
- Data on the presence or progression of bone marrow lesions and meniscus pathology are especially lacking.
- No strong evidence was found for the presence or absence of an association between PA and the presence or progression of features of OA on MRI among subjects without OA.

knee OA (17). By using MRI, several studies have found that both meniscus extrusion and greater meniscus volume were risk factors for early progress of OA (18,19). Thus, MRI may be a sensitive and promising technique to detect potential structural changes caused by PA (20). By systematically reviewing all these studies, the current study evaluated the association between PA and early features of knee OA on MRI among subjects free of knee OA.

MATERIALS AND METHODS

The protocol has been registered in the International Prospective Register of Systematic Reviews database (PROSPERO: CRD42020218996). Searches were conducted of electronic databases (Medline [all], Ovid, Embase, Web of Science Core Collection, Cumulated Index to Nursing and Allied Health Literature [CINAHL], EBSCOhost, and Cochrane Central Register of Controlled Trials) from their earliest date until October 29, 2020. The Medial Subject Heading (MeSH) list is shown in Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25083.

Selection criteria. *Primary research of any study design.* This systematic review included controlled trials, prospective and retrospective studies, and cross-sectional studies. There was no limitation on language.

One reviewer (DX) conducted title and abstract screening for all citations, while either of 3 reviewers (MVM, SMAB-Z, or JR) independently screened the citations for verification. Then all researchers conducted screening for full-text articles based on PICO (population, intervention, comparator, outcomes), as outlined below.

Population/participant. Subjects were without radiographic knee OA (Kellgren/Lawrence grades <2) and with no or minimal, nonchronic knee symptoms (joint pain, aching, and/or stiffness) at baseline. The mean age of reported subjects was between 35 and 80 years. There was no limitation on sex or other potential risk OA factors.

Intervention (all types of PA) and comparator. Intervention included self-reported PA (questionnaire) or any objective measurement of PA, with no limitation on minimum duration. The measurements of PA levels for cross-sectional studies were assessments of the history of PA levels among the study population. The comparator was no exposure of PA or a lower level of PA (e.g., a varying level of PA).

Outcomes of interest. Outcomes were all cross-sectional and longitudinal measures of meniscus, cartilage, BMLs, osteophytes, and effusion-synovitis on knee MRI. We excluded the outcomes that are currently not well recognized as typical features of OA (e.g., patella bone volume and subchondral bone volume) (21,22). Studies that measured MRI features immediately after PA were also excluded.

Data extraction (selection and coding). Data extraction was carried out by one reviewer (DX) and independently verified by the second reviewer (JR). All reviewers made a final agreement on selected information and data.

Information was extracted on the following: 1) study title, authors, publication year, country, and study design; 2) participants, including total number and key baseline characteristics (age, population description, body mass index [BMI], percentage of female subjects); 3) physical activity type, recording method (questionnaire/objective measurement), intensity, session frequency, duration of exposure, and score range; 4) knee joint MRI outcome data at baseline and follow-up; 5) adjusted odds ratios or any association coefficient for development and/or progression to MRI features for varying levels of PA; and 6) confounders used in the analyses.

Data synthesis. Due to the substantial heterogeneity within studies, a narrative synthesis was conducted. Moreover, results were analyzed with a focus on the direction of the association (harmful/protective/no) of PA with MRI features rather than on the magnitude of the association. The synthesis included collating and summarizing outcomes from separate features of OA on MRI and knee joint sublocations (i.e., tibia, femur, and patella for both medial and lateral compartments). Within each MRI outcome sublocation, types of PA and their associations with outcomes were summarized.

Risk of bias (RoB) assessment. The studies selected for inclusion in this systematic review were evaluated by 2 researchers (DX and JR) to avoid any discrepancies. Cochrane Collaboration's RoB Tool for randomized controlled trials was accepted as a standard tool (23). The Risk of Bias in Nonrandomized Studies of Interventions (ROBINS-I) tool for observational studies (24) was also used. All RoB graphs were generated by a free access tool in McGuinness and Higgins (25).

RESULTS

In total, 2,322 articles were retrieved from the databases. After the records were screened by title and abstract, a total of 107 articles were selected for further screening. In the end, 18 studies met the inclusion criteria (for reasons of exclusion, see Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25083), which included 2 randomized controlled trials (RCTs) (26,27) and 16 observational studies (28–43). The mean age of subjects in the selected studies ranged from 35.0 to 57.8 years. The characteristics of the selected studies are summarized in Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25083.

Results of RoB assessment. One RCT showed a low-level of risk bias, while the other showed bias with some concerns. In all, 14 observational studies had a moderate risk of bias, and 2 had a serious risk of bias. All details of subdomains are shown in Supplementary Figures 2 and 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25083.

Impact of PA on cartilage volume and thickness. One RCT study and 10 observational studies described the association between PA and cartilage volume or thickness. In the RCT, subjects were randomized over endurance training, strength training, or a control group. Among the 10 observational studies, of which several explored multiple exposures, the exposures varied between a composite score of the amount of PA (n = 2), light PA (n = 2), vigorous PA (n = 7), PA to improve aerobic capacity (n = 1), and occupational activities involving knee bending (n = 11). See Supplementary Table 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25083, for an overview of all exposures.

One of 10 observational studies found that more frequent PA was significantly associated with greater loss or lower current cartilage volume or thickness. In contrast, 3 observational studies found that greater PA was significantly associated with less loss or higher current cartilage volume. One observational study found that more frequent PA was significantly associated with lower cartilage loss in high baseline cartilage volume but greater cartilage volume loss in low baseline cartilage volume. Five observational studies and the RCT study found that PA was not associated with any outcome of cartilage volume. All detailed results are shown in Supplementary Table 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25083.

Impact of PA on cartilage defects. One RCT study and 9 observational studies measured the effect of PA on cartilage defects. The exposure in the RCT study was randomly assigned unilateral high-impact exercise, and outcomes were compared to the contralateral leg. Among the 9 observational studies, the exposures varied between light PA (n = 1), vigorous PA (n = 8), and occupational PA involving knee bending (n = 9). See Supplementary Table 2, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25083, for an overview of all exposures.

Among 9 observational studies, 4 studies found a significant association between PA and cartilage defects. All 4 studies showed that PA was associated with a greater risk of cartilage defects. The RCT study and 5 observational studies found that PA was not associated with any outcome of cartilage defects. The details are shown in Supplementary Table 4, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25083.

Impact of PA on cartilage T2 values. One RCT study and 2 observational studies measured the association between PA and cartilage T2 values. The RCT study measured 12 sublocations, but in none of them was there a significant T2 difference between the unilateral high-impact exercise leg and the contralateral leg. Among the 2 observational studies, exposures were occupational PA involving knee bending (n = 1) and a categorical measure of intensity of PA (n = 2). See Supplementary Table 2, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25083, for an overview of all exposures.

One observational study did not show any significant association between PA and T2 values. The other observational study found that more frequent vigorous PA was related to a significantly higher T2 value. It also showed that occupational PA involving knee bending was associated with significantly higher T2 values. Details are shown in Supplementary Table 5, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25083.

Impact of PA on BMLs. One RCT and 2 observational studies assessed the association between PA and BMLs. The RCT did not observe an association between randomly assigned unilateral high-impact exercise and change in BMLs over 6 months compared to the contralateral knee (27). In one observational study, vigorous PA was assessed twice and was not associated with the presence of BMLs (37). One cross-sectional study also did not find an association between participation in marathons and BML grade (43). Details were shown in Supplementary Tables 2 and 6, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25083.

Impact of PA on meniscus pathologies. Two observational studies reported on the association between PA and meniscus pathologies. One study found that more frequent occupational PA involving knee bending was associated with a greater risk of progression overall and medial meniscus score (30). However, this study did not observe an association between occupational PA involving knee bending and meniscal lesions or

meniscus tears in a cross-sectional design. One cohort study found that PA (composite score of amount of PA) was not associated with meniscus extrusion in a population at risk for OA (28). Details are shown in Supplementary Tables 2 and 6, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25083.

DISCUSSION

To assess the impact of PA level on features of OA on MRI among populations without OA, this systematic review summarized the evidence of 2 RCTs and 16 observational studies. The finding of this review indicated that in most cases PA was not associated with features of OA on MRI. Most studies reported on the association between PA and cartilage. However, these associations were generally conflicting. Similar to radiographic findings in some studies (44–46), the diverse effects of PA could be due to, among other things, the different outcome measures (47), populations, and study designs. Moreover, there was little evidence on the association between PA and BMLs or meniscal pathologies.

The results of this study indicated that both light and vigorous PA might be important for maintenance of cartilage thickness/volume but also could lead to cartilage volume loss over time. The inconsistency in results may be explained by cartilage volume being affected by many confounding factors. The study from Teichtahl et al suggested an interaction between baseline cartilage volume and PA, which indicates that the protective role of PA might be dependent on cartilage condition (29). In addition, previous research has indicated that cartilage swelling appears to precede volume loss in early OA (48,49). In all selected studies, the condition of cartilage prior to initiating PA was unknown. Although cartilage loss is one of the major characters of OA progression, it will take years to observe an obvious change of cartilage volume/thickness. Most of the selected studies were of a cross-sectional design or with a short follow-up period, which could further explain the inconsistent results.

Although there were some possible concerns that PA, especially vigorous PA, was related to the presence and/or progression of cartilage defects, more than one-half of the selected studies showed no association between PA and cartilage defects. Cartilage injury may be of various etiologies, including acute traumatic injuries and early posttraumatic degenerative changes. Abnormal forces across the knee joint can also lead to cartilage damage and subsequent degeneration. Vigorous PA may cause cartilage injuries, which consequently increase the risk of OA progression. However, based on the current literature, we could not conclude that any specific PA type was associated with cartilage defects. This finding is supported by a recently published review that reported that no new cartilage lesions were observed after running (50).

Only 1 study reported that light or vigorous PA was associated with cartilage T2 values. From a compositional perspective, light PA could be protective to cartilage, while vigorous PA might be detrimental to cartilage. T2 relaxation time measurements in the knee are sensitive to initial cartilage degeneration and reflect the histologic changes of the cartilage matrix, particularly affecting water and collagen content as well as tissue anisotropy (51–53). Furthermore, T2 changes could predict the onset of radiographic OA (17) because the compositional measures enable early detection of changes in cartilage composition (50). If vigorous PA causes cartilage damage, the change in cartilage content could be detected by T2 at a very early phase. Nevertheless, owing to the very low number of studies available in the literature, the direct association between PA and change of cartilage T2 values is still debatable.

Since 1 of only 2 available studies found that more frequent PA was associated with the progression of meniscus pathologies over 3 years, there is still a lack of evidence for the association between PA and meniscus pathologies. Previous research has indicated that among patients with mild-to-moderate OA, PA and dietary interventions that reduced BMI were associated with less meniscus extrusion progression (54). Overall, the number of available studies was too low to draw strong conclusions.

To our knowledge, this is the first systematic review of the evidence on the association between features of OA on MRI among subjects without OA. We included observational studies in our review to obtain more information. There were some limitations of this review. First, there were only 2 RCTs included. The number of observational studies was also low, which means that the results remain inconclusive. Second, some eligible studies included participants with potential structural changes visible on MRI only at baseline, which may confound the association between PA and any subsequent structural changes. However, obtaining evidence for the association between PA and structural features of OA among nonsymptomatic individuals and those not diagnosed with OA, irrespective of the presence of features of OA on MRI, might be more appreciated in clinical practice, as it is not feasible or advised to screen for features of OA on MRI when prescribing PA for individuals without a diagnosis of knee OA. Third, from this study, we could not indicate a threshold for safe levels of PA. Because the exposure of most included studies combined several types of PA, we were not able to present any results for specific types of PA. Fourth, in some studies, the sample size might have been too small to find significant associations. Finally, many studies were from the same country or the same population. Although the population characteristics showed some differences, it is still highly possible that these studies include the same population, which may limit generalizability.

In conclusion, in the sparse and diverse evidence available, no strong evidence was found for the presence or absence of an association between PA and the presence or progression of features of OA on MRI in subjects without OA. Therefore, more research is required before PA in general and also specific forms of PA can be deemed safe for knee joint structures.

ACKNOWLEDGMENT

The authors acknowledge Dr. Wichor M. Bramer for contributing to the search strategy of this systematic review.

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. Runhaar 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. Xu, Van Middelkoop, Bierma-Zeinstra, Runhaar.

Acquisition of data. Xu, Van Middelkoop, Bierma-Zeinstra, Runhaar.

Analysis and interpretation of data. Xu, Runhaar.

REFERENCES

1. Nelson AE, Allen KD, Golightly YM, et al. A systematic review of recommendations and guidelines for the management of osteoarthritis: the Chronic Osteoarthritis Management Initiative of the U.S. Bone and Joint Initiative. Semin Arthritis Rheum 2014;43:701–12.
2. Golightly YM, Allen KD, Caine DJ. A comprehensive review of the effectiveness of different exercise programs for patients with osteoarthritis. Phys Sportsmed 2012;40:52–65.
3. Waterman MB, White PH. Utilizing the National Physical Activity Plan to create a disease-specific approach: environmental and policy strategies to increase physical activity among adults with arthritis. J Phys Act Health 2014;11:487–8.
4. Felson DT, Niu J, Clancy M, et al. Effect of recreational physical activities on the development of knee osteoarthritis in older adults of different weights: the Framingham Study. Arthritis Rheum 2007;57:6–12.
5. Hannan MT, Felson DT, Anderson JJ, et al. Habitual physical activity is not associated with knee osteoarthritis: the Framingham Study. J Rheumatol 1993;20:704–9.
6. Rogers LQ, Macera CA, Hootman JM, et al. The association between joint stress from physical activity and self-reported osteoarthritis: an analysis of the Cooper Clinic data. Osteoarthritis Cartilage 2002;10:617–22.
7. Roos EM, Dahlberg L. Positive effects of moderate exercise on glycosaminoglycan content in knee cartilage: a four-month, randomized, controlled trial in patients at risk of osteoarthritis. Arthritis Rheum 2005;52:3507–14.
8. Peeters GM, Pisters MF, Mishra GD, et al. The influence of long-term exposure and timing of physical activity on new joint pain and stiffness in mid-age women. Osteoarthritis Cartilage 2015;23:34–40.
9. Spector TD, Harris PA, Hart DJ, et al. Risk of osteoarthritis associated with long-term weight-bearing sports: a radiologic survey of the hips and knees in female ex-athletes and population controls. Arthritis Rheum 1996;39:988–95.
10. McAlindon TE, Wilson PW, Aliabadi P, et al. Level of physical activity and the risk of radiographic and symptomatic knee osteoarthritis in the elderly: the Framingham Study. Am J Med 1999;106:151–7.
11. Kujala UM, Kaprio J, Sarna S. Osteoarthritis of weight bearing joints of lower limbs in former elite male athletes. BMJ 1994;308:231–4.
12. Buck RJ, Wirth W, Dreher D, et al. Frequency and spatial distribution of cartilage thickness change in knee osteoarthritis and its relation to clinical and radiographic covariates: data from the Osteoarthritis Initiative. Osteoarthritis Cartilage 2013;21:102–9.
13. Hunter DJ, March L, Sambrook PN. The association of cartilage volume with knee pain. Osteoarthritis Cartilage 2003;11:725–9.
14. Cicuttini FM, Wluka AE, Forbes A, et al. Comparison of tibial cartilage volume and radiologic grade of the tibiofemoral joint. Arthritis Rheum 2003;48:682–8.
15. Ding C, Garnero P, Cicuttini F, et al. Knee cartilage defects: association with early radiographic osteoarthritis, decreased cartilage volume, increased joint surface area and type II collagen breakdown. Osteoarthritis Cartilage 2005;13:198–205.
16. Mosher TJ, Dardzinski BJ. Cartilage MRI T2 relaxation time mapping: overview and applications. Semin Musculoskelet Radiol 2004;8:355–68.
17. Liebl H, Joseph G, Nevitt MC, et al. Early T2 changes predict onset of radiographic knee osteoarthritis: data from the Osteoarthritis Initiative. Ann Rheum Dis 2015;74:1353–9.
18. Van der Voet JA, Runhaar J, van der Plas P, et al. Baseline meniscal extrusion associated with incident knee osteoarthritis after 30 months in overweight and obese women. Osteoarthritis Cartilage 2017;25:1299–303.
19. Xu D, van der Voet J, Hansson NM, et al. Association between meniscal volume and development of knee osteoarthritis. Rheumatology (Oxford) 2021;60:1392–9.
20. Mahmoudian A, Lohmander LS, Mobasheri A, et al. Early-stage symptomatic osteoarthritis of the knee: time for action. Nat Rev Rheumatol 2021;17:621–32.
21. Hunter DJ, Guermazi A, Lo GH, et al. Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). Osteoarthritis Cartilage 2011;19:990–1002.
22. Peterfy CG, Guermazi A, Zaim S, et al. Whole-Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis. Osteoarthritis Cartilage 2004;12:177–90.
23. Sterne JA, Savovic J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019;366:l4898.
24. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016;355:i4919.
25. McGuinness LA, Higgins JP. Risk-of-bias VISualization (robvis): an R package and shiny web app for visualizing risk-of-bias assessments. Res Synth Methods 2021;12:55–61.
26. Cotofana S, Ring-Dimitriou S, Hudelmaier M, et al. Effects of exercise intervention on knee morphology in middle-aged women: a longitudinal analysis using magnetic resonance imaging. Cells Tissues Organs 2010;192:64–72.
27. Hartley C, Folland JP, Brooke-Wavell K, et al. High impact exercise increased femoral neck bone density with no adverse effects on imaging markers of knee osteoarthritis in postmenopausal women. J Bone Miner Res 2020;35:53–63.
28. Zhang F, Bierma-Zeinstra SM, Oei EH, et al. Factors associated with meniscal body extrusion on knee MRI in overweight and obese women. Osteoarthritis Cartilage 2017;25:694–9.
29. Teichtahl AJ, Wang Y, Heritier S, et al. The interaction between physical activity and amount of baseline knee cartilage. Rheumatology (Oxford) 2016;55:1277–84.
30. Virayavanich W, Alizai H, Baum T, et al. Association of frequent knee bending activity with focal knee lesions detected with 3T magnetic resonance imaging: data from the Osteoarthritis Initiative. Arthritis Care Res (Hoboken) 2013;65:1441–8.
31. Lin W, Alizai H, Joseph GB, 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.
32. Teichtahl AJ, Wluka AE, Wang Y, et al. Effect of long-term vigorous physical activity on healthy adult knee cartilage. Med Sci Sports Exerc 2012;44:985–92.

33. Hovis KK, Stehling C, Souza RB, et al. Physical activity is associated with magnetic resonance imaging-based knee cartilage T2 measurements in asymptomatic subjects with and those without osteoarthritis risk factors. Arthritis Rheum 2011;63:2248–56.
34. Teichtahl AJ, Wluka A, Wang YY, et al. Occupational activity is associated with knee cartilage morphology in females. Maturitas 2010;66: 72–6.
35. Teichtahl AJ, Wluka AE, Forbes A, et al. Longitudinal effect of vigorous physical activity on patella cartilage morphology in people without clinical knee disease. Arthritis Rheum 2009;61:1095–102.
36. Wijayaratne SP, Teichtahl AJ, Wluka AE, et al. The determinants of change in patella cartilage volume: a cohort study of healthy middle-aged women. Rheumatology (Oxford) 2008;47:1426–9.
37. Racunica TL, Teichtahl AJ, Wang Y, et al et al. Effect of physical activity on articular knee joint structures in community-based adults. Arthritis Rheum 2007;57:1261–8.
38. Hanna FS, Bell RJ, Davis SR, et al. Factors affecting patella cartilage and bone in middle-aged women. Arthritis Rheum 2007;57:272–8.
39. Hanna F, Teichtahl AJ, Bell R, et al. The cross-sectional relationship between fortnightly exercise and knee cartilage properties in healthy adult women in midlife. Menopause 2007;14:830–4.
40. Wang Y, Ding C, Wluka AE, et al. Factors affecting progression of knee cartilage defects in normal subjects over 2 years. Rheumatology (Oxford) 2006;45:79–84.
41. Hanna F, Ebeling PR, Wang Y, et al. Factors influencing longitudinal change in knee cartilage volume measured from magnetic resonance imaging in healthy men. Ann Rheum Dis 2005;64:1038–42.
42. Cicuttini FM, Wluka AE, Wang Y, et al. Compartment differences in knee cartilage volume in healthy adults. J Rheumatol 2002;29: 554–6.
43. Shellock FG, Morris E, Deutsch AL, et al. Hematopoietic bone marrow hyperplasia: high prevalence on MR images of the knee in asymptomatic marathon runners. AJR Am J Roentgenol 1992;158:335–8.
44. Urquhart DM, Tobing JF, Hanna FS, et al. What is the effect of physical activity on the knee joint? A systematic review. Med Sci Sports Exerc 2011;43:432–42.
45. Lo GH, McAlindon TE, Kriska AM, et al. Football increases future risk of symptomatic radiographic knee osteoarthritis. Med Sci Sports Exerc 2020;52:795–800.
46. Gates LS, Perry TA, Golightly YM, et al. Recreational physical activity and risk of incident knee osteoarthritis: an international meta-analysis of individual participant–level data. Arthritis Rheumatol 2022;74: 612–22.
47. Roemer FW, Felson DT, Stefanik JJ, et al. Heterogeneity of cartilage damage in Kellgren and Lawrence grade 2 and 3 knees: the MOST study. Osteoarthritis Cartilage 2022;30:714–23.
48. Calvo E, Palacios I, Delgado E, et al. Histopathological correlation of cartilage swelling detected by magnetic resonance imaging in early experimental osteoarthritis. Osteoarthritis Cartilage 2004;12:878–86.
49. McBride A, Khan HI, Aitken D, et al. Does cartilage volume measurement or radiographic osteoarthritis at baseline independently predict ten-year cartilage volume loss? BMC Musculoskelet Disord 2016; 17:54.
50. Khan MC, O'Donovan J, Charlton JM, et al. The influence of running on lower limb cartilage: a systematic review and meta-analysis. Sports Med 2022;52:55–74.
51. Li X, Cheng J, Lin K, et al. Quantitative MRI using T1ρ and T2 in human osteoarthritic cartilage specimens: correlation with biochemical measurements and histology. Magn Reson Imaging 2011;29:324–34.
52. Marik W, Apprich S, Welsch GH, et al. Biochemical evaluation of articular cartilage in patients with osteochondrosis dissecans by means of quantitative T2- and T2-mapping at 3T MRI: a feasibility study. Eur J Radiol 2012;81:923–7.
53. Nishioka H, Hirose J, Nakamura E, et al. T1ρ and T2 mapping reveal the in vivo extracellular matrix of articular cartilage. J Magn Reson Imaging 2012;35:147–55.
54. Munugoda IP, Beavers DP, Wirth W, et al. The effect of weight loss on the progression of meniscal extrusion and size in knee osteoarthritis: a post-hoc analysis of the Intensive Diet and Exercise for Arthritis (IDEA) trial. Osteoarthritis Cartilage 2020; 28:410–7.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 75, No. 9, September 2023, pp 1914–1924
DOI 10.1002/acr.25089

Clinically Relevant Subgroups Among Athletes Who Have Ruptured Their Anterior Cruciate Ligaments: A Delaware-Oslo Cohort Study

Elanna K. Arhos,[1] Ryan T. Pohlig,[1] Stephanie Di Stasi,[2] May Arna Risberg,[3] Lynn Snyder-Mackler,[1] and Karin Grävare Silbernagel[1]

Objective. To identify subgroups of individuals with anterior cruciate ligament (ACL) injuries based on patient characteristics, self-reported outcomes, and functional performance at baseline, and to associate subgroups with long-term outcomes after ACL rupture.

Methods. A total of 293 participants (45.7% male, mean ± SD age 26.2 ± 9.4 years, days from injury 58 ± 35) were enrolled after effusion, pain, and range of motion impairments were resolved and quadriceps strength was at least 70% of the uninvolved limb. Mixture modeling was used to uncover latent subgroups without a prior group classification using probabilistic assignment. Variables include demographics, functional testing, and self-reported outcome measures. Radiographic evidence of osteoarthritis (OA; i.e., Kellgren/Lawrence grade of ≥1) in the involved knee at 5 years after injury was the primary outcome of interest. Chi-square tests assessed differences in the presence of radiographic OA in the involved knee between subgroups at 5 years after ACL rupture. Secondary outcomes of interest included radiographic OA in the uninvolved knee, return to preinjury sport by 2 years, operative status, and clinical OA (classified using Luyten et al criteria) at 5 years.

Results. Four distinct subgroups exist after ACL rupture (younger good self-report, younger poor self-report, older poor self-report, older good self-report) with 30%, 31%, 47%, and 53%, respectively, having involved knee OA. The percentage of radiographic OA was not significantly different between the groups ($P = 0.059$).

Conclusion. The prevalence of OA in all subgroups is highly concerning. These results suggest there are unique subgroupings of individuals that may guide treatment after ACL rupture and reconstruction by providing support for developing a patient-centered approach.

INTRODUCTION

Anterior cruciate ligament (ACL) ruptures are one of the most common traumatic knee joint injuries in adolescents and young adults. Posttraumatic osteoarthritis (OA) in the knee joint is one of many concerning long-term outcomes facing individuals who have torn their ACL. Recent data suggest that 50–80% of individuals develop posttraumatic OA within 10 years of ACL reconstruction (1–3). Most who undergo ACL reconstruction are young and active (4), leaving them at a high risk of developing posttraumatic OA in young adulthood. These data also suggest that there are some individuals who are successful in avoiding some of the most devastating long-term outcomes, suggesting a need for early identification of individuals who are most at risk. Identifying relevant clinical characteristics of patients who may be on a trajectory to developing posttraumatic OA is a critical step toward early detection of at-risk individuals and may provide insights into prevention.

Immediate and long-term outcomes after ACL reconstruction are highly variable. When considering the metrics of return to sport, 65% of individuals return to their preinjury level of sport, with only 55% of athletes returning to a competitive level of sport (5). Overall reinjury rates are estimated at 15%, increasing to 23% for individuals younger than 25 who return to sport (6).

Supported by the NIH (grants R37-HD-037985 and F31-AR-078580) and by the Foundation for Physical Therapy Research (PODS level II scholarship).

[1]Elanna K. Arhos, PT, DPT, PhD, Ryan T. Pohlig, PhD, Lynn Snyder-Mackler, PT, ScD, Karin Grävare Silbernagel, PT, PhD, ATC: University of Delaware, Newark; [2]Stephanie Di Stasi, PT, PhD: Ohio State University Wexner Medical Center, Columbus; [3]May Arna Risberg, PT, PhD: Oslo University Hospital and Norwegian School of Sport Sciences, Oslo, Norway.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.25089&file=acr25089-sup-0001-Disclosureform.pdf.

Address correspondence via email to Elanna K. Arhos, PT, DPT, PhD, at earhos@udel.edu.

Submitted for publication June 20, 2022; accepted in revised form January 12, 2023.

SIGNIFICANCE & INNOVATIONS

- While the older good self-report group had the highest prevalence of osteoarthritis (OA) at 5 years, the prevalence of OA at 5 years in the 2 younger subgroups is highly concerning.
- Using subgroup analyses to relate clinical characteristics to subsequent development of posttraumatic OA is an important step in identifying associations between subgroups and long-term outcomes and providing appropriate targets for rehabilitation.
- The 4 subgroups uncovered may assist in targeting clinical treatments that are individualized after anterior cruciate ligament rupture.

Previous work has identified individuals who are copers, those resuming prior activity levels with dynamic knee stability, and noncopers, those who demonstrate dynamic knee instability and poor clinical presentation (7). Copers have better outcomes after ACL rupture compared to noncopers across functional tests and patient-reported outcome measures (8–10). Collectively, these differences suggest that there may be homogenous subgroups among ACL injured individuals, which may help explain the heterogeneity seen in long-term outcomes (11).

The objectives of rehabilitation after ACL injury are similar across patients: restore the range of motion and minimize effusion, restore quadriceps strength, and when ready, return to sport or recreational activity. Clinical test batteries assist in ensuring that patients do not return to sport or previous activity level until the risk of re-rupture is minimized. However, there are no test batteries or clinical prediction rules to assist in identifying the risk for posttraumatic OA or other long-term deficits. Further, the presence of subgroups may identify individuals who are at greater or lesser risk for negative long-term outcomes, enabling insight into targeted treatments.

The primary purpose of this study was to identify whether subgroups of ACL-injured individuals exist based on personal characteristics, self-reported outcomes, and functional performance measures. The secondary purpose was to determine whether associations exist between these latent subgroups and long-term outcomes, including 1) as the primary long-term outcome, the development of radiographic posttraumatic OA of the involved knee, and 2) as additional outcomes, radiographic posttraumatic OA of the uninvolved knee, return to preinjury sport level by 2 years, operative status (i.e., has the participant undergone ACL reconstruction by 2 years), and clinical OA (classified using Luyten et al criteria [12]). Identifying subgroups based on commonly measured clinical characteristics is an important step in pinpointing rehabilitation strategies for each group, moving ACL rehabilitation toward a more patient-centered approach.

PATIENTS AND METHODS

This study was an analysis of 293 patients (Figure 1) enrolled in the Delaware-Oslo ACL prospective cohort study. Patients included were recruited between 2006 and 2012 from both the University of Delaware in Newark, Delaware, and the Norwegian Sports Medicine Clinic in Oslo, Norway, with outcomes previously reported in the 5 past years (13,14). Individuals were screened for outliers among all variables using histograms and boxplots, and 7 individuals were removed (2 based on days from surgery, and 5 based on age).

Participants. Participants were included in the parent cohort study if they had an ACL rupture, achieved a quiet knee based on a clinical examination (i.e., minimal to no pain or effusion) (15), were age 13–60 years, and participated in level I and II sports (16) (e.g., cutting, jumping, pivoting) for ≥50 hours a year prior to injury. Injuries were verified using magnetic resonance imaging (MRI) and increased anterior knee joint laxity measured with a KT-1,000 arthrometer (MED Metric). Participants with a previous history of ACL rupture were included, but participants with any other previous injuries or surgeries to either knee, bilateral injuries, concomitant grade III ligament injuries, repairable menisci on MRI, full-thickness articular cartilage damage, or fracture were excluded. All participants provided informed consent, and the study was approved by the Institutional Review Board at the University of Delaware or the Regional Committee for Medical Research Ethics South East Norway.

Treatment algorithm. Participants underwent baseline testing when effusion was resolved, and they could hop on the involved knee without pain (mean ± SD days from injury 58 ± 35). Participants were classified as potential copers or noncopers at baseline based on previously established criteria (7). They underwent a 5-week program of neuromuscular and strength training prior to the decision for ACL reconstruction or continued nonoperative management (17). Patients who were managed nonoperatively continued progressive rehabilitation for another 3–4 months and were assessed at follow-ups the same as the operative group.

Assessments and outcome variables. Variables selected for assessments and outcomes were collected in the parent cohort study and based on prior literature as those that predicted success after ACL rupture or had an association with posttraumatic OA (18,19). Measures included in the model were baseline patient characteristics (age, sex, preinjury level, concomitant injuries, and a history of previous ACL rupture), days from injury to baseline evaluation, and body mass index (BMI). Further variables included at baseline were quadriceps strength, single and triple hop for distance, the Knee Outcome Survey–Activity of Daily Living Scale (KOS-ADLS), a global rating scale (GRS) of

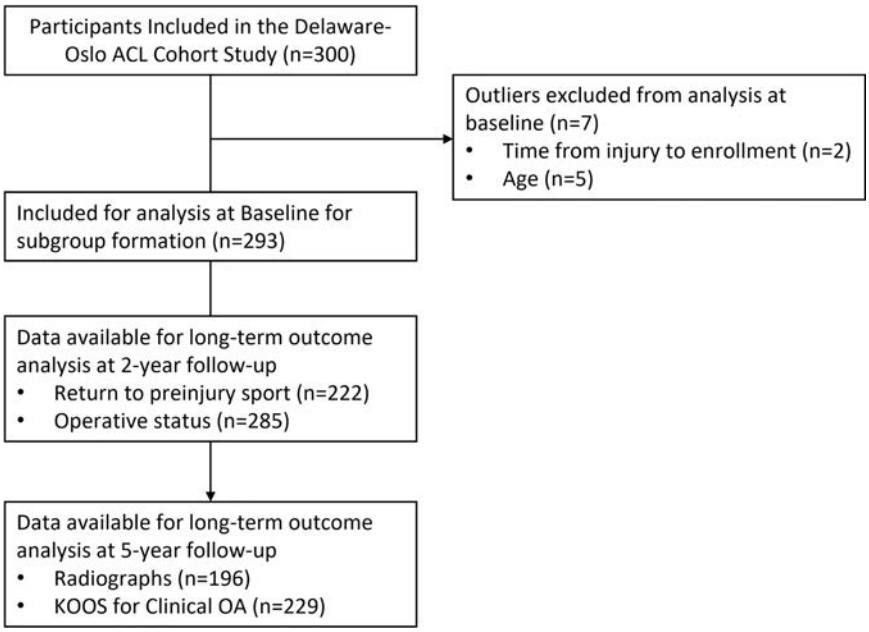

Figure 1. Delaware-Oslo Anterior Cruciate Ligament (ACL) Cohort Study consort diagram for data available at baseline subgroup formation and long-term outcomes. OA = osteoarthritis; KOOS = Knee Injury and Osteoarthritis Outcome Score.

perceived function, and the International Knee Documentation Committee (IKDC) Subjective Knee Form.

Quadriceps strength was measured differently at the 2 sites, in Delaware and in Oslo. In Delaware, strength was measured using an electromechanical dynamometer (Kin-com, DJO Global, or System 3, Biodex) during a maximal voluntary isometric contraction knee extension test. Participants were seated with hips and knees flexed to 90 degrees, and the dynamometer's axis of rotation aligned with the axis rotation of the knee joint. The leg was strapped in at the upper thigh, pelvis, and shank to minimize accessory motion during testing. Participants completed 3 submaximal practice trials, followed by 3 maximal effort trials on the uninvolved limb first, then on the involved limb. In Oslo, strength was assessed using an isokinetic dynamometer (Biodex 6000, Biodex Medical Systems). Participants performed 4 submaximal practice trials, then 5 recorded maximal effort repetitions for the uninvolved limb first, then the involved limb. Quadriceps strength is reported as a limb symmetry index (LSI), calculated as the involved extremity maximum torque divided by the uninvolved extremity maximum torque, expressed as a percentage.

Single-hop testing consisted of 4-hop tests (single, crossover, triple, 6-meter timed) (20,21). We only included the single-hop for distance and triple-hop for distance in the model, as hop scores for triple hop, crossover hop, and timed hop were highly correlated (all $r \geq 0.95$). Each hop test consisted of 2 practice trials for familiarization, followed by 2 recorded trials. Uninvolved extremities were tested first, followed by the involved extremities. Hop tests were also reported using LSI scores, calculated from the average of 2 trials per extremity (involved/uninvolved × 100).

A variety of valid, reliable, and responsive self-reported outcome measures were used to assess self-reported knee function at baseline. The KOS-ADLS assesses knee function during activities of daily living (22). A higher number represents less limitation in knee function in daily life, with 100% indicating no limitation. GRS is a single item rating from 0% to 100% that rates overall knee function compared to the knee function prior to injury (23). A score closer to 100% indicates better perceived function. Finally, the IKDC measures knee-specific symptoms, function, and sports activities. The IKDC is scored from 0 to 100, with scores closer to 100 indicating higher subjective reports of knee function (24,25).

As participants were measured at 5 years from baseline, we wanted to capture individuals who were on the trajectory for early radiographic OA beyond those who already had the definite presence of osteophytes. The presence of joint characteristics consistent with the development of radiographic OA in the involved knee, therefore, was operationally defined as a Kellgren/Lawrence (K/L) grade of ≥1. Patients returned 5 years from baseline, after either ACL reconstruction or nonoperative management, for standardized bilateral posteroanterior bent knee radiographs. Radiographs were taken at 5 years only, and there were no radiographs collected at baseline. Participants in Delaware were assessed using the Lyon-Schuss protocol (26), where the radiograph beam was adjusted for each image to align with the medial tibial plateau. Participants were positioned with a 30-degree knee flexion angle with pelvis, thigh, and patella against the film cassette and coplanar with the tips of the great toes. In Oslo, a fixed flexion protocol was used with a 10-degree caudal beam angulation and a SynaFlexer Positioning Frame (Synarc) to make knee

alignment and angulation reproducible (27,28). Levels of OA in the tibiofemoral joint were graded by an experienced radiologist with high intrarater reliability (κ = 0.77) using the K/L system in the tibiofemoral joint (29).

Secondary outcomes included radiographic contralateral knee OA (K/L grade ≥1) at 5 years. The presence of clinical knee OA was determined at 5 years using Luyten et al criteria, which require 2 of 4 Knee Injury and Osteoarthritis Outcome Score subscales to score ≤85%, and consistent with our previous publications (12,30). New injuries to the ipsilateral and contralateral knee were reported at the 5-year follow-up. Return to the preinjury sport level by 2 years was assessed using the question "Has the subject returned to at least preinjury level?" with a dichotomous yes/no, and operative status. Operative status was defined as undergoing ACL reconstruction or remaining nonoperatively managed by 2 years after baseline.

Statistical analysis. We identified the number of latent subgroups present at baseline using mixture modeling (31,32), which can include both continuous and categorical variables (see Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25089) (Mplus). The model included the 13 previously described variables for subgroup identification, and the long-term outcome (radiographic knee OA) was included as an auxiliary variable (33) using the automatic Bolck, Croon, and Hagenaars (34,35) procedure. Individuals were assigned to a latent class based on their highest posterior probability. Missing data were handled using Mplus' maximum-likelihood estimator.

The number of subgroups was determined based on multiple factors, including the fit criteria of Akaike's information criterion (AIC) (36), Bayesian information criterion (BIC) (37), and sample-size adjusted BIC (37), and evaluating class homogeneity by examining entropy and tests of model comparison (38,39). Lower scores are better for AIC, BIC, and adjusted BIC, while a higher entropy (between 0 and 1) indicates a better separation of the classes with a high level of cohesion within classes (40). Vuong-Lo-Mendell-Rubin (VLMR), Lo-Mendell-Rubin likelihood ratio, and the bootstrap likelihood ratio tests were used to determine whether a model with k classes fit better than a model with k – 1 classes (39). Significant values (i.e., *P* values less than or equal to 0.050) indicate that a model with k classes fits better than a model with 1 class less. Finally, clinical relevance and class sizes (≥5% of the cohort in each group) (41) were evaluated by expert opinion to ensure that the differences in group membership were clinically meaningful.

Variables used to form the subgroups and the primary long-term outcome were compared across subgroups within the mixed model. Secondary long-term outcomes were compared using a chi-square test for categorical variables and analysis of variance for continuous variables, with a Bonferroni correction applied to post hoc testing (SPSS, version 26). These comparisons were done after subgroup enumeration to prevent any influence on class performance. All variables were assessed for normality using boxplots and histograms prior to comparison across subgroups.

RESULTS

Model fit statistics. A total of 293 participants (45.7% male, mean ± SD age 26.2 ± 9.4 years, 58 ± 35 days from injury) were included in this study. The best fitting model identified 4 latent subgroups (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25089) based on information criteria, class size, and clinical relevance. While a 5-subgroup model had the highest entropy and the lowest AIC, BIC, and sample-size adjusted BIC, one of the classes only consisted of 8 individuals (<5% of the sample) (40). Therefore, a 4-subgroup model was chosen, as its entropy (0.82) was nearly identical (0.83), and the AIC, BIC, sample-size adjusted BIC, and VLMR likelihood ratio test were all lower than 2- and 3-group models (36–39). The VLRM *P* value was only significant in a 2-group model and uninformative in the other models (39).

Group formation. The specific patient demographics at baseline are reported as probability weighted results (Table 1). Subgroups primarily differed on age ($P < 0.001$) and self-reported outcomes (i.e., IKDC, KOS-ADLS, global rating score; $P < 0.001$) at baseline (Table 1). Group 1 (younger good self-report, n = 99 [34%]) and 2 (younger poor self-report, n = 119 [41%]) were on average age <25 years, and group 3 (older poor self-report, n = 48 [16%]) and 4 (older good self-report, n = 27 [9%]) were on average age >30 years.

Latent subgroups. The 2 younger subgroups were significantly younger than the 2 older subgroups (mean ± SD age 22.7 ± 0.9 and 24.6 ± 1.3, respectively, versus 31.3 ± 2.7 and 36.3 ± 3.0 years, respectively; $P < 0.001$). The younger good self-report group and older poor self-report group had higher percentages of male participants (62% and 69%, respectively, versus 47% and 33%; $P < 0.03$), the older good self-report group had a higher percentage of female participants (67%; $P < 0.01$), and the younger poor self-report group was evenly split. The older poor self-report group had a higher BMI than the other subgroups (mean ± SD 26.2 ± 0.9; $P = 0.003$). The younger good self-report group had the best functional and self-reported outcomes ($P < 0.02$) (Table 1), while the older poor self-report group had the poorest functional and self-reported outcomes. The younger poor self-report group had the second-best functional outcomes, but the older good self-report group had the second-best self-reported outcomes (Figure 2).

Table 1. Comparisons of participant characteristics and subgroup comparisons (1–4)*

		Younger self-report		Older self-report		*P* values						
	Total	1: good (n = 99)	2: poor (n = 119)	3: poor (n = 48)	4: good (n = 27)		1 vs. 2	1 vs. 3	1 vs. 4	2 vs. 3	2 vs. 4	3 vs. 4
Baseline demographics												
Age, years	26.2 ± 9.4	22.7 ± 0.9	24.6 ± 1.3	31.3 ± 2.7	36.3 ± 3.0	<0.001†	0.30	<0.001†	<0.001†	<0.001†	<0.001†	0.020†
F:M ratio	134:159	38:61	63:56	15:33	18:9	0.004†	0.03†	0.40	0.01†	0.01†	0.20	0.003†
BMI, kg/m^2	24.7 ± 4.0	24.4 ± 0.5	24.4 ± 0.4	26.2 ± 0.9	24.7 ± 1.0	0.040†	0.97	0.03†	1.0	0.07	0.99	0.18
Days from injury	56.6 ± 30.6	50.5 ± 3.1	56.4 ± 3.8	62.6 ± 7.2	69.3 ± 12.2	0.020†	0.67	0.11	0.04†	0.48	0.18	0.87
Preinjury, level 1:2	203:90	86:13	87:32	30:18	0:27	<0.001†	0.01†	<0.001†	<0.001†	0.18	<0.001†	<0.001†
Concomitant injury, no. (%)	147 (50)	46 (47)	66 (55)	29 (60)	6 (22)	0.006†	0.19	0.11	0.023†	0.56	0.002†	0.002†
Previous ACL tear, no. (%)	24 (8)	11 (11)	8 (7)	5 (10)	0 (0)	0.24	–	–	–	–	–	–
Potential coper, no. (%)‡	173 (59)	87 (88)	69 (58)	1 (2)	16 (59)	<0.001†	<0.001†	<0.001†	<0.001†	<0.001†	0.87	<0.001†
Function at baseline												
Quadriceps strength	89.4 ± 11.0	94.0 ± 1.3	88.2 ± 1.6	85.1 ± 1.5	85.1 ± 2.1	<0.001†	<0.001†	<0.001†	<0.001†	0.22	0.22	0.99
Single hop	89.7 ± 11.9	94.7 ± 1.1	89.8 ± 1.5	80.4 ± 3.23	85.0 ± 2.7	<0.001†	0.02†	<0.001†	<0.001†	<0.001†	0.13	0.30
Timed hop	94.0 ± 9.3	98.5 ± 1.0	93.5 ± 1.1	87.3 ± 2.8	88.8 ± 2.6	<0.001†	<0.001†	<0.001†	<0.001†	0.002†	0.02†	0.99
Self-reported outcome measures at baseline												
IKDC	70.7 ± 12.55	81.6 ± 2.0	66.7 ± 1.7	53.6 ± 1.7	77.6 ± 2.3	<0.001†	<0.001†	<0.001†	0.01†	<0.001†	<0.001†	<0.001†
KOS-ADLS	84.6 ± 10.6	93.3 ± 1.2	83.2 ± 1.8	66.9 ± 1.4	90.1 ± 1.4	<0.001†	<0.001†	<0.001†	0.02†	<0.001†	<0.001†	<0.001†
Global rating score	78.4 ± 14.0	86.7 ± 1.7	75.8 ± 1.4	62.9 ± 4.1	86.0 ± 2.3	<0.001†	<0.001†	<0.001†	1.0	<0.001†	<0.001†	<0.001†
Secondary outcomes at 5 years: new injury												
Ipsilateral ACL tear, no. (%)	23 (9.7)	12 (12.2)	8 (10.7)	3 (7.5)	0 (0)	0.31	–	–	–	–	–	–
Contralateral ACL tear, no. (%)	15 (6.4)	7 (9.3)	6 (6.3)	1 (2.5)	1 (4.2)	0.51	–	–	–	–	–	–
All second injuries, no. (%)	54 (23.1)	20 (24.2)	23 (26.7)	8 (20)	3 (12.5)	0.50	–	–	–	–	–	–

* Values are the mean ± SD unless indicated otherwise. ACL = anterior cruciate ligament; BMI = body mass index; F:M = female:male; IKDC = International Knee Documentation Committee; KOS-ADLS = Knee Outcome Survey–Activity of Daily Living Subscale.
† Statistically significant.
‡ Potential coper is defined as KOS-ADLS scores ≥80%, global rating scale of perceived function scores ≥60%, symmetry on the timed hop ≥80%, and ≤1 episode of knee giving way during activities of daily living.

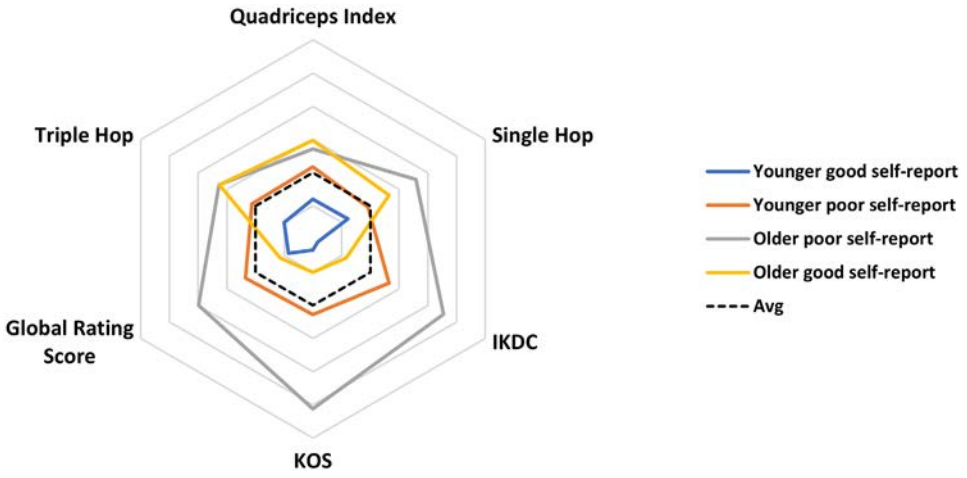

Figure 2. Comparison of functional performance and self-reported outcomes among subgroups (colored lines) and the group average (broken black line). Variables have been standardized and adjusted so that lines closer to the center represents better function or outcome. IKDC = International Knee Documentation Committee; KOS = Knee Outcome Survey.

Long-term outcomes. The younger good self-report group had the lowest percentage of involved knee radiographic OA (30%) while the older poor self-report group and older good self-report group had higher incidences (47% versus 53%), though not statistically significant ($P = 0.073$) (Table 2 and Figures 2 and 3).

Subgroups were statistically different in the development of uninvolved knee radiographic OA ($P = 0.004$) (Table 2). Pairwise comparisons revealed that the younger good self-report group had a lower prevalence of uninvolved knee radiographic OA than both the older poor self-report group ($P = 0.031$) and the older good self-report group ($P = 0.001$). Differences were also identified between the younger poor self-report group and the older good self-report group ($P = 0.006$), where the younger good self-report group had the lowest percentage of uninvolved radiographic OA (17%), and the older good self-report group had the greatest (58%).

The development of clinical OA was statistically different between subgroups ($P = 0.017$). The younger good self-report group had the lowest rate of clinical OA (11%), which was significantly lower than the younger poor self-report group (25%; $P = 0.019$) and the older poor self-report group (33%; $P = 0.007$). There was a significant difference among subgroups in operative status at 2 years. The older good self-report group had a significantly lower percentage of individuals who underwent operative management (44%) compared to the older poor self-report (79%; $P = 0.004$), younger poor self-report group (74%; $P = 0.011$), and the younger good self-report subgroups (72%; $P = 0.019$). No significant difference was found between subgroups in new injuries at 5 years, including ipsilateral and contralateral ACL reruptures ($P > 0.31$) (Table 1).

DISCUSSION

The purpose of this study was 2-fold: 1) to identify whether latent subgroups of ACL-injured individuals exist based on patient

Table 2. Long-term outcomes based on group membership*

	Younger self-report		Older self-report		
	1: good (n = 99)	2: poor (n = 119)	3: poor (n = 48)	4: good (n = 27)	P†
Radiographic OA involved	19/64 (30)	24/77 (31)	17/36 (47)	10/19 (53)	0.073
K/L grade 1	9	15	1	3	–
K/L grade 2	9	8	13	7	–
K/L grade 3	1	1	3	0	–
Radiographic OA uninvolved	11/65 (17)	18/76 (24)	13/36 (36)	11/19 (58)	0.004‡
K/L grade 1	11	18	13	11	–
K/L grade 2	0	0	0	0	–
K/L grade 3	0	0	0	0	–
Clinical OA	8/71 (11)	24/95 (25)	13/39 (33)	6/24 (25)	0.017‡
Return to preinjury sport level	45/73 (62)	53/91 (58)	25/38 (66)	13/20 (65)	0.013‡
Operative status	68/95 (72)	86/117 (74)	37/47 (79)	12/26 (44)	0.039‡

* Values are the number/total number available at time point (% yes) unless indicated otherwise.
† Adjusted *P* value reported. *P* value is for chi-square analysis for the presence of radiographic osteoarthritis (OA) between subgroups; it does not take Kellgren/Lawrence (K/L) level into account.
‡ Statistically significant.

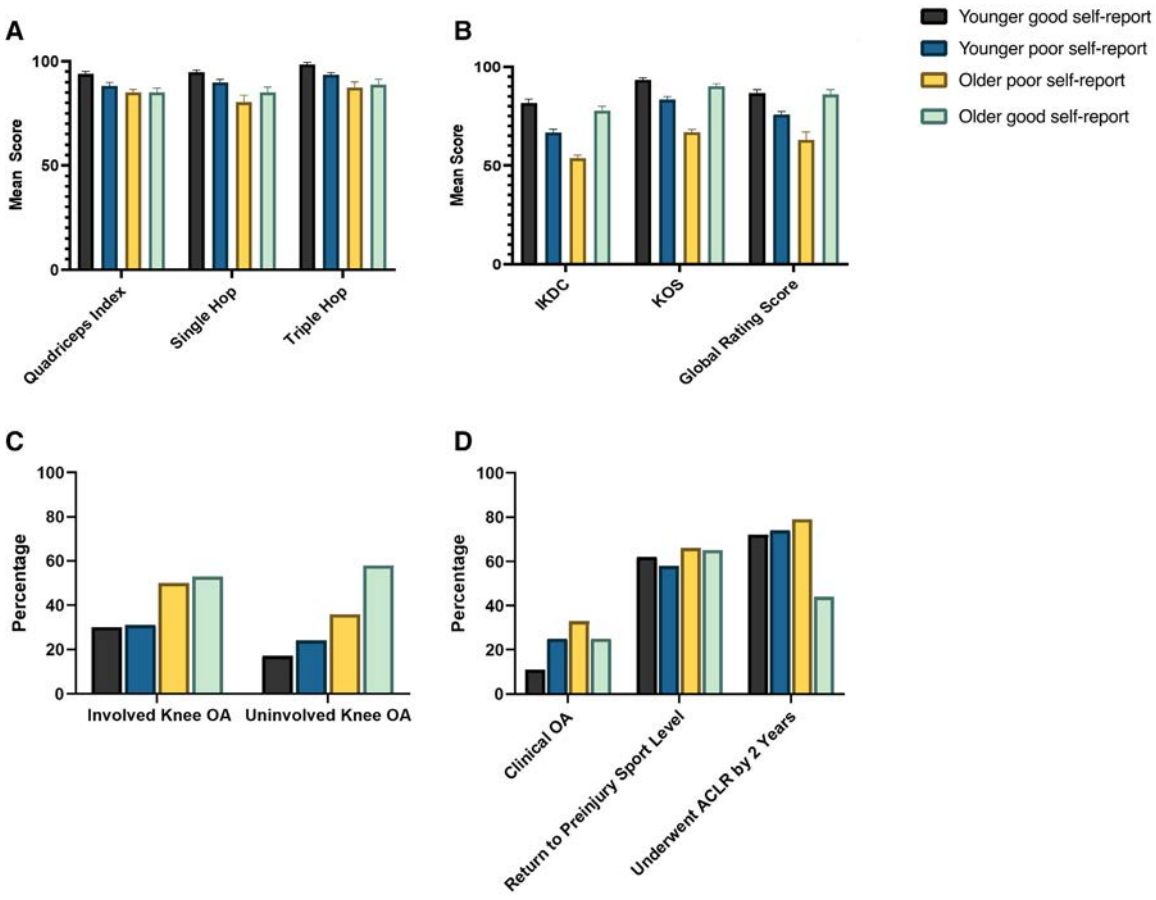

Figure 3. Group differences at baseline in function and self-reported outcomes (mean ± SD), primary outcome at 5 years, and secondary outcomes at 2 and 5 years between subgroups. **A**, Functional outcomes; **B**, Patient-reported outcome measures; **C**, Radiographic knee osteoarthritis (OA); **D**, Secondary outcomes. ACLR = anterior cruciate ligament reconstruction; IKDC = International Knee Documentation Committee; KOS = Knee Outcome Survey. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25089/abstract.

characteristics, self-reported outcome measures, and functional performance at baseline shortly after their ACL ruptures, and 2) to determine associations between subgroups and posttraumatic OA and clinically relevant long-term outcomes 2–5 years after ACL injury. We identified 4 subgroups at baseline within our population of individuals after ACL rupture (Table 1 and Figure 2): younger good self-report, younger poor self-report, older poor self-report, and older good self-report. The latent subgroups found in this study demonstrated distinct characteristics that may provide insight into both variability in patient outcomes and clinical rehabilitation targets for patients within each subgroup. Each latent subgroup demonstrated differences in the prevalence of uninvolved knee radiographic OA and clinical OA at 5 years, the percentage undergoing operative management by 2-years, and potential coper status at baseline.

The younger good self-report and younger poor self-report subgroups were the largest subgroups (34% and 41% of the cohort, respectively). The younger good self-report group was the highest performing group on all functional and self-reported outcome measures and were predominately classified as potential copers (88%) at baseline (Table 1). Both young subgroups had a comparable majority who underwent operative management (younger good self-report: 72%, younger poor self-report: 74%). Long-term, the younger good self-report group had the lowest percentage of involved and uninvolved radiographic OA, and the lowest percentage of clinical OA.

The younger poor-self report group was the closest to the group average in all baseline characteristics (Figure 2). The younger poor-self report group had acceptable outcomes on all functional measures at baseline, ranging from mean ± SD 88.2 ± 1.6 LSI for quadriceps strength up to 93.5 ± 1.1 LSI for the triple hop. Self-reported outcome measures, however, were second to lowest in this group. The lowest mean ± SD score was 66.7 ± 1.7 for IKDC and the highest was 83.2 ± 1.8 for the KOS-ADLS. These data indicate that although the patients were on the cusp of normal return-to-sport values for function at baseline, they had substantial knee-related symptoms that may have ultimately hindered their successful return to preinjury activity levels at 2 years after ACL rupture.

The older poor self-report and older good self-report subgroups made up smaller percentages (16 and 9%, respectively) of the sample. Notably, the older poor self-report group shared

similar rates of individuals who chose operative management (79%) as the 2 younger subgroups. The older good self-report group, conversely, had the lowest percentage of individuals choosing operative management (44%). Like their larger, younger counterpart subgroups, the older subgroups differed primarily on self-reported outcome measures at baseline (Table 1). The older poor self-report group had the lowest scores across all self-reported outcome measures, representing the group with the poorest self-assessed function and functional performance. The older poor self-report group scored significantly lower on the self-reported outcomes than the older good self-report group. Interestingly, the older poor self-report group's functional test outcomes were not significantly different from the older good self-report group. The older poor self-report group also had the highest percentage of people who reported early clinical knee OA. Further, the older poor self-report group had the highest percentage of individuals who chose operative management (79%) and the lowest percentage of potential copers (2%). The older poor self-report group having the lowest self-reported outcome measures may partially explain the high percentage of operative management, as these individuals may have had knee-related symptoms preventing them from success with nonoperative treatment.

The older good self-report group had the highest percentage of both involved and uninvolved knee radiographic OA. The uninvolved knees in this group had a higher percentage of radiographic OA than the involved knee, suggesting that the ACL injury may not be the main factor in this group. Further, all K/L grades in the uninvolved knee for all groups were at K/L grade 1, suggesting a relatively early-stage disease process (Table 2). The older good self-report group had no level I athletes to begin with and was predominately female compared to the other subgroups. They also had the highest percentage of individuals who chose nonoperative management (54%). The self-reported outcome measures of the older good self-report group, however, exceeded those of the younger poor self-report group and the older poor self-report group and were the second highest in the sample, but also had the lowest quadriceps LSI at baseline. Clinically, this subgroup may represent individuals who may benefit from education on the risk of the development of posttraumatic OA at baseline, and the importance of maintaining quadriceps strength to support long-term knee joint health (30,42). Future work assessing the qualitative reason for selecting to reduce the level of sport after ACL rupture is needed to confirm our speculation.

Although the oldest subgroup may be expected to have the highest percentage of individuals with radiographic changes in both the involved and uninvolved knees, the percentages of individuals meeting our definition of knee OA in the younger subgroups is highly concerning. At a mean age of 22 years, our youngest subgroup, the younger good self-report group, demonstrated radiographic changes in 30% of ipsilateral and 17% of contralateral knees at 5 years after ACL rupture. These numbers are consistent with literature suggesting that anywhere from 30% to 90% of individuals develop knee OA within 10 years of ACL rupture (1,43,44). The individuals in the older 2 subgroups that have radiographic OA data were an average age of 38 and 43 years, respectively, at 5 years, falling far below the age range of idiopathic OA, which ranges between ages 55 and 64 years (45). Our results stress the need for widespread patient education regarding the risk of developing OA after knee joint injury for all patients after ACL rupture, regardless of subgroup (46). Properly understanding long-term risks may in turn affect decision-making with respect to a return to activity.

While age seems to be a differentiating factor among the subgroups, there may be other underlying mechanisms related to lifestyle that further affect the long-term outcomes. When considering the older poor self-report and older good self-report subgroups, the subgroups with the smallest number of individuals and older ages, lifestyle changes may explain some of the long-term outcomes. The older poor self-report and older good self-report subgroups had the lowest percentage of level I athletes at baseline, which may explain why they also had the highest percentage of individuals returning to preinjury sport level, as the preinjury level was inherently not as demanding on the knee. The individuals in these subgroups, being older, may want to balance knee limitations and an active lifestyle. Qualitative research on how goals change after ACL reconstruction has suggested a shift in some patients from return to sports participation as a primary goal to return to an active everyday life (47). Even among young athletes, a common theme of "balancing physical activity and future knee health" emerges as individuals consider their ACL injury in terms of long-term knee health (48).

Clinically, continuing to assess self-reported outcome measures throughout the course of rehabilitation is important. Not only does assessing self-reported outcome measures give a snapshot of where the patients feel they are, but often cases appear where the self-reported outcome measures and functional performance do not line up. We do not know what caused individuals to report their knee outcomes as lower than their measured functional outcomes. This phenomenon was particularly evident in the younger poor self-report group and older poor self-report group. These subgroups had the lowest scores on self-reported outcome measures, but their means on functional testing were not the lowest of the 4 subgroups. In fact, the younger poor self-report group functionally was the closest to the average of the total study sample (i.e., all subgroups combined) and had the second highest functional outcomes after the younger good self-report group. This mismatch in self-report function and functional outcomes may be explained by recent data that suggest an association between psychological factors (e.g., kinesiophobia) and a return to preinjury sport after ACL reconstruction (49). While psychological factors were not directly measured at baseline in the current study, literature does suggest a relationship between psychological factors and a number of functional outcomes, including

RTS (50,51) and second injury (52). This literature, however, is conflicting, with data suggesting that both high and low fear have negative relationships with outcomes (52,53). Self-reported function, specifically psychological factors, is an important next step in understanding the presence of subgroups in individuals after ACL rupture.

The results of our study suggest that there are subgroupings of individuals that may guide treatment after ACL rupture and reconstruction by providing support for developing a patient-centered approach. While returning to preinjury sport level may be a goal for some individuals, symptom management and returning to a generally active lifestyle may be the goal for others as they transition away from previous sport participation. This analysis provides support for developing an individual-based approach, where all aspects of baseline evaluation are incorporated to inform treatment decisions, including assessing multiple domains of self-reported outcome measures, function, patient age, and most importantly patient goals. Treatment should also include education on long-term outcomes after ACL rupture (posttraumatic OA), but also on outcomes most relevant to patients themselves and their individual goals. Trajectories of self-reported function 5 years after treatment have been assessed in the Delaware-Oslo cohort using the IKDC score to assess factors relating to the response after ACL injury and treatment (11). The current article differs, as it uses a variety of demographic information and functional and self-report outcomes to form baseline subgroups, and does not assess trajectories but rather determines baseline subgroup associations with 2–5-year outcomes.

There are limitations to consider when interpreting the data presented in this study. First, patients may fit into >1 subgroup clinically, and therefore treatment should continue to be multimodal and not just target one specific area. Both participants and variables included in this analysis were limited by the inclusion criteria and study design of the parent study, so results may not be generalizable to the broader patient population. Strength testing did differ slightly between sites, so data were reported using limb symmetry measures to ensure that strength data are comparable. Inclusion criteria were stringent, and individuals with more extensive concomitant injuries were excluded. A return to sport was defined as the first exposure to level I or II sport and did not necessarily mean full match play. Only the self-reported outcome component of the Luyten et al (12) criteria was applied to the sample, and the full criteria have not yet been validated. Similarly, the term "clinical knee OA" was used to describe the partial application of the Luyten et al (12) criteria in our sample to be consistent with previous published work from our cohort (30). However, this term may also be described as early knee OA symptoms and is consistent with the heterogeneity in early OA definitions for this population described by the most recent OPTIKNEE consensus (54).

Future research should work to develop definition and classification criteria to best identify individuals with posttraumatic knee OA at an early disease stage. Radiographs were only assessed at 5 years, and therefore we do not know the K/L grade of the knee joint at baseline. Finally, radiographic OA was defined as K/L grade ≥1, which is not defined as definite osteophytes like in grade 2. However, K/L grade ≥1 has been proposed as an alternative cutoff due to the demonstration of early joint disease and association with the ultimate progression of radiographic features (55,56). Finally, this study was a secondary analysis of a cohort study and was not originally powered to detect differences between subgroups within the larger group; therefore caution should be used when interpreting and applying results.

Four distinct subgroups were identified at baseline with clinically meaningful differences in long-term outcomes: younger good self-report, younger poor self-report, older poor self-report, and older good self-report. The younger good self-report group had the highest function, self-reported outcomes, and number of potential copers at baseline, along with the lowest percentage of involved and uninvolved radiographic OA and clinical OA long term. The younger poor self-report group was the closest to the total sample average in all variables at baseline and had the second lowest percentage of involved and uninvolved knee radiographic OA. The older poor self-report group had the lowest percentage of potential copers at baseline, the highest percentage of individuals returning to preinjury sport level at 2 years, the highest percentage of individuals choosing operative management, and the highest percentage of individuals with clinical OA at 5 years. Finally, the older good self-report group had the lowest percentage of individuals who chose operative management but the highest percentage of involved and uninvolved knee radiographic OA.

ACKNOWLEDGMENTS

The authors acknowledge Håvard Moksnes, Ingrid Eitzen, Annika Storevold, Ida Svege, Espen Selboskar, Karin Rydevik, Marte Lund, Kristine Roberg Hytten, Andrew Lynch, Mathew Failla, Airelle Giordano, and Angela Smith for their assistance with data collection for this study. They also acknowledge the Norwegian Sports Medicine Clinic, Nimi (www.nimi.no), for supporting the Norwegian Research Center for Active Rehabilitation with rehabilitation facilities and research staff. Additionally, the authors acknowledge the University of Delaware Physical Therapy Clinic rehabilitation resources and staff, and Martha Callahan and the Center for Human Research Coordination staff.

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. Arhos 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. Arhos, Pohlig, Silbernagel.

Acquisition of data. Risberg, Snyder-Mackler,

Analysis and interpretation of data. Arhos, Pohlig, Di Stasi, Risberg, Snyder-Mackler, Silbernagel.

REFERENCES

1. Barenius B, Ponzer S, Shalabi A, et al. Increased risk of osteoarthritis after anterior cruciate ligament reconstruction: a 14-year follow-up study of a randomized controlled trial. Am J Sports Med 2014;42: 1049–57.
2. Lohmander LS, Englund PM, Dahl LL. The long-term consequence of anterior cruciate ligament and meniscus injuries: osteoarthritis. Am J Sports Med 2007;35:1756–69.
3. Janssen RP, du Mée AW, van Valkenburg J, et al. Anterior cruciate ligament reconstruction with 4-strand hamstring autograft and accelerated rehabilitation: a 10-year prospective study on clinical results, knee osteoarthritis and its predictors. Knee Surg Sports Traumatol Arthrosc 2013;21:1977–88.
4. Beck NA, Lawrence JT, Nordin JD, et al. ACL tears in school-aged children and adolescents over 20 years. Pediatrics 2017;139: e20161877.
5. Ardern CL, Taylor NF, Feller JA, et al. Fifty-five percent return to competitive sport following anterior cruciate ligament reconstruction surgery: an updated systematic review and meta-analysis including aspects of physical functioning and contextual factors. Br J Sports Med 2014;48:1543–52.
6. Wiggins ME, Grandhi RK, Schneider DK, et al. Risk of secondary injury in younger athletes after anterior cruciate ligament reconstruction: a systematic review and meta-analysis. Am J Sports Med 2016;44: 1861–76.
7. Fitzgerald GK, Axe MJ, Snyder-Mackler L. A decision-making scheme for returning patients to high-level activity with nonoperative treatment after anterior cruciate ligament rupture. Knee Surg Sports Traumatol Arthrosc 2000;8:76–82.
8. Eastlack ME, Axe MJ, Snyder-Mackler L. Laxity, instability, and functional outcome after ACL injury: copers versus noncopers. Med Sci Sports Exerc 1999;31:210–5.
9. Eitzen I, Eitzen TJ, Holm I, et al. Anterior cruciate ligament-deficient potential copers and noncopers reveal different isokinetic quadriceps strength profiles in the early stage after injury. Am J Sports Med 2010;38:586–93.
10. Rudolph KS, Axe MJ, Buchanan TS, et al. Dynamic stability in the anterior cruciate ligament deficient knee. Knee Surg Sports Traumatol Arthrosc 2001;9:62–71.
11. Pedersen M, Grindem H, Berg B, et al. Four distinct 5-year trajectories of knee function emerge in patients who followed the Delaware-Oslo ACL Cohort treatment algorithm. Am J Sports Med 2022;50: 2944–52.
12. Luyten FP, Bierma-Zeinstra S, Dell'Accio F, et al. Toward classification criteria for early osteoarthritis of the knee. Semin Arthritis Rheum 2018;47:457–63.
13. Pedersen M, Grindem H, Johnson JL, et al. Clinical, functional, and physical activity outcomes 5 years following the treatment algorithm of the Delaware-Oslo ACL Cohort study. J Bone Joint Surg Am 2021;103:1473–81.
14. Pedersen M, Grindem H, Berg B, et al. Low rates of radiographic knee osteoarthritis 5 years after ACL reconstruction or rehabilitation alone: the Delaware-Oslo ACL Cohort study. Orthop J Sport Med 2021;9: 23259671211027530.
15. Sturgill LP, Snyder-Mackler L, Manal TJ, et al. Interrater reliability of a clinical scale to assess knee joint effusion. J Orthop Sport Phys Ther 2009;39:845–9.
16. Daniel DM, Stone ML, Dobson BE, et al. Fate of the ACL-injured patient: a prospective outcome study. Am J Sports Med 1994;22: 632–44.
17. Eitzen I, Moksnes H, Snyder-Mackler L, et al. A progressive 5-week exercise therapy program leads to significant improvement in knee function early after anterior cruciate ligament injury. J Orthop Sport Phys Ther 2010;40:705–21.
18. Wellsandt E, Axe MJ, Snyder-Mackler L. Poor performance on single-legged hop tests associated with development of posttraumatic knee osteoarthritis after anterior cruciate ligament injury. Orthop J Sport Med 2018;6:2325967118810775.
19. Patterson B, Culvenor AG, Barton CJ, et al. Poor functional performance 1 year after ACL reconstruction increases the risk of early osteoarthritis progression. Br J Sports Med 2020;54:546–53.
20. Barber SD, Noyes FR, Mangine RE, et al. Quantitative assessment of functional limitations in normal and anterior cruciate ligament-deficient knees. Clin Orthop Relat Res 1990;255:204–14.
21. Noyes FR, Barber SD, Mangine RE. Abnormal lower limb symmetry determined by functional hop tests after anterior cruciate ligament rupture. Am J Sports Med 1991;19:513–8.
22. Irrgang JJ, Snyder-Mackler L, Wainner RS, et al. Development of a patient-reported measure of function of the knee. J Bone Joint Surg Am 1998;80:1132–45.
23. Hopper DM, Goh SC, Wentworth LA, et al. Test-retest reliability of knee rating scales and functional hop tests one year following anterior cruciate ligament reconstruction. Phys Ther Sport 2002;3:10–8.
24. Irrgang JJ, Anderson AF, Boland AL, et al. Development and validation of the International Knee Documentation Committee Subjective Knee Form. Am J Sports Med 2001;29:600–13.
25. Irrgang JJ, Anderson AF, Boland AL, et al. Responsiveness of the International Knee Documentation Committee Subjective Knee Form. Am J Sports Med 2006;34:1567–73.
26. Le Graverand MP, Vignon EP, Brandt KD, et al. Head-to-head comparison of the Lyon Schuss and fixed flexion radiographic techniques: long-term reproducibility in normal knees and sensitivity to change in osteoarthritic knees. Ann Rheum Dis 2008;67:1562–6.
27. Hunter DJ, Altman RD, Cicuttini F, et al. OARSI clinical trials recommendations: knee imaging in clinical trials in osteoarthritis. Osteoarthritis Cartilage 2015;23:698–715.
28. Kothari M, Guermazi A, von Ingersleben G, et al. Fixed-flexion radiography of the knee provides reproducible joint space width measurements in osteoarthritis. Eur Radiol 2004;14:1568–73.
29. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957;16:494–502.
30. Arhos EK, Thoma LM, Grindem H, et al. Association of quadriceps strength symmetry and surgical status with clinical osteoarthritis 5 years after anterior cruciate ligament rupture. Arthritis Care Res (Hoboken) 2022;74:386–91.
31. McLachlan GJ, Peel D. Finite mixture models. New York (NY): John Wiley; 2000.
32. Oberski D. Mixture models: latent profile and latent class analysis. In: Robertson J, Kaptein M, editors. Modern statistical methods for HCI. Switzerland: Springer Cham; 2016. p. 275–287.
33. Asparouhov T, Muthen B. Auxiliary variables in mixture modeling: using the BCH method in Mplus to estimate a distal outcome model and an arbitrary secondary model. Mplus web notes 21, version 11. 2021. URL: https://www.statmodel.com/examples/webnotes/webnote21.pdf.
34. Bakk Z, Vermunt JK. Robustness of stepwise latent class modeling with continuous distal outcomes. Struct Equ Modeling 2016;23: 20–31.
35. Vermunt JK. Latent class modeling with covariates: two improved three-step approaches. Polit Anal 2010;18:450–69.

36. Akaike H. Factor analysis and AIC. Psychometrika 1987;52:317–32.
37. Schwarz G. Estimating the dimension of a model. Ann Stat 1978;6: 461–4.
38. Geiser C. Data analysis with Mplus. New York: Guilford Press; 2013.
39. Henson JM, Reise SP, Kim KH. Detecting mixtures from structural model differences using latent variable mixture modeling: a comparison of relative model fit statistics. Struct Equ Modeling 2007;14: 202–26.
40. Celeux G, Soromenho G. An entropy criterion for assessing the number of clusters in a mixture model. J Classif 1996;13:195–212.
41. Yang CC. Evaluating latent class analysis models in qualitative phenotype identification. Comput Stat Data Anal 2006;50:1090–104.
42. Øiestad BE, Juhl CB, Eitzen I, et al. Knee extensor muscle weakness is a risk factor for development of knee osteoarthritis: a systematic review and meta-analysis. Osteoarthritis Cartilage 2015;23:171–7.
43. Lie MM, Risberg MA, Storheim K, et al. What's the rate of knee osteoarthritis 10 years after anterior cruciate ligament injury? An updated systematic review. Br J Sports Med 2019;53:1162–7.
44. Lohmander LS, Östenberg A, Englund M, et al. 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.
45. Losina E, Weinstein AM, Reichmann WM, et al. Lifetime risk and age of diagnosis of symptomatic knee osteoarthritis in the US. Arthritis Care Res (Hoboken) 2013;65:703–11.
46. Feucht MJ, Cotic M, Saier T, et al. Patient expectations of primary and revision anterior cruciate ligament reconstruction. Knee Surg Sport Traumatol Arthrosc 2016;24:201–7.
47. Heijne A, Silbernagel KG, Lundberg M. "I don't opt out of things because I think I will get a sore knee, but I don't expose myself to stupid risks either." Patients' experiences of a second ACL injury: an interview study. Knee Surg Sport Traumatol Arthrosc 2022;30: 2244–50.
48. Truong LK, Mosewich AD, Miciak M, et al. Balance, reframe, and overcome: the attitudes, priorities, and perceptions of exercise-based activities in youth 12–24 months after a sport-related ACL injury. J Orthop Res 2022;40:170–81.
49. Baez SE, Hoch MC, Hoch JM. Psychological factors are associated with return to pre-injury levels of sport and physical activity after ACL reconstruction. Knee Surg Sport Traumatol Arthrosc 2020;28: 495–501.
50. Ardern CL, Österberg A, Tagesson S, et al. The impact of psychological readiness to return to sport and recreational activities after anterior cruciate ligament reconstruction. Br J Sports Med 2014;48:1613–9.
51. Ardern CL, Taylor NF, Feller JA, et al. Sports participation 2 years after anterior cruciate ligament reconstruction in athletes who had not returned to sport at 1 year: a prospective follow-up of physical function and psychological factors in 122 athletes. Am J Sports Med 2015;43:848–56.
52. McPherson AL, Feller JA, Hewett TE, et al. Psychological readiness to return to sport is associated with second anterior cruciate ligament injuries. Am J Sports Med 2019;47:857–62.
53. Paterno MV, Flynn K, Thomas S, et al. Self-reported fear predicts functional performance and second ACL injury after ACL reconstruction and return to sport: a pilot study. Sports Health 2018;10:228–33.
54. Whittaker JL, Culvenor AG, Juhl CB, et al. OPTIKNEE 2022: consensus recommendations to optimise knee health after traumatic knee injury to prevent osteoarthritis. Br J Sports Med 2022;56:1393–405.
55. Thorstensson CA, Petersson IF, Jacobsson LT, et al. Reduced functional performance in the lower extremity predicted radiographic knee osteoarthritis five years later. Ann Rheum Dis 2004;63:402–7.
56. Hart DJ, Spector TD. Kellgren & Lawrence grade 1 osteophytes in the knee: doubtful or definite? Osteoarthritis Cartilage 2003;11:149–50.

Arthritis Care & Research
Vol. 75, No. 9, September 2023, pp 1925–1938
DOI 10.1002/acr.25085

American College of Rheumatology
Empowering Rheumatology Professionals

Absence of Improvement With Exercise in Some Patients With Knee Osteoarthritis: A Qualitative Study of Responders and Nonresponders

Rana S. Hinman,[1] Sarah E. Jones,[1] Rachel K. Nelligan,[1] Penelope K. Campbell,[1] Michelle Hall,[1] Nadine E. Foster,[2] Trevor Russell,[3] and Kim L. Bennell[1]

Objective. To compare the perceptions of patients about why they did, or did not, respond to a physical therapist–supported exercise and physical activity program.

Methods. This was a qualitative study within a randomized controlled trial. Twenty-six participants (of 40 invited) with knee osteoarthritis sampled according to response (n = 12 responders, and 14 nonresponders based on changes in both pain and physical function at 3 and 9 months after baseline) to an exercise and physical activity intervention. Semistructured individual interviews were conducted. Inductive thematic analysis was undertaken within each subgroup using grounded theory principles. A deductive approach compared themes and subthemes across subgroups. Findings were triangulated with quantitative data.

Results. (Sub)themes common to responders and nonresponders included the intervention components that facilitated engagement, personal attitudes and expectations, beliefs about osteoarthritis and exercise role, importance of adherence, and perceived strength gains with exercise. In contrast to responders who felt empowered to self-manage, nonresponders accepted responsibility for lack of improvement in pain and function with exercise, acknowledging that their adherence to the intervention was suboptimal (confirmed by quantitative adherence data). Nonresponders believed that their excess body weight (supported by quantitative data) contributed to their outcomes, encountered exercise barriers (comorbidities, stressors, and life events), and perceived that the trial measurement tools did not adequately capture their response to exercise.

Conclusion. Responders and nonresponders shared some similar perceptions of exercise. However, along with perceived limitations in trial outcome measurements, nonresponders encountered challenges with excess weight, comorbidities, stressors, and life events that led to suboptimal adherence and collectively were perceived to contribute to nonresponse.

INTRODUCTION

Over 260 million people globally have knee osteoarthritis (OA) (1), a condition that accounts for a considerable proportion of global disability. Joint pain and physical dysfunction are common features of knee OA and the main reasons that drive people to seek care from health professionals (2,3). There is no cure for knee OA, and arthroplasty is typically reserved for patients with end-stage disease whose joint pain has not been adequately relieved by appropriate nonsurgical approaches.

Clinical guidelines advocate nondrug nonsurgical strategies (4–6) focused on self-management. Exercise and physical activity are recommended as standard care for all patients with OA throughout the course of the disease (7). Muscle weakness is

The views expressed herein are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

Supported by the National Health and Medical Research Council (grant 1157977). Dr. Hinman's work was supported by the National Health and Medical Research Council (Senior Research fellowship grant 1154217). Dr. Hall's work was supported by the National Health and Medical Research Council (Investigator grant 1172928). Dr. Foster's work was supported by the NIHR (research professorship grant NIHRRP-011-015). Dr. Bennell's work was supported by the National Health and Medical Research Council (Investigator grant 1174431).

[1]Rana S. Hinman, BPhysio (Hons), PhD, Sarah E. Jones, BSc (Hons), PhD, Rachel K. Nelligan, BPhysio, PGCertPhysio, PhD, Penelope K. Campbell, BAppSci (FoodSci&Nutr), Michelle Hall, BSc (Hons), MSc, PhD, Kim L. Bennell, BAppSci (Physio), PhD: The University of Melbourne, Melbourne, Victoria, Australia; [2]Nadine E. Foster, BSc (Hons), DPhil: The University of Queensland and Metro North Health, Brisbane, Queensland, Australia; [3]Trevor Russell, BPhysio, PhD: The University of Queensland, Brisbane, Queensland, Australia.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.25085&file=acr25085-sup-0001-Disclosureform.pdf.

Address correspondence via email to Rana S. Hinman, BPhysio (Hons), PhD, at ranash@unimelb.edu.au.

Submitted for publication September 27, 2022; accepted in revised form December 29, 2022.

SIGNIFICANCE & INNOVATIONS

- Nonresponders accepted responsibility for lack of improvement in pain and function with exercise, acknowledging that their adherence was suboptimal.
- Nonresponders believed that their excess body weight contributed to their outcomes, encountered exercise barriers (such as comorbidities, stressors, and life events), and perceived that trial measurement tools did not adequately capture their response to exercise.
- Qualitative design provides richer information over the few existing quantitative studies that explore moderators of exercise response and have yielded little information to date.

common in knee OA (8,9), and knee extensor weakness may increase the risk of worsening of knee pain and deterioration in physical functioning over time (10,11). Thus muscle strengthening is an important component of exercise management (12). Given that walking >6,000 steps/day protects against functional decline in patients with knee OA (13), and >50% of men and nearly 80% of women with or at risk of knee OA do not meet public health recommendations for physical activity (14), promotion of general physical activity is also advocated.

Although meta-analyses show, on average, moderate improvements in pain and function with exercise in knee OA (15,16), it is widely recognized that symptomatic improvements are not achieved by all patients (i.e., some patients are nonresponders to exercise) (17). Indeed, our own clinical trials in individuals with knee OA who underwent physical therapist–supported strengthening exercise and physical activity show that only 40–60% of participants reported global improvements in pain and physical function immediately after intervention (18–20). It is not clear why some patients with knee OA respond to exercise while others do not. A range of factors are barriers to exercise participation and have been postulated to play a role in determining treatment response, including but not limited to beliefs about OA and the role of exercise, capability to exercise, adherence, therapeutic relationship with the clinician, motivation, self-efficacy, health status, and mood (17,21–25). The limited research exploring factors associated with exercise response is largely confined to quantitative secondary analyses of randomized controlled trials (RCTs) (24). Qualitative studies that increase our understanding of factors that patients perceive to influence response to exercise-based treatment may facilitate earlier identification of individuals at risk of nonresponse and may help guide a more personalized patient-centered approach to exercise management. The aim of this qualitative study, therefore, was to compare the perceptions of patients about why they did, or did not, respond to a physical therapist–supported exercise and physical activity program.

PATIENTS AND METHODS

Design. We used a qualitative design based on an interpretivist paradigm in which knowledge about a phenomenon is developed by gathering perceptions and interpretations of participants who experience it (26). The phenomenon of interest in this study was symptomatic response and nonresponse to an exercise intervention for knee OA. This study was nested within our noninferiority RCT (27) comparing exercise and physical activity delivered via in-person consultations with a physical therapist to videoconferencing with a physical therapist for patients with knee OA (ANZCTR: 12619001240134). The Consolidated Criteria for Reporting Qualitative Research checklist (28) guided reporting of this study. Institutional human ethics approval was obtained, and participants provided consent.

Participants. Patients with knee OA were recruited for this qualitative study from both arms of the RCT. Recruitment procedures and selection criteria for the RCT have been published (27). Briefly, participants were recruited from the community in Victoria, Queensland, and New South Wales, Australia via advertisements and our volunteer database. Participants met National Institute for Health and Care Excellence (29) clinical criteria for OA (age ≥45 years; activity-related knee joint pain and morning knee stiffness ≤30 minutes), among other eligibility criteria, in order to participate in the RCT.

Purposive sampling from both trial arms was based upon RCT global rating of change data, which classified trial participants as responders or nonresponders. Global rating of change in knee pain and physical function were each rated by participants at 3 months and 9 months postrandomization using 7-point Likert scales (response options from "much worse" to "much better" compared to baseline). For each outcome at each time point, participants recording "moderately better" or "much better" were classified as improved, with all others classified as not improved. This approach is consistent with how we plan to analyze and interpret this outcome in our overarching noninferiority trial (27). For this qualitative study, participants classified as improved for both pain and function at 3 months, and who maintained improvement on both at 9 months, were deemed responders. Conversely, participants classified as not improved on both pain and function at 3 months, and who remained not improved on both at 9 months, were deemed nonresponders. Purposive sampling (August 2021 to May 2022) recruited each subgroup from participants completing the trial via email invitation.

Quantitative data were extracted from the RCT, including information on age, sex, body mass index, geographic location, employment status, comorbidities, number of consultations attended, consultation modality (in person or videoconferencing), and self-reported adherence to strengthening exercises and the physical activity plan.

Interventions. Briefly, the exercise and physical activity program (27) involved 5 individual consultations over 3 months with 1 of 15 trial physical therapists (in person or via videoconferencing). Physical therapists prescribed an individualized strengthening program (5–6 home exercises, 3 times/week) selected from an "Exercise Booklet" of 37 exercises. Participants were provided with exercise resistance bands. Review and modification of the strengthening program occurred at each consultation. An individualized physical activity plan was also devised aiming to increase physical activity to, or maintain it at, recommended levels (30). Participants were provided a wearable activity tracker and individualized step goals, which were reviewed and modified at each consultation. Participants were encouraged to use a "Knee Plan and Log Book" for recording adherence and monitoring progress. Education about OA and its management occurred at all sessions, supplemented by an "Osteoarthritis Information" booklet. Participants were advised how to independently progress their program and were encouraged to continue with it, after consultations ended, until 9 months postrandomization.

Interviews. A semistructured interview guide was developed (Table 1) informed by similar research in chronic whiplash (31) and broader research on barriers and facilitators to exercise in OA (22,23). The interview explored participants' perceptions about why they did/did not respond to the exercise intervention, including beliefs about OA and exercise, mood and psychological factors, lifestyle and other health problems, and the therapeutic relationship with the physical therapist. Individual telephone interviews were conducted by SEJ, a female PhD-qualified nonclinician researcher trained in qualitative methodologies, who was otherwise unknown to participants and was not involved in the RCT. All interviews were conducted with participants in their own home or workplace ~7 weeks (on average) after completion of a 9-month RCT outcome assessment and lasted ~45 minutes. Interviews were audio-recorded and transcribed verbatim by an external provider. Audio-recordings and transcripts were deidentified, with transcripts assigned gender-matched pseudonyms to maintain confidentiality. All data were stored on a password-protected university server.

Statistical analysis. Reflexive thematic analysis, applying principles of grounded theory, was performed separately for responders and nonresponders. After transcription, interviews were read by SEJ to familiarize with the data. Transcripts were then re-read, and open coding, using an inductive approach, was used to identify topics and patterns of ideas. To demonstrate credibility and confirmability, inductive analysis was also performed by a second researcher (RKN; a physical therapist with qualitative research experience not involved in the RCT). Both researchers independently organized codes into categories before discussing identified topics and patterns. Transcripts were also read by RSH (a physical therapist who leads the RCT and an experienced qualitative researcher). Topics identified by both SEJ and RKN were reviewed and discussed with RSH, and in collaboration, final axial coding was performed (32), whereby closely related codes were examined and collated to generate themes and subthemes within the subgroups. In this latter stage, a deductive approach compared and contrasted themes and subthemes relevant to the research question across the responders and nonresponders. Agreement was strong among researchers (SEJ, RKN, and RSH), therefore additional input into theme generation was not sought. Themes and subthemes are presented across responders and nonresponders with exemplary quotes (33). To ensure trustworthiness, findings were triangulated with relevant quantitative data from the RCT. For example, when a subtheme related to available quantitative data from the RCT, we considered if the quantitative data supported the qualitative finding for responders and nonresponders. Data management was supported using NVivo 12 software (QSR International) to organize, store, and index manual coding by SEJ and RKN and for cross-referencing with quotes.

Analysis occurred concurrently with data collection. Data collection within each subgroup ceased when the authors felt confident of having achieved, or at least closely approached, inductive thematic saturation (34). Specifically, inductive thematic saturation was determined via consensus between the authors who independently coded all manuscripts (SEJ and RKN) and was confirmed by a third author (RSH). Inductive thematic saturation was defined as the point in coding where there were mounting instances of the same codes, and where new codes did not lead to new theme generation nor new insights related to the research question.

RESULTS

Twelve (of 16 invited) responders (mean ± SD age 57 ± 7 years) and 14 (of 24 invited) nonresponders (mean ± SD age 67 ± 9 years) were interviewed (Table 2). Almost all participants attended 100% of their physical therapist consultations, except for 1 responder and 1 nonresponder (each attending 4 of 5). Themes and subthemes, with exemplary quotes, are summarized in Table 3. Figure 1 summarizes the similarities and differences across subgroups.

Theme 1 (facilitators). Responders and nonresponders spoke about accountability and monitoring, finding that tracking steps, recording exercise, and reporting to their physical therapist was motivating. When consultations stopped and external accountability ceased, it was more difficult to maintain exercise and physical activity habits, although the activity tracker helped keep participants accountable. Most participants believed that individualization, guided by the physical therapist, ensured that their program suited their needs. Nonresponders reflected that individualization was important when challenges were encountered to ensure that the exercise and physical activity

Table 1. Interview guide*

- Overall experiences
  1. Tell me about your overall experience with the PEAK program?
     - Prompts
       - What motivated you to volunteer?
       - What treatment were you expecting to receive when you signed up?
       - Was there anything included in the PEAK program that surprised you/was unexpected?
  2. At 3 and 9 months, your trial survey indicated that you experienced an [improvement/no improvement] in pain and function. Do you feel that you were a person who had "responded" to the treatment program? By responder I mean someone in whom the treatment "worked"
     - Prompts
       - What worked for you? What didn't work for you? Why? Why not?
  3. At 3 and 9 months, your trial survey indicated that you experienced an [improvement/no improvement] in pain and function. How did this outcome compare to what you were expecting to achieve when you signed up?
     - Prompts
       - What benefits/positive effects were you expecting?
       - What side/negative effects did you expect?
  4. Why do you think you [improved/did not improve]? Think about the exercise program, the physical activity plan, the physiotherapist, your own knee problems, your general health and your lifestyle when you answer.
     - Prompts
       - Responders
         - Were there other things that contributed to this improvement?
         - What was the most helpful/best part of the program?
         - What wasn't so good?
       - Nonresponders
         - Was something missing from the PEAK program?
         - Was something included that wasn't good for you?
         - What do you think would have helped you more?
- Osteoarthritis/chronic knee pain
  5. What do you think are the most effective treatments for chronic knee pain caused by osteoarthritis?
     - Prompts
       - What treatments have you tried?
       - What worked/didn't work?
  6. What do you think causes your knee pain?
     - Prompts
       - How much control do you have over your knee pain?
  7. How does your knee pain make you feel?
     - Prompts
       - How well do you cope with your knee pain? Why/why not?
- Exercise and physical activity
  8. How important do you think strengthening exercise is for your knee problems? Why?
     - Prompts
       - Have you done much exercise over your lifetime?
       - What have your past experiences with exercise been like?
       - Do you consider strengthening exercise to be "therapeutic" (provide definition of 'therapeutic' as: 'like medicine')? Why/why not?
  9. How did you feel about the strengthening exercises that were prescribed for you by the PEAK physiotherapist?
     - Prompts
       - How did your body respond to the strengthening exercise?
       - How closely did you follow the dosage/frequency/repetitions prescribed by the physio (including after the physio consults had ended)? Why/Why not?
       - What was easy/challenging about the exercises? What did you like/not like?
       - How well did the strengthening program suit your needs? How did the program change over time?
       - How much did you log/record your exercise sessions?
       - How motivated were you to exercise?
  10. How important do you think general physical activity is for your knee problems? Why?
     - Prompts
       - How important do you think physical activity is for your general health and well-being?
       - Do you consider general physical activity to be "therapeutic" (provide definition of 'therapeutic' as 'like medicine')? Why/why not?
  11. How did you feel about the physical activity program prescribed for you by the PEAK physiotherapist?
     - Prompts
       - How did your body respond to the general physical activity recommended by the physio?
       - How closely did you follow the physical activity plan (including after the physio consults had ended)? Why/Why not?
       - What was easy/challenging about the physical activity plan? What did you like/not like?
       - How well did the physical activity plan suit your needs? How did the plan change over time?
       - How much did you use the activity tracker?
       - How motivated were you to be physically active?

(Continued)

Table 1. (*Cont'd*)

Mood/psychological factors
12. Did anything about your mood, and how you felt, contribute to the [lack of improvement/amount of improvement] in pain and function with the PEAK program? (e.g., stress, anxiety, etc.)
Prompts
Did participating in the exercise program affect how you felt? How?
Did you talk about how you felt with your physiotherapist? How did that help/not help?
Does stress usually affect your knee pain? How so? Why do you think this happens?
Lifestyle and other health problems
13. Do you think anything related to your lifestyle and circumstances contributed to the fact that you [didn't get improvement/got improvement] in pain and function with the PEAK program?
Prompts
Sleep? Support from family/friends? Availability of time? Where you live?
14. To what extent do you think your general health influenced the fact that you [didn't get improvement/got improvement] in pain and function with the PEAK program?
Prompts
How would you rate your general health?
How much did other health problems limit you from exercising/being physically active?
Did you/physio have to modify your program to accommodate other health problems/needs?
Therapeutic relationship
15. How did you feel about your physiotherapist?
Prompts
Did you like them? Why/why not?
Did you trust them and the advice they gave you? Why/why not?
Did you understand what the physio wanted you to do?
How motivated were you to listen to the physio and follow their advice?
Do you think the physio had a good understanding of your knee problems and your goals?
16. How did you feel about the number of consultations you had with your physio?
Prompts
Would you have liked more? Less? Over what timeframe?
Longer consults? Shorter?
Closing comments
17. Is there anything else you wish to add?

* PEAK = Physiotherapy Exercise and Physical Activity for Knee Osteoarthritis.

program was achievable. Responders and nonresponders described a positive therapeutic relationship with their physical therapist. Participants generally trusted their physical therapist, perceived them as knowledgeable, and felt they understood their knee problems.

Theme 2 (personal attitudes and expectations). Both subgroups felt that there was nothing to lose by participating in the exercise and physical activity program. Participants described personal traits rooted in acceptance and realism with respect to their knee problems and the outcomes they expected from exercise. Participants did not expect a cure for their pain, rather, they largely had realistic functional goals they hoped to achieve. While most responders held a "no pain, no gain" attitude towards their exercises, some nonresponders were more resigned about their knee and less accepting of any pain encountered with exercise. Participants largely had perceived good health, despite most reporting comorbidities (Table 2). Unique to nonresponders was accepting responsibility. Nonresponders felt they were responsible for their nonresponse to exercise, acknowledging that adherence to exercise and physical activity goals was not as good as it should have been. This converges with quantitative data (Table 2) showing that one-half of nonresponders self-reported less than the prescribed number of exercise sessions at 3 months (compared with only 1 responder) and variable adherence to physical activity.

Theme 3 (osteoarthritis beliefs). Responders and nonresponders believed that exercise is important for managing knee OA. Responders believed that exercise played a crucial role in relieving their knee OA symptoms, in particular the strengthening exercises. While nonresponders also believed that exercise was important, belief systems were based more on logic rather than actual experiences. The subtheme "It's degenerative" arose from both subgroups, reflecting participant beliefs that their knees were worn out and OA was an inevitable aspect of aging associated with cartilage and bone breakdown. Only nonresponders reflected on their body weight as a contributor to OA, believing that extra weight likely contributed to their knee problems and lack of response with exercise. These perceptions converge with quantitative data (Table 2) showing a higher prevalence of obesity among nonresponders (n = 7 [50%]) compared to responders (n = 3 [25%]). Some nonresponders reflected on how past experiences with weight loss had helped their knee, and some described weight loss that they had observed as a result of the exercise and physical intervention, which they considered a positive outcome.

Table 2. Participant characteristics (n = 26)*

Pseudonym	Sex	Age, years	Weight†	Location‡	Employment	Comorbidity§	Consult mode	Exercise adherence (3 mo.)¶	Exercise adherence (9 mo.)¶	Physical activity adherence (3 mo.)#	Physical activity adherence (9 mo.)#
Responders (n = 12)**											
Bill	Male	60–64	Overweight	Major city	Retired		IP	3	3	8	7
Christine	Female	55–59	Obese	Inner regional	Work full time	Back pain	IP	3	1	8	6
Elizabeth	Female	70–74	Healthy	Major city	Retired	Heart disease	VC	3	3	10	10
Francis	Male	45–49	Overweight	Major city	Work full time		VC	3	2	9	9
James	Male	50–54	Overweight	Inner regional	Work full time	Back pain	VC	2	2	10	7
Judith	Female	55–59	Overweight	Major city	Work part time		VC	4	5	10	9
Lesley	Male	50–54	Healthy	Major city	Work full time		VC	3	0	9	2
Gregory	Male	55–59	Overweight	Major city	Work full time	Heart disease; back pain	VC	3	2	10	6
Nancy	Female	60–64	Obese	Major city	Retired	Lung disease; depression; bradycardia	IP	3	4	6	4
Patricia	Female	50–54	Obese	Inner regional	Work part time	Back pain; asthma	VC	4	0	10	2
Valerie	Female	70–74	Healthy	Major city	Retired	Back pain	VC	3	1	10	8
Melissa	Female	50–54	Overweight	Major city	Work full time		VC	3	2	10	8
Nonresponders (n = 14)**											
Beverley	Female	60–64	Obese	Outer regional	Unable to work due to health	Hypertension; cancer; back pain	IP	3	0	10	2
Edward	Male	75–79	Overweight	Major city	Unable to work due to health	Hypertension; back pain; gall stones	VC	7	4	7	8
Catherine	Female	60–64	Overweight	Major city	Retired	Depression; back pain	VC	0	0	8	0
Joseph	Male	60–64	Obese	Major city	Work full time	Heart disease; back pain	IP	0	0	5	1
Joyce	Female	75–79	Healthy	Major city	Retired	Back pain	IP	2	4	7	6
Kathleen	Female	55–59	Obese	Major city	Retired	Diabetes mellitus	VC	3	2	9	3
Patrick	Male	75–79	Overweight	Major city	Retired		VC	1	0	9	1
Matilda	Female	70–74	Overweight	Inner regional	Retired	Hypertension; back pain	VC	2	0	10	3
Lisa	Female	60–64	Obese	Major city	Work part time		VC	3	3	9	7
Susan	Female	60–64	Obese	Major city	Retired		IP	0	0	9	1
David	Male	80–84	Healthy	Major city	Retired	Hypertension; back pain; leukemia	VC	3	0	9	8
Wendy	Female	55–59	Overweight	Major city	Unemployed		VC	2	2	7	5
Michelle	Female	70–74	Severely obese	Inner regional	Unable to work due to health	Hypertension; back pain; shortness of breath	IP	3	0	10	8
Cynthia	Female	55–59	Obese	Major city	Work full time		VC	3	0	7	7

* IP = in person; VC = video conferencing.
† Classified according to the World Obesity classification system using body mass index calculated from self-reported height and weight.
‡ Metropolitan and regional classification, based on residential postcode, in accordance with Australian Standard Geographical Classification.
§ Reported using the Self-Administered Comorbidity Questionnaire.
¶ Self-reported number of strengthening exercise sessions performed over the prior week (in whole numbers), noting that participants were encouraged to perform exercises 3 times per week.
# Self-reported using a numerical rating scale (with terminal descriptors of 'strongly disagree' [score = 0] and 'strongly agree' [score = 10]) responding to the statement, "I followed the physical activity plan that my PEAK [Physiotherapy Exercise and Physical Activity for Knee Osteoarthritis] trial physiotherapist helped me to develop."
** Responders = participants who were improved for both pain and function at 3 months and who maintained improvement at 9 months; nonresponders = participants who were not improved on both of the pain and function variables at 3 months and remained not improved on both variables at 9 months.

Table 3. Themes, subthemes, and exemplary quotes across responders and nonresponders*

Themes and subthemes	Responders	Nonresponders
Theme 1: Facilitators		
Accountability and monitoring	Bill: "I guess the fact that I suppose I knew somebody was going to be marking my homework so to speak meant that there was that element as well. If I skip a day or whatever, what's [my physical therapist] going to say?" Lesley: "Just having that consultation and someone working with you along the way, there's a sense of obligation to yourself and to the other person." Gregory: "That was kind of a good little challenge to have your little [activity tracker] on your arm and see how many steps and you know if you need to go for an extra walk, well, I would"	Joyce: "When I was filling out the booklet all the time, that was an incentive to make sure I kept doing the exercises. And I've noticed since I haven't got to fill it out all the time, I'm not doing them 3 times a week, I've let it slip…I was using the stepper all the time – yes, I used the stepper every day. Even when I wasn't doing the exercise and that thing, I was putting it on every day and measuring my daily steps. That was good actually, that was a real motivator." Michelle: "I found going to [the physio], it made you do it." David: "Well, it just gave you something to work on. Yeah, look, it was something that I used to do every night before I went to bed, I'd fill it in…I thought that was good."
Individualization	Bill: "I guess certainly the physio treatment and the tailoring of the selection of the exercises and the ability to have that reviewed on a regular basis and ratchet it up accordingly [helped achieve results]." Nancy: "I was able to negotiate away from the ones that were awkward or difficult for me, to the ones that were easier or more present or physically possible in my house." Melissa: "You know, I think there was a couple that, they were hard to do and we just adjusted them and then worked towards the harder ones…I liked that you could build on it, so you weren't expected to just, you know, do a mammoth effort in the beginning."	David: "With [the physio], we sort of worked on trying to not do it – not go down as far, and so just played that one by ear as to a point where you were continuing to do them, but not to the full degree that it was originally required." Michelle: "When I first started, it hurt a lot; the first lot of exercises. And it made it that every time I took a step it felt like someone stabbed me in the front of the kneecap with a knife; it was that sharp. And then I went back to [the physio] for my next visit and he changed one of the exercises because it was irritating the knee…and when he changed that one, even though we still did the same exercise but minus the band, it was much better…" Kathleen: "She changed a couple of them because I just said to her, "I can't squat down on that chair, it just doesn't happen," and she did change a couple of them around for me…"
Positive therapeutic relationship	Bill: "Certainly if it's anything to do with the knee I'll seek [my PEAK physical therapist] out again and if it's to do with anything else is a very high probability that I'll seek him out again." Lesley: "I think he was really good. He was very easy to talk to, get along with. I think he explained everything really well." Melissa: "[My physical therapist] was really good. Really approachable and listened, and really took onboard whatever I said as well."	Patrick: "I thought she was excellent. She was good at looking at what was happening and trying to change the program to fit, and I thought that she had a very positive approach…" David: "I must say, in the past I haven't been all that fussy about physios, because I didn't feel that they were as hands-on as I would like, but he seemed to be – yeah, sort of easy to talk to and understood the problems quite well." Joyce: "I thought she was very good. I thought she was very professional. And, yes, I trusted her with what she was telling me to do."
Theme 2: Personal attitudes and expectations		
Nothing to lose	Patricia: "…it was only ever going to do nothing or improve things…So for me it was just like if you do this and it makes you feel better, that's awesome. If you do this and it doesn't make any difference, well, it's actually still making me get off my arse and do something. So that's good." Melissa: "I guess when I signed up, I didn't have an expectation. I thought, you know, anything's better than nothing." Bill: "I was fairly confident that would give at least some benefit."	Joyce: "I didn't expect great improvements. All I wanted was to either maintain it, get a little bit better, but not get worse." Susan: "I was hopeful but I wasn't unrealistic. So I didn't expect, I did not expect a miracle." Kathleen: "I think there's nothing negative that can happen, even if it didn't get better, it's not a negative thing because you tried."
Acceptance and realism	Christine: "There's always some pain to have a gain [laughs]. So sometimes doing the exercises, yeah, I would find that there'd be some sort of pain…Yes, it hurts, yes, it's uncomfortable but hopefully it will keep it moving and going and whatever." Judith: "And so my aim was to…get back to closer to 15 to 20,000 steps a day – which I achieved. So doing the exercises and strengthening the knee, I was able to	Patrick: "You ignore pain, that's the thing too. So it comes and goes. You treat it, you get an antiinflammatory, you get a massage, do what you can and just keep going." Lisa: "Oh look, I've had it for so long, I just, it's just part of life. It's a limiter, but just puts boundaries on things." Susan: "It's just there. It's just part of me now. So, I don't feel necessarily…oh, I guess, what I feel down about is when I'm with other people and they go for a

(Continued)

Table 3. (*Cont'd*)

Themes and subthemes	Responders	Nonresponders
	get back to all of that again…Realistically I knew it was not going to get me back to running." Lesley: "So the exercises were 1 aspect of it, but the other aspect of it is just understanding my limits, and maybe tapering my expectations a little bit as well."	big long walk, and I say, "I have to sit down, I'll wait for you here." So, that's pretty annoying, but other than that, I just have to accommodate it into my life."
Perceived good health	Lesley: "…so I would say my general health was very good. If I'm giving myself a rating out of 10, I would say my general health was a 8." Elizabeth: "I would say it's certainly above average…I have atrial fibrillation which I manage, and then I've got the osteoarthritis…other people will say we can't believe that you're 70 and your energy and activity and stuff like that."	Wendy: "My general health is fine…I mean, like my knee issues is something that's been there for a long, long time, so I don't think my general health affected that program at all; and I'm in good health, touch wood." Kathleen: "My general health is pretty good. I don't have any major issues." Patrick: "Again, I think my general health is pretty good. The only thing I've got is osteoarthritis in the knees, which I tend to ignore and work around."
Accepting responsibility	NA	Joyce: "I really just think it was – I really think it was because I wasn't doing what I needed to do. I wasn't – it was mainly in the walking and things like that, I just wasn't doing it." Catherine: "Virtually I guess what I'm saying, the problem was probably 100% my lack of 100% commitment rather than any fault of the study." Matilda: "So I think I get – when I say I didn't benefit from it, it's more that I probably didn't adhere to the exercise program well enough to benefit from it. But I think if I did, I probably would. Does that make sense?" Joseph: "I think for me it's more disappointment for not following it through like I should have followed it through I guess…At the end of the day when I did turn up it was really good."
Theme 3: Osteoarthritis beliefs		
Exercise is important	Judith: "…It's the answer. If you can't strengthen those muscles, you're not going to see any improvement. Without strengthening those muscles, you're just going to become a couch potato." Gregory: "… Increasing the muscle strength helps – I don't know what the term is but it helps support the knee, helps the function in the knee." Patricia: "But I'm pretty confident that it was, you know, and it was whatever the damage that I had could only be improved by doing strengthening exercises."	Catherine: "You just can't pop a pill for relief. You have to do other things, other logical things like the exercises and strengthen and what have you." Joseph: "I know exercise is correct. That's obviously just to strengthen what you have got there and it does work. As far as I don't know, massage or manipulation or TENS machines, braces and that I don't know if that makes any difference but I agree with, well, just exercise in general. I know that works." Matilda: "If you don't keep your legs stronger – mine aren't strong enough. I know that, and the more strength you lose in the muscles around – that support your knees, the more limited you become in what your capabilities are and what you can do, and so you lose some. Without the strength in your legs, you lose your life. You lose a desire to go and do things."
It's degenerative	James: "…the bones now have become weak at the very end of the leg bones, where they would normally be cushioned on the meniscus, so they've become soft. And, yeah, it's just they're more tender, that's my understanding of it." Lesley: "…there must be a link surely that things have to wear out. Just like your car wears out after a certain amount of kilometres." Gregory: "The right knee was – worn out, wearing away on the inner side of the joint just because of the structure of my legs so that's what caused it. My knees are plain old worn out."	Patrick: "And as I said, they're stuffed anyway. There's no cartilage in either knee so there's only so much you can do. So it's a maintenance – it's not an improvement program, it's a maintenance program." Catherine: "I just think osteoarthritis is just a part of life. It's incurable, if that's the word, and you've got to live with it and therefore manage it." Michelle: "…they've worn out. I did 50 years of hairdressing. I've done a lot of heavy work in my time like concreting and stuff like that. A lot of gym work which most probably wasn't real brilliant for them. So I reckon they're just worn out."
Body weight as contributor	NA	Joseph: "I think that was half my problems with my knee is because I'm overweight so that doesn't help in the first place." Cynthia: "I think if I lost 20 kilos, which should be my ultimate game, maybe 30, I suspect my knees would improve out of sight. And I mean, I lost 10

(Continued)

Table 3. (*Cont'd*)

Themes and subthemes	Responders	Nonresponders
		kilos and actually after losing 10, my knees did feel a bit better. So if I lost another 20, I probably think they'd be a lot better." Michelle: "And then I lost – I can't remember how much it was now, it was like a fair bit in 6 months – like 8 kilos or something – and I think nothing else would – well, apart from my physical fitness would have been pretty poor – so I think those 2 things. If I could have got more weight off more quickly, I reckon I might have seen more benefits, you know?"
Theme 4: Self-efficacy		
Exercise adherence	Bill: "I was very, very diligent with the exercise program certainly through the first 3 months. I didn't miss a single day. Did all the exercises as required and to the level required." Judith: "I felt, as I worked on the exercises – like religiously, doing them every second day." Valerie: "I must admit, towards the end, I did flag off a little bit. Mainly because my knee was feeling so good."	Michelle: "I followed exactly what I had to do. Yes, 100 percent." Catherine: "Probably because I didn't do the exercises. I did some exercises with trepidation for fear of causing my back pain...There was times that I just didn't do the exercises due to other factors, whatever, at the time." Joseph: "The first time I went back it was I pretty much did them all. And then the second time I went back I did, I don't know, three-quarters. And the third time I went back I did half and sort of dwindled away so by the fifth time I went back it was hardly anything."
Empowered for self-management	Elizabeth: "But we covered the book and it got me to a stage where I was comfortable doing my exercise and there was nothing of concern really..." James: "Well, now I understand it is very important, understanding that a little bit of pain is OK and how to deal and manage that pain and understand that some pain to do with any sort of physical activity is OK and I'm not doing any further damage..." Melissa: "It got me where I had a wide variety of different exercises that I could do and I felt supported and I knew what I needed to do, so I didn't really need more [physio consults]."	NA
Theme 5: Barriers		
Comorbid health conditions	NA	Beverley: "I thought it was doing me OK but then no, I just couldn't deal with it anymore...I ended up having to have injections in my hips afterwards. Because I do have bursitis in my hips, so it actually created more problems for me." Catherine: "I have had for some years, quite a long time ago diagnosed with depression and I am on medication for that that works but different issues in life that come up are still very hard to live with and I did have a very severe bout of depression, and that happens." Joyce: "Well I think it was the overriding factor – my general health with my hips and my back and that were the major contributor for me easing up and not doing as much as I should have been."
Stressors and life events	NA	Kathleen: "I persevered with it until a couple of months ago, because I had a lot of bad news in the family and things – stress just took over." Michelle: "Kept it up right up until most probably just before I left. And then I was packing. So it was more a time factor, and I was – the packing – I had to do most of it myself. And I was really struggling...So there really was no thought of, "Oh yes, I must do my exercises," because by the end of the day I could barely move. So yes, I dropped off a fair bit in that few months leading up to moving." Matilda: "Other things happened in my life that changed as well about the same time. That always complicates outcomes, and they were not something

(Continued)

Table 3. (*Cont'd*)

Themes and subthemes	Responders	Nonresponders
		that could be avoided. Just some of my physical activity ceased due to other issues, other people's injuries, actually. So there was a bit of a sudden change in lifestyle."
Theme 6: Outcomes		
Strength gains	Gregory: "It did in terms of resistance and reps and stuff, yes, it progressed a lot. I got a lot stronger, definitely, yeah." Judith: "I felt all the muscles leading into my knee really – so you know my quad, my hammy, I felt all of that starting to build strength, which was taking a bit of pressure off the poor old knee as well." Lesley: "You could just definitely feel the strengthening in the quads and things like that, that were actually taking the load off the knee a little bit."	Joseph: "I must admit at one stage I did feel better. When I was doing the exercises initially I did feel stronger and more flexible and that side of things." David: "Well, I do feel that it did – I got stronger in the legs, and that was a help." Michelle: "I felt so much stronger. I could barely walk, I'd use a walker – inside the house, just to go to the sink. And when I first went there I could barely walk and I was doing just a few hundred steps a day…And then I worked up to 6,000 and – I could walk to the sink without my walker. So I definitely got improvement as far as strength went."
Measurement limitations	NA	Michelle: "When I filled out that final part of that survey, I was just recovering from being really, really sick…And yes, but unfortunately that survey asked me how I felt in the last 2 weeks and I wrote on it that I had been very sick so I didn't think it was really fair to have to write how I felt in that 2 weeks." Catherine: "But if I'm having to mow a lawn, lift a lawnmower into a car, which I did and ended up with 2 bulging discs in my back, you're not doing the survey justice at those points in time." Matilda: "I don't think it helped with the pain, put it that way, because that's not – but that wasn't my issue anyway…And so, my knee issue was more learning to live with the dysfunction and how to avoid falling over, learning how to use my legs differently…So that's why, I suppose, the questions, when you do a survey, questions are always limited in terms of how you can answer them because of the way they're worded…they've been worded for a particular answer, and you can't always give that answer."

* NA = not applicable (as the subtheme did not arise within the subgroup); PEAK = Physiotherapy Exercise and Physical Activity for Knee Osteoarthritis; TENS = transcutaneous electrical nerve stimulation.

Theme 4 (self-efficacy). Both subgroups felt that exercise adherence was important in determining intervention outcome. Responders were quite adherent to the self-directed exercise program and physical activity plan, although for some, adherence declined once their knee pain had reduced. This converges with quantitative data (Table 2), which shows that all but 1 responder performed the recommended number of exercise sessions (or more) at 3 months, and that 7 (58%) reported reduced adherence at 9 months. Nonresponders described more variable patterns of exercise adherence (convergent with data in Table 2), often not doing as many exercise sessions as prescribed or only adhering for a short duration. Nonresponders were more likely to reduce or stop exercise in response to any aggravation in knee pain or when they encountered barriers that made adherence a challenge (e.g., life stressors). Only responders spoke of being empowered for self-management, believing that they could continue, or start up again, with their exercise and physical activity program in the future if they needed to, having developed knowledge and confidence to self-manage.

Theme 5 (barriers). Among nonresponders only, barriers to the exercise and physical activity intervention emerged. Many described negative impacts of comorbid health conditions (such as cardiovascular disease, cancers, balance issues, and fibromyalgia as well as acute illnesses such as the flu) on their capacity to undertake the strengthening exercises and/or physical activity goals. Exercise often aggravated existing comorbid musculoskeletal health problems, such as hip or back pain, causing participants to stop exercising. Another barrier was stressors and life events. Nonresponders described circumstances that caused significant emotional or mental stress and/or impacted on their ability to adhere to their exercise and activity program. These included caring responsibilities, significant health problems affecting loved ones, moving house, and the impacts of COVID-19.

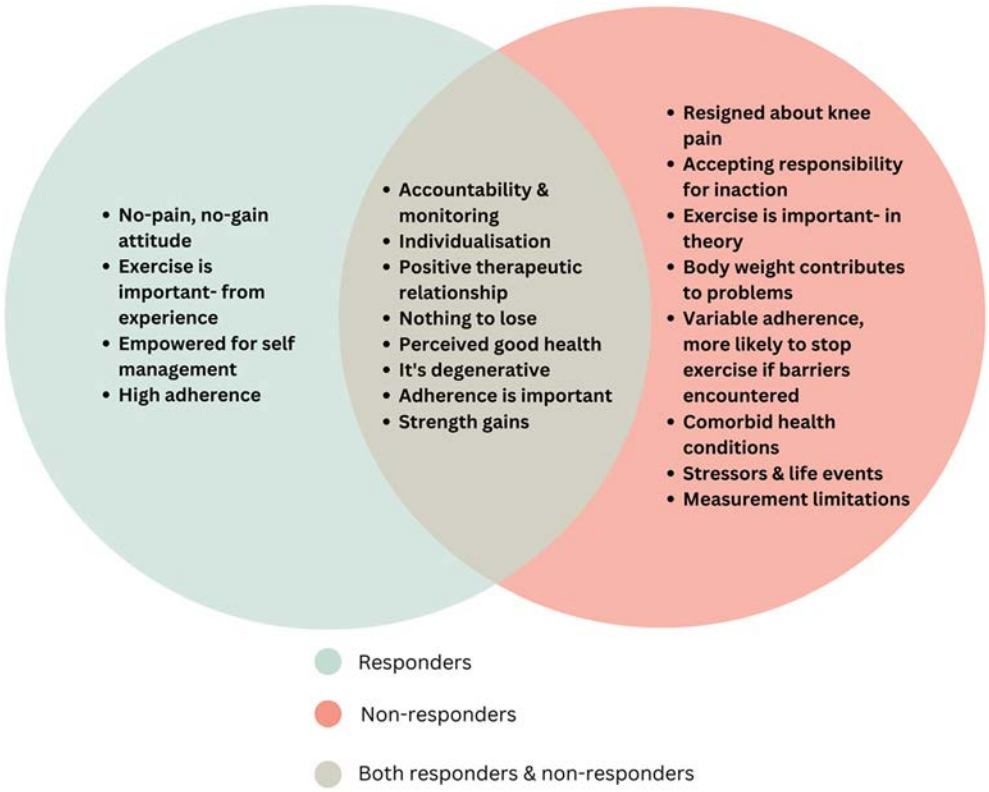

Figure 1. Similarities and differences across responders and nonresponders.

Theme 6 (outcomes). Participants discussed beneficial effects of exercise on parameters other than pain and function, with both responders and nonresponders describing strength gains. For nonresponders, improvements in knee strength were viewed as a beneficial intervention outcome. Nonresponders often felt that the methods used to measure intervention outcomes in the trial were inadequate. This subtheme of measurement limitations was unique to nonresponders. There was dissatisfaction with the time-limited recall period for surveys, with nonresponders often reporting that the recall period was not reflective of their outcomes at other time points and therefore may not have captured the intervention benefits they gained. Others felt that the measures of pain and physical function (primary outcomes of the trial upon which response was determined) did not capture their main knee OA-related problems, and that had the surveys asked about other issues (e.g., falls, weight loss) or more personally relevant activities (e.g., ability to ride a bike, drive long distances, ski, do martial arts), then beneficial intervention outcomes may have been observed.

DISCUSSION

We compared the perceptions of patients about why they did, or did not, respond to a physical therapist–supported exercise and physical activity program. We found similarities across both subgroups, including the intervention components that facilitated engagement, personal attitudes and expectations, beliefs about OA and the role of exercise, the importance of adherence, and perceived strength gains from exercise. There were also key differences. Responders felt empowered to self-manage, while nonresponders accepted responsibility for lack of improvement in pain and function, acknowledging that their adherence to the intervention was suboptimal. Nonresponders believed that their excess body weight contributed to their outcomes, encountered numerous barriers to exercise and physical activity, and felt that the measures for determining intervention outcomes in the trial had limitations.

Our results show that a complex and interrelated array of factors are perceived to contribute to responsiveness to an exercise and physical activity intervention among patients with knee OA. Our work builds on qualitative work in younger people with chronic whiplash, which also showed that therapeutic relationship, self-efficacy, acceptance, exercise experiences, and beliefs are important (31). We found that responders and nonresponders to exercise shared some similar perceptions and beliefs. Both subgroups found the intervention components engaging and highlighted strong therapeutic relationships with their physical therapist. This is important because prior research has shown that strong therapeutic alliance facilitates adherence to exercise in people with knee pain (22). A mixed-methods Cochrane Review (35) recommends that optimal delivery of exercise for OA should include information about the safety and value of

exercise, prescribe exercise tailored to an individual's abilities and needs, challenge inappropriate OA beliefs, and provide support. However, we found no evidence that nonresponders differed from responders in their perceptions about any of these factors, which suggests that the factors did not explain exercise nonresponse in our sample.

Nonresponders accepted responsibility for not achieving improvements in pain and function, acknowledging suboptimal adherence to exercise and physical activity goals. It is important to note that waning exercise adherence was also described by responders, consistent with research showing that trajectories of adherence typically decline over time (36). Although nonresponders recognized the importance of exercise adherence, adherence was variable, often because it aggravated knee pain or because other barriers (such as comorbidities, stressors, life events) were encountered. Nonresponders also believed that extra body weight contributed to their knee problems and lack of success with exercise. Our prior qualitative work in individuals with knee OA and comorbid obesity (37) found numerous physical (such as the complexity of performing weight-bearing exercise movements) and psychological (such as fear of pain) challenges to participating in strengthening exercises among this cohort. Our findings support secondary analyses of RCTs (24) in knee OA showing that presence of obesity and/or anxiety/depression predict poorer pain and function outcomes with physical therapist–led exercise interventions.

Our responders to exercise felt empowered to self-manage their OA because of knowledge they had gained about OA and exercise safety and the confidence they developed to independently continue or resume exercising when required. This was not identified in nonresponder data. Although few exercise RCTs in patients with OA have evaluated exercise effect moderators, there is some evidence that individuals with higher pain self-efficacy at baseline are more likely to report greater improvements in pain with physical therapist–supported exercise and online pain coping skills training (38). Emerging evidence also suggests that increases in self-efficacy may partially explain why pain and physical function improve with exercise (39). Research is warranted to identify effective strategies for improving self-efficacy to manage pain with knee OA, including engaging in exercise and physical activity. Nonresponders to exercise therapy may require additional strategies, such as educational approaches based on empowerment (40) or psychological approaches (41) for pain coping and stress management.

We observed a discrepancy, similar to that in individuals with chronic whiplash (31), between response to exercise and physical activity measured by our RCT outcome measures and how participants perceived that they responded. Although nonresponders did not improve with respect to knee pain or function, they felt they had responded in other beneficial ways to the exercise and the physical activity program. Some described weight loss, while others described improved ability to perform personally important tasks and activities, including driving, mowing the lawn, walking inside the home, and cycling. Along with responders, many described improved muscle strength, which was not measured in the RCT. Researchers should consider incorporating participant-specific outcomes instead of, or in addition to, generic or disease-specific measures when evaluating exercise response. Measurement tools, such as the patient-specific functional scale (42), that allow participants to nominate a specific activity for reassessment may improve detection of exercise response.

Our findings may guide researchers and clinicians to better identify patients at risk of nonresponse to exercise therapy and to consider strategies to optimize outcomes for this subgroup. Proactively screening for comorbidities, in particular musculoskeletal problems, and using this information to individualized exercise programs at the outset is warranted. Careful monitoring of the impact of exercise on knee pain and comorbidities may alert the clinician to problems as soon as they arise. Clinicians should be aware that aggravation in knee pain can deter some people from exercising and should teach patients how to positively deal with pain flares. Stressors and life events are unavoidable challenges encountered by many patients. Clinicians should be sensitive to the adverse impacts that these circumstances have on individuals and support the patient as much as possible to reengage with exercise as soon as is feasible. Recognizing the importance of accountability and monitoring, patients should be encouraged to monitor exercise adherence (e.g., using logbooks, diaries, wearables, mobile apps), and clinicians should regularly review adherence so that problems can be identified in a timely manner. Evidence-based strategies that may boost adherence include additional physical therapy consultations over a longer time frame (booster sessions) (43), use of messages (email, SMS) to remind and motivate people to exercise and support adherence (44), or digital exercise programming systems with or without remote clinician support (45,46).

Strengths of this study include its qualitative design, which provides richer information over the few existing quantitative studies that explore moderators of exercise response and have yielded little information to date (47). Our study was nested within an RCT, which provided unique opportunities for purposive sampling of responders and nonresponders and allowed triangulation with quantitative data. Our interview guide encouraged participants to explore a range of biopsychosocial factors that may have contributed to exercise (non)response. There are also limitations. Fourteen other participants were eligible and invited to participate but did not respond or declined. It is not clear if their data would have changed our findings. There are alternative approaches to determining responder status in OA RCTs (48), and our findings may not necessarily apply to alternative approaches. Our findings cannot be transferred to individuals of non–English-speaking backgrounds, as our RCT and this qualitative study limited inclusion to those who could speak English. Our study design does

not allow conclusions about causation of exercise non(response), thus findings are hypothesis generating and should be tested in RCTs by evaluating moderators (e.g., presence of comorbidities, overweight/obesity) of exercise effects.

In conclusion, responders and nonresponders shared some similar perceptions of a physical therapist–supported exercise and physical activity program. However, along with perceived limitations in trial outcome measurements, nonresponders also encountered challenges with excess weight, comorbidities, stressors, and life events that led to suboptimal exercise adherence and collectively were perceived to contribute to nonresponse.

ACKNOWLEDGMENTS

The authors acknowledge the contributions of the trial physical therapists who delivered care to the participants in this study, and the participants themselves for taking part in the interviews.

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. Hinman 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. Hinman, Campbell, Hall, Foster, Russell, Bennell.

Acquisition of data. Jones.

Analysis and interpretation of data. Hinman, Jones, Nelligan.

REFERENCES

1. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392:1789–858.
2. Paskins Z, Sanders T, Hassell AB. What influences patients with osteoarthritis to consult their GP about their symptoms? A narrative review. BMC Fam Pract 2013;14:195.
3. Teo PL, Bennell KL, Lawford B, et al. Patient experiences with physiotherapy for knee osteoarthritis in Australia: a qualitative study. BMJ Open 2021;11:e043689.
4. Kolasinski SL, Neogi T, Hochberg MC, et al. 2019 American College of Rheumatology/Arthritis Foundation guideline for the management of osteoarthritis of the hand, hip, and knee. Arthritis Care Res (Hoboken) 2020;72:149–62.
5. Bannuru RR, Osani MC, Vaysbrot EE, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthritis Cartilage 2019;27:1578–89.
6. The Royal Australian College of General Practitioners. Guideline for the management of knee and hip osteoarthritis. 2nd ed. East Melbourne, Vic: RACGP; 2018. URL: https://www.racgp.org.au/download/Documents/Guidelines/Musculoskeletal/guideline-for-the-management-of-knee-and-hip-oa-2nd-edition.pdf.
7. Rausch Osthoff AK, Niedermann K, Braun J, et al. 2018 EULAR recommendations for physical activity in people with inflammatory arthritis and osteoarthritis. Ann Rheum Dis 2018;77:1251–60.
8. Deasy M, Leahy E, Semciw AI. Hip strength deficits in people with symptomatic knee osteoarthritis: a systematic review with meta-analysis. J Orthop Sports Phys Ther 2016;46:629–39.
9. Bennell K, Hunt M, Wrigley T, et al. Role of muscle in the genesis and management of knee osteoarthritis. Rheum Dis Clin North Am 2008;34:731–54.
10. De Rooij M, van der Leeden M, Heymans MW, et al. Prognosis of pain and physical functioning in patients with knee osteoarthritis: a systematic review and meta-analysis. Arthritis Care Res (Hoboken) 2016;68:481–92.
11. Culvenor AG, Ruhdorfer A, Juhl C, et al. Knee extensor strength and risk of structural, symptomatic, and functional decline in knee osteoarthritis: a systematic review and meta-analysis. Arthritis Care Res (Hoboken) 2017;69:649–58.
12. Holden MA, Button K, Collins NJ, et al. Guidance for implementing best practice therapeutic exercise for people with knee and hip osteoarthritis: what does the current evidence base tell us? Arthritis Care Res (Hoboken) 2021;73:1746–53.
13. White DK, Tudor-Locke C, Zhang Y, et al. Daily walking and the risk of incident functional limitation in knee osteoarthritis: an observational study. Arthritis Care Res (Hoboken) 2014;66:1328–36.
14. Chang AH, Song J, Lee J, et al. Proportion and associated factors of meeting the 2018 physical activity guidelines for Americans in adults with or at risk for knee osteoarthritis. Osteoarthritis Cartilage 2020;28:774–81.
15. Fransen M, McConnell S, Harmer AR, et al. Exercise for osteoarthritis of the knee. Cochrane Database Syst Rev 2015;1:CD004376.
16. Uthman OA, van der Windt DA, Jordan JL, et al. Exercise for lower limb osteoarthritis: systematic review incorporating trial sequential analysis and network meta-analysis. BMJ 2013;347:f5555.
17. Moseng T, Dagfinrud H, van Bodegom-Vos L, et al. Low adherence to exercise may have influenced the proportion of OMERACT-OARSI responders in an integrated osteoarthritis care model: secondary analyses from a cluster-randomised stepped-wedge trial. BMC Musculoskelet Disord 2020;21:236.
18. Bennell KL, Nelligan RK, Kimp AJ, et al. What type of exercise is most effective for people with knee osteoarthritis and co-morbid obesity? The TARGET randomized controlled trial. Osteoarthritis Cartilage 2020;28:755–65.
19. Bennell KL, Campbell PK, Egerton T, et al. Telephone coaching to enhance a home-based physical activity program for knee osteoarthritis: a randomized clinical trial. Arthritis Care Res (Hoboken) 2017;69:84–94.
20. Hinman RS, Campbell PK, Lawford BJ, et al. Does telephone-delivered exercise advice and support by physiotherapists improve pain and/or function in people with knee osteoarthritis? Telecare randomised controlled trial. Br J Sports Med 2020;54:790–7.
21. Lawford BJ, Bennell KL, Campbell PK, et al. Association between therapeutic alliance and outcomes following telephone-delivered exercise by a physical therapist for people with knee osteoarthritis: secondary analyses from a randomized controlled trial. JMIR Rehabil Assist Technol 2021;8:e23386.
22. Moore AJ, Holden MA, Foster NE, et al. Therapeutic alliance facilitates adherence to physiotherapy-led exercise and physical activity for older adults with knee pain: a longitudinal qualitative study. J Physiother 2020;66:45–53.
23. Dobson F, Bennell KL, French SD, et al. Barriers and facilitators to exercise participation in people with hip and/or knee osteoarthritis: synthesis of the literature using behavior change theory. Am J Phys Med Rehabil 2016;95:372–89.
24. Legha A, Burke DL, Foster NE, et al. Do comorbidities predict pain and function in knee osteoarthritis following an exercise intervention, and do they moderate the effect of exercise? Analyses of data from

three randomized controlled trials. Musculoskeletal Care 2020;18:3–11.
25. Holden MA, Burke DL, Runhaar J, et al. Subgrouping and TargetEd Exercise pRogrammes for knee and hip OsteoArthritis (STEER OA): a systematic review update and individual participant data meta-analysis protocol. BMJ Open 2017;7:e018971.
26. Thanh NC, Thanh T. The interconnection between interpretivist paradigm and qualitative methods in education. Am J Ed Sci 2015;1:24–7.
27. Hinman RS, Kimp AJ, Campbell PK, et al. Technology versus tradition: a non-inferiority trial comparing video to face-to-face consultations with a physiotherapist for people with knee osteoarthritis. Protocol for the PEAK randomised controlled trial. BMC Musculoskelet Disord 2020;21:522.
28. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19:349–57.
29. National Clinical Guideline Centre (UK). Osteoarthritis: care and management in adults. NICE clinical guidelines. London: National Institute for Health and Care Excellence (UK); 2014.
30. Australian Government, Department of Health and Aged Care. Physical activity and exercise guidelines for all Australians. 2019. URL: https://www.health.gov.au/topics/physical-activity-and-exercise/physical-activity-and-exercise-guidelines-for-all-australians.
31. Griffin AR, Moloney N, Leaver A, et al. Experiences of responsiveness to exercise in people with chronic whiplash: a qualitative study. Musculoskelet Sci Pract 2021;54:102380.
32. Charmaz K. Grounded theory. Objectivist and constructivist methods. In: Denzin N, Lincoln Y, editors. Handbook of qualitative research. 2nd ed. Thousand Oaks (CA): Sage; 2000. p. 509–37.
33. Mays N, Pope C. Rigour and qualitative research. BMJ 1995;311:109–12.
34. Saunders B, Sim J, Kingstone T, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52:1893–907.
35. Hurley M, Dickson K, Hallett R, et al. Exercise interventions and patient beliefs for people with hip, knee or hip and knee osteoarthritis: a mixed methods review. Cochrane Database Syst Rev 2018;4:CD010842.
36. Nicolson PJ, Hinman RS, Kasza J, et al. Trajectories of adherence to home-based exercise programs among people with knee osteoarthritis. Osteoarthritis Cartilage 2018;26:513–21.
37. Lawford BJ, Bennell KL, Allison K, et al. Challenges with strengthening exercises for individuals with knee osteoarthritis and comorbid obesity: a qualitative study with patients and physical therapists. Arthritis Care Res (Hoboken) 2022;74:113–25.
38. Lawford BJ, Hinman RS, Kasza J, et al. Moderators of effects of internet-delivered exercise and pain coping skills training for people with knee osteoarthritis: exploratory analysis of the IMPACT randomized controlled trial. J Med Internet Res 2018;20:e10021.
39. Lima YL, Lee H, Klyne DM, et al. How do non-surgical interventions improve pain and physical function in people with osteoarthritis? A scoping review of mediation studies. Arthritis Care Res (Hoboken) 2023;75:467–81.
40. Egerton T, Bennell KL, McManus F, et al. Comparative effect of two educational videos on self-efficacy and kinesiophobia in people with knee osteoarthritis: an online randomised controlled trial. Osteoarthritis Cartilage 2022;30:1398–410.
41. Zhang L, Fu T, Zhang Q, et al. Effects of psychological interventions for patients with osteoarthritis: a systematic review and meta-analysis. Psychol Health Med 2018;23:1–17.
42. Stratford P, Gill C, Westaway M, et al. Assessing disability and change on individual patients: a report of a patient specific measure. Physiother Can 1995;47:258–63.
43. Nicolson PJ, Bennell KL, Dobson FL, et al. Interventions to increase adherence to therapeutic exercise in older adults with low back pain and/or hip/knee osteoarthritis: a systematic review and meta-analysis. Br J Sports Med 2017;51:791–9.
44. Bennell K, Nelligan RK, Schwartz S, et al. Behavior change text messages for home exercise adherence in knee osteoarthritis: randomized trial. J Med Internet Res 2020;22:e21749.
45. Bennell KL, Marshall CJ, Dobson F, et al. Does a web-based exercise programming system improve home exercise adherence for people with musculoskeletal conditions? A randomized controlled trial. Am J Phys Med Rehabil 2019;98:850–8.
46. Lambert TE, Harvey LA, Avdalis C, et al. An app with remote support achieves better adherence to home exercise programs than paper handouts in people with musculoskeletal conditions: a randomised trial. J Physiother 2017;63:161–7.
47. Quicke J, Runhaar J, van der Windt DA, et al. Moderators of the effects of therapeutic exercise for people with knee and hip osteoarthritis: a systematic review of sub-group analyses from randomised controlled trials. Osteoarthritis Cartil Open 2020;2:100113.
48. Pham T, van der Heijde D, Altman RD, et al. OMERACT-OARSI initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. Osteoarthritis Cartilage 2004;12:389–99.

Arthritis Care & Research
Vol. 75, No. 9, September 2023, pp 1939–1948
DOI 10.1002/acr.25096

American College of Rheumatology
Empowering Rheumatology Professionals

Reliability of Wearable Sensors for Assessing Gait and Chair Stand Function at Home in People With Knee Osteoarthritis

Michael J. Rose,[1] Tuhina Neogi,[1] Brian Friscia,[1] Kaveh A. Torabian,[1] Michael P. LaValley,[1] Mary Gheller,[1] Lukas Adamowicz,[2] Pirinka Georgiev,[2] Lars Viktrup,[3] Charmaine Demanuele,[2] Paul W. Wacnik,[2] and Deepak Kumar[1]

Objective. To assess the reliability of wearable sensors for at-home assessment of walking and chair stand activities in people with knee osteoarthritis (OA).

Methods. Baseline data from participants with knee OA (n = 20) enrolled in a clinical trial of an exercise intervention were used. Participants completed an in-person laboratory visit and a video conference–enabled at-home visit. In both visits, participants performed walking and chair stand tasks while fitted with 3 inertial sensors. During the at-home visit, participants self-donned the sensors and completed 2 sets of acquisitions separated by a 15-minute break, when they removed and redonned the sensors. Participants completed a survey on their experience with the at-home visit. During the laboratory visit, researchers placed the sensors on the participants. Spatiotemporal metrics of walking gait and chair stand duration were extracted from the sensor data. We used intraclass correlation coefficients (ICCs) and the Bland-Altman plot for statistical analyses.

Results. For test–retest reliability during the at-home visit, all ICCs were good to excellent (0.85–0.95). For agreement between at-home and laboratory visits, ICCs were moderate to good (0.59–0.87). Systematic differences were noted between at-home and laboratory data due to faster task speed during the laboratory visits. Participants reported a favorable experience during the at-home visit.

Conclusion. Our method of estimating spatiotemporal gait measures and chair stand duration function remotely was reliable, feasible, and acceptable in people with knee OA. Wearable sensors could be used to remotely assess walking and chair stand in participant's natural environments in future studies.

INTRODUCTION

Osteoarthritis (OA) is a leading cause of chronic pain and disability among adults (1). For people with knee OA, assessment of movement patterns during daily activities like walking and chair stand are considered clinically important functional outcomes (2). Alterations in movement patterns during these activities in people with knee OA are related to worse functional outcomes and disease progression (3–5). For example, individuals with knee OA walk with greater stride duration and lower cadence compared to controls (6), and while getting up from a chair, people with knee OA take longer compared to controls (7). Hence, standardized tests of gait and chair stand function are recommended as core outcomes for clinical trials of interventions for people with knee OA (8,9). However, assessments of gait and chair stand patterns are usually performed in tightly controlled laboratory environments that require expensive and time-consuming motion capture technologies. These visits not only can be burdensome for participants (e.g., due to travel to site) and researchers alike, but also may not yield data that reflect movement patterns used in a person's natural environments (10).

The COVID-19 pandemic has accelerated the adoption and implementation of telehealth (the practice of remote, virtual health care) (11) as well as the use of digital health technologies for remote assessment of participants in clinical trials (12). Wearable inertial sensors offer the possibility of remotely assessing gait

ClinicalTrials.gov identifier: NCT04243096.

Supported by Pfizer and Eli Lilly. Drs. Neogi and Kumar's work was supported by the NIH/National Institute of Arthritis and Musculoskeletal and Skin Diseases (grants K24-AR-070892 and K01-AR-069720, respectively).

[1]Michael J. Rose, BS, Tuhina Neogi, MD, PhD, Brian Friscia, MS, Kaveh A. Torabian, MS, Michael P. LaValley, PhD, Mary Gheller, MS, Deepak Kumar, PT, PhD: Boston University, Boston, Massachusetts; [2]Lukas Adamowicz, MS, Pirinka Georgiev, BS, Charmaine Demanuele, PhD, Paul W. Wacnik, PhD: Pfizer, Cambridge, Massachusetts; [3]Lars Viktrup, MD, PhD: Eli Lilly, Indianapolis, Indiana.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.25096&file=acr25096-sup-0001-Disclosureform.pdf.

Address correspondence via email to Deepak Kumar, PT, PhD, at kumard@bu.edu.

Submitted for publication July 6, 2022; accepted in revised form January 31, 2023.

SIGNIFICANCE & INNOVATIONS

- Given the importance of daily activities as outcomes for interventions, standardized assessment of walking and chair stand in a person's natural environments is important for knee osteoarthritis (OA) research and clinical practice.
- In this study, we observed good to excellent reliability in remotely assessing walking gait and chair stand activities using wearable inertial sensors at home in adults with knee OA; the agreement between at-home and in-person laboratory assessments showed a small bias, explained by participants walking faster in the laboratory environment. Participants were highly accepting of the at-home visit.
- Our approach could be used to monitor gait and chair stand activities reliably and remotely in individuals' natural environments at a lower cost, reduced participant and researcher burden, and greater ecological validity.

and chair stand movements for people with knee OA using standardized tests in a person's natural environment (13–18). However, determining whether such measures are reliable in individuals' home environments is important (i.e., whether they report consistent measurements from repeated tests) and how they agree with measures collected in well-controlled laboratory environments before large-scale use in clinical trials. While prior studies have reported excellent reliability in controlled laboratory environments (19), only a few studies have been conducted in individuals' homes and none in people with knee OA (20–22).

Hence, our objective was to examine the reliability of wearable sensor metrics of walking gait and chair stand using standardized tests in participants' homes and to examine agreement between these metrics collected in the laboratory and at-home. We hypothesized that wearable-sensor–derived walking gait and chair stand measures collected in a person's home environment would show good to excellent reliability in repeated measures, and that at-home data would agree with measures collected in a laboratory environment.

PATIENTS AND METHODS

Participants. We used data from a subset of participants (n = 20) enrolled in our single-arm clinical trial of an exercise intervention in people with knee OA. Participants were recruited from the community using print and online advertising and targeted social media strategies. Key inclusion criteria were age ≥ 50 years, body mass index of ≤40 kg/m^2, physician-diagnosed knee OA, score of ≥3 of 12 on weight-bearing questions (walking on flat surface, going up and down stairs, standing upright) from the Knee Injury and Osteoarthritis Outcome Score (KOOS) pain subscale in the index knee (23), and an ability to walk for 20 minutes without assistance. Key exclusion criteria included contraindications to exercise, other pain in lower back or legs that is greater than knee pain, any knee surgery in the previous 6 months, joint replacement in either hip or ankle, previous knee osteotomy, partial or total knee replacement in either knee, glucocorticoid or hyaluronic acid injections in either knee in the previous 3 months, other health conditions that may affect motor function, and receiving physical therapy for knee OA within the past 6 months. All study procedures were approved by a Boston University Institutional Review Board. Participants signed an informed consent prior to any study procedures.

For each participant, a "study leg" was identified as the leg with the diagnosis of knee OA provided by the physician, or the more painful leg in case both knees were diagnosed with OA (9). In cases where individuals had knee OA diagnosed in both knees and equal pain scores, the study leg was chosen at random.

Data collection. All assessments took place prior to initiation of the study intervention. All participants completed the KOOS questionnaire and provided information on their education, employment status, and family income. Participants also self-identified their sex assigned at birth and race from a fixed set of categories that included options to not provide this information. For race, participants could also select "unknown," or select multiple options. Participants completed 2 study visits, an in-person laboratory visit and a remote at-home visit. The order of these visits was randomized across participants, half completing the laboratory visit first and half completing the at-home visit first, with participants completing both visits between 1 and 20 days of each other. Participants were asked to wear their same daily walking shoes during both visits, and we used the same equipment and instructions for each task across both visits. During the visits, timestamps for the start and end of each trial of each task were recorded by a researcher using a custom system designed in REDCap electronic data capture tool, hosted at Boston University (24). Participants who were randomized to complete their at-home visit first were required to attend an in-person session to complete the informed consent procedures and then took home the equipment for the remote at-home visit. Participants who were randomized to complete the in-person laboratory visit first took home the equipment following the first in-person visit.

Regardless of the sequence, all participants received the same equipment for the at-home visit. Specifically, we provided participants with a wearable system consisting of 3 inertial sensors and docking station with charging cable, 2 cones connected by a 7-meter rope, and an armless chair. We used Opal inertial sensors (APDM). Each sensor contains an accelerometer (±16 g), gyroscope (±2,000 degrees/second), and magnetometer (±8 Gauss) and measures 43.7 × 39.7 × 13.7 mm (length × width × height) and weighs approximately 25 grams (Figure 1). Sensors were initialized to collect data in logging mode (data stored

Figure 1. Opal inertial sensor (left) and example of 3-sensor system as worn by participants. Two sensors were placed with straps around the shoes (center), and 1 sensor was placed on the lumbar and buckled around the waist (right).

onboard the sensor) at 128 Hz. The initialization process also synchronized the internal sensor clock with the computer clock in our laboratory. We placed the sensors in "standby" mode to not deplete the battery. Sensors were provided to the participants in a briefcase provided by the manufacturer with sufficient padding for protection. Participants were also provided verbal and written instructions on the use of sensors. Additionally, participants were provided with a tablet computer (Galaxy Tab S5e, Samsung). A video of walking and chair stand tasks along with instructions was saved on the tablet computer for participants to review (see Supplementary Video, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25096).

During the at-home study visit, researchers guided the participants via video conference. Participants were asked to place the sensors into the docking station that was connected to a power supply. This process activated the sensors so that they were no longer in standby mode and started logging data, time-synchronized to each other. Then we guided the participants to place 3 sensors on their body. One sensor was placed on the dorsum of each foot, and 1 was placed on the lower back as per manufacturer's guidelines (Figure 1). The sensors were secured to each area using straps attached to the devices. After participants donned the sensors, we guided the participants through 2 trials each of a standardized walking task and a chair stand task. For the standardized walking task, we asked the participants to walk at their self-selected, comfortable pace for 2 laps of a 7-meter path defined by the previously provided cones and rope for a total walking distance of 28 meters. We selected 7 meters for each direction because the manufacturer of the inertial sensors recommended a minimum of 7 meters for extraction of gait metrics. All but 2 participants had enough room for the 7-meter walking course; those 2 individuals walked as far as they could in a straight line before turning around. Before beginning the walking test, participants were instructed to stand still for 30 seconds for sensor initialization, as stated in the device manual provided by the manufacturer.

For the standardized chair stand task, we asked the participants to stand up from the provided chair 5 times as quickly as possible with arms crossed across the chest, similar to a traditional 5-timed sit-to-stand test (25). After the first set of tests, participants removed the sensors, waited 15-minutes, redonned the sensors, and performed 2 more trials of each task. At the end of the at-home visit, participants completed a REDCap survey evaluating the ease of using the devices, their comfort with performing the tests, and the likelihood they would participate in a similar visit in the future. Participants returned the equipment to us either at their first intervention visit or their in-person laboratory baseline data collection, depending on the randomization of their at-home visit. During the in-person laboratory visit, participants performed 2 trials of the same tasks as in the at-home visit, this time with a researcher placing the sensors on the participants.

Data processing. On return of the sensors, we downloaded the data from the devices. We extracted spatiotemporal walking gait metrics for both legs using the manufacturer provided

software (MoveoExplorer, version 1.0.0.201904110002) (26,27). The algorithm excluded turns using a validated method that uses the angular velocity signal around the vertical axis from the lower back sensor (28,29). For each gait trial, metrics were averaged across all strides. Step and stride duration were reported in seconds; stance duration, swing duration, double-support duration, and terminal double-support duration were reported in percent of gait cycle time; gait speed was in meters per second; cadence was in steps per minute; and stride length was in meters. For the chair stand test, the duration of chair stand (in seconds) was extracted from the sensors as an average of all chair stands detected. All data underwent visual inspection for errors or inconsistencies. A significant time shift was noticed in some participants' data (16 at-home and 9 laboratory recordings), due to a drift in the clocks of all sensors, which were time-synchronized with each other (see Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25096). The maximum drift was 24 seconds. To correct for this discrepancy, a researcher visually inspected the raw sensor data for a walking trial from each participant and manually recorded the time difference between when a participant began walking during the gait test and when we expected them to begin walking (30 seconds after the initial timestamp was recorded due to a required 30-second still period for sensor initialization). The time offset correction was then applied to all timestamps for all sensors of that participant's visit.

All at-home visits were conducted 1–9 days from the initialization and provision of sensors to participants, with no association observed between the amount of time between initialization and data collection with sensor clock drift. Three sets of walking gait and chair stand measures were generated, i.e., the mean of the first 2 trials from the at-home visit, the mean of the second 2 trials from the at-home visit, and the mean of the 2 trials from the laboratory visit. While gait data were extracted for both legs, since our analyses are within-person, we are only reporting gait data from the study leg as defined above. Gait data for the contralateral leg as well as data reported as right and left legs can be found in Supplementary Tables 1–3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25096.

Sample size estimation. The sample size was estimated using a prior study in healthy adults (n = 32, ages 65–85 years) (30). In the cited study, the test–retest reliability based on the correlation coefficient (ICC) for gait speed between 2 laboratory visits derived from an inertial sensor placed on the lower back was 0.85 (95% confidence interval 0.71–0.92) (31). We computed an ICC H_0 of 0.42 by generating a null distribution based on 10,000 permutations of the gait speed across participants between visit 1 and visit 2 and taking the 99th percentile of this distribution as ICC H_0. Using this ICC H_0 and a 1-tailed test with alpha = 0.05 and power = 0.8, we would need at least 11 participants to detect an ICC Of 0.85. We estimated that 12 participants would be needed, assuming 10% attrition. We overenrolled to ensure that we had sufficient power for all measures of interest and anticipating that not all participants would be able to complete all tasks. We used an R package, ICC.Sample.Size, to perform the sample size calculations (32).

Statistical analysis. We calculated test–retest reliability ICCs (95% confidence intervals) for each measure of function between repeated measures obtained in the at-home visit. We also computed agreement between the laboratory visit and the first set of measures from the at-home visit using ICCs. For all of these, we used ICC (2,1) based on absolute agreement in a 2-way random-effects model (33). We interpreted ICCs as poor (<0.5), moderate (0.5–0.74), good (0.75–0.90), and excellent (>0.90) (33). We also calculated Pearson's correlation between the first and second set of measures from the remote at-home visit. We report the standard error of measurement (SEM) and minimum detectable change (MDC) (equations 1 and 2) for assessment of walking gait and chair stand tasks at home as calculated similarly in previous studies (26,34–36). We also report SEM%, and MDC% (equations 3 and 4), calculated as shown below, where x is the mean for all observations (37).

$$SEM = SD \times \sqrt{1 - ICC} \quad (1)$$

$$MDC = SEM \times 1.96 \times \sqrt{2} \quad (2)$$

$$SEM\% = 100 \times (SEM/\bar{x}) \quad (3)$$

$$MDC\% = 100 \times (MDC/\bar{x}) \quad (4)$$

These values are independent of the units of measurement and allow comparison of inherent error between measures.

For agreement between the laboratory and at-home visits, we additionally performed Bland-Altman analyses to observe any possible biases and paired *t*-tests to determine whether differences were significant. ICCs and Bland-Altman plots were created using R statistical software (version 4.1.1) using the packages psych and BlandAltmanLeh, respectively, and Pearson's correlations were calculated in Matlab (MathWorks).

RESULTS

We enrolled 20 participants from the parent study for this substudy. Data from 2 of the 20 participants were unusable due to challenges with participants following our instructions on how to use the sensor, resulting in battery depletion and lack of data collection in some sensors. Participant characteristics are shown in Table 1.

Test–retest reliability during at-home visit. Figure 2 shows scatter plots for selected gait measures and the chair stand duration from the first and second set of trials in the at-home visit. During the at-home visit, the test–retest reliability of walking speed (ICC = 0.85) and chair stand duration (ICC = 0.89) were good, and reliabilities of all other gait measures were excellent (ICC >0.9) when reported for the study leg (Table 2). Similar reliability was noted for the contralateral leg, left leg, and right leg (see Supplementary Tables 1–3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25096). Pearson's correlation ranged from 0.81 to 0.97 for these tests (Table 2). SEM, SEM%, MDC, and MDC% for the wearable sensor–derived gait and chair stand metrics from the at-home visit are shown in Table 3. Participant feedback indicated that, in general, participants were highly accepting of the at-home visit (see Supplementary Table 4, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25096).

Agreement between at-home and laboratory visits. Figure 3 shows the Bland-Altman plots for selected gait measures and the chair stand duration from the first set of trials in the at-home visit and the trials from the laboratory visit. For the at-home versus laboratory visits, agreement was moderate (ICC >0.5 to <0.75) for gait speed, stride length, cadence, and chair stand duration, and good (ICC >0.75 to <0.9) for stride duration, step duration, stance, swing, double support, and terminal double support (Table 2). Mean differences between values derived in the laboratory versus at home are also shown in Table 2. Small systematic differences were noted in the Bland-Altman plots in all metrics, likely due to participants walking faster during the laboratory visit compared to the remote at-home visit (Figure 3).

Table 1. Participant characteristics*

Characteristic	Value (n = 20)
Age, years	70.5 ± 4.7
Body mass index, kg/m^2	30.6 ± 4.7
KOOS pain, 0–100	60.2 ± 10.6
KOOS ADL, 0–100	68.8 ± 14.2
KOOS weight-bearing pain	
More painful knee, 0–12	5.15 ± 1.35
Less painful knee, 0–12	3.15 ± 2.06
Study leg = left leg, no. (%)	9 (45)
Unilateral knee OA, no. (%)	10 (50)
Had previous knee injury, no. (%)	13 (65)
Sex assigned at birth, no. (%)	
Female	17 (85)
Male	3 (15)
White, no. (%)†	19 (95)
Education, no. (%)	
Without a college degree	2 (10)
Undergraduate	2 (10)
Graduate	13 (65)
Doctorate	3 (15)
Annual income, no. (%)	
< $50,000	4 (20)
$50,000–$150,000	6 (30)
> $150,000	4 (20)
Did not report	6 (30)
Currently employed, no. (%)	11 (55)

* Values are the mean ± SD unless indicated otherwise. ADL = activities of daily living; KOOS = Knee Injury and Osteoarthritis Outcome Score; OA = osteoarthritis.
† For privacy reasons, the race of 1 participant who did not self-identify as White is not reported.

DISCUSSION

We evaluated the reliability of wearable inertial sensor–derived walking gait and chair stand metrics collected remotely in a person's home environment and compared those measures to those obtained in a research laboratory environment. Our results show good to excellent test–retest reliability of these measures obtained at home. We also observed moderate to good agreement for the gait and chair stand measures across laboratory and home environments. Our findings suggest that, in future studies in people with knee OA, wearable sensors could be used for standardized assessment of gait and chair stand function in individual's homes.

To our knowledge, there is limited prior information on the reliability of gait or chair stand metrics from wearable sensors in a person's home environment. In individuals with multiple sclerosis, ICC values similar to ours were reported (0.91–0.95) for mean spatiotemporal gait metrics derived from inertial sensors in a smartphone collected during multiple at-home daily self-administered gait assessments (20). A meta-analysis of studies on reliability of gait metrics from wearable sensors in a laboratory environment in healthy adults reported good to excellent reliability for stride time, stride length, stance time, and swing time metrics (0.85–0.92) (19). In people with knee OA, a prior study that used treadmill walking to study the reliability of gait and sensor metrics found good to excellent reliability for step length, single-leg support time, and ground reaction force first and second peaks walking on an instrumented treadmill at 2 different speeds and inclinations (38). The reliability of raw acceleration waveforms from wearable sensors on shank and thigh (36) and the foot and lower back (35) during treadmill walking has also been reported to be acceptable (ICC >0.75) in people with knee OA. Thus, our approach for collecting these measures remotely can be used to reliably measure walking gait and chair stand movement patterns in people with knee OA in their home environments. Importantly, implementation of our approach in future studies should consider the use of similar rigorous methods, including written and video instructions for participants and a guided at-home virtual visit.

For chair stand, a study in healthy young adults reported good to excellent reliability for measuring acceleration during chair stand using a single inertial sensor embedded in smart glasses (21). Another study reported ICCs of 0.84 and 0.87 for chair stand duration measured using a single hip-worn inertial sensor during a

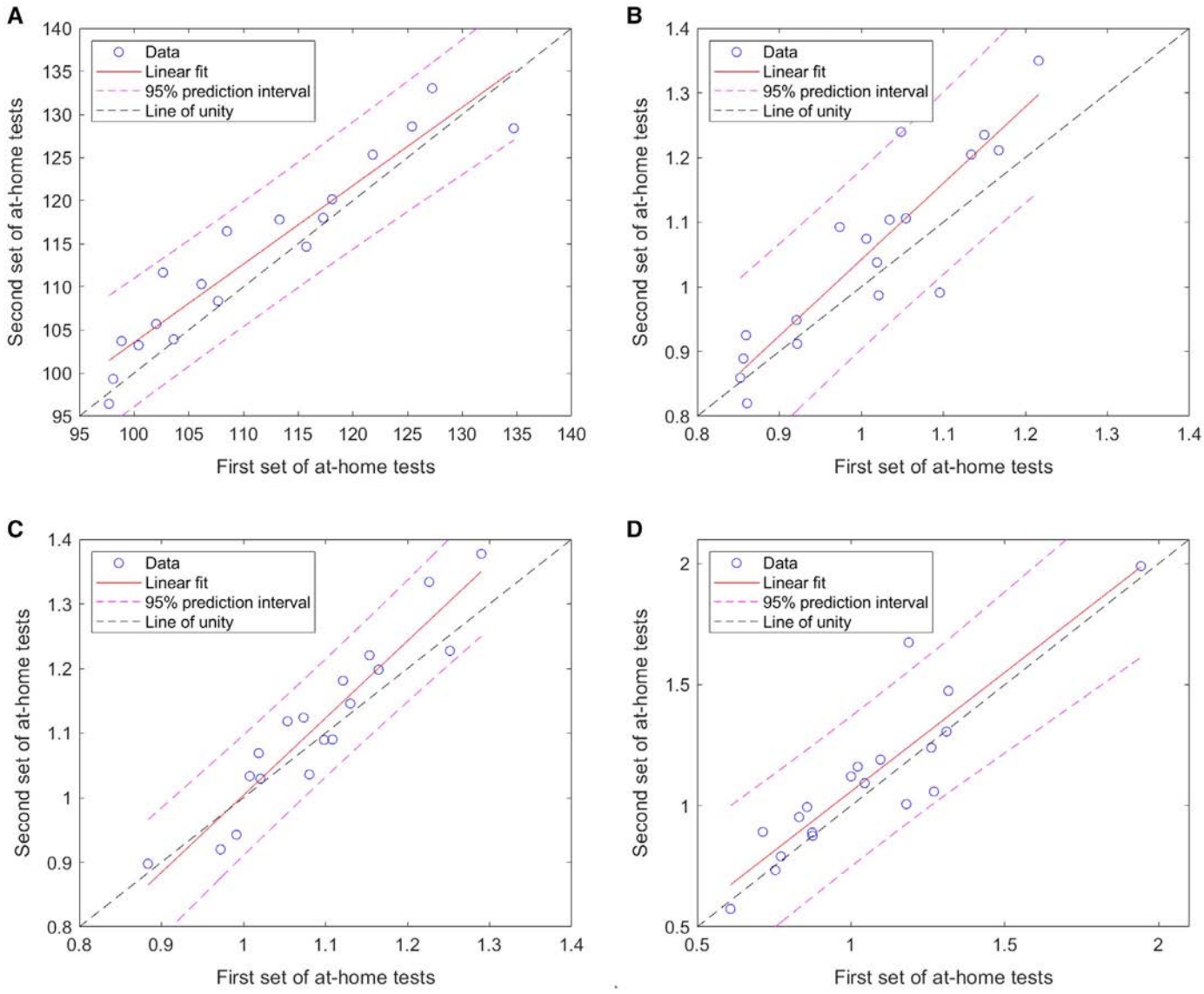

Figure 2. Scatter plots for selected gait metrics from the left leg, showing collected data, linear fit between the data, 95% prediction interval, and the line of unity between the 2 sets of at-home tests. **A**, Cadence, steps/minute (r = 0.95); **B**, Gait speed, meters/second (r = 0.90); and **C**, Stride length, meters (r = 0.94). **D**, Chair stand duration, seconds (r = 0.90), derived from the first and second set of trials during the remote visit.

5-timed sit-to-stand test performed as quickly as possible in young and old healthy adults (22). Our findings are similar to these prior studies. However, in the prior study using a hip-worn inertial sensor, the reliability for chair stand duration was worse when the task was performed at a self-selected pace (ICCs of 0.25 and 0.66 in young and old, respectively) (22). Hence, while our results suggest that wearable sensors may be used to remotely assess chair stand performance at home in future studies in people with knee OA, our findings may only be generalizable to the standardized sit-to-stand test performed as quickly as possible.

As expected with inertial sensors, temporal gait metrics outperformed spatial metrics, with stride length and gait speed (which is derived from stride length) having the lowest ICC values compared to the temporal metrics in both testing scenarios. Step and stride length measures are typically extracted from inertial sensors using single or double inverted pendulum biomechanical models that require signal integration and can accumulate errors (27,39). A meta-analysis of published studies supports this finding where temporal metrics (i.e., step time and stride time) were identified as having the strongest body of evidence for excellent validity and reliability (19). Importantly, our findings were consistent irrespective of knee pain severity in this population (see results for study leg and contralateral leg). We also report data for left and right leg, which can serve as a reference for future studies in healthy populations or where a distinction based on study and contralateral leg is not needed.

Table 2. Test–retest reliability data*

Sensor variables	Home 1†	Home 2†	Home 1 vs. Home 2 ICC (2,1)†	Home 1 vs. Home 2 r	Laboratory‡	Home 1 vs. laboratory ICC (2,1)‡	Home 1 vs. laboratory *P*
Gait speed, meters/second	1.01 ± 0.11	1.05 ± 0.15	0.85 (0.62, 0.93)	0.90	1.06 ± 0.15	0.63 (0.31, 0.82)	0.01§
Cadence, steps/minute	111 ± 11	114 ± 11	0.92 (0.77, 0.97)	0.97	111 ± 8	0.74 (0.51, 0.87)	0.61
Stride length, meters	1.09 ± 0.10	1.11 ± 0.13	0.91 (0.80, 0.96)	0.94	1.14 ± 0.12	0.59 (0.28, 0.79)	0.02§
Stride duration, seconds	1.09 ± 0.11	1.07 ± 0.10	0.92 (0.75, 0.97)	0.95	1.08 ± 0.08	0.75 (0.53, 0.88)	0.37
Step duration, seconds	0.55 ± 0.06	0.53 ± 0.06	0.92 (0.73, 0.97)	0.95	0.54 ± 0.04	0.78 (0.59, 0.89)	0.25
Stance, % GCT	61.57 ± 1.71	61.43 ± 1.97	0.93 (0.86, 0.97)	0.94	61.21 ± 2.03	0.81 (0.67, 0.91)	0.22
Swing, % GCT	38.43 ± 1.71	38.57 ± 1.97	0.96 (0.91, 0.98)	0.94	38.79 ± 2.03	0.81 (0.67, 0.91)	0.22
Double support, % GCT	23.5 ± 3.2	23.0 ± 3.6	0.95 (0.89, 0.98)	0.97	22.6 ± 4.0	0.87 (0.67, 0.94)	0.01§
Terminal double support, % GCT	11.66 ± 1.91	11.34 ± 2.14	0.94 (0.86, 0.97)	0.96	11.02 ± 2.08	0.84 (0.64, 0.92)	0.04§
Chair stand duration, seconds	1.05 ± 0.31	1.11 ± 0.34	0.89 (0.76, 0.95)	0.90	0.98 ± 0.21	0.66 (0.40, 0.83)	0.12

* Values are the mean ± SD unless indicated otherwise. Gait data are reported for study leg. ICC = intraclass correlation coefficient; GCT = gait cycle time.
† Gait data during remote home visit were available from 18 participants (1 participant did not wear the sensors correctly and sensors for 1 participant lost power). Chair stand data during home visit were available from 18 participants (2 participants were not able to perform the task).
‡ Chair stand data during laboratory visit were available from 16 participants (1 person could not perform the task, sensors did not detect any chair stands for 2 participants, and data for 1 participant were of poor quality).
§ Statistically significant difference between laboratory and home values ($P < 0.05$).

Data collected at home may have greater ecological validity than those collected in the laboratory. Bland-Altman plots of at-home minus laboratory gait data showed a small bias toward participants walking with a greater cadence, faster gait speed, and longer stride length in the laboratory compared to their home environments. Further, gait speed and stride length were significantly faster on average for laboratory values compared to at-home. These results are consistent with previous literature showing that both healthy individuals and those with Parkinson's disease tend to walk faster in a laboratory or hospital environment, respectively, than in free-living and home settings (10,30,40). Given the importance of gait speed for predictions of functional decline and other outcomes in people with knee OA, future studies may consider measuring gait speed in a person's natural environment (41,42). Overall, these results suggest that, while still comparable, caution should be used in instances where data are collected in different environments and aggregated.

In our study, feedback from the participants showed high acceptability of our approach. Previous work found assessing gait to be feasible using wearable sensors both in the clinic and in the homes of individuals with mild Alzheimer's disease (43). Participants reported the sensors as very easy to apply and completely comfortable to wear. The participants also noted that the level of commitment for the visit (45–60 minutes) was very manageable and that they would be very likely to participate in a similar visit again. Hence, our approach could be used to collect such data over repeated visits in clinical trials of interventions for people with knee OA.

Table 3. Standard error of measurement (SEM) and minimum detectable change (MDC) values for at-home data*

Sensor variables	MDC	SEM	MDC%	SEM%
Gait speed, meters/second	0.15	0.06	14.8	5.34
Cadence, steps/minute	8.51	3.07	7.58	2.73
Stride length, meters	0.10	0.04	8.84	3.18
Step duration, seconds	0.04	0.02	8.25	2.98
Stride duration, seconds	0.08	0.03	7.42	2.68
Stance, % GCT	1.34	0.48	2.17	0.78
Swing, % GCT	1.34	0.48	3.47	1.25
Double support, % GCT	2.08	0.75	8.93	3.22
Terminal double support, % GCT	1.36	0.49	11.83	4.27
Chair stand duration, seconds	0.25	0.09	23.15	8.35

* GCT = gait cycle time.

We have provided important information that could be used to design and implement future studies. Consistent inertial sensor placement on the body is known to be important (44,45). While half of our participants performed the at-home assessment after wearing the sensors on their own without any prior familiarity, all participants were provided instructions and were also guided by researchers in real-time via video. While ICC provides a relative measure of reliability, the SEM and MDC (Table 3) provide measures of absolute reliability. MDC represents the minimum amount of change that needs to take place to overcome error in the measurement (shown as SEM in Table 3). These values can be used in future studies where measured changes larger than the MDC can be considered a real change for a given participant (for example, a change >0.15 meters/second in gait speed). The SEM% and MDC% are similar to SEM and MDC but are independent of units of measurement. With SEM% being <10 for all measures and MDC% being <10 for most measures, these values are sensitive and could be used to detect small effects of interventions in future studies. Notably, MDC is not the same as minimally important change that is considered clinically meaningful (46).

There are some limitations to this study that should be considered. In our cohort, 95% of participants (19 of 20) self-identified their race as White, compared to the greater knee OA population. Individuals from minoritized races or ethnicities report greater challenges with technology in general (47) and specifically for health-related purposes (48,49), which may limit the current generalizability of our results to a more diverse population. We also had a larger proportion of women in our cohort than what is

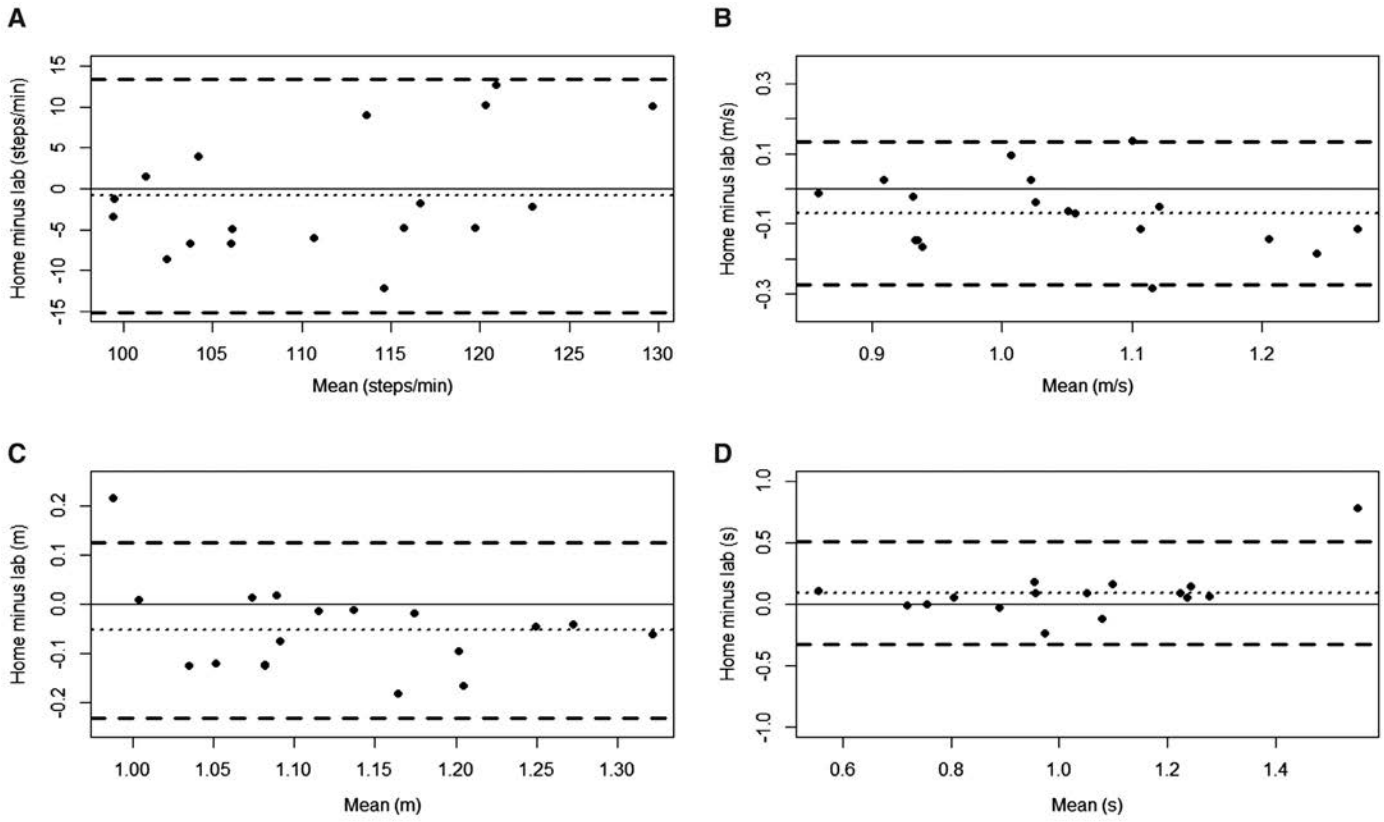

Figure 3. Bland-Altman plots from the remote and laboratory visits: **A**, Cadence (OA leg); **B**, Gait speed (OA leg); **C**, Stride length (OA leg); **D**, Chair stand duration. Black dots indicate data points. The dotted line shows mean difference, and the dashed line shows 95% limits of agreement. min = minute; m/s = meters/second; OA = osteoarthritis.

reflective of knee OA patients. Our sample size was small but justified a priori. Additionally, we scheduled only a 15-minute gap between the 2 at-home collections, as that was sufficient to demonstrate test–retest reliability because the participants removed and rewore the sensors. However, implementation of our approach in future studies will likely include larger gaps between visits (e.g., baseline and follow-up in a clinical trial), and having a longer delay between the collections may yield different results. The gap between the laboratory and home visits was variable, ranging from 1 to 20 days due to scheduling challenges across participants. However, given the chronic nature of OA pathology, this variability is unlikely to influence our results. Finally, we did not record the types of walking surfaces during the at-home visits, which could partially explain the differences in gait parameters between home versus laboratory visits.

In this cohort of people with knee OA who had moderate pain and disability, our method of estimating spatiotemporal gait measures and chair stand duration remotely was reliable, feasible, and participant accepted. Wearable sensors could be used to remotely monitor gait and chair stand function in participant's natural environments at a lower cost, reduced participant and researcher burden, and greater ecological validity, overcoming many limitations of laboratory visits. Hence, our approach could be used in future longitudinal studies or clinical trials of people with knee OA.

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. Kumar 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. Neogi, Demanuele, Wacnik, Kumar.
Acquisition of data. Rose, Friscia, Torabian, Gheller, Georgiev.
Analysis and interpretation of data. Rose, Torabian, LaValley, Adamowicz, Viktrup, Demanuele, Kumar.

ROLE OF THE STUDY SPONSOR

The study was conducted as a collaboration between Boston University, Pfizer, and Eli Lilly. Investigators from Pfizer and Eli Lilly were involved in the study design, data analysis and interpretation, and drafting of the manuscript. Publication of this article was not contingent upon approval by Pfizer and Eli Lilly. Boston University is the study sponsor.

REFERENCES

1. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. Lancet 2019;393: 1745–59.

2. Dobson F, Hinman RS, Roos EM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. Osteoarthritis Cartilage 2013; 21:1042–52.
3. Purser JL, Golightly YM, Feng Q, et al. Association of slower walking speed with incident knee osteoarthritis-related outcomes. Arthritis Care Res (Hoboken) 2012;64:1028–35.
4. Segal NA, Boyer ER, Wallace R, et al. Association between chair stand strategy and mobility limitations in older adults with symptomatic knee osteoarthritis. Arch Phys Med Rehabil 2013;94:375–83.
5. Hart HF, Gross KD, Crossley KM, et al. Step rate and worsening of patellofemoral and tibiofemoral joint osteoarthritis in women and men: the Multicenter Osteoarthritis Study. Arthritis Care Res (Hoboken) 2020;72:107–13.
6. Mills K, Hunt MA, Ferber R. Biomechanical deviations during level walking associated with knee osteoarthritis: a systematic review and meta-analysis. Arthritis Care Res (Hoboken) 2013;65:1643–65.
7. Sonoo M, Iijima H, Kanemura N. Altered sagittal plane kinematics and kinetics during sit-to-stand in individuals with knee osteoarthritis: a systematic review and meta-analysis. J Biomech 2019;96:109331.
8. Dobson F, Hinman RS, Roos EM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. Osteoarthritis Cartilage 2013; 21:1042–52.
9. McAlindon TE, Driban JB, Henrotin Y, et al. OARSI clinical trials recommendations: design, conduct, and reporting of clinical trials for knee osteoarthritis. Osteoarthritis Cartilage 2015;23:747–60.
10. Takayanagi N, Sudo M, Yamashiro Y, et al. Relationship between daily and in-laboratory gait speed among healthy community-dwelling older adults. Sci Rep 2019;9:3496.
11. Olayiwola JN, Magaña C, Harmon A, et al. Telehealth as a bright spot of the COVID-19 pandemic: recommendations from the Virtual Frontlines ("Frontweb"). JMIR Public Health Surveill 2020;6:e19045.
12. Gaba P, Bhatt DL. The COVID-19 pandemic: a catalyst to improve clinical trials. Nat Rev Cardiol 2020;17:673–5.
13. Tao W, Liu T, Zheng R, et al. Gait analysis using wearable sensors. Sensors (Basel) 2012;12:2255–83.
14. Mancini M, King L, Salarian A, et al. Mobility lab to assess balance and gait with synchronized body-worn sensors. J Bioeng Biomed Sci 2011;Suppl 1:007.
15. Kluge F, Gaßner H, Hannink J, et al. Towards mobile gait analysis: concurrent validity and test-retest reliability of an inertial measurement system for the assessment of spatio-temporal gait parameters. Sensors (Basel) 2017;17:1522.
16. Hellmers S, Fudickar S, Lau S, et al. Measurement of the chair rise performance of older people based on force plates and IMUs. Sensors (Basel) 2019;19:1370.
17. Millor N, Lecumberri P, Gómez M, et al. An evaluation of the 30-s chair stand test in older adults: frailty detection based on kinematic parameters from a single inertial unit. J Neuroeng Rehabil 2013;10:86.
18. Janssen WG, Bussmann JB, Horemans HL, et al. Validity of accelerometry in assessing the duration of the sit-to-stand movement. Med Biol Eng Comput 2008;46:879–87.
19. Kobsar D, Charlton JM, Tse CT, et al. Validity and reliability of wearable inertial sensors in healthy adult walking: a systematic review and meta-analysis. J Neuroeng Rehabil 2020;17:62.
20. Bourke AK, Scotland A, Lipsmeier F, et al. Gait characteristics harvested during a smartphone-based self-administered 2-minute walk test in people with multiple sclerosis: test-retest reliability and minimum detectable change. Sensors (Basel) 2020;20:5906.
21. Hellec J, Chorin F, Castagnetti A, et al. Sit-to-stand movement evaluated using an inertial measurement unit embedded in smart glasses: a validation study. Sensors (Basel) 2020;20:5019.
22. Regterschot GR, Zhang W, Baldus H, et al. Test–retest reliability of sensor-based sit-to-stand measures in young and older adults. Gait Posture 2014;40:220–4.
23. Roos EM, Roos HP, Lohmander LS, et al. Knee Injury and Osteoarthritis Outcome Score (KOOS): development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998;28:88–96.
24. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap): a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42:377–81.
25. Mong Y, Teo TW, Ng SS. 5-repetition sit-to-stand test in subjects with chronic stroke: reliability and validity. Arch Phys Med Rehabil 2010;91: 407–13.
26. Washabaugh EP, Kalyanaraman T, Adamczyk PG, et al. Validity and repeatability of inertial measurement units for measuring gait parameters. Gait Posture 2017;55:87–93.
27. Salarian A, Russmann H, Vingerhoets FJ, et al. Gait assessment in Parkinson's disease: toward an ambulatory system for long-term monitoring. IEEE Trans Biomed Eng 2004;51:1434–43.
28. Salarian A, Horak FB, Zampieri C, et al. iTUG, a sensitive and reliable measure of mobility. IEEE Trans Neural Syst Rehabil Eng 2010;18: 303–10.
29. Pearson S, Mancini M, El-Gohary M, et al. Turn detection and characterization with inertial sensors. Proceedings of the International Congress on Sports Science Research and Technology Support–icSPORTS. Vilamoura, Portugal; 2013. URL: https://www.scitepress.org/Link.aspx?doi=10.5220/0004647000190022.
30. Czech MD, Psaltos D, Zhang H, et al. Age and environment-related differences in gait in healthy adults using wearables. NPJ Digit Med 2020;3:127.
31. Czech MD, Patel S. GaitPy: an open-source Python package for gait analysis using an accelerometer on the lower back. J Open Source Software 2019;4:1778.
32. Zou GY. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. Stat Med 2012;31: 3972–81.
33. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15: 155–63.
34. Kesar TM, Binder-Macleod SA, Hicks GE, et al. Minimal detectable change for gait variables collected during treadmill walking in individuals post-stroke. Gait Posture 2011;33:314–7.
35. Kobsar D, Osis ST, Phinyomark A, et al. Reliability of gait analysis using wearable sensors in patients with knee osteoarthritis. J Biomech 2016;49:3977–82.
36. Turcot K, Aissaoui R, Boivin K, et al. Test-retest reliability and minimal clinical change determination for 3-dimensional tibial and femoral accelerations during treadmill walking in knee osteoarthritis patients. Arch Phys Med Rehabil 2008;89:732–7.
37. Flansbjer UB, Holmbäck AM, Downham D, et al. Reliability of gait performance tests in men and women with hemiparesis after stroke. J Rehabil Med 2005;37:75–82.
38. Pacifico D, Visscher R, List R, et al. Discriminant validity and reproducibility of spatiotemporal and kinetic parameters during treadmill walking in patients with knee osteoarthritis. Gait Posture 2020;80:77–9.
39. Zhao Q, Zhang B, Wang J, et al. Improved method of step length estimation based on inverted pendulum model. Int J Distrib Sens Netw 2017;13: 1550147717702914.
40. Gaßner H, Sanders P, Dietrich A, et al. Clinical relevance of standardized mobile gait tests. Reliability analysis between gait recordings at hospital and home in Parkinson's disease: a pilot study. J Parkinsons Dis 2020;10:1763–73.

41. Master H, Neogi T, Callahan LF, et al. The association between walking speed from short- and standard-distance tests with the risk of all-cause mortality among adults with radiographic knee osteoarthritis: data from three large United States cohort studies. Osteoarthritis Cartilage 2020;28:1551–8.

42. White DK, Neogi T, Zhang Y, et al. Association of slow gait speed with trajectories of worsening depressive symptoms in knee osteoarthritis: an observational study. Arthritis Care Res (Hoboken) 2017;69: 209–15.

43. McArdle R, Morris R, Hickey A, et al. Gait in mild Alzheimer's disease: feasibility of multi-center measurement in the clinic and home with body-worn sensors: a pilot study. J Alzheimers Dis 2018;63:331–41.

44. Vanegas M, Stirling L. Characterization of inertial measurement unit placement on the human body upon repeated donnings. IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN). June 9-12, 2015, Cambridge, Massachusetts.

45. Marques CJ, Bauer C, Grimaldo D, et al. Sensor positioning influences the accuracy of knee rom data of an e-rehabilitation system: a preliminary study with healthy subjects. Sensors (Basel) 2020;20; 2237.

46. De Vet HC, Terwee CB, Knol DL, et al. When to use agreement versus reliability measures. J Clin Epidemiol 2006;59:1033–9.

47. Hill JH, Burge S, Haring A, et al. Communication technology access, use, and preferences among primary care patients: from the Residency Research Network of Texas (RRNeT). J Am Board Family Med 2012;25:625–34.

48. Mitchell UA, Chebli PG, Ruggiero L, et al. The digital divide in health-related technology use: the significance of race/ethnicity. Gerontologist 2018;59:6–14.

49. Yoon H, Jang Y, Vaughan PW, et al. Older adults' internet use for health information: digital divide by race/ethnicity and socioeconomic status. J Appl Gerontol 2020;39:105–10.

Arthritis Care & Research
Vol. 75, No. 9, September 2023, pp 1949–1954
DOI 10.1002/acr.25084

AMERICAN COLLEGE of RHEUMATOLOGY
Empowering Rheumatology Professionals

BRIEF REPORT

Changes in Tophus Composition During Urate-Lowering Therapy: A Dual-Energy Computed Tomography Study

Leanne Chen,[1] Gregory D. Gamble,[1] Anne Horne,[1] Jill Drake,[2] Anthony J. Doyle,[3] Till Uhlig,[4] Lisa K. Stamp,[2] and Nicola Dalbeth[1]

Objective. The gouty tophus is an organized structure composed of monosodium urate (MSU) crystals and chronic inflammatory soft tissue. This dual-energy computed tomography (DECT) study aimed to determine whether the composition of the tophus changes during urate-lowering therapy.

Methods. Serial DECT scans from 32 people with gout were obtained over 2 years of allopurinol therapy, dose-escalated to serum urate of <0.36 mmoles/liter. Up to 5 index tophi were selected for each patient, with 103 separate tophi included in the analysis. Using manual outlining methods of conventional CT and DECT scans, the same index tophi were serially measured for total tophus volume and urate volume. For each tophus, the soft tissue volume was then calculated by subtracting the urate volume from the total tophus volume.

Results. The mean ± SD serum urate reduced from 0.43 ± 0.03 mmoles/liter at baseline to 0.31 ± 0.02 mmoles/liter at year 2. The mean ± SD total tophus volume reduced over the 2-year period from 5.17 ± 5.55 cm^3 to 2.61 ± 2.73 cm^3 ($P < 0.0001$). Greater reductions in tophus urate volumes than tophus soft tissue volumes were observed; the tophus urate volume decreased by 70.6%, and tophus soft tissue volume decreased by 37.8% ($P < 0.0001$). The mean tophus urate:soft tissue ratio reduced from 0.15 at baseline to 0.05 at year 2 ($P < 0.001$).

Conclusion. The composition of the tophus is dynamic and changes during urate-lowering therapy for gout management. The soft tissue component of the tophus is slower to respond and may persist without measurable MSU crystal deposition.

INTRODUCTION

In people with gout, tophi typically present as nodules within subcutaneous or musculoskeletal tissues and cause cosmetic concern, restricted joint movement, and joint damage (1,2). Microscopically, the tophus is an organized structure composed of monosodium urate (MSU) crystals and chronic inflammatory soft tissue (3). Developments in computed tomography technology have allowed analysis of the composition of the tophus, using integrated analyses of manual outlining methods for conventional computed tomography (CT) images and urate volume analysis for dual-energy CT (DECT) images (4). This analysis has shown that the MSU crystal component of the tophus is variable between tophi of similar physical sizes (4,5).

Long-term urate-lowering therapy dosed to achieve a serum urate below 0.36 mmoles/liter (6 mg/dl) leads to regression of tophi, which can be measured using physical methods such as Vernier calipers (6). However, we do not know whether the MSU crystal and soft tissue components of the tophus reduce in a similar manner during serum urate lowering to a target of <0.36 mmoles/liter, or whether the composition of the tophus changes during therapy. The aim of this study was to describe how the MSU crystal and soft tissue components of the tophus change over time during urate-lowering therapy.

ACTRN: 12611000845932.
Supported by the Health Research Council of New Zealand (grants11-203 and 15-576).
[1]Leanne Chen, BMedImag (Hons), Gregory D. Gamble, MSc, Anne Horne, MBChB, Nicola Dalbeth, MD: University of Auckland, Auckland, New Zealand; [2]Jill Drake, RN, Lisa K. Stamp, PhD: University of Otago, Christchurch, New Zealand; [3]Anthony J. Doyle, FRANZCR: Auckland District Health Board and University of Auckland, Auckland, New Zealand; [4]Till Uhlig, PhD: Center for Treatment of Rheumatic and Musculoskeletal Diseases, Diakonhjemmet Hospital, Oslo, Norway.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.25084&file=acr25084-sup-0001-Disclosureform.pdf.
Address correspondence via email to Nicola Dalbeth, MD, at n.dalbeth@auckland.ac.nz.
Submitted for publication June 30, 2022; accepted in revised form December 29, 2022.

SIGNIFICANCE & INNOVATIONS

- While tophus regression is recognized to occur during urate-lowering therapy, whether the composition of the tophus also changes is unknown.
- This study used integrated analyses of manual outlining methods with conventional computed tomography (CT) images and urate volume analysis with dual-energy CT to analyze changes in the composition in 103 individual tophi during allopurinol treatment.
- This analysis showed changes in the tophus composition during urate-lowering therapy, with a shift to a lower urate crystal component and higher soft tissue component.

PATIENTS AND METHODS

Study participants. Serial foot and ankle DECT scans from people with gout were obtained from an imaging substudy during allopurinol dosing (7). Participants in the study were recruited from primary and secondary health care settings and public advertising in Auckland, New Zealand. Ethical approval for the study was obtained from the New Zealand Ministry of Health Multi-Regional Ethics Committee (approval number MEC/11/06/060). Written informed consent was obtained from all participants, and the studies were undertaken in accordance with the declaration of Helsinki. All participants had gout according to the American College of Rheumatology 1977 classification of gout (8), were receiving at least creatinine clearance–based allopurinol dose for ≥1 month, and had a serum urate of >0.36 mmoles/liter at the time of screening. For the tophus composition analysis, participants were included if they received allopurinol dose escalation; achieved the target serum urate (<0.36 mmoles/liter) during the study, had completed DECT scans at all time points (baseline, year 1, and year 2), and had imaging evidence of tophus at baseline. The participant characteristics, including demographic information, in the tophus composition analysis (n = 32) were similar to those of the entire imaging cohort (n = 87), with the exception of fewer tophi at baseline and a slightly longer gout disease duration (7).

DECT scanning. DECT scans of both ankles and feet were obtained using a 128-detector row Siemens Somaton Definition Flash scanner (7). The patients were positioned feet-first supine with their feet plantarflexed, and the scans were obtained craniocaudally, with the scan range encompassing approximately 5 cm above both ankle joints and the distal ends of both big toes. The axial images were acquired at 128 × 0.6 mm with a pitch of 0.7 and a 30-cm field of view, and the radiograph tubes were operated at 80kV/260mA and 140kV/130mA. The acquired images were reconstructed using a bone algorithm, a 512-mm matrix and a slice thickness of 0.75 mm with 0.5-mm increments and were stored in a picture archiving and communication system as 0.75-mm and 3-mm sliced images.

Tophus analysis. The CT scans were analyzed using a Siemens MultiModality Workspace workstation with Syngo MMCP VE 36A 2009 software. We generated 3-dimensional DECT models of these scans using the gout setting within the DECT application and the 80kV and 140kV scans collected from each patient. Based on these models, up to 5 index tophi were selected from each baseline scan. A tophus was considered for analysis if it was "a well-defined opaque structure denser than adjacent soft tissue but less than surrounding bone" on conventional CT (2) with visible urate deposition on DECT. If >5 tophi were present, the tophi with the largest volume of urate deposition that could be demarcated were selected. The same index tophi were assessed at each timepoint: baseline, year 1, and year 2. Nail beds were excluded from tophus assessments due to potential artifacts.

The total tophus volume for each tophus was calculated on the 140kV scans, which were reconstructed using a tophus algorithm with window width = 600 and window level = 200 (9). Using the ROI tool in the Volume application, the reader drew freehand around each tophus (region of interest) in multiple 2-dimensional slices, ensuring that the most superior and inferior aspects of the tophus were included and omitting adjacent bone (see Supplementary Figure 1A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25084). A subsequent 3-dimensional model of the tophus was generated, with the upper and lower Hounsfield Unit (HU) evaluation limits set to 3,071 and –1,024, respectively, and the total tophus volume was automatically calculated (see Supplementary Figure 1B). During original validation work, this method of tophus volume measurement was highly reproducible, with both interobserver and intraobserver intraclass correlation coefficients >0.98 (9). This method of tophus measurement also correlated highly with physical measurement using Vernier calipers ($r = 0.91$, $P < 0.0001$) (9).

The tophus urate volume for each tophus was then calculated on DECT scans (4). For the 80kV images, fluid was set at 50 HU, the ratio for urate at 1.28, minimum HU 150 and smoothing range 5. For the 140kV images, fluid was set at 50 HU and maximum HU at 500. DECT urate volume within each tophus was measured using automated assessment software on the Volume application of the workstation, with upper evaluation limit –1 HU and lower evaluation limit –1,024 HU (see Supplementary Figures 1C and 1D, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25084).

Using the recorded measurements, the soft tissue volume of each tophus was calculated by subtracting the tophus urate volume from the total tophus volume (4). Additionally, the tophus urate:soft tissue ratio was calculated by dividing the tophus urate volume by the tophus soft tissue volume for each tophus.

All scans in known order were analyzed by a single reader (LC) who was blinded to serum urate results. Prior to the commencement of analysis, the reader underwent training and then calibration exercises with a rheumatologist with experience in CT and DECT scoring (ND), using a separate DECT data set. In calibration exercises in which readers were blinded to each other's scores, interreader intraclass correlation coefficients between the reader and the rheumatologist for total tophus volume was 0.90 (95% confidence interval [95% CI] 0.45–0.99) and for tophus urate volume was 0.99 (95% CI 0.98–1.00).

Statistical analysis. Data were analyzed using Prism (version 9, GraphPad) and SAS (version 9.4). A mixed-models approach to repeated measures was used to fit general linear models with random intercepts (GLIMMIX) for participant and tophus nested within participant to account for clustering (unstructured covariance) within individuals to describe the change from baseline (first study visit) of the urate, soft tissue, tophus, and ratio of the urate volume/soft tissue volumes with robust CI estimates. Statistical significance was confirmed by a P value less than 0.05. The dependent variable was the change from baseline, and the baseline values were used as covariates. Confirmatory analyses were performed on the absolute values over time.

RESULTS

A total of 32 participants with a total of 103 tophi were included in the analysis. The mean ± SD number analyzed per participant was 3.2 ± 1.1. Clinical features of the 32 participants are shown in Table 1. All were male, with a mean disease duration of >20 years. The mean ± SD serum urate at baseline was 0.43 ± 0.03 mmoles/liter, and this level reduced to 0.31 ± 0.02 mmoles/liter at year 2.

Total tophus volumes, tophus urate volumes, and tophus soft tissue volumes at each time point are shown in Table 1 and in Supplementary Figure 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25084. Within the analyzed tophi, the total tophus volume reduced over the 2-year period from mean ± SD 5.17 ± 5.55 cm^3 to 2.61 ± 2.73 cm^3 ($P < 0.0001$).

Both the tophus urate volumes and tophus soft tissue volumes reduced over the 2-year study period ($P < 0.001$). However, there was a change in the composition of the tophus, with greater reductions in tophus urate volume than soft tissue volume (Figures 1 and 2). Overall, the tophus urate volume decreased by 70.6% (95% CI 62.8–78.4), and tophus soft tissue volume decreased by 37.8% (95% CI 31.4–44.1) ($P < 0.0001$ for comparison between tophus urate volume and tophus soft tissue volume). The tophus urate:soft tissue ratio reduced from 0.15 (95% CI 0.12–0.19) at baseline to 0.05 (95% CI 0.02–0.09) at year 2 ($P < 0.001$). Analysis of the tophus urate:soft tissue ratio in individual tophi showed an increased ratio in 14 (13.6%), no change in 3 (2.9%), and reduction in 86 (83.4%). Analysis of tophi according to the median total tophus volume at baseline showed similar reductions in tophus urate:soft tissue ratios for small and large tophi at baseline, with no interaction between baseline total tophus volume and change in ratio ($P_{time} < 0.0001$, $P_{baseline\ volume} = 0.90$, $P_{time\times\ baseline\ volume\ interaction} = 0.39$).

At year 2, 3 of 103 tophi had disappeared with no urate or soft tissue component, and a further 25 of 103 had no measurable urate component but had a persistent soft tissue component (with mean ± SD soft tissue volume of 1.9 ± 2.2 cm^3).

DISCUSSION

This study, using integrated analysis of conventional and DECT tophus images, has provided new insights into the process of tophus regression in the setting of urate-lowering therapy.

Table 1. Clinical features of the 32 participants at baseline, and changes in allopurinol dose, serum urate, and DECT volumes over the 2-year period*

Variable	Baseline	Year 1	Year 2
Age, years	59 ± 13	–	–
Male, no. (%)	32 (100)	–	–
Ethnicity, no. (%)			
Asian	1 (3)	–	–
Māori	2 (6)	–	–
New Zealand European	13 (41)	–	–
Pacific Peoples	16 (46)	–	–
Disease duration, years	22 ± 10	–	–
Allopurinol dose, mg/day	298 ± 49	377 ± 60	425 ± 62
Serum urate, mmoles/liter	0.43 ± 0.03	0.35 ± 0.03	0.31 ± 0.02
Total tophus volume, cm^3 (n = 103)	5.17 ± 5.55	3.96 ± 4.17†	2.61 ± 2.73‡
Tophus urate volume, cm^3 (n = 103)	0.68 ± 1.29	0.31 ± 0.67‡	0.11 ± 0.29‡
Tophus soft tissue volume, cm^3 (n = 103)	4.50 ± 4.66	3.66 ± 3.78§	2.50 ± 2.63‡

* Values are the mean ± SD unless stated otherwise. DECT = dual-energy computed tomography.
† $P < 0.001$, compared with baseline values.
‡ $P < 0.0001$, compared with baseline values (post hoc Tukey honestly significant difference adjusted P values).
§ $P < 0.01$, compared with baseline values.

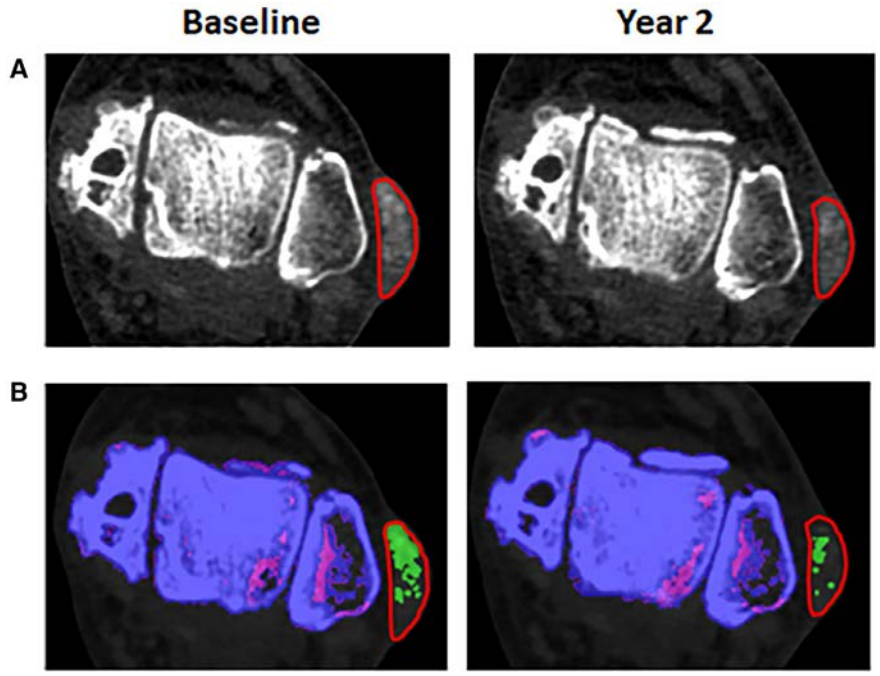

Figure 1. Example of change in tophus composition at baseline and year 2. A tophus near the left lateral malleolus is shown. **A**, Tophus outline on conventional computed tomography (CT). **B**, Tophus outline on dual-energy CT images. Red line shows total tophus outline.

We demonstrated that within the tophus, both urate and soft tissue volumes decrease during serum urate lowering to the treatment target of <0.36 mmoles/liter. However, tophus urate volumes reduce more rapidly than tophus soft tissue volumes, leading to altered composition of the tophus over time.

The results are consistent with previous studies demonstrating that the composition of the tophus is not uniform (3,10), and that for tophi of similar physical size, the urate volume within the tophus can vary, even within the same person (5). In addition, this study shows that the composition of the tophus is dynamic over time, and that resolution of the chronic inflammatory tissue takes longer than MSU crystals. The presence of fibrovascular tissue in the tophus may contribute to this slow resolution (3). MSU crystals below the level of detection by DECT may also lead to the persistence of chronic inflammatory soft tissue.

The study methodology, integrating data from conventional and DECT scans, allowed us to analyze the composition of individual index tophi over time. Our findings build on prior studies

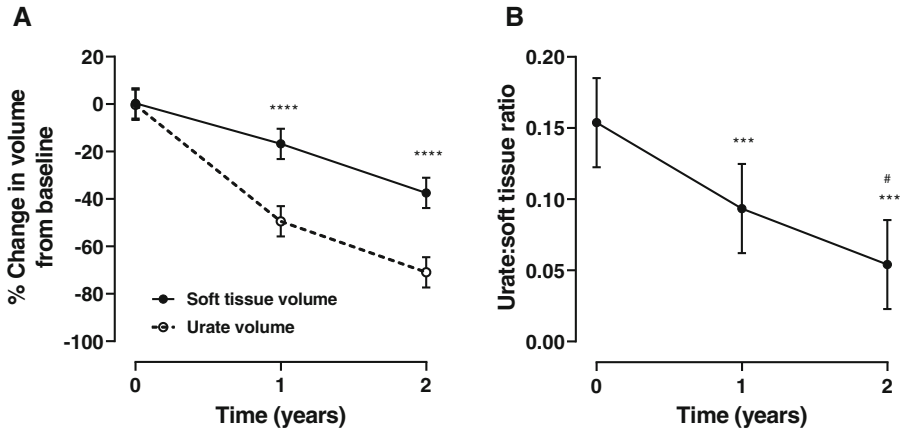

Figure 2. Comparison of change in urate and soft tissue volumes within the analyzed tophi. **A**, Percentage change from baseline in urate volume and soft tissue volume. The percentage change in soft-tissue volume $P_{time} < 0.0001$, and the percentage change in urate volume $P_{time} < 0.0001$. **B**, Urate volume:soft tissue volume ratio over time; $P_{time} < 0.0001$. Data are presented as the mean (95% CI). Symbols are for post hoc Tukey honestly significant difference adjusted *P* values. *** = $P < 0.001$ compared to baseline; **** = $P < 0.0001$ for comparison; # = $P < 0.05$ compared to year 1.

reporting changes in tophi during urate-lowering therapy. Using ultrasound, Hammer et al reported reduced length and width, but not depth of measured tophi over 1 year of urate-lowering therapy (11). These investigators postulated that the morphology of the tophi may change with treatment, making defining the outer margins of the tophus by ultrasound increasingly difficult. In a case report of a patient with tophaceous gout treated with pegloticase, DECT urate volume in 3 index tophi was undetectable, with persistent tophi measured by Vernier calipers and ultrasound (12). These authors suggested that the persistence of tophi by physical measurement and ultrasound was due to the persistence of the non-MSU crystal component of the tophus.

Our findings have some implications for clinical practice. EULAR recommends that patients with tophi should have a lower serum urate target, to allow more rapid dissolution of crystals (13). Some patients may have persistent tophi despite intensive serum urate lowering, with subcutaneous nodules slow to resolve (14). Our findings suggest that at least some of these lesions may not have active MSU crystal deposition, only representing residual soft tissue. Whether such lesions benefit from ongoing intensive urate-lowering therapy is unclear, or whether a less intensive serum urate target is appropriate. DECT may play a role in clinical practice to assess the MSU crystal burden when considering serum urate targets in patients with tophaceous gout who have received intensive serum urate lowering long-term. We have previously reported, using mediation analysis, that both the soft tissue component and the MSU crystal component of the tophus contribute to bone erosion scores (4), and whether persistence of the soft tissue component contributes to ongoing bone erosion, gout flares, and functional impact is unknown; this question will be the focus of future research.

We acknowledge the study limitations. DECT may not detect crystals that are suspended in liquid or that have low density (15). This possibility is unlikely to be a major limitation, as most tophi have tightly packed sheets of MSU crystals (3), but changes in the density of MSU crystal deposits during urate-lowering therapy may underestimate the total urate volume when assessed by DECT. Additionally, very small urate deposits (<0.01 cm^3) cannot be measured by DECT, and we cannot be certain that the tophi without visible urate deposits on DECT had complete clearance of MSU crystals. Finally, DECT does not allow analysis of different components of the tophus soft tissue, so how changes in the corona zone and fibrovascular zone of the tophus change with urate-lowering therapy is unknown (3).

In summary, this DECT study demonstrates that the composition of the tophus changes during urate-lowering therapy for gout management. While both urate and soft tissue volumes reduce over time, the chronic inflammatory soft tissue within the tophus is slower to respond than the MSU crystals. Following long-term urate-lowering therapy, tophus lesions may consist of soft tissue only, without measurable MSU crystal deposition.

ACKNOWLEDGMENT

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

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. Dalbeth 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. Gamble, Doyle, Uhlig, Stamp, Dalbeth.
Acquisition of data. Chen, Horne, Drake, Dalbeth.
Analysis and interpretation of data. Chen, Gamble, Horne, Doyle, Uhlig, Stamp, Dalbeth.

REFERENCES

1. Aati O, Taylor WJ, Horne A, et al. Toward development of a Tophus Impact Questionnaire: a qualitative study exploring the experience of people with tophaceous gout. J Clin Rheumatol 2014;20:251–5.
2. Dalbeth N, Clark B, Gregory K, et al. Mechanisms of bone erosion in gout: a quantitative analysis using plain radiography and computed tomography. Ann Rheum Dis 2009;68:1290–5.
3. Dalbeth N, Pool B, Gamble GD, et al. Cellular characterization of the gouty tophus: a quantitative analysis. Arthritis Rheum 2010;62:1549–56.
4. Sapsford M, Gamble GD, Aati O, et al. Relationship of bone erosion with the urate and soft tissue components of the tophus in gout: a dual energy computed tomography study. Rheumatology (Oxford) 2017;56:129–33.
5. Dalbeth N, Aati O, Gao A, et al. Assessment of tophus size: a comparison between physical measurement methods and dual-energy computed tomography scanning. J Clin Rheumatol 2012;18:23–7.
6. Perez-Ruiz F, Calabozo M, Pijoan JI, et al. Effect of urate-lowering therapy on the velocity of size reduction of tophi in chronic gout. Arthritis Rheum 2002;47:356–60.
7. Dalbeth N, Billington K, Doyle A, et al. Effects of allopurinol dose escalation on bone erosion and urate volume in gout: a dual-energy computed tomography imaging study within a randomized, controlled trial. Arthritis Rheumatol 2019;71:1739–46.
8. Wallace SL, Robinson H, Masi AT, et al. Preliminary criteria for the classification of the acute arthritis of primary gout. Arthritis Rheum 1977;20:895–900.
9. Dalbeth N, Clark B, Gregory K, et al. Computed tomography measurement of tophus volume: comparison with physical measurement. Arthritis Rheum 2007;57:461–5.
10. Larmon WA, Kurtz JF. The surgical management of chronic tophaceous gout. J Bone Joint Surg Am 1958;40-A:743–72.
11. Hammer HB, Karoliussen L, Terslev L, et al. Ultrasound shows rapid reduction of crystal depositions during a treat-to-target approach in gout patients: 12-month results from the NOR-Gout study. Ann Rheum Dis 2020;79:1500–5.
12. Modjinou DV, Krasnokutsky S, Gyftopoulos S, et al. Comparison of dual-energy CT, ultrasound and surface measurement for assessing tophus dissolution during rapid urate debulking. Clin Rheumatol 2017;36:2101–7.
13. Richette P, Doherty M, Pascual E, et al. 2016 updated EULAR evidence-based recommendations for the management of gout. Ann Rheum Dis 2017;76:29–42.

14. Mandell BF, Yeo AE, Lipsky PE. Tophus resolution in patients with chronic refractory gout who have persistent urate-lowering responses to pegloticase. Arthritis Res Ther 2018; 20:286.

15. Ahn SJ, Zhang D, Levine BD, et al. Limitations of dual-energy CT in the detection of monosodium urate deposition in dense liquid tophi and calcified tophi. Skeletal Radiol 2021;50: 1667–75.

Arthritis Care & Research
Vol. 75, No. 9, September 2023, pp 1955–1966
DOI 10.1002/acr.25095

American College of Rheumatology
Empowering Rheumatology Professionals

Familial Risk of Gout and Interaction With Obesity and Alcohol Consumption: A Population-Based Cohort Study in Korea

Kyoung-Hoon Kim,[1] In Ah Choi,[2] Hyun Jung Kim,[3] Heather Swan,[3] Sayada Zartasha Kazmi,[3] Gahwi Hong,[3] Young Shin Kim,[3] Seeun Choi,[3] Taeuk Kang,[4] Jaewoo Cha,[3] Jungmin Eom,[3] Kyeong Uoon Kim,[5] Hoo Jae Hann,[6] and Hyeong Sik Ahn[3]

Objective. Population-based studies of the familial aggregation of gout are scarce, and gene/environment interactions are not well studied. This study was undertaken to evaluate the familial aggregation of gout as well as assess interactions between family history and obesity or alcohol consumption on the development of gout.

Methods. Using the Korean National Health Insurance database, which includes information regarding familial relationships and risk factor data, we identified 5,524,403 individuals from 2002 to 2018. Familial risk was calculated using hazard ratios (HRs) with 95% confidence intervals (95% CIs) to compare the risk in individuals with and those without affected first-degree relatives. Interactions between family history and obesity/alcohol consumption were assessed on an additive scale using the relative excess risk due to interaction (RERI).

Results. Individuals with a gout-affected first-degree relative had a 2.42-fold (95% CI 2.39, 2.46) increased risk of disease compared to those with unaffected first-degree relatives. Having both a family history of gout and being either overweight or having moderate alcohol consumption was associated with a markedly increased risk of disease, with HRs of 4.39 (95% CI 4.29, 4.49) and 2.28 (95% CI 2.22, 2.35), respectively, which exceeded the sum of their individual risks but was only statistically significant in overweight individuals (RERI 0.96 [95% CI 0.85, 1.06]). Obese individuals (RERI 1.88 [95% CI 1.61, 2.16]) and heavy drinkers (RERI 0.36 [95% CI 0.20, 0.52]) had a more prominent interaction compared to overweight individuals and moderate drinkers, suggesting a dose-response interaction pattern.

Conclusion. Our findings indicate the possibility of an interaction between gout-associated genetic factors and obesity/alcohol consumption.

INTRODUCTION

Gout is a common form of inflammatory arthritis that is caused by chronic deposition of monosodium urate (MSU) crystals, which form in the presence of increased urate concentrations and leads to substantial morbidity associated with excruciating pain. Genetic predisposition is known to play a role in the pathogenesis of gout, and the incidence of gout varies according to ethnicity, with higher estimates reported in North American and European populations compared to Asians (1). Genetic studies have identified hundreds of genetic variants that affect serum urate concentrations, with the 3 most prominent loci encoding urate transporters SLC2A9, SLC22A12, and ABCG2 (2,3). Findings from twin studies show that the heritability of serum urate is 45–73%, and the heritability of gout is estimated to be 0–43% (2,4).

Case–control and cross-sectional studies conducted since the 1980s have shown that ~10% of gout patients have a family history of disease and that the first-degree relatives of patients are at a 2-fold increased risk of developing the disease themselves. However, current evidence for the familial aggregation of gout is limited, due to the small number of studies and their limited scope. Most existing studies included a few hundred participants and were unable to calculate incidence and familial risk due to

[1]Kyoung-Hoon Kim, PhD: Health Insurance Review and Assessment Service, Wonju-si, Republic of Korea; [2]In Ah Choi, MD, PhD: Chungbuk National University, Chungcheongbuk-do, Republic of Korea; [3]Hyun Jung Kim, MPH, PhD, Heather Swan, BA, Sayada Zartasha Kazmi, MD, Gahwi Hong, BS, Young Shin Kim, MD, Seeun Choi, BSc, Jaewoo Cha, MPH, Jungmin Eom, BScN, Hyeong Sik Ahn, MD, PhD: Korea University, Seoul, Republic of Korea; [4]Taeuk Kang, MPH, PhD: Sungshin Women's University Woonjung Green Campus, Seoul, Republic of Korea; [5]Kyeong Uoon Kim, PhD, RN: Seojeong University, Yangju-si, Republic of Korea; [6]Hoo Jae Hann, MD, PhD: Ewha Womans University, Seoul, Republic of Korea.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.25095&file=acr25095-sup-0001-Disclosureform.pdf.

Address correspondence via email to Hyeong Sik Ahn, MD, PhD, at iebm.ku@gmail.com.

Submitted for publication August 11, 2022; accepted in revised form January 26, 2023.

SIGNIFICANCE & INNOVATIONS

- In our population-based cohort study, we found an increased familial risk of gout that was highest among those with an affected brother, followed by father, sister, and mother.
- Familial risk adjusted for lifestyle and biologic risk factors decreased only slightly, suggesting that a genetic component is the predominant driver in the familial aggregation of gout.
- There was a statistically significant interaction between family history of gout and overweight and heavy alcohol consumption, as the combined effect of these 2 factors was associated with a markedly increased risk of disease, which was even further elevated among obese individuals.
- Our findings indicate the possibility of an interaction between gout-associated genes and obesity/alcohol consumption; heavy drinkers and obese individuals who have a family history of disease should be considered a high-risk group, and weight loss or cessation of alcohol consumption should be advised.

their study design; therefore, their estimates may suffer from imprecision. A large-scale population-based study from Taiwan reported a 2-fold increased familial risk of gout; however, due to its cross-sectional design, this study calculated familial risk as prevalence rather than incidence (5).

Several factors are associated with the development of gout, including obesity (6,7) and alcohol intake (8,9), in addition to factors such as increasing age (1), male sex, and ethnicity (10). Accordingly, the familial aggregation of gout may be influenced by both these factors along with genetic factors. However, the relative contribution of environmental and genetic factors in the familial clustering of gout is not well studied.

Given that both genetic and environmental factors contribute to the complex mechanism underlying gout pathogenesis, it may be possible that these factors have an interactive relationship, where the presence of obesity or alcohol consumption in individuals with a family history of disease yields a greater or lesser impact compared to non–genetically predisposed persons. Although some studies have been performed on gene–environment interactions in the context of gout, including on the influence of diuretic use and alcohol intake, these studies were often small scale and yielded non-replicated findings (11–14). Therefore, evidence supporting these associations remains inconclusive. Population-level epidemiologic studies of gene–environment interactions in gout are currently unavailable.

In the current study, we aimed to assess the familial aggregation of gout and to evaluate the relative contribution of family history and obesity or alcohol consumption using the National Health Insurance (NHI) database, which records health care utilization of the entire Korean population, in addition to the National Health Screening Program (NHSP) database, which includes screening information in up to 80% of the Korean adult population.

PATIENTS AND METHODS

Data sources. In the current study, we utilized the NHI and NHSP databases to obtain information regarding risk factors and family relationships. Linking these databases enabled us to estimate familial risk and assess the effect of obesity and alcohol consumption on gout development. The NHI is a government-operated service that provides mandatory insurance to South Korea's entire population of >50 million people. The NHI includes both inpatient and outpatient health care utilization data, which each medical facility must submit to the NHI for reimbursement purposes. The NHI database also includes comprehensive information regarding the family relationships of each enrolled individual and their dependents, which enabled us to identify family relationships.

The NHSP is a screening program through which all insured persons and their dependents are offered biannual health checkups that include a questionnaire regarding lifestyle habits and metabolic profile, the results of which are thereafter recorded in the database. At the NHSP checkup, each participant must fill out a standardized questionnaire related to their medical and lifestyle habits. The NHSP questionnaire includes questions regarding drinking status, including frequency of alcohol consumption and standard drink amount. Anthropometric measurements are also taken during the checkup, as well as basic tests, such as testing for liver enzymes, lipid parameters, radiography of the chest, fasting blood glucose levels, and creatinine.

Assessing family relationships. Employed or self-employed individuals can become NHI beneficiaries by paying a percentage of their income. Thereafter, the children and spouse of beneficiaries are eligible to enroll in the program as dependents after registration of birth or marriage. A detailed description of the familial relationship data from the NHI has been published elsewhere (15). All Korean residents are assigned a unique personal identification number at birth, which is used until death for a number of general administrative purposes, including voting, identification, etc., and is managed by the central government. The NHI also uses this number for personal identification. For familial relationships, birth or marriage are reported to the government in a formal process that involves this identification number, and therefore, the familial information may be considered accurate. Individuals with an identifiable biological mother and father were included. An individual was defined as the biological offspring of a married couple if they were registered as a dependent at birth. We excluded individuals with single parents, stepparents, or half-siblings and children who were not registered as a dependent at birth.

Assessing obesity and alcohol consumption. We acquired information regarding body mass index (BMI) and alcohol intake using NHSP data. From anthropometric measurements taken at the NHSP health checkup, BMI was classified as underweight (<18.5 kg/m^2), normal (18.5 to <25 kg/m^2), overweight (25 to <30 kg/m^2), and obese (≥30 kg/m^2). With regard to alcohol consumption, individuals were categorized according to standardized guidelines (16) as either a non-drinker, moderate drinker (<2 times per week or <5 drinks on any day [men]; <2 times per week or <4 drinks on any day [women]), or heavy drinker (≥2 times per week and ≥5 drinks on any day [men]; ≥2 times per week and ≥4 drinks on any day [women]). Occasional or social alcohol consumption could not be included. Based on a literature review of gout-associated risk factors, we selected other relevant lifestyle and biologic factors that were available in the NHSP data, including blood pressure, fasting blood glucose levels, smoking status, and cholesterol levels (see Supplementary Methods, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25095). We were unable to include diuretic use, a known gout risk factor, as it is not available in NHSP data.

Identification of gout case diagnosis. Gout patients were defined as those who were assigned an International Statistical Classification of Diseases and Related Health Problems, Tenth Revision code of M10 as a primary diagnosis, were prescribed allopurinol or febuxostat, and visited the clinic more than once in the following year. To identify patients with gout using the diagnostic code, we developed several case algorithms based on frequencies of clinic visits and drug prescriptions with gout as the principal diagnosis. The verification process for the identified cases is shown in Supplementary Methods (http://onlinelibrary.wiley.com/doi/10.1002/acr.25095). From this process, the case algorithm had a sensitivity and specificity of 96% and 98.5%, respectively.

Statistical analysis. The study population was followed up from January 1, 2002 until a diagnosis of gout, death, or the end of the follow-up period on December 31, 2018, whichever came first. We calculated person-years in each subject in the study. In individuals with gout-affected family members, the beginning of follow-up was the date of gout diagnosis in the family. In unaffected individuals, the index date was the start of the study period, on January 1, 2002 (see Supplementary Figure 1, at http://onlinelibrary.wiley.com/doi/10.1002/acr.25095). Study subjects contributed person-years only when they were considered still at risk (i.e., alive and living in Korea without a diagnosis of gout).

During the study period, starting from the date of diagnosis of the first gout case in the family, family members were considered "exposed" and were identified as "with affected first-degree relatives" or "familial group" and accumulated person-time of "being affected." If a second member was diagnosed, they were regarded as the first "familial case" and the remaining unaffected family members were "exposed" to 2 "family cases" from the date of diagnosis of the second case and were identified as "with >1 affected first-degree relative." In families with no gout cases, all family members were identified as "without affected first-degree relative" or the "non-familial group," and they accumulated person-time of "being unaffected." Meanwhile, if any of them were diagnosed, that person was regarded as a "non-familial case" and the rest of the family members accumulated person-time of being affected.

We used Cox proportional hazards regression models to estimate the familial risk of gout, and hazard ratios (HRs) with 95% confidence intervals (95% CIs) were calculated. Familial risk was also calculated according to family relationship, categorized as father, mother, or sibling. The proportional hazard assumption was tested using the Schoenfeld assumption and scale Schoenfeld residuals. Separate sex-stratified familial analyses were conducted.

Cox proportional hazards regression was used to examine the association of risk factors in gout, and HRs with 95% CIs were computed. The independent variables were alcohol use and obesity, as well as other confounding factors, while the dependent variable was the development of gout. In order to account for missing data, we excluded each missing value in the univariate analysis, which was a complete case analysis. For the multivariate analysis, we replaced the missing data with the most frequent values in each column. We also imputed mean/median data and used multiple imputation, and since the results were similar for all 3 imputation methods, we chose imputation using the most frequent variables. Among the total 5.5 million individuals identified, 30,875 had missing lifestyle factor data, accounting for ~0.6%.

To examine the contribution of risk factors to the familial aggregation of gout, familial risk before and after controlling for risk factors was examined. Familial risk was adjusted for age and sex using a Cox model. Then, it was adjusted again for potential risk factors, including smoking, BMI, hypertension, and hyperglycemia, in another Cox model.

Interactions were examined on an additive scale using the assumption that family history and obesity or alcohol consumption are independent of one another. Under the null hypothesis, the risk difference related with 1 exposure (e.g., familial risk) is constant across levels of other exposures (e.g., alcohol consumption) and vice versa. We analyzed whether the combined presence of family history and a given risk factor exceeded the sum of their separate risks. This was done using categorical variables, where each category was coded as a dichotomous variable, and 4 disjoint categories were created for the combinations of obesity/alcohol consumption and family history of gout. Incidence was calculated using multivariate analysis for individuals in each group. HRs were estimated by comparing the incidence in each group to that of a reference group, defined as

individuals with neither family history nor high BMI or alcohol intake.

In the presence of an interaction, having both a family history and a risk factor would increase the risk for gout more than expected. The amount of interaction as a departure from additivity was represented by relative excess risk due to interaction (RERI) and the corresponding 95% CI. When RERI is zero, it indicates that there is no interaction between the 2 exposures, while any deviation suggests an interaction (17). Interaction analyses were performed separately according to BMI and alcohol consumption subcategories (moderate versus non-drinker and heavy versus non-drinker; overweight versus normal weight and obese versus normal weight). All statistical analyses were performed using Stata version 15.0. Personal details were protected, and all data were anonymized. This study was approved by the Korea University Institutional Review Board.

RESULTS

Demographic characteristics of the risk population and risk factors. Using the study database, we identified 5,524,403 individuals with a biological mother and father, comprising 1.7 million families. Demographic variables and lifestyle and biologic characteristics of individuals with and those without affected first-degree relatives are shown in Table 1. We found no significant difference in terms of smoking, drinking, BMI, blood sugar, blood pressure, and cholesterol between the 2 groups.

Familial risk analysis. Among individuals with affected first-degree relatives, 29,391 cases developed gout during the study period, with an incidence of 31.33 per 10,000 person-years (95% CI 30.97, 31.69) (Table 2). Among individuals without affected first-degree relatives, 103,719 cases developed gout, with an incidence of 11.92 per 10,000 person-years (95% CI 11.85, 11.99). The age- and sex-adjusted HR of developing gout in individuals with an affected relative compared to those without affected first-degree relatives was 2.42 (95% CI 2.39, 2.46).

The incidence of gout in offspring with an affected mother was 23.93 (95% CI 23.12, 24.76) and the incidence with an affected father was 29.70 per 10,000 person-years (95% CI 29.28, 30.14) (Table 2). The corresponding HRs were 1.68 (95% CI 1.61, 1.75) and 2.33 (95% CI 2.29, 2.38), respectively. Gout incidence among individuals with an affected brother or sister was 35.34 (95% CI 34.32, 36.40) and 22.44 per 10,000 person-years (95% CI 19.55, 25.75), with respective HRs of 3.00 (95% CI 2.90, 3.12) and 1.97 (95% CI 1.67, 2.33). In individuals with >1 affected first-degree relative, the incidence was 63.78 per 10,000 person-years (95% CI 61.52, 66.13), with an HR of 4.71 (95% CI 4.50, 4.94).

Sex-specific familial risk. The familial risk of gout according to sex is shown in Figure 1. Overall, male patients with a family history of disease had an increased risk of gout compared to female patients, with HRs of 5.43 (95% CI 5.23, 5.63) and 2.98 (95% CI 2.43, 4.13), respectively. According to family relationship, male patients with an affected father (HR 2.46 [95% CI 2.42, 2.50]), mother (HR 3.60 [95% CI 2.63, 4.82]), and sibling (HR 3.25 [95% CI 3.15, 3.35]) had a higher risk of disease compared to female patients.

Risk of obesity and alcohol consumption on gout. The association between risk factors and gout development in the overall study population is shown in Table 3; among the examined variables, we found that high BMI and heavy alcohol consumption were significantly associated with disease development. The HRs for overweight and obese were 3.27 (95% CI 3.08, 3.48) and 5.33 (95% CI 5.00, 5.67), respectively. Regarding alcohol consumption, the HR for heavy drinking was 1.32 (95% CI 1.29, 1.36), compared to non-drinkers, while moderate drinking was not associated with an increased risk of disease (HR 1.00 [95% CI 0.97, 1.02]).

To assess the relative contribution of risk factors to gout, we adjusted the HR for lifestyle and biologic factors. The adjusted HR decreased slightly from 2.42 (95% CI 2.39, 2.46) to 2.29 (95% CI 2.25, 2.32), suggesting that the impact of these factors may be limited in familial aggregation.

Evaluation of the interaction between obesity/alcohol consumption and familial risk. Our analyses of the interaction between obesity and familial risk in gout development are shown in Figure 2. Overweight individuals with a family history of gout had a markedly increased risk of disease compared to individuals in the general population who do not have either genetic risk or high BMI, with a corresponding HR of 4.39 (95% CI 4.29, 4.49). The combined effect of being overweight and a family history of gout was higher than the sum of their separate effects (HR 4.39 versus 3.43), which indicates a statistically significant interaction (RERI 0.96 [95% CI 0.85, 1.06]). Obese individuals with a family history had a substantially increased risk of disease (HR 6.62 [95% CI 6.35, 6.91]), which was higher compared to overweight individuals, indicating a dose-response interactive relationship (RERI 1.88 [95% CI 1.61, 2.16]).

The sex-specific interaction analyses showed that the risk of gout in overweight male patients with a family history (HR 4.44 [95% CI 4.43, 4.55]) or obese male patients (HR 6.67 [95% CI 6.39, 6.96]) with a family history was markedly higher compared to female patients (overweight HR 2.64 [95% CI 2.17, 3.21], obese HR 4.53 [95% CI 3.43, 5.97]). Interactions were also more pronounced in male patients (RERI 1.84 [95% CI 1.56, 2.12]) compared to female patients (RERI 1.70 [95% CI 0.44, 2.96]).

Drinkers with a family history of gout had an increased risk of disease. In moderate drinkers, the combined effect with a family history of gout was similar to the sum of their individual effects

Table 1. Demographic data in the total study population and their association with lifestyle risk factors in gout*

	Subjects with an affected first-degree relative (n = 545,447)	Subjects without an affected first-degree relative (n = 4,978,956)	Standardized difference
Sex			
Male	351,809 (64.5)	3,217,360 (64.6)	
Female	193,638 (35.5)	1,761,596 (35.4)	
Year of birth			0.16
Up to 1971	120,531 (22.1)	976,909 (19.6)	
1972–1981	238,775 (43.8)	1,933,466 (38.8)	
1982–1991	172,132 (31.6)	1,873,014 (37.6)	
1992–2001	14,009 (2.6)	195,567 (3.9)	
Alcohol consumption			0.044
Non-drinker	97,665 (17.9)	976,473 (19.6)	
<2 times per week or <5 drinks (men) or <4 drinks (women) on any day	346,828 (63.6)	3,112,500 (62.5)	
≥2 times per week and ≥5 drinks (men) or ≥4 drinks (women) on any day	80,026 (14.7)	708,517 (14.2)	
BMI, kg/m^2			0.097
<18.5	32,415 (5.9)	358,419 (7.2)	
18.5–<25	336,218 (61.6)	3,208,604 (64.4)	
25–<30	145,508 (26.7)	1,181,253 (23.7)	
≥30	31,306 (5.7)	230,680 (4.6)	
Blood pressure, mm Hg			0.073
SBP <120 and DBP <80 (reference)	232,331 (42.6)	2,258,352 (45.4)	
SBP ≥120 and SBP <130/DBP <80	63,620 (11.7)	597,050 (12.0)	
SBP ≥130 and SBP <140 or DBP ≥80 and DBP <90	188,667 (34.6)	1,655,264 (33.2)	
SBP ≥140 and SBP <180/DBP ≥90	59,481 (10.9)	459,036 (9.2)	
SBP ≥180 and/or DBP ≥120	1,293 (0.24)	8,739 (0.2)	
Fasting blood glucose, mg/dl			0.026
<110	504,662 (92.5)	4,639,225 (93.2)	
110 to <126	26,438 (4.9)	223,094 (4.5)	
≥126	14,230 (2.6)	115,630 (2.3)	
Cholesterol, mg/dl			0.062
<200	365,975 (67.1)	3,475,386 (69.8)	
200 to <240	135,939 (24.9)	1,160,541 (23.3)	
≥240	42,616 (7.8)	333,431 (6.7)	
Smoking, pack/year			0.059
Non-smoker	253,339 (46.5)	2,399,954 (48.2)	
<5	93,721 (17.2)	895,058 (18.0)	
5 to <10	85,428 (15.7)	758,290 (15.2)	
10 to <20	78,205 (14.3)	643,410 (12.9)	
20 to <30	25,358 (4.6)	202,470 (4.1)	
≥30	9,396 (1.7)	79,774 (1.6)	
Physical activity, times/week			0.013
None	256,696 (47.1)	2,345,919 (47.1)	
1–2	200,904 (36.8)	1,852,306 (37.2)	
≥3	86,032 (15.8)	765,371 (15.4)	

* Except where indicated otherwise, values are the number (%) of individuals. BMI = body mass index; DBP = diastolic blood pressure; SBP = systolic blood pressure.

(HR 2.28 [95% CI 2.22, 2.35] versus 2.24), and the interaction was not statistically significant (RERI 0.05 [95% CI –0.07, 0.17]) (Figure 3). Heavy drinkers with a family history demonstrated an even higher magnitude of excess risk (HR 2.95 [95% CI 2.83, 3.09]) with statistical significance (RERI 0.36 [95% CI 0.20, 0.52]). Sex-specific analyses showed that interactions were more prominent in heavy drinking compared to moderate drinking among both male and female subjects, although the magnitude of excess risk was higher in male subjects (HR 2.99 [95% CI 2.86, 3.13]) compared to female subjects (HR 2.09 [95% CI 1.52, 2.84]).

Influence of risk factors on gout among familial and non-familial cases. An assessment of the association between gout development and obesity and alcohol consumption was performed separately in the familial group and non-familial groups (see Table 3). The magnitude of the risk estimate for obesity was higher in the familial group compared to non-familial group, with HRs of 5.50 (95% CI 4.69, 6.46) and 5.36 (95% CI 5.01, 5.74), respectively. The magnitude of the risk estimates for heavy alcohol consumption was similar between the familial group and non-familial group, with HRs of 1.28 (95% CI 1.20, 1.36) and 1.34 (95% CI 1.30, 1.38).

Table 2. Familial risk of gout among first-degree relatives of affected patients*

	Subjects with an affected first-degree relative						
	Father	Mother	Brother	Sister	>1 first-degree relative	Total	Subjects without an affected first-degree relative
No. (%) of subjects at risk	359,992	78,919	73,126	5,216	27,667	545,447	4,978,956
Male	235,063 (65)	53,606 (68)	42,448 (58)	3,301 (63)	16,965 (61)	351,809 (64)	3,217,360 (65)
Female	124,929 (35)	25,313 (32)	30,678 (42)	1,915 (37)	10,702 (39)	193,638 (36)	1,761,596 (35)
No. of cases	18,426	3,263	4,431	203	2,944	29,391	103,719
Person-years	6,203,141	1,363,805	1,253,669	90,470	461,583	9,381,037	87,014,604
Incidence/10,000 person-years (95% CI)	29.70 (29.28, 30.14)	23.93 (23.12, 24.76)	35.34 (34.32, 36.40)	22.44 (19.55, 25.75)	63.78 (61.52, 66.13)	31.33 (30.97, 31.69)	11.92 (11.85, 11.99)
Crude HRs (95% CI)	2.33 (2.29, 2.37)	1.80 (1.72, 1.87)	2.66 (2.56, 2.76)	1.82 (1.54, 2.15)	4.38 (4.18, 4.58)	2.40 (2.36, 2.44)	1.00
HRs (95% CI) adjusted for age and sex	2.33 (2.29, 2.38)	1.68 (1.61, 1.75)	3.00 (2.90, 3.12)	1.97 (1.67, 2.33)	4.71 (4.50, 4.94)	2.42 (2.39, 2.46)	1.00
HRs (95% CI) adjusted for age, sex, and risk factors	2.21 (2.17, 2.25)	1.62 (1.55, 1.69)	2.78 (2.68, 2.88)	1.95 (1.65, 2.30)	4.23 (4.04, 4.44)	2.29 (2.25, 2.32)	1.00

* 95% CI = 95% confidence interval; HR = hazard ratio.

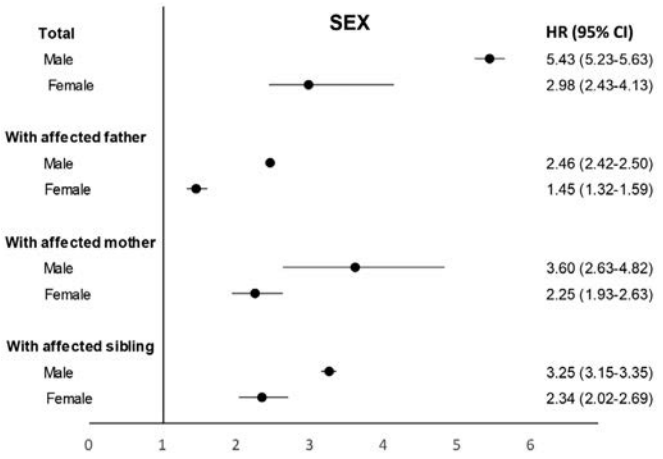

Figure 1. Familial risk of gout according to sex. 95% CI = 95% confidence interval; HR = hazard ratio.

DISCUSSION

Using the NHI and NHSP databases, we identified 5,524,403 individuals with a blood-related first-degree relative. Overall, individuals with affected first-degree relative had a 2.42-fold increased risk of disease. Although obesity and alcohol consumption were significantly associated with disease risk, it appears that a genetic component is the primary driver of familial aggregation. Our findings indicate the possibility of a dose-dependent gene–environment interaction, as the combination of both a family history of gout and either high BMI or heavy alcohol consumption was associated with a markedly increased risk of disease, which was even further elevated among obese individuals.

Although the importance of a hereditary factor in gout has long been recognized, only a few studies of familial risk have been performed. We identified 5 small-scale studies from the 1940s and 1970s that demonstrated a higher risk of disease among the family members of gout-affected patients, including small-scale case series studies from the UK and US that demonstrated familial incidence estimates of 36% (18) and 6–18% (19,20), respectively. Similarly, a Hungarian study showed that 8.5% of patients had a gout-affected family member (21). However, these studies typically enrolled up to a few hundred participants from specialized medical centers, which may have in turn led to the inclusion of more selected cases that are not representative of the general population. Moreover, these studies often relied on self-reported questionnaires in order to acquire information regarding family relationships and disease, which may be prone to selective recall bias.

A recent large-scale Taiwanese study demonstrated an overall prevalence risk ratio of 2.91 (95% CI 2.49, 3.42) among first-degree relatives of gout patients (5). However, this cross-sectional study calculated familial risk by comparing the gout point prevalence between first-degree relatives of gout-affected patients and the general population. In contrast, our cohort design enabled us to concurrently follow up the first-degree relative of gout patients after diagnosis as well as follow up unaffected first-degree relatives, and consequently we were able to provide the time-related incidence pattern and incidence ratio as familial risk, rather than prevalence.

Although the familial aggregation of gout is influenced by both genetic and lifestyle/biologic factors, our findings suggest that a genetic predisposition is the predominant driver of familial aggregation. After adjusting for lifestyle and biologic characteristics, the magnitude of familial risk reduced slightly from 2.42 to 2.29, suggesting a limited contribution of these factors on familial aggregation. We found that the magnitude of familial risk increased with increasing genetic relatedness, as risk was highest among individuals with >1 affected first-degree relative, followed by siblings, then offspring. These findings suggest that a genetic component plays a substantial role in the familial aggregation of gout.

High BMI and heavy alcohol consumption were associated with an increased risk of gout in both the familial group and non-familial group. Overweight individuals with a family history of disease had a markedly increased risk of disease, and their combined risk was significantly higher than the sum of their individual risks (HR 4.39 versus 3.43) and was even higher among obese individuals (HR 6.62 versus 4.74). This trend was observed among both male and female patients, but the interaction was more pronounced in male patients (RERI 1.84 [95% CI 1.56, 2.12]). Our findings contrast a recent US cohort study that found that interactions between excess adiposity and gout genetic predisposition were stronger among female subjects than male subjects (22). Our findings suggest a dose-dependent interactive relationship in which genetic factors and obesity potentiate each other rather than operating independently. Moreover, in the separate risk analyses for the familial group and non-familial group, the magnitude of risk associated with obesity was higher in the familial group compared to the non-familial group (HR 5.50 versus 5.36).

The combined effect of heavy drinking with a family history of gout was also higher than the sum of their separate effects (HR 2.95 versus 2.60), suggesting an interaction, although the magnitude was lower than that of obesity, and interactions were not observed for moderate drinking (HR 2.28 versus 2.24). These findings support the notion that the impact of alcohol consumption and obesity is more pronounced among genetically predisposed persons. Based on our findings, we emphasize that obese individuals or heavy drinkers with a family history of gout should undergo genetic counseling, which includes risk communication and management, especially for obesity and alcohol consumption. Genetic counseling should include informing the family members of gout patients that these factors may further increase their chance of developing gout, as well as the importance of weight management and alcohol moderation for disease

Table 3. Gout risk factor analyses in the total study population and in the familial group and non-familial group*

	Total	Subjects with an affected first-degree relative	Subjects without an affected first-degree relative
Sex			
Male	1.00	1.00	1.00
Female	0.13 (0.13, 0.14)	0.10 (0.09, 0.10)	0.14 (0.14, 0.15)
BMI, kg/m^2			
<18.5 (reference)	1.00	1.00	1.00
18.5 to <25	1.68 (1.58, 1.78)	2.03 (1.74, 2.37)	1.62 (1.52, 1.73)
25 to <30	3.27 (3.08, 3.48)	3.68 (3.15, 4.30)	3.22 (3.01, 3.43)
≥30	5.33 (5.00, 5.67)	5.50 (4.69, 6.46)	5.36 (5.01, 5.74)
Alcohol consumption, drinks/week			
Non-drinker	1.00	1.00	1.00
<2 times per week or <5 drinks (men) or <4 drinks (women) on any day	1.00 (0.97, 1.02)	1.07 (1.01, 1.13)	0.98 (0.95, 1.01)
≥2 times per week and ≥5 drinks (men) or ≥4 drinks (women) on any day	1.32 (1.29, 1.36)	1.28 (1.20, 1.36)	1.34 (1.30, 1.38)
Blood pressure, mm Hg			
SBP <120/DBP <80 (reference)	1.00	1.00	1.00
SBP ≥120 and SBP <130/DBP <80	1.06 (1.03, 1.08)	1.07 (1.01, 1.13)	1.05 (1.03, 1.08)
SBP ≥130 and SBP <140 or DBP ≥80 and DBP <90	1.24 (1.22, 1.26)	1.23 (1.19, 1.28)	1.24 (1.22, 1.27)
SBP ≥140 and SBP <180 or DBP ≥90	1.58 (1.55, 1.62)	1.55 (1.48, 1.62)	1.60 (1.56, 1.63)
SBP ≥180 and/or DBP ≥120	2.66 (2.47, 2.86)	2.46 (2.10, 2.89)	2.73 (2.51, 2.96)
Fasting blood glucose, mg/dl			
<110 (reference)	1.00	1.00	1.00
110 to <126	1.03 (1.00, 1.05)	1.03 (0.98, 1.09)	1.02 (1.00, 1.05)
≥126	0.79 (0.76, 0.82)	0.78 (0.72, 0.84)	0.79 (0.76, 0.82)
Total cholesterol, mg/dl			
<200 (reference)	1.00	1.00	1.00
200 to <240	1.22 (1.20, 1.24)	1.25 (1.21, 1.29)	1.21 (1.19, 1.23)
≥240	1.48 (1.45, 1.51)	1.50 (1.44, 1.57)	1.48 (1.45, 1.51)
Smoking status, pack/year			
Non-smoker (reference)	1.00	1.00	1.00
<5	0.99 (0.97, 1.01)	1.01 (0.97, 1.06)	0.98 (0.96, 1.00)
5 to <10	1.03 (1.01, 1.05)	1.00 (0.96, 1.04)	1.03 (1.01, 1.06)
10 to <20	1.02 (1.00, 1.04)	0.99 (0.95, 1.04)	1.03 (1.01, 1.06)
20 to <30	1.07 (1.04, 1.11)	1.02 (0.96, 1.09)	1.09 (1.06, 1.13)
≥30	1.24 (1.19, 1.29)	1.21 (1.10, 1.34)	1.24 (1.18, 1.30)
Physical activity, times/week			
None	1.00	1.00	1.00
1–2	1.00 (0.97, 1.02)	1.01 (0.98, 1.04)	1.01 (0.99, 1.03)
≥3	1.08 (1.06–1.10)	1.08 (1.04, 1.13)	1.08 (1.06, 1.10)

* Values are the hazard ratio (HR) (95% confidence interval [95% CI]). DBP = diastolic blood pressure; SBP = systolic blood pressure.

prevention. Screening for hyperuricemia and gout in high-risk patients should also be considered.

Few studies have investigated interactions between gout-associated genes and obesity (23–26). For instance, Brandstatter et al (23) identified 4 single-nucleotide polymorphisms (SNPs) located within *SLC2A9* associated with uric acid levels that are modified by BMI. In addition, a T-allele of a SNP (rs2544390) located in *LRP2* demonstrated a nonadditive interaction with alcohol consumption on the risk of gout among Japanese, Maori, and Pacific Islander populations, though this was not observed in European populations (12,13). In individuals of European descent, an interaction was observed between alcohol consumption and *GCKR* and *A1CF* genes, in which alcohol exposure negated or fully suppressed the genetic effect (14). With respect to other variables, gene–environment interactions have also been observed between renal urate transporter genes *SLC22A11* and *SLC2A9* and thiazide or loop diuretics among hypertensive patients (11), as well as between glucose transporter gene *ABCG2* and sweetened beverages (27). However, relatively few studies have investigated gene–environment interactions in the context of gout, most of which were limited in terms of size and power and yielded inconsistent and nonreplicated findings.

To date, genome-wide association studies have identified hundreds of loci associated with serum urate levels, including *SLC2A9, ABCG2, PDZK1, SLC22A11*, and *SLC17A1*, which encode renal urate transporters or regulators, as well as *GCKR*, *R3HDM2-INHBC*, and *RREB1* (3,11). In East Asian populations, *SLC2A9*, *SLC22A12*, and *SCL2A12* demonstrated significant genome-wide associations with serum urate levels, and 28 loci

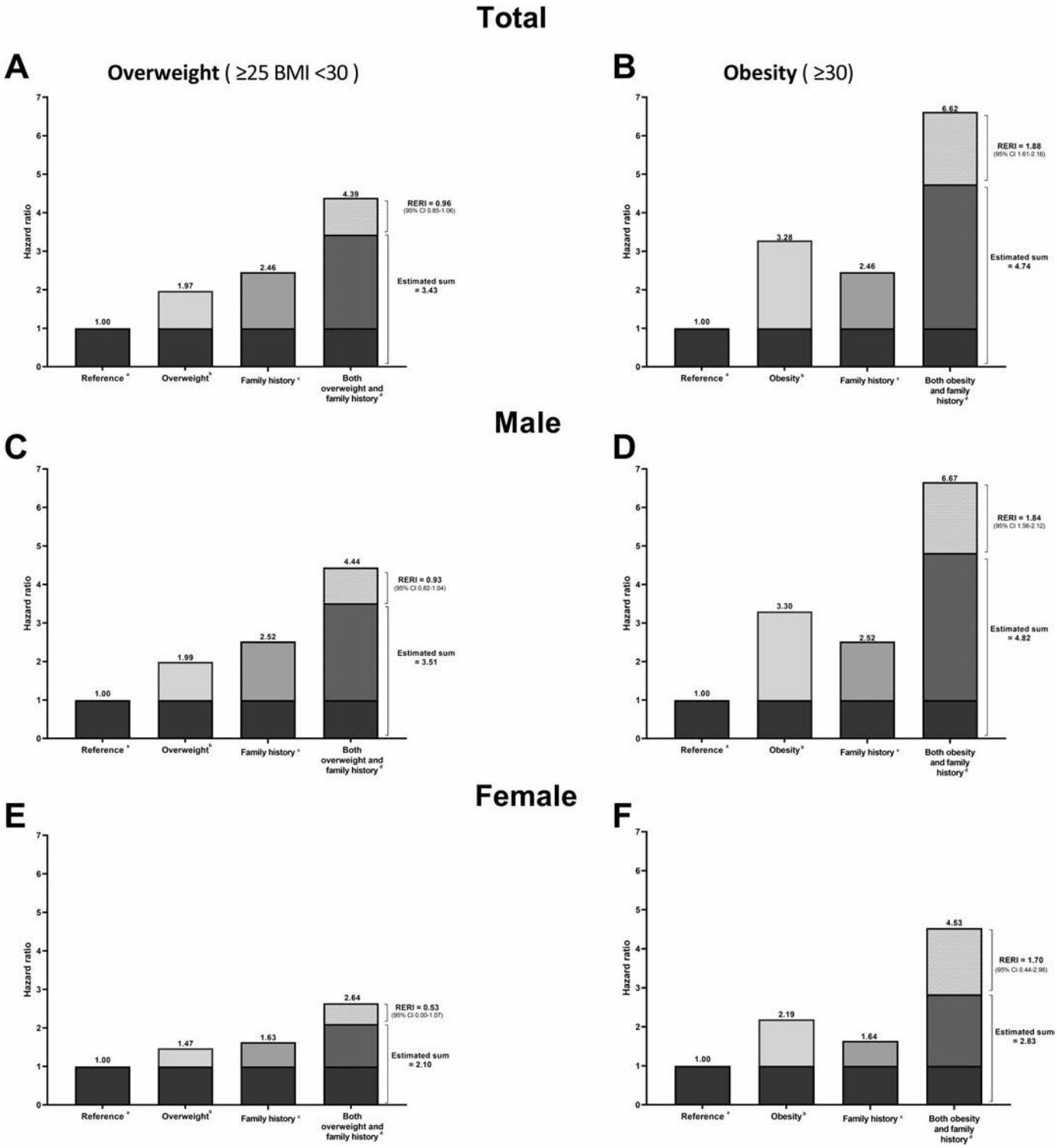

Figure 2. Combined effect of family history and being either overweight or obese overall (**A** and **B**, respectively) or stratified by male subjects (**C** and **D**, respectively) or female subjects (**E** and **F**, respectively) on the risk of gout compared to those with a normal weight. a = Subjects with no family history of gout nor high body mass index (BMI). b = Individual effect of the risk factor BMI in subjects with compared to those without a high BMI. c = Individual effect of family history in those with compared to those without a family history of gout. d = Combined effect of family history and high BMI. 95% CI = confidence interval; RERI = relative excess risk due to interaction.

have been identified among European populations (2). Although relatively little is known regarding the genetic contribution to MSU deposition or gouty inflammation, a genome-wide association study identified an association between the gene *ALDH2*, which is a well-known variation of alcohol-metabolizing enzymes, and crystal-induced inflammation as well as hyperuricemia (28).

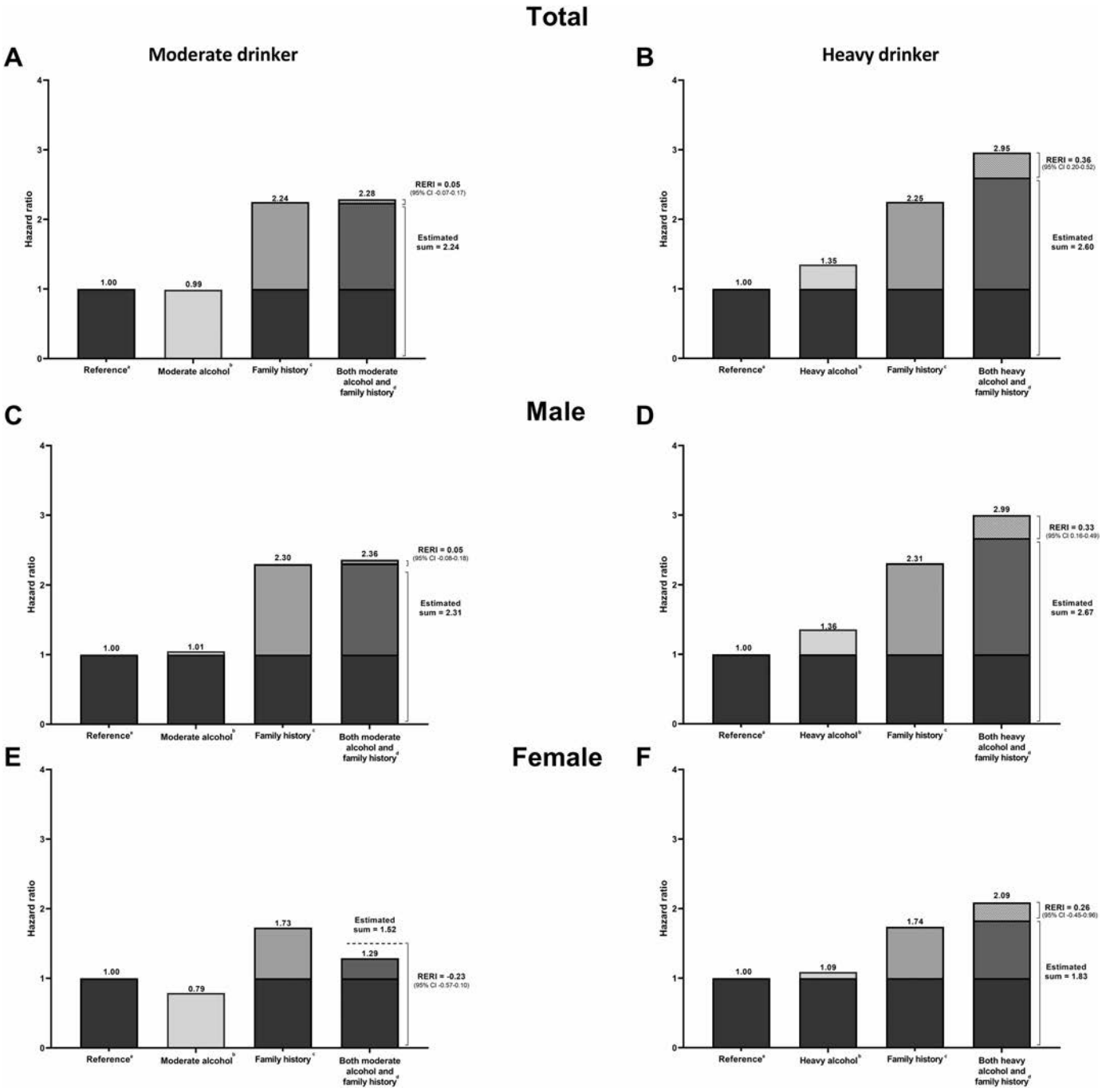

Figure 3. Combined effect of family history and either moderate or heavy drinking overall (**A** and **B**, respectively) or stratified by male subjects (**C** and **D**, respectively) or female subjects (**E** and **F**, respectively) on the risk of gout compared to non-drinkers. a = Subjects with no family history of gout nor alcohol intake. b = Individual effect of the risk factor of alcohol intake in subjects with compared to those without alcohol intake. c = Individual effect of family history in subjects with compared to those without a family history of gout. d = Combined effect of family history and alcohol intake. 95% CI = 95%confidence interval; RERI = relative excess risk due to interaction.

Our epidemiologic findings represent the average effect of an interaction between these genes and obesity/alcohol consumption, and therefore our analysis highlights the need for further studies to assess the interaction of specific genes with obesity and alcohol intake, especially at the genome-wide level. Given that our study database included the entire Korean population, our findings may be considered applicable to individuals of Korean descent. While some studies report genetic variation in gout risk between Asian and European populations (12,13), gout-related genes are likely to overlap, and therefore the

generalizability of our findings would likely extend to individuals with other ethnic backgrounds; although, this should be investigated in future studies.

The mechanisms by which alcohol consumption confers gout risk include increasing the production of adenosine triphosphate necessitated for alcohol metabolism and by reducing urinary excretion due to the elevation of blood lactate produced by the oxidation of ethanol (12). In obesity, it has been suggested that circulating molecules such as saturated fatty acids and cholesterol crystals may activate *TLR2* and *TLR4* signaling pathway and activate the *NLRP3* inflammasome to cause obesity-induced inflammation. It is possible that the genes involved in these biologic mechanisms related to alcohol consumption and obesity might also be involved in the pathogenesis of gout and may mediate interactions. For instance, studies have suggested that genetic factors in gout are related to the function of proinflammatory cytokines, such as *TNF*, *PRKG2*, and *TGFB* (29). Therefore, future studies are needed to elucidate the relationship between specific alcohol/obesity–related genes and gout development, which may explain the interactive relationship between the 2 factors.

Our study is not without limitations. We aimed to demonstrate gene–environment interactions based on the assumption that familial and genetic factors are interchangeable. However, the familial risk of gout is likely influenced not only by shared genetics, but also by non-genetic factors such as diet, diuretic use, or socioeconomic factors like education or occupation, which could not be accounted for in our study (30). Additionally, biologically based risk factors such as obesity are known to have a genetic component, but the genetic influence of obesity could also not be considered. However, after adjusting for risk factors, the magnitude of familial risk only slightly decreased from 2.42-fold to 2.29-fold, indicating a limited contribution of these factors to familial aggregation. Since these findings suggest that genetic factors are the primary determinants of familial aggregation, we believe this limitation may be minimal.

Another limitation of our study was the use of administrative data, which may raise particular concerns regarding the validity of gout diagnosis. However, in order to maximize diagnostic accuracy, case algorithms were developed based on the number of hospital visits, and we selected a combination of algorithms with 98.5% and 96% sensitivity. An additional weakness was the length of the follow-up period, which may not have been sufficiently long enough to cover all familial occurrences and may therefore have resulted in the omission of some gout cases.

We found a 2.42-fold increased risk of disease among individuals with gout-affected first-degree relative, and our findings suggest that a genetic component is the primary driver in familial aggregation. Heavy drinkers and obese individuals who have a family history of gout should be considered at high-risk, and weight loss or cessation of alcohol consumption should be advised. Screening for hyperuricemia and gout in high-risk patients should also be considered.

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. Ahn 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. K. H. Kim, I. Choi, Swan, K. U. Kim, Ahn.
Acquisition of data. H. Kim, Hong, Y. Kim, Kang, Cha, Ahn.
Analysis and interpretation of data. K. H. Kim, Swan, Kazmi, Y. Kim, S. Choi, Eom, Hann, Ahn.

REFERENCES

1. Dehlin M, Jacobsson L, Roddy E. Global epidemiology of gout: prevalence, incidence, treatment patterns and risk factors. Nat Rev Rheumatol 2020;16:380–90.
2. Major TJ, Dalbeth N, Stahl EA, et al. An update on the genetics of hyperuricaemia and gout. Nat Rev Rheumatol 2018;14:341–53.
3. Leask MP, Merriman TR. The genetic basis of urate control and gout: insights into molecular pathogenesis from follow-up study of genome-wide association study loci. Best Pract Res Clin Rheumatol 2021;35:101721.
4. Krishnan E, Lessov-Schlaggar CN, Krasnow RE, et al. Nature versus nurture in gout: a twin study. Am J Med 2012;125:499–504.
5. Kuo CF, Grainge MJ, See LC, et al. Familial aggregation of gout and relative genetic and environmental contributions: a nationwide population study in Taiwan. Ann Rheum Dis 2015;74:369–74.
6. Evans PL, Prior JA, Belcher J, et al. Obesity, hypertension and diuretic use as risk factors for incident gout: a systematic review and meta-analysis of cohort studies. Arthritis Res Ther 2018;20:136.
7. Larsson SC, Burgess S, Michaelsson K. Genetic association between adiposity and gout: a Mendelian randomization study. Rheumatology (Oxford) 2018;57:2145–8.
8. Nieradko-Iwanicka B. The role of alcohol consumption in pathogenesis of gout. Crit Rev Food Sci Nutr 2022;62:7129–37.
9. Choi HK, Atkinson K, Karlson EW, et al. Alcohol intake and risk of incident gout in men: a prospective study. Lancet 2004;363:1277–81.
10. MacFarlane LA, Kim SC. Gout: a review of nonmodifiable and modifiable risk factors. Rheum Dis Clin North Am 2014;40:581–604.
11. McAdams-DeMarco MA, Maynard JW, Baer AN, et al. A urate gene-by-diuretic interaction and gout risk in participants with hypertension: results from the ARIC study. Ann Rheum Dis 2013;72:701–6.
12. Hamajima N, Naito M, Okada R, et al. Significant interaction between LRP2 rs2544390 in intron 1 and alcohol drinking for serum uric acid levels among a Japanese population. Gene 2012;503:131–6.
13. Rasheed H, Phipps-Green A, Topless R, et al. Association of the lipoprotein receptor-related protein 2 gene with gout and non-additive interaction with alcohol consumption. Arthritis Res Ther 2013;15:R177.
14. Rasheed H, Stamp LK, Dalbeth N, et al. Interaction of the GCKR and A1CF loci with alcohol consumption to influence the risk of gout. Arthritis Res Ther 2017;19:161.
15. Kim HJ, Shah SC, Hann HJ, et al. Familial risk of inflammatory bowel disease: a population-based cohort study in South Korea. Clin Gastroenterol Hepatol 2021;19:2128–37.e15.
16. National Institute on Alcohol Abuse and Alcoholism. Drinking levels defined. URL: https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking.
17. Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction. Int J Epidemiol 2012;41:514–20.

18. Grahame R, Scott JT. Clinical survey of 354 patients with gout. Ann Rheum Dis 1970;29:461–8.
19. Neel JV. The clinical detection of the genetic carriers of inherited disease. Medicine (Baltimore) 1947;26:115–53.
20. Talbott JH, Yü Ta-F, Seegmiller JE. Gout and uric acid metabolism. New York (NY): Stratton Intercontinental Medical Book Corporation; 1976. p. 303.
21. Mituszova M, Judak A, Poor G, et al. Clinical and family studies in Hungarian patients with gout. Rheumatol Int 1992;12:165–8.
22. McCormick N, Yokose C, Lu N, et al. Impact of adiposity on risk of female gout among those genetically predisposed: sex-specific prospective cohort study findings over >32 years. Ann Rheum Dis 2022;81:556–63.
23. Brandstatter A, Kiechl S, Kollerits B, et al. Sex-specific association of the putative fructose transporter SLC2A9 variants with uric acid levels is modified by BMI. Diabetes Care 2008;31:1662–7.
24. Li WD, Jiao H, Wang K, et al. A genome wide association study of plasma uric acid levels in obese cases and never-overweight controls. Obesity (Silver Spring) 2013;21:E490–4.
25. Huffman JE, Albrecht E, Teumer A, et al. Modulation of genetic associations with serum urate levels by body-mass-index in humans. PLoS One 2015;10:e0119752.
26. Reynolds RJ, Vazquez AI, Srinivasasainagendra V, et al. Serum urate gene associations with incident gout, measured in the Framingham Heart Study, are modified by renal disease and not by body mass index. Rheumatol Int 2016;36:263–70.
27. Pascart T, Liote F. Gout: state of the art after a decade of developments. Rheumatology (Oxford) 2019;58:27–44.
28. Kawamura Y, Nakaoka H, Nakayama A, et al. Genome-wide association study revealed novel loci which aggravate asymptomatic hyperuricaemia into gout. Ann Rheum Dis 2019;78:1430–7.
29. Lai HM, Chen CJ, Su BY, et al. Gout and type 2 diabetes have a mutual inter-dependent effect on genetic risk factors and higher incidences. Rheumatology (Oxford) 2012;51:715–20.
30. Kapetanovic MC, Hameed M, Turkiewicz A, et al. Prevalence and incidence of gout in southern Sweden from the socioeconomic perspective. RMD Open 2016;2:e000326.

Arthritis Care & Research
Vol. 75, No. 9, September 2023, pp 1967–1975
DOI 10.1002/acr.25086

American College of Rheumatology
Empowering Rheumatology Professionals

Mind the Mood: Momentary Depression and Anxiety Moderate the Correspondence Between Subjective and Objective Cognitive Functioning in Fibromyalgia

Anson E. Kairys, Thomas R. Valentine, Daniel Whibley, and Anna L. Kratz

Objective. Subjective cognitive dysfunction (SCD) affects 55–75% of individuals with fibromyalgia (FM), but those reporting cognitive difficulties often lack corresponding objective deficits. Symptoms of depression and anxiety are prevalent in FM and may account for part of this discrepancy. This study was undertaken to investigate whether momentary (within-day, across 7 days) changes in mood moderate the relationship between within-the-moment SCD and mental processing speed performance.

Methods. A total of 50 individuals with FM (mean age 44.8 years, mean education 15.7 years, 88% female, 86% White) completed momentary assessments of subjective cognitive functioning, depressive and anxious symptoms, and a test of processing speed. Assessments were completed 5 times per day for 8 consecutive days on a study-specific smartphone application.

Results. Momentary ratings of SCD were positively associated with mean reaction time ($P < 0.001$) and variability of processing speed ($P = 0.02$). Depressive symptoms moderated the relationship between SCD and processing speed, with lower correspondence when depressive symptoms were higher ($P = 0.03$). A similar moderating effect was demonstrated for both depression ($P = 0.02$) and anxiety ($P = 0.03$) on the association between SCD and variability in processing speed performance.

Conclusion. Individuals with FM may have more accurate self-perception of momentary changes in mental processing speed during periods of less pronounced mood symptoms based on their corresponding objective processing speed performance. However, during moments of heightened depression and anxiety, we found increasingly less correspondence between SCD and objective performance, suggesting that psychological symptoms may play an important role in self-perception of cognitive dysfunction in FM as it relates to mental processing speed.

INTRODUCTION

Fibromyalgia (FM) is a chronic musculoskeletal pain disorder commonly accompanied by symptoms of depression, anxiety, fatigue, and cognitive dysfunction (1–4). Subjective cognitive dysfunction (SCD) refers to an individual's perception of a reduction in their cognitive capacity and is reported in ~55–75% of adults with FM (5). SCD encompasses a wide range of cognitive domains including aspects of attention, mental processing speed, executive functioning, and memory (2,6,7) with significant negative effects on daily functioning, occupational outcomes, and quality of life, making it one of the most troubling symptoms in individuals with FM (2,6–12).

Despite the prominent impact of SCD, individuals with FM do not consistently demonstrate deficits regarding objective measures of cognition (6,12–15). Rather, there is often a "cognitive discrepancy" (i.e., an overestimation or underestimation of subjective cognitive functioning relative to objective performance) observed in FM (12,16–18). One plausible explanation for this observed cognitive discrepancy is the influence of mood symptoms, such as depression and anxiety, on an individual's perception of their cognitive abilities (15,19).

Symptoms of depression and, to a lesser extent, anxiety are common in individuals with FM (4). Prior research suggests that symptoms of depression and anxiety can contribute to cognitive biases that can negatively affect perception of cognitive function,

Supported by the NIH/National Institute of Arthritis and Musculoskeletal and Skin Diseases (grant K01-AR-064275). Dr. Valentine's work was supported during manuscript preparation by the Mentor-Based Postdoctoral Fellowship Program in Rehabilitation Research from the National Multiple Sclerosis Society (grant MB-1706-27943).

Anson E. Kairys, PhD, Thomas R. Valentine, PhD, Daniel Whibley, PhD, Anna L. Kratz, PhD: University of Michigan, Ann Arbor.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.25086&file=acr25086-sup-0001-Disclosureform.pdf.

Address correspondence via email to Anson E. Kairys, PhD, at aekairys@med.umich.edu.

Submitted for publication September 15, 2022; accepted in revised form January 3, 2023.

SIGNIFICANCE & INNOVATIONS

- This is the first study to evaluate the influence that momentary changes in mood have on the relationship between objective and subjective cognitive functioning using momentary ecological assessments.
- Individuals with fibromyalgia may be more accurate in their perception of momentary changes in mental processing speed (i.e., better correspondence between subjective and objective measures) when mood symptoms are minimal.
- Increasing mood symptoms (i.e., higher ratings of depression and anxiety) lead to a larger discrepancy between perceived cognitive functioning and objective cognitive performance, suggesting psychological symptoms have an influence on cognitive self-appraisal.

such as negative self-directed thinking patterns (20), underestimation of true capabilities (21), and excessive worry and hopelessness during cognitive challenges (22). However, it remains unclear whether SCD corresponds with objective cognitive dysfunction and to what extent mood symptoms contribute to any discrepancy between SCD and objective performance.

A potential limiting factor in previous research is that the majority of studies utilized single time point assessments and cross-sectional designs to assess cognition in FM, which fail to account for the evidence that cognitive functioning and mood symptoms tend to fluctuate within individuals across short periods of time, even within a single day. These approaches also do not allow for examination of the effects of transient mood states on the association between perceived and objective cognition. Thus, more intensive approaches to data collection that use multiple within-day assessments may shed additional light on this perplexing issue.

Overall, the possible moderating influence of depressed and anxious mood on the relationship between SCD and objective cognitive performance is understudied in FM. This limitation poses challenges in understanding the nature of SCD and in identifying focused interventions targeting SCD in FM. Therefore, the goal of the present study was to assess the influence of momentary changes in self-reported symptoms of depression and anxiety on the relationship between SCD and objective cognitive performance in individuals with FM. To this end, we utilized a microlongitudinal study design wherein adults with FM completed self-report measures of cognitive functioning (SCD), depression and anxiety, and completed an objective measure of mental processing speed multiple times a day. We hypothesized that there would be a significant, negative association between momentary ratings of SCD and processing speed. Further, we expected a weaker relationship between SCD and objective performance at times when symptoms of depression and anxiety were higher.

MATERIALS AND METHODS

Participants. Participants with FM who fulfilled the 2016 American College of Rheumatology survey criteria (23), were ≥18 years old, and had at least a sixth-grade reading level in English were eligible for inclusion in this study. Individuals were excluded if they had 1) a comorbid neurologic disorder, learning disability, or cognitive impairment; 2) current alcohol or recreational drug use dependence or prolonged (≥5 years) history of substance dependance; 3) visual or hearing impairment that would preclude cognitive assessment; 4) a diagnosis of untreated obstructive sleep apnea; or 5) atypical sleep-wake pattern (e.g., night shift workers).

Study procedures. All study procedures were approved by the Medical Institutional Review Board at the University of Michigan prior to study initiation. Participants were recruited through existing patient registries, community groups, placement of fliers in health centers and community settings, and advertisement on a university-based recruitment website (umhealthreserach.org). This paper addresses one of the primary study aims; previous articles from the study have shown that ambulatory measures are able to detect cognitive dysfunction in FM relative to individuals without FM (24), that cognitive test performance is worse when participants are distracted (25), and subjective and objective cognitive functioning is worse in those with FM when pain intensity is high (26).

Procedures for this study have been described previously (24,25). Study participation involved a ~90-minute baseline visit followed by an 8-day home monitoring period (i.e., a 1-day run-in period followed by 7 days of data collection). At the baseline visit, enrolled participants completed a series of self-report measures and standardized cognitive testing (baseline self-report and cognitive testing data were reported previously [24]) and were given data collection devices. At the conclusion of the home monitoring period, participants returned the devices via a postage-paid return box to the laboratory for data processing. Participants were compensated ≤$175 for full completion of the study. Participants were issued a ZTE Axon 7 mini smartphone, with a 5.2" display (1,080 × 1,920 pixels), programmed with a customized study-specific application to administer Ecological Momentary Assessment (EMA) measures and ambulatory cognitive tests. Participants were instructed to begin the first of the 5 daily EMA and cognitive testing sessions upon waking. For the following 4 sessions, the smartphone was programmed to play an audible alert to prompt the respondent to complete EMA and cognitive assessments. Alerts were programmed on a quasi-random schedule based on each individual's typical waking time, with scheduled intervals between prompts ranging between 3 and 4.5 hours (27).

Measures. *Baseline self-report measures and ambulatory assessments.* Participants completed surveys regarding demographic characteristics, medication use, and validated symptoms. Results of the additional symptoms surveys have been reported

previously (24). A study-specific smartphone app was programmed to administer EMA measures and cognitive tests in a single assessment/testing session.

SCD. Two items from the Patient-Reported Outcomes Measurement Information System applied general concerns item bank (28) were used and adapted for momentary assessment. The items "How slow is your thinking right now?" rated on a scale of 0–100 (where 0 = very fast, and 100 = very slow) and "How foggy is your thinking right now?" rated on a scale of 0–100 (where 0 = very clear, and 100 = very foggy) were averaged to produce an aggregate score, where higher scores indicate worse SCD.

Objective cognitive functioning. Participants completed a test of processing speed (symbol search) at each assessment time point. During the task, participants were shown a row of 4 symbol pairs at the top of the screen and 2 symbol pairs at the bottom of the screen. Participants were instructed to decide which symbol pair at the bottom matched a symbol pair at the top and to select the matching pair as quickly as possible by touching their response on the screen. Stimuli were presented until a response was provided. A lure stimulus wherein only 1 of the symbols in a pair matched 1 of the symbols presented at the top, but the pair did not match, was presented during 75% of the trials. Each testing session contained 16 trials. Reaction time (milliseconds) and accuracy were recorded.

Accuracy during each session was used to gauge participant effort during the symbol search task. Indiscriminate selection of responses with little or no effort would be consistent with accuracy rates of ~50%. Intentional poor performance (i.e., "faking bad") would likewise be expected to correspond with low accuracy and could be expected to play a role in cases where accuracy was <50%. To ensure adequate task engagement, accuracy of <70% was used as a conservative cutoff point to indicate poor task engagement, which is consistent with validation procedures used in the development of this task (27). The mean ± SD reaction times were calculated for each testing session. The SD reaction time was considered because within-person variability has been identified as an independent indicator of poor cognitive functioning and as a risk factor of future cognitive decline (29–31).

Mood/affect. A subset of items from the Profile of Mood States (32), adapted for use as a momentary measure, was used to assess mood/affect. Participants were prompted with, "Right now, I feel…" and rated each mood item on a 5-point scale, ranging from 0 (not at all) to 4 (extremely). Momentary depressed mood was assessed with 3 items: sad, hopeless, and discouraged. Momentary symptoms of anxiety were assessed with 3 items: anxious, on edge, and uneasy. For depressed and anxious mood, the 3 items were averaged to produce a single scale score.

Data analysis. *Preliminary analyses.* Descriptive statistics were generated for sociodemographic and study variables. Since the first day of home monitoring was a training/run-in day, data from day 1 were excluded from all analyses. Individually averaged variables for symbol search performance (mean ± SD response time), depression, and anxiety were generated by averaging each participant's scores across the assessment period. Person-centered variables for symbol search performance (mean ± SD response time), depression, and anxiety were generated by subtracting each participant's score for the assessment period (average of 16 trials for symbol search performance variables) from their individually averaged score.

Primary analyses. First, multilevel models (MLMs) tested the within-person association between momentary changes in symbol search performance and SCD. MLMs are able to model both between- and within-person variance and retain all cases (regardless of missing data within-person). Person-centered symbol search mean ± SD response times were included in separate models. Models were adjusted for individually averaged symbol search performance (to control for between-person variance), within-day time point (ordinal variable, to control for within-day variation in associations), age, and education. Next, MLMs tested momentary depression and anxiety as moderators of the within-person momentary association between symbol search performance and SCD. The models included the person-centered symbol search performance and psychological symptom (depression and anxiety) variables and interaction terms for each combination of person-centered symbol search performance variable and person-centered psychological symptom variable. Models were adjusted for individually averaged symbol search performance and psychological symptoms, time point, age, and education. Maximum likelihood estimation accounted for missing data. *P* values less than 0.05 were considered significant. All analyses were performed using SPSS version 26 software.

RESULTS

Fifty participants with FM were enrolled in and completed study activities. Descriptive statistics for sociodemographic characteristics and study variables are shown in Table 1. Participants were a mean ± SD of 44.88 ± 13.95 years old with a mean ± SD of 15.70 ± 2.03 years of education. The majority were female (88.0%) and White (86.0%).

Results of objective cognitive functioning assessments. At the within-person level, moments of slower processing speed (higher symbol search mean response time) were associated with more severe SCD ($\beta = 0.003$, $P < 0.001$) (Table 2). Additionally, moments of higher variability in processing speed (symbol search SD of response times) were associated with more severe SCD ($\beta = 0.002$, $P = 0.020$).

Analysis of effort on the ambulatory symbol search task. Accuracy on the symbol search task suggested good

Table 1. Descriptive statistics for sociodemographic characteristics and study variables in 50 individuals with fibromyalgia*

		Possible range	Observed range
Age, years	44.88 ± 13.95	–	20–70
Female sex, no. (%)	44 (88.0)	–	–
Race/ethnicity, no. (%)			
White	43 (86.0)	–	–
African American/Black	5 (10.0)	–	–
Biracial/multiracial	2 (4.0)	–	–
Education, years	15.70 ± 2.03	–	10–21
Symbol search†			
Mean response time, msec	2,444.19 ± 752.39	–	11,08.23–44,00.40
SD of response time, msec	1,027.99 ± 344.84	–	240.69–1769.45
EMA			
Subjective cognitive dysfunction	49.04 ± 16.65	0–100	3.64–94.82
Depression	0.66 ± 0.77	0–4	0.00–3.56
Anxiety	0.78 ± 0.64	0–4	0.00–2.45

* Except where indicated otherwise, values are the mean ± SD. EMA = Ecological Momentary Assessment.
† Individual averaged.

effort. Accuracy was >70% for 1,784 of 1,813 of sessions (98.4%) (range 43.75–100.00%, median 100.00, mean ± SD 95.81 ± 6.83). Eight individuals were identified as having had ≥1 session with <70% accuracy. Of these, 3 participants had multiple sessions with low accuracy (range 5–12 sessions) and were identified as possible cases of low effort. No reaction time variables were calculated for low-accuracy sessions. Sensitivity analyses, excluding the 3 participants who demonstrated repeated low accuracy/effort, were conducted for all ambulatory cognition analyses. The results with/without these 3 individuals did not change the magnitude or significance of any results. Therefore, results in the full sample are reported, aside from the several sessions with low accuracy scores.

Moderating role of momentary depression. Momentary depression significantly moderated the within-person association between momentary symbol search mean response time and SCD (β = –0.003, *P* = 0.03) (Table 3 and Figure 1A).

Table 2. Results of multilevel models testing the within-person association between momentary symbol search performance (mean and SD of response times) and subjective cognitive dysfunction in individuals with fibromyalgia*

	Estimate/β†	SE	*P*	95% CI
Subjective cognitive dysfunction				
Random effect				
Intercept	241.54	52.17	<0.0001‡	158.17, 368.84
AR1	0.22	0.03	<0.0001‡	0.16, 0.27
Residual	202.75	7.69	<0.0001‡	188.21, 218.40
Fixed effect				
Between-person variables				
Intercept	40.29	18.60	0.04‡	2.87, 77.71
Symbol search mean response time	0.01	0.004	0.03‡	0.001, 0.02
Within-person variable				
Symbol search mean response time	0.003	0.001	<0.001‡	0.001, 0.004
Subjective cognitive dysfunction				
Random effect				
Intercept	257.72	55.56	<0.0001‡	168.91, 393.22
AR1	0.22	0.03	<0.0001‡	0.16, 0.28
Residual	204.05	7.75	<0.0001‡	189.40, 219.82
Fixed effect				
Between-person variables				
Intercept	37.44	19.58	0.06	–1.96, 76.85
Symbol search response time SD	0.01	0.01	0.17	–0.005, 0.03
Within-person variable				
Symbol search response time SD	0.002	0.001	0.02‡	0.000, 0.004

* All models were adjusted for age, education, and within-day time point of the assessment. 95% CI = 95% confidence interval; AR1 = autoregressive; SCD = subjective cognitive dysfunction; SE = standard error.
† For random effects, values are the covariance parameter estimate, and for fixed effects, values are the unstandardized β value.
‡ Statistically significant.

Table 3. Results of multilevel models testing momentary depression symptoms as a moderator of the within-person association between momentary symbol search performance (mean and SD of response times) and subjective cognitive dysfunction in individuals with fibromyalgia*

	Estimate/β†	SE	*P*	95% CI
Subjective cognitive dysfunction				
Random effect				
Intercept	203.87	45.61	<0.0001‡	131.50, 316.06
AR1	0.21	0.03	<0.0001‡	0.16, 0.27
Residual	178.70	6.78	<0.0001‡	165.89, 192.50
Fixed effect				
Between-person variables				
Intercept	17.84	18.59	0.34	−19.64, 55.31
Symbol search mean response time	0.01	0.003	0.03‡	0.001, 0.01
Depression	−0.74	4.09	0.86	−8.99, 7.51
Anxiety	8.82	6.51	0.18	−4.31, 21.94
Within-person variables				
Symbol search mean response time	0.003	0.001	<0.0001‡	0.001, 0.004
Depression	6.70	0.88	<0.0001‡	4.97, 8.43
Anxiety	−0.63	0.82	0.44	−2.24, 0.97
Symbol search mean response time × depression	−0.003	0.002	0.03‡	−0.01, −0.00
Subjective cognitive dysfunction				
Random effect				
Intercept	225.06	50.20	<0.0001‡	145.35, 348.47
AR1	0.22	0.03	<0.0001‡	0.16, 0.28
Residual	179.92	6.85	<0.0001‡	166.99, 193.85
Fixed effect				
Between-person variables				
Intercept	17.12	19.75	0.39	−22.69, 56.94
Symbol search response time SD	0.01	0.01	0.35	−0.01, 0.02
Depression	−0.08	4.32	0.99	−8.78, 8.63
Anxiety	6.65	6.76	0.33	−6.98, 20.29
Within-person variables				
Symbol search response time SD	0.003	0.001	0.003‡	0.001, 0.004
Depression	6.67	0.89	<0.0001‡	4.93, 8.40
Anxiety	−0.57	0.82	0.49‡	−2.18, 1.03
Symbol search response time SD × depression	−0.005	0.002	0.02‡	−0.01, −0.001

* All models were adjusted for age, education, and within-day time point of the assessment. 95% CI = 95% confidence interval; AR1 = autoregressive; SCD = subjective cognitive dysfunction; SE = standard error.
† For random effects, values are the covariance parameter estimate, and for fixed effects, values are the unstandardized β value.
‡ Statistically significant.

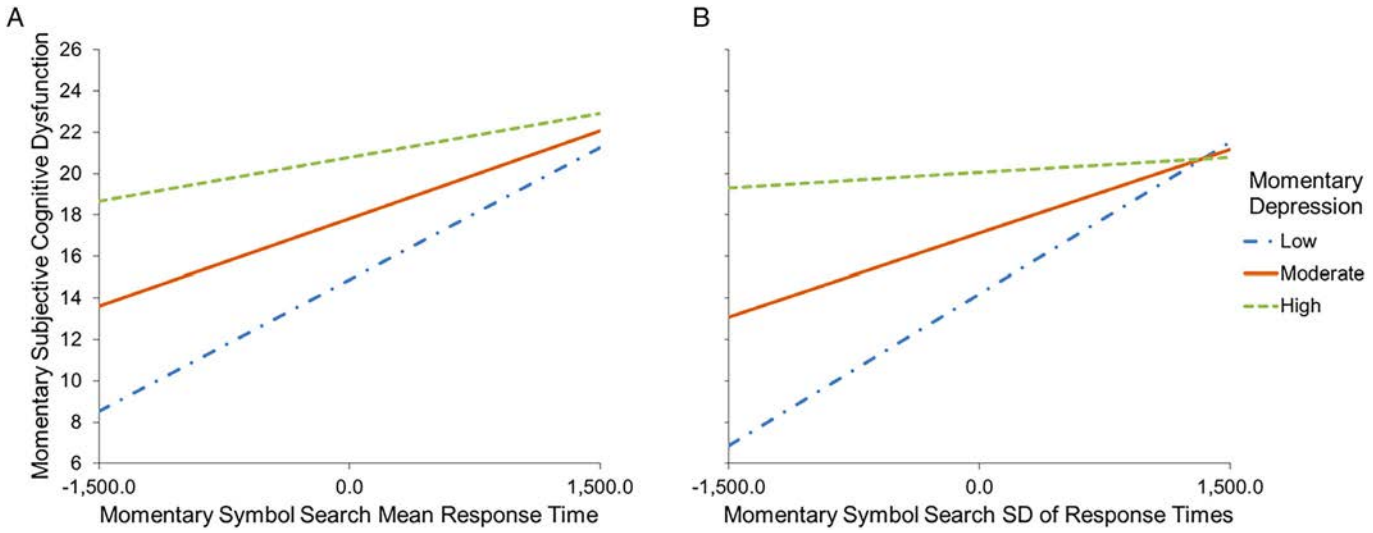

Figure 1. **A**, Simple slopes depicting momentary depression ratings as a moderator of the within-person association between momentary symbol search mean response time and subjective cognitive dysfunction. **B**, Simple slopes depicting momentary depression symptoms as moderator of the within-person association between momentary symbol search SD of response times and subjective cognitive dysfunction.

Specifically, the correspondence between moments of slower processing speed and more severe SCD was strongest when depressive symptoms were lower. In contrast, when depressive symptoms were higher than usual, the correspondence between SCD and reaction time was weaker. In moments of more severe depression symptoms, SCD was relatively high across the range of symbol search mean response times. Depression also significantly moderated the within-person association between momentary symbol search SD of response times and SCD (β = –0.005, *P* = 0.02) (Figure 1B). That is, moments of higher variability in processing speed were related to more severe SCD, but this association was weaker (smaller positive association) when depression symptoms were higher. In moments of more severe depression symptoms, SCD was relatively high regardless of symbol search SD of response times.

Moderating role of momentary anxiety. There was no significant moderating effect of momentary anxiety ratings on the association between momentary symbol search mean response time and SCD (*P* = 0.34). However, momentary anxiety ratings significantly moderated the within-person association between symbol search SD of response time and SCD (β = –0.004, *P* = 0.02) (Table 4 and Figure 2). Specifically, moments of higher variability in processing speed were related to more severe SCD, but this association was weaker (smaller positive association) when anxiety symptoms were higher.

DISCUSSION

SCD is prominent in FM, and thus far our understanding of the factors influencing the relationship between SCD and objective cognitive performance is limited. This is the first study to use a microlongitudinal design to assess the moderating role of momentary level of depression and anxiety on the association between SCD and processing speed performance in FM.

Table 4. Results from multilevel models testing whether momentary anxiety symptoms moderate the within-person association between momentary symbol search performance (mean and SD of response times) and subjective cognitive dysfunction in individuals with fibromyalgia*

	Estimate/β†	SE	*P*	95% CI
Subjective cognitive dysfunction				
Random effect				
Intercept	203.99	45.63	<0.0001‡	131.58, 316.24
AR1	0.21	0.03	<0.0001‡	0.15, 0.27
Residual	178.99	6.79	<0.0001‡	166.17, 192.80
Fixed effect				
Between-person variables				
Intercept	18.12	18.59	0.34	–19.36, 55.61
Symbol search mean response time	0.01	0.003	0.03‡	0.001, 0.01
Depression	–0.73	4.09	0.86	–8.98, 7.51
Anxiety	8.83	6.51	0.18	–4.30, 21.96
Within-person variables				
Symbol search mean response time	0.003	0.001	<0.0001‡	0.002, 0.004
Depression	6.69	0.89	<0.0001‡	4.95, 8.42
Anxiety	–0.52	0.82	0.52	–2.13, 1.08
Symbol search mean response time × anxiety	–0.001	0.001	0.34	–0.004, 0.001
Subjective cognitive dysfunction				
Random effect				
Intercept	225.70	50.33	<0.0001‡	145.78, 349.42
AR1	0.22	0.03	<0.0001‡	0.16, 0.27
Residual	179.62	6.82	<0.0001‡	166.73, 193.50
Fixed effect				
Between-person variables				
Intercept	17.65	19.77	0.38	–22.22, 57.52
Symbol search response time SD	0.01	0.01	0.35	–0.01, 0.02
Depression	–0.06	4.32	0.99	–8.77, 8.66
Anxiety	6.53	6.77	0.34	–7.12, 20.18
Within-person variables				
Symbol search response time SD	0.003	0.001	0.003‡	0.001, 0.004
Depression	6.63	0.89	<0.0001‡	4.89, 8.37
Anxiety	–0.39	0.82	0.64	–2.00, 1.22
Symbol search response time SD × anxiety	–0.004	0.002	0.02‡	–0.01, –0.001

* All models were adjusted for age, education, and within-day time point of the assessment. 95% CI = 95% confidence interval; AR1 = autoregressive; SCD = subjective cognitive dysfunction; SE = standard error.
† For random effects, values are the covariance parameter estimate, and for fixed effects, values are the unstandardized β value.
‡ Statistically significant.

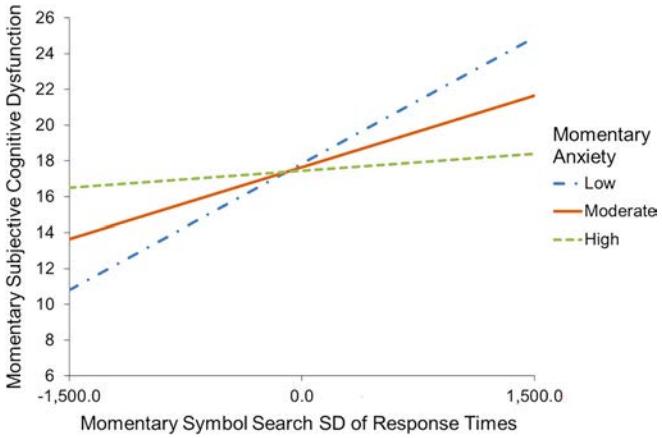

Figure 2. Simple slopes depicting momentary depression symptoms as a moderator of the within-person association between momentary symbol search SD of response times and subjective cognitive dysfunction. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25086/abstract.

The highest correspondence between SCD and processing speed performance (mean ± SD reaction time) occurred when symptoms of depression and anxiety were at their lowest, suggesting that individuals were able to gauge their cognitive performance more accurately when affective symptoms were minimally influencing their self-perception. However, when momentary ratings of depression and anxiety were higher, there was an increased discrepancy between SCD and mental processing speed as well as more variability in performance, supporting our hypothesis regarding the moderating influence of depression and anxiety on the association between SCD and objective cognitive performance.

Landro et al (16) studied a group of individuals with chronic nonmalignant pain and found that self-reports of cognitive functioning were largely consistent with objective neuropsychological assessment. In contrast, Tesio et al found that while self-perception and objective performance correlated in some domains (e.g., working memory), there were poor correlations in the majority of cognitive domains assessed (33). Our findings expand on that previous research in several important ways. First, a key difference is in our study design, which used multiple momentary assessments rather than a single assessment point, allowing for examination of within-person changes. Second, our momentary assessments captured the changes in state experienced by participants occurred in "real time," and are not as subject to recall biases as other measures that rely on recollection of cognitive symptoms over the past 7 days (33). Finally, rather than relying on a broad measure of SCD, we selected 2 specific questions, one which corresponded directly to the processing speed task at hand (i.e., "How slow is your thinking right now?") and another that represented a general cognitive complaint (i.e., "How foggy is your thinking?").

These findings can be interpreted within the context of theoretical frameworks that suggest that having symptoms of depression can make a person more prone to certain cognitive biases that affect perceptions of self, including perceived cognitive functioning (20–22). Although this study did not study depressed individuals, research in individuals with depression has shown that more severe depressive symptoms are related to underestimation of cognitive abilities due to mood-related biases such as negative self-schemas and negative perceptions of thoughts and behaviors (34). Evidence that depressive symptoms play a role in cognitive biases related to perceived cognitive functioning is further supported by research showing that after depressive symptoms have remitted, individuals tend to overestimate their own cognitive abilities (35). This framework aligns well with our findings showing that SCD becomes more disparate from objective functioning when depressive symptoms are greater, supporting the notion that depression may influence perception of cognitive performance within the domain of mental processing speed (17). Because we have only considered processing speed in this study, it will be important for future work to determine whether heightened depressive symptoms also influence perception of other cognitive abilities.

Anxiety can also negatively impact processing speed and self-perception of cognitive abilities; however, the influence of anxiety on the correspondence between SCD and objective performance may be somewhat different than that found in depression. In one of the few studies examining the moderating effects of mood symptoms on the relationship between SCD and cognitive performance, Baker et al (36) found that individuals with chronic pain who reported more severe symptoms of anxiety demonstrated better correspondence between SCD and objective performance, compared to individuals with milder symptoms. Other research suggests that when anxiety symptoms are severe, there is a negative impact on processing speed performance; however, anxiety can actually prove beneficial for rapid responding when symptoms are mild (37).

While individual's experience of mood symptoms and psychological distress are often considered to be primary factors in self-perceived cognitive difficulties, it will be important to differentiate the different effects depression and anxiety may have on the subjective/objective cognitive discrepancy. Future research should also seek to determine potential methods for determining at what point an individual's mood symptoms may be the primary factor influencing cognition beyond other FM symptoms. Then, focused intervention may be developed and administered for those individuals identified to be at the highest risk of developing cognitive symptoms to help avoid the negative impact these symptoms may have on daily functioning.

These factors should be considered in the context of treatment of cognitive and psychological symptoms in FM and suggest that focused treatment on mood symptoms may lead to a more accurate self-appraisal of cognitive functioning. However, longitudinal data are needed to support this hypothesis. Cognitive

behavioral therapy has been demonstrated to be effective in reducing both pain catastrophizing and pain severity and has effects on brain functioning associated with these symptoms (38). Thus, a more individually tailored approach to such an intervention that capitalizes on the daily experience of the individual may help alleviate aspects of SCD. Utilizing an EMA-informed approach to interventions such as these could lead to the identification of what emotional states, life events, or diurnal factors most strongly impact variability in SCD and cognitive performance in FM.

This research has several limitations that potentially limit the generalizability of our findings. Despite the various cognitive difficulties reported by individuals with FM (9), our objective assessment was limited to mental processing speed. Processing speed was selected since it is often impaired in depression and anxiety and represents a foundational cognitive skill underlying most other cognitive functions, and therefore is likely to be sensitive to momentary fluctuations. Additionally, we attempted to achieve the highest possible concordance between descriptions used in our subjective ratings and the cognitive domain assessed (i.e., "How slow is your thinking right now?"). Future research will be strengthened by incorporating assessments of multiple cognitive domains to extend these findings. Given that our primary research questions focused on the moderating effects of mood symptoms, we did not evaluate the effects of momentary pain in our analyses. However, recent population-level data suggests that pain and mood symptoms may differentially influence cognitive symptoms (5). Nonetheless, it will remain important to understand these variables and their various interactions in order to develop a more thorough understanding of the myriad factors influencing cognitive appraisal in FM.

Finally, it is important to consider task engagement and effort when interpreting performance on cognitive assessments. While this study did not include a standalone measure of effort or performance validity, we instituted a conservative cutoff score (70% accuracy) as a means of detecting poor effort or engagement on the processing speed task. Further, trials which did not meet that cutoff were not analyzed for reaction time or included in analysis. Sensitivity analyses also determined that there was no significant impact of including the "valid" trials on an individual's accuracy (trials that were >70%) who may have had several trials below our designated cutoff for accuracy. Future research will benefit from including standalone measures to ensure adequate effort on task performance. However, in a research context, individuals with FM typically perform within normal limits on standalone measures of task engagement, suggesting that when disability or other medicolegal aspects are not involved, there may be less concern regarding performance validity in the context of FM research (39). Finally, our own data comparing individuals with and those without FM demonstrated that rates of instances of poor accuracy were nearly identical in the 2 groups, suggesting that individuals with FM do not demonstrate higher rates of poor effort on the specific tests used in this study (24).

In conclusion, these data suggest that individuals with FM can formulate a more realistic self-appraisal of their objective cognitive ability when symptoms of depression and anxiety are less prominent. This highlights the importance of thorough mental health assessment of psychological symptoms during evaluation of SCD in FM. Identifying mood symptoms during routine clinical care may assist patients with accessing necessary, cost-effective interventions. Research in this area will likely continue to benefit from studying SCD using within-person study designs to adequately account for the numerous interactions among the comorbid symptoms observed in FM.

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. Kairys 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. Kratz.

Acquisition of data. Kratz.

Analysis and interpretation of data. Kairys, Valentine, Whibley, Kratz.

REFERENCES

1. Wolfe F, Smythe HA, Yunus MB, et al. The American College of Rheumatology 1990 criteria for the classification of fibromyalgia. Arthritis Rheum 1990;33:160–72.
2. Bennett RM, Jones J, Turk DC, et al. An internet survey of 2,596 people with fibromyalgia. BMC Musculoskeletal Disord 2007;8:27.
3. Lawrence RC, Felson DT, Helmick CG. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Arthritis Rheum 2008;58:26–35.
4. Kleykamp BA, Ferguson MC, McNicol E, et al. The prevalence of psychiatric and chronic pain comorbidities in fibromyalgia: an ACTTION systematic review. Semin Arthritis Rheum 2021;51:166–74.
5. Wolfe F, Rasker JJ, Ten Klooster P, et al. Subjective cognitive dysfunction in patients with and without fibromyalgia: prevalence, predictors, correlates, and consequences. Cureus 2021;13:e20351.
6. Katz RS, Heard AR, Mills M, et al. The prevalence and clinical impact of reported cognitive difficulties (fibrofog) in patients with rheumatic disease with and without fibromyalgia. J Clin Rheumatol 2004;10:53–8.
7. Glass JM, Park DC. Cognitive dysfunction in fibromyalgia. Curr Rheumatol Rep 2001;3:123–7.
8. Arnold LM, Crofford LJ, Mease PJ, et al. Patient perspectives on the impact of fibromyalgia. Patient Educ Couns 2008;73:114–20.
9. Glass JM. Cognitive dysfunction in fibromyalgia syndrome. J Musculoskelet Pain 2010;18:367–72.
10. Kravitz HM, Katz RS. Fibrofog and fibromyalgia: a narrative review and implications for clinical practice. Rheumatol Int 2015;35:1115–25.
11. Williamson J, Larner AJ. Cognitive dysfunction in patients with fibromyalgia. Br J Hospital Med (Lond) 2016;77:116.
12. Gelonch O, Garolera M, Valls J, et al. Cognitive complaints in women with fibromyalgia: are they due to depression or to objective cognitive dysfunction? J Clin Exp Neuropsychol 2017;39:1013–25.
13. Bell T, Trost Z, Buelow MT, et al. Meta-analysis of cognitive performance in fibromyalgia. J Clin Exp Neuropsychol 2018;40:698–714.

14. Pidal-Miranda M, González-Villar AJ, Carrillo-de-la-Peña MT, et al. Broad cognitive complaints but subtle objective working memory impairment in fibromyalgia patients. PeerJ 2018;6:e5907.
15. Suhr JA. Neuropsychological impairment in fibromyalgia: relation to depression, fatigue, and pain. J Psychosom Res 2003;55:321–9.
16. Landrø NI, Fors EA, Våpenstad LL, et al. The extent of neurocognitive dysfunction in a multidisciplinary pain centre population. Is there a relation between reported and tested neuropsychological functioning? Pain 2013;154:972–7.
17. Walitt B, Čeko M, Khatiwada M, et al. Characterizing "fibrofog": subjective appraisal, objective performance, and task-related brain activity during a working memory task. Neuroimage Clin 2016;11:173–80.
18. Gelonch O, Garolera M, Valls J, et al. Executive function in fibromyalgia: comparing subjective and objective measures. Compr Psychiatry 2016;66:113–22.
19. Galvez-Sánchez CM, Muñoz Ladrón de Guevara C, Montoro CI, et al. Cognitive deficits in fibromyalgia syndrome are associated with pain responses to low intensity pressure stimulation. PloS One 2018;13: e0201488.
20. Gotlib IH, Joormann J. Cognition and depression: current status and future directions. Annu Rev Clin Psychol 2010;6:285–312.
21. Szu-Ting Fu T, Koutstaal W, Poon L, et al. Confidence judgment in depression and dysphoria: the depressive realism vs. negativity hypotheses. J Behav Ther Exp Psychiatry 2012;43:699–704.
22. Robinson O, Vytal K, Cornwell B, et al. The impact of anxiety upon cognition: perspectives from human threat of shock studies. Front Hum Neurosci 2013;7:203.
23. Wolfe F, Clauw DJ, Fitzcharles MA, et al. 2016 revisions to the 2010/2011 fibromyalgia diagnostic criteria. Semin Arthritis Rheum 2016;46:319–29.
24. Kratz AL, Whibley D, Kim S, et al. Fibrofog in daily life: an examination of ambulatory subjective and objective cognitive function in fibromyalgia. Arthritis Care Res (Hoboken) 2020;72:1669–77.
25. Kratz AL, Whibley D, Kim S, et al. The role of environmental distractions in the experience of fibrofog in real-world settings. ACR Open Rheumatol 2020;2:214–21.
26. Whibley D, Williams DA, Clauw DJ, et al. Within-day rhythms of pain and cognitive function in people with and without fibromyalgia: synchronous or syncopated? Pain 2022;163:474–82.
27. Sliwinski MJ, Mogle JA, Hyun J, et al. Reliability and validity of ambulatory cognitive assessments. Assessment 2018;25:14–30.
28. Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Med Care 2007;45 Suppl 1:S3–11.
29. Cerino ES, Katz MJ, Wang C, et al. Variability in cognitive performance on mobile devices is sensitive to mild cognitive impairment: results from the Einstein Aging Study. Front Digit Health 2021;3:758031.
30. Strauss E, Bielak AA, Bunce D, et al. Within-person variability in response speed as an indicator of cognitive impairment in older adults. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2007; 14:608–30.
31. Hultsch DF, MacDonald SW, Dixon RA. Variability in reaction time performance of younger and older adults. J Gerontol B Psychol Sci Soc Sci 2002;57:P101–15.
32. McNair DM, Lorr M, Droppleman LF. Revised manual for the profile of mood states. San Diego (CA): Educational Industrial Testing Service; 1992.
33. Tesio V, Torta DM, Colonna F, et al. Are fibromyalgia patients cognitively impaired? Objective and subjective neuropsychological evidence. Arthritis Care Res (Hoboken) 2015;67:143–50.
34. Baeza-Velasco C, Guillaume S, Olié E, et al. Decision-making in major depressive disorder: subjective complaint, objective performance, and discrepancy between both. J Affect Disord 2020;270:102–7.
35. Petersen JZ, Porter RJ, Miskowiak KW. Clinical characteristics associated with the discrepancy between subjective and objective cognitive impairment in depression. J Affect Disord 2019;246:763–74.
36. Baker KS, Gibson SJ, Georgiou-Karistianis N, et al. Relationship between self-reported cognitive difficulties, objective neuropsychological test performance and psychological distress in chronic pain. Eur J Pain 2018;22:601–13.
37. Bierman EJ, Comijs HC, Jonker C, et al. Effects of anxiety versus depression on cognition in later life. Am J Geriatr Psychiatry 2005; 13:686–93.
38. Lazaridou A, Kim J, Cahalan CM, et al. Effects of Cognitive-Behavioral Therapy (CBT) on brain connectivity supporting catastrophizing in fibromyalgia. Clin J Pain 2017;33:215–21.
39. Iverson GL, Le Page J, Koehler BE, et al. Test of Memory Malingering (TOMM) scores are not affected by chronic pain or depression in patients with fibromyalgia. Clin Neuropsychol 2007;21:532–46.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 75, No. 9, September 2023, pp 1976–1985
DOI 10.1002/acr.25088

Development and Validation of a Simulation Model for Treatment to Maintain Remission in Antineutrophil Cytoplasmic Antibody–Associated Vasculitis

Zachary S. Wallace,[1] John H. Stone,[1] Xiaoqing Fu,[2] Peter A. Merkel,[3] Eli M. Miloslavsky,[1] Yuqing Zhang,[1] Hyon K. Choi,[1] and Emily P. Hyle[1]

Objective. Fixed and tailored rituximab retreatment strategies to maintain remission in antineutrophil cytoplasmic antibody (ANCA)–associated vasculitis (AAV) are associated with tradeoffs. The current study was undertaken to develop a simulation model (AAV-Sim) to project clinical outcomes with these strategies.

Methods. We developed the AAV-Sim, a microsimulation model of clinical events among individuals with AAV initiating treatment to maintain remission. Individuals transition between health states of remission or relapse and are at risk for severe infection, end-stage renal disease, or death. We estimated transition rates from published literature, stratified by individual-level characteristics. We performed validation using the mean average percent error (MAPE) and the coefficient of variation of root mean square error (CV-RMSE). In internal validation, we compared model-projected outcomes over 28 months with outcomes observed in the Rituximab versus Azathioprine in ANCA-Associated Vasculitis 2 (MAINRITSAN2) trial, which compared fixed versus tailored retreatment. In external validation, we compared outcomes with fixed rituximab retreatment from the AAV-Sim to outcomes from the MAINRITSAN1 trial and an observational study.

Results. The AAV-Sim projected outcomes similar to those in the MAINRITSAN2 trial, including minor (AAV-Sim 6.0% fixed versus 7.3% tailored; MAINRITSAN2 6.2% versus 8.6%; MAPE 3% and 15%) and major relapse (AAV-Sim 3.5% versus 5.5%; MAINRITSAN2 3.7% versus 7.4%; MAPE 5% and 26%), severe infection (AAV-Sim 19.4% versus 11.1%; MAINRITSAN2 19.8% versus 10.2%; MAPE 2% and 9%), and relapse-free survival (AAV-Sim 84.8% versus 82.3%; MAINRITSAN2 86% versus 84%; CV-RMSE 2.3% and 2.5%). Similar performance was observed in external validation.

Conclusion. The AAV-Sim projected a range of clinical outcomes for different treatment approaches that were validated against published data. The AAV-Sim has the potential to inform management guidelines and research priorities.

INTRODUCTION

Treatment of antineutrophil cytoplasmic antibody (ANCA)–associated vasculitis (AAV) has evolved substantially in recent decades. However, given the rarity of AAV and competing research priorities, clinical trials and large observational studies are unlikely to address all current and future knowledge gaps regarding treatment strategies. For example, rituximab is one of the most frequently used medications to maintain remission (1) and is recommended over alternative strategies by recent guidelines (2). Additionally, fixed rituximab retreatment is conditionally recommended over retreatment guided by ANCA titers or CD19+ B cell counts (e.g., tailored strategy) (2). This latter recommendation is based, in part, on the Rituximab versus Azathioprine in ANCA-Associated Vasculitis 2 (MAINRITSAN2) trial, in which fixed rituximab retreatment had a trend toward a reduced relapse rate. However, fixed versus tailored retreatment also had a trend toward a higher rate of serious infection (3). Overall, the levels of evidence supporting these recommendations range from very low to moderate (2), and additional data could strengthen guideline recommendations and inform clinical trial development.

Dr. Wallace's work was supported by the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases grants K23-AR-073334 and R03-AR-078938) and the Rheumatology Research Foundation (K supplement).

[1]Zachary S. Wallace, MD, MSc, John H. Stone, MD, MPH, Eli M. Miloslavsky, MD, Yuqing Zhang, DSc, Hyon K. Choi, MD, DrPH, Emily P. Hyle, MD, MSc: Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts; [2]Xiaoqing Fu, MS: Massachusetts General Hospital, Boston; [3]Peter A. Merkel, MD, MPH: University of Pennsylvania, Philadelphia.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.25088&file=acr25088-sup-0001-Disclosureform.pdf.

Address correspondence via email to Zachary S. Wallace, MD, MSc, at zswallace@mgh.harvard.edu.

Submitted for publication June 21, 2022; accepted in revised form January 10, 2023.

SIGNIFICANCE & INNOVATIONS

- Treatment of antineutrophil cytoplasmic antibody–associated vasculitis (AAV) has evolved substantially in recent years with multiple strategies that are associated with tradeoffs in outcomes now available for maintenance of remission.
- Simulation models can produce scientific evidence that is complementary to data from clinical trials and cohort studies and can be used by providers and other stakeholders but have not been robustly developed for AAV.
- We developed and validated a novel state-transition microsimulation model, the AAV simulation model, to project clinical outcomes among individuals with AAV in remission.

Strategies to maintain remission, including fixed versus tailored rituximab retreatment, are associated with tradeoffs regarding risks of AAV relapse and severe infection, which affect morbidity, mortality, quality of life, and costs. Simulation models can produce scientific evidence that is complementary to data from clinical trials and cohort studies and can be used by providers and other stakeholders (4). Simulation modeling has important advantages: data from multiple sources can be incorporated as input parameters, model structure and inputs can be revised in real time as new data and strategies emerge, and model outcomes can account for various clinical outcomes, quality of life, and costs. Additionally, uncertain input parameters may be varied to evaluate the impact of estimate uncertainty on outcomes. Finally, models can project outcomes over longer time horizons than typical clinical trials or observational studies. In other conditions, simulation models have informed care and treatment guidelines (4,5).

Therefore, simulation models will be useful to address unmet needs for AAV care. Models can be applied to identify patient subgroups (e.g., by ANCA type, manifestation, genetic profile) or scenarios (e.g., a pandemic in which infection risks and vaccination needs impact preferred strategies) in which one strategy may yield superior outcomes. Additionally, models can be used to identify influential areas of uncertainty where research funding can generate new data to inform decisions that may have a substantial impact on practice. Simulation models may have particular value in the study of rare diseases where both funding and trial sample sizes are limited compared to common diseases.

The objective of the current project was to develop and validate a novel state-transition microsimulation model, the AAV simulation model (AAV-Sim), to project clinical outcomes among individuals with AAV in remission. For the purposes of model development and validation, we simulated rituximab-based retreatment strategies to maintain remission with the capacity to simulate additional strategies in future studies.

MATERIALS AND METHODS

Analytic overview. We developed the AAV-Sim model (using TreeAge Pro Healthcare 2020 software) to project clinical outcomes (i.e., relapse, infection, end-stage renal disease [ESRD], death) among patients with AAV who were in their first remission after induction therapy. Clinical experts contributed to the development of the model structure and assessed its face validity. We derived model input parameters from cohort studies, clinical trials, and nationwide registries. We verified model structure, assessed the face validity of input parameters, and examined model outcomes using data from the MAINRITSAN2 trial (3), which compared fixed with tailored rituximab retreatment guided by serum ANCA and blood CD19+ B cell test results. We then performed 2 external model validations by populating the model with cohort characteristics from the MAINRITSAN1 trial (fixed rituximab retreatment versus azathioprine for maintenance of remission) (6) and an observational study reporting outcomes of long-term B cell depletion in AAV (7).

Model structure. The AAV-Sim model is an individual-level, microsimulation model with a monthly time step (Figure 1). Certain assumptions are necessary when modeling a complex condition but can be varied and updated as new data become available. At model start, all individuals are in remission and draw for demographic (i.e., age, sex) and disease-specific characteristics (i.e., renal involvement [yes/no], ANCA type [proteinase 3 (PR3)– or myeloperoxidase (MPO)–ANCA]) from user-defined distributions; we selected these individual-level factors given their impact on outcomes (e.g., ESRD, relapse, mortality risk). We defined renal involvement as a history of AAV-related renal disease with an estimated glomerular filtration rate (eGFR) of ≥45 ml/minute because a tailored strategy may not be appealing in patients with severe renal disease (see Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25088). Each month, individuals can transition between remission or relapse, stratified by major/minor severity, and are at risk each month for severe infection, ESRD, or death.

We defined major relapse as organ- or life-threatening disease. The risk of relapse in the model is influenced by a patient's ANCA type and treatment strategy (e.g., fixed versus tailored treatment). Individuals remain in the major relapse health state for up to 6 months; they can exit temporarily for 1 month (due to severe infection) or permanently (due to ESRD or death). At 6 months (inclusive of any temporary exits for infection), they return to the remission health state. Individuals can experience a 2-month minor relapse (6,8), during which they face a monthly risk of ESRD, severe infection, or death but not major relapse to reflect the way these data were reported in clinical trials. After 2 months in minor relapse, they return to remission and are again at risk for relapse.

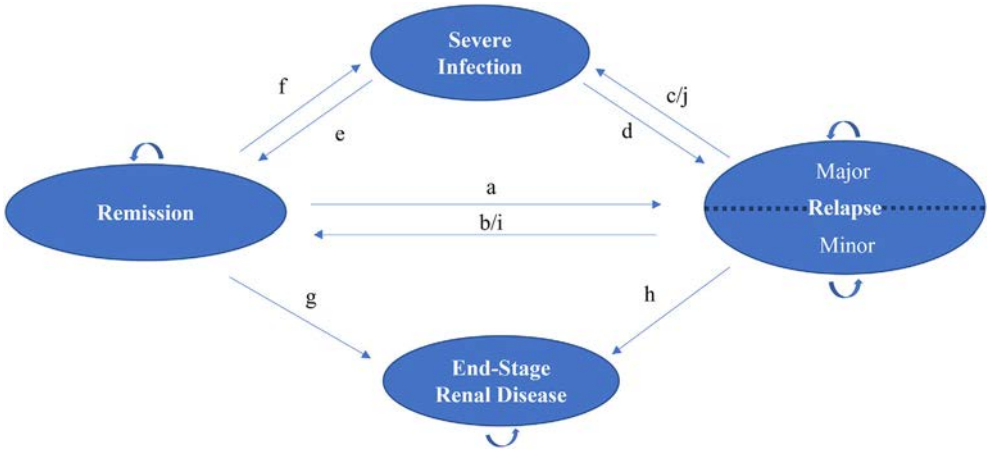

Figure 1. Schematic of health states and transitions in the antineutrophil cytoplasmic antibody–associated vasculitis simulation model (AAV-Sim). This figure represents a simplified depiction of health states and transitions (shown in sequence by lower-case letters) in the AAV-Sim. The ovals represent health states: relapse (stratified by major and minor relapse); remission; end-stage renal disease; and severe infection. All individuals are in remission at model start. The arrows indicate monthly transitions by which individuals can progress to a different state or remain in the same health state at each monthly time step. Individuals spend 6 months in relapse after experiencing a major relapse; they then return to remission and are again at risk for major or minor relapse. Individuals spend 2 months in relapse after experiencing a minor relapse, after which they return to remission and are again at risk for major or minor relapse. From remission, individuals can experience a relapse and transition to the relapse health state. Once individuals develop end-stage renal disease, they remain in that health state until death. Death (not pictured) can occur from any health state. Transitions a, c, j, and f are treatment strategy dependent.

We defined severe infection as an infection requiring hospitalization, and ESRD as permanent renal replacement therapy (i.e., hemo- or peritoneal dialysis, or renal transplant). The monthly risk of ESRD is stratified according to whether a patient has renal involvement at model start or not. Individuals without renal involvement at model start do not develop renal manifestations of AAV in the future based on prior data indicating that this is rare (9,10) and would be unlikely to strongly influence outcomes in the modeled 28-month follow-up that patients enter with preserved renal function. Individuals with ESRD do not experience major or minor relapse based on prior literature suggesting that this is infrequent (11). Each health state has a distinct monthly mortality rate.

Cohort characteristics and AAV natural history model input parameters. The characteristics of the cohort reflect the distributions of age (mean 60.6 years), sex (58% male), PR3-ANCA+ (58%), and renal involvement (71%) in the MAINRITSAN2 trial (3) (Table 1). The monthly probability of relapse and severe infection during remission are strategy dependent (Table 1). The monthly probability of severe infection during major relapse is based on results from the RITAZAREM trial (12), which enrolled relapsing patients (Table 1 and Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25088). The monthly probability of ESRD among individuals with prior renal involvement during major relapse was derived from the Mass General Brigham (MGB) AAV cohort and published literature (Table 1 and Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25088) (13).

The AAV-Sim simulates individuals who have achieved remission and have an eGFR of ≥45 ml/minute for which data are limited regarding ESRD risk. Among individuals with and without a history of renal involvement, we assumed that the monthly probability of ESRD during remission and major relapse is age dependent, derived from the MGB AAV cohort and rates of ESRD due to non-glomerulonephritis causes observed in the general population, respectively (see Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25088) (14–16). ESRD risk during minor relapse is the same as during remission among those with and without renal involvement at model start.

Although AAV cohort studies have demonstrated higher mortality among patients with AAV than general population comparators (17), data are limited regarding the risk of noninfectious and non-ESRD death among patients with AAV who survived their initial remission induction treatment. We used data from the MGB AAV cohort and National Vital Statistics report to determine age- and sex-specific monthly probabilities of death (see Supplementary Appendix A). We derived distinct age- and sex-specific mortality rates for patients with ESRD from the US Renal Data System (14–16) and for patients hospitalized with severe infection from the Nationwide Inpatient Sample (see Supplementary Appendix A) (18).

Strategies. Using the AAV-Sim, we can project the outcomes associated with different strategies to maintain remission. The monthly probability of minor and major relapse and severe infection during remission vary based on the strategy: fixed (base case strategy) or tailored (alternative strategy) retreatment. In the

Table 1. Model input parameters*

Input parameter	Base case value	Fixed schedule	Tailored treatment	Ref.
Baseline cohort characteristics				3
Age, mean ± SD years	60.5 ± 13.0	–	–	
Male, %	58	–	–	
PR3-ANCA, %	58	–	–	
Renal involvement, %	72	–	–	
Relapse, monthly probability				3
PR3-ANCA major relapse	–	0.0017	0.00351	
PR3-ANCA minor relapse	–	0.0029	0.0047	
MPO-ANCA major relapse	–	0.0009	0.0017	
MPO-ANCA minor relapse	–	0.0015	0.0024	
Severe infection, monthly probability				
Major relapse	–	0.00944	0.00944	12
Remission and minor relapse	–	0.00785	0.00419	3
ESRD, monthly probability				
Major relapse				
Prior renal involvement	0.000851	–	–	†
No prior renal involvement	Age-stratified‡	–	–	14
Remission and minor relapse				
Prior renal involvement	Age-stratified‡	–	–	†
No prior renal involvement	Age-stratified‡	–	–	14
Mortality, monthly probability				
Major relapse	Age- and sex-stratified‡	–	–	42
Remission and minor relapse	Age- and sex-stratified‡	–	–	42
End-stage renal disease	Age- and sex-stratified‡	–	–	14
Severe infection	Age-stratified‡	–	–	18

* ANCA = antineutrophil cytoplasmic antibody; ESRD = end-stage renal disease; MPO = myeloperoxidase; PR3 = proteinase 3; Ref. = reference citation.
† Details regarding source data are available in Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25088.
‡ Monthly probabilities are reported in the Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25088.

fixed strategy, individuals receive rituximab every 6 months. In the tailored strategy, individuals are retreated with rituximab when they experience a rise in their ANCA titer or a repopulation of CD19+ B cells. Therefore, in the model, individuals who do not have a relapse face a monthly probability of experiencing a repopulation in their B cell count (i.e., any count >10 when the prior test had no detectable B cell count), ANCA turning from negative to positive (i.e., any titer ≥20), or a doubling of their ANCA titer, if previously detectable. If the patient experiences any of these events, they receive rituximab and enter a post-rituximab remission health state for 3 months, during which they face a probability of ESRD, severe infection, or death without a probability of relapse to reflect the decreased risk of relapse immediately following rituximab administration; the probability of experiencing these outcomes during these 3 months is based on their assigned treatment strategy and baseline demographic and disease features.

Strategy-specific model input parameters. Based on estimates that patients with PR3-ANCA+ AAV have a 2-fold higher risk of relapse than MPO-ANCA+ AAV (19–22), we derived PR3- and MPO-ANCA–specific probabilities of minor and major relapse from the MAINRITSAN2 trial (Table 1 and Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25088). We estimated strategy-specific monthly probabilities of severe infection during remission and relapse using data from the MAINRITSAN2 trial (3). Because minor relapses in trials are often treated with only small increases in immunosuppression (8,23), we assumed the risk of severe infection to be equal in minor relapse and remission.

In the tailored strategy, retreatment with rituximab is determined by changes in B cell counts or ANCA titers, as described above. To estimate the probability of B cell repopulation or ANCA titer changes, we used data from patients randomized to rituximab in the Rituximab in AAV (RAVE) trial (see Supplementary Appendix A) (24).

Model outcomes. We projected 6 primary outcomes using the AAV-Sim: 1) minor relapse; 2) major relapse; 3) relapse-free survival; 4) severe infection; 5) ESRD; and 6) all-cause mortality. Secondary outcomes included: 1) number of rituximab infusions triggered by a rising ANCA level or B cell repopulation; 2) cause-specific death; 3) ANCA titer rise; and 4) B cell repopulation. Table 2 shows outcomes assessed in validation.

Internal and external validation. For internal validation, we used the AAV-Sim to project outcomes over 28 months using fixed and tailored strategies among patients in remission

Table 2. Primary outcomes projected in the antineutrophil cytoplasmic antibody (ANCA)–associated vasculitis simulation model (AAV-Sim) and observed in clinical trials and cohorts used for validation of the AAV-Sim*

Outcome assessed	Internal validation, MAINRITSAN 2 (time horizon 28 months)	External validation†	
		MAINRITSAN 1 (time horizon 28 months)	Pendergraft et al (time horizon 12 months)
Minor relapse	✔	✔	✘
Major relapse	✔	✔	✘
Severe infection	✔	✔	✘
ESRD	✔	✘	✘
Death	✔	✔	✘
Relapse-free survival	✔	✔	✔

* ESRD = end-stage renal disease; MAINRITSAN = Rituximab versus Azathioprine in ANCA-Associated Vasculitis (trial). ✔ = reported in trial/study; ✘ = not assessed or reported in trial/study.
† MAINRITSAN1 and Pendergraft et al (7) were not used to estimate input parameters.

at baseline (Table 1). We simulated a population of 50,000 individuals with AAV with clinical characteristics (i.e., age, sex, ANCA type, and renal involvement) of patients enrolled in the MAINRITSAN2 trial (Table 1 and Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25088). In the MAINRITSAN2 trial, rituximab was not administered after month 18. Therefore, for the purposes of external validation, we assumed that individuals with a rising ANCA titer or B cell repopulation had a 2-fold higher risk of flare once rituximab was held (see Supplementary Appendix A).

We performed 2 external validations of our model using demographic characteristics, disease-specific features, and study design from the MAINRITSAN1 trial and the study by Pendergraft et al (7) to compare projected outcomes to those observed in these studies (see Supplementary Appendix A) (6,7). We also conducted a 2-way sensitivity analysis to evaluate the impact of age and risk of infection on relapse-free survival (see Supplementary Appendix A).

Statistical analyses. We used mean average percent error (MAPE) to assess how the cumulative incidence of model-projected outcomes compared with observed outcomes in the MAINRITSAN2 and MAINRITSAN1 trials (see Supplementary Appendix A). To compare relapse-free survival curves for model-projected outcomes with those observed in MAINRITSAN 2, MAINRITSAN1, and the observational study by Pendergraft et al (7), we used the coefficient of variation of root mean square error (CV-RMSE) (see Supplementary Appendix A).

RESULTS

Model outcomes. We first evaluated the outcomes projected by the AAV-Sim over 28 months among patients treated with a fixed strategy (Table 3). At 28 months, 6.0% of patients in the fixed strategy would experience at least 1 minor relapse, and 3.5% would experience at least 1 major relapse. We projected that at least 1 severe infection would occur in 19.4% of individuals treated with the fixed strategy in the AAV-Sim. Given the exclusion of individuals with significant kidney disease, we projected the incidence of ESRD to be rare (0.2%). We projected that 5.9% of patients would die and 84.8% would attain relapse-free survival with the fixed strategy over the 28-month period. We also projected cause-specific death: noninfectious causes in patients without ESRD (4.3%), severe infection (1.6%), and ESRD

Table 3. Internal validation comparing antineutrophil cytoplasmic antibody (ANCA)–associated vasculitis simulation model (AAV-Sim) projected outcomes with those observed in the MAINRITSAN2 trial over 28 months according to treatment strategy*

	Fixed strategy			Tailored strategy		
	AAV-Sim, %	MAINRITSAN2†	MAPE, %	AAV-Sim, %	MAINRITSAN2†	MAPE, %
Minor relapse	6.0	6.2 (1.0–11.4%)	3	7.3	8.6 (2.5–14.8%)	15
Major relapse	3.5	3.7 (0.0–7.8%)	5	5.5	7.4 (1.7–13.1%)	26
≥1 severe infection	19.4	19.8 (11.1–28.4%)	2	11.1	10.2 (3.4–16.4%)	9
ESRD	0.24	1.2 (0.0–3.6%)	1	0.24	0 (0–0%)	NR
Relapse-free survival	84.8	86 (79.2–94.2%)	1	82.3	84 (76.1–92.3%)	2
Survival	94.1	96.3 (89.6–99.2%)	2	94.8	98.8 (93.3–99.9%)	4

* Values are the percentage (95% confidence interval [95% CI]) unless indicated otherwise. ESRD = end-stage renal disease; MAINRITSAN2 = Rituximab versus Azathioprine in ANCA-Associated Vasculitis 2 (trial); MAPE = mean average percent error; NR = not reported (due to the low number of observed outcomes in the MAINRITSAN2 trial).
† 95% CIs are reported for outcomes observed in the MAINRITSAN2 trial to demonstrate the accuracy of AAV-Sim projections.

(0.03%). Most deaths due to infection occurred during remission (1.6%), as opposed to during active disease (0.02%).

Internal validation using MAINRITSAN2 outcomes. *Proportion with minor and major relapses.* We then evaluated the outcomes projected by the AAV-Sim over 28 months among patients treated with a tailored strategy and compared the projected outcomes with each strategy (fixed versus tailored) to those observed in the MAINRITSAN2 trial (Table 3). When a tailored strategy was used, more minor and major relapses were projected compared with the fixed strategy (minor relapse 7.3% versus 6.0%; major relapse 5.5% versus 3.5%). The projected proportion of individuals with at least 1 major relapse was similar in the AAV-Sim compared with the MAINRITSAN2 trial in each treatment strategy (fixed strategy 3.5% in AAV-Sim versus 3.7% [95% confidence interval (95% CI) 0.0–7.8%] in MAINRITSAN2, MAPE 5%; tailored strategy 5.5% in AAV-Sim versus 7.4% [95% CI 1.7–13.1%] in MAINRITSAN2, MAPE 26%).

Proportion with severe infections. Among individuals treated with the fixed strategy in the AAV-Sim, at least 1 severe infection would occur more frequently than with the tailored strategy (Table 3), which was similar to observed results in the MAINRITSAN2 trial (fixed strategy 19.4% in AAV-Sim versus 19.8% [95% CI 11.1–28.4%] in MAINRITSAN2, MAPE 2%; tailored strategy 11.1% in AAV-Sim versus 10.2% [95% CI 3.4–16.4%], MAPE 9%).

Cumulative incidence of ESRD. The projected incidence of ESRD in both treatment strategies was similar to that observed in the MAINRITSAN2 trial (fixed strategy 0.2% in AAV-Sim versus 1.2% [95% CI 0.0–3.6%] in MAINRITSAN2, MAPE 1%; tailored strategy 0.2% in AAV-Sim versus 0% in MAINRITSAN2; MAPE cannot be calculated because 0 is in the denominator) (Table 3).

Cumulative incidence of death and relapse-free survival. Results were similar when comparing overall survival projected by the AAV-Sim to that observed in the MAINRITSAN2 trial (Table 3) for fixed (94.1% in AAV-Sim versus 96.3% [95% CI 89.6–99.2%] in MAINRITSAN2, MAPE 2%) and tailored strategies (94.8% in AAV-Sim versus 98.8% [95% CI 93.3–99.9%] in MAINRITSAN2, MAPE 4%). Relapse-free survival was similar to observed results over the 28 months of the MAINRITSAN2 trial in both the fixed and tailored strategies (fixed strategy 84.8% in AAV-Sim versus 86.0% [95% CI 79.2–94.2%] in MAINRITSAN2, MAPE 1%; tailored strategy 82.2% in AAV-Sim versus 84.0% [95% CI 76.1–92.3%] in MAINRITSAN2, MAPE 2%). When comparing goodness of fit, we found that our projected outcomes from the AAV-Sim were similar to those observed in the MAINRITSAN2 trial for both strategies (RMSE 2.1 and 2.2, and CV-RMSE 2.3% and 2.5%, respectively) (Figure 2).

Secondary outcomes. Using the AAV-Sim, we projected that the proportion of patients experiencing noninfectious death without a history of ESRD at 28 months would be 4.3% with the tailored strategy, similar to projections in the fixed strategy (4.3%). The projected proportion of deaths due to infection was higher in the fixed than the tailored strategy (1.6% versus 0.9%), with the majority occurring during remission as opposed to relapse in both treatment strategies, respectively (1.6% versus 0.9%). The proportion of projected deaths from ESRD was small in both strategies (0.03% versus 0.04%, respectively).

In the tailored strategy, we projected that 95.4% of simulated individuals would experience at least 1 repopulation of peripheral B cells and 89.6% would experience at least 1 ANCA rise or seroconversion over 28 months. The projected mean ± SD number of rituximab infusions was 3.9 ± 0.9, which was similar to the median number of infusions (3 [interquartile range 2–4]) administered to this group in the MAINRITSAN2 trial. In a 2-way sensitivity analysis varying the age and infection risk with fixed strategy, we projected that differences in relapse-free survival were sensitive to infection risk, especially at older ages (see Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25088).

External validation. *MAINRITSAN 1 external validation.* We then compared the outcomes projected in the AAV-Sim with the fixed strategy to those observed in the fixed strategy of the MAINRITSAN1 trial (Table 4). At 28 months, the AAV-Sim projected a similar proportion of patients experiencing minor and major relapse when compared with the MAINRITSAN1 trial (minor relapse 6.6% in AAV-Sim versus 11.0% [95% CI 4.0–21.5%] in MAINRITSAN 1, MAPE 40%; major relapse 4.0% in AAV-Sim versus 5.0% [95% CI 1.1–14.6%] in MAINRITSAN 1, MAPE 21%). The proportion of individuals experiencing at least 1 severe infection was also similar among those receiving fixed retreatment in the AAV-Sim (19.4%) and MAINRITSAN1 (19.0% [95% CI 10.1–31.9%], MAPE 2%). There were no deaths observed in the MAINRITSAN 1 trial among patients randomized to fixed retreatment, whereas 5.9% of individuals in the AAV-Sim were projected to die over 28 months in the fixed strategy. At 28 months, 89.2% of individuals in the AAV-Sim experienced relapse-free survival compared with 86.1% (95% CI 74.2–93.7%) of patients in the MAINRITSAN1 trial (RMSE 2.5, CV-RMSE 2.7%) (Figure 2B).

Pendergraft et al (7) external validation. The relapse-free survival associated with the fixed strategy in the AAV-Sim at 12 months was similar to that observed in the observational study of fixed B cell retreatment by Pendergraft et al (major relapse-free survival 94.7% in AAV-Sim versus 97.0% in Pendergraft et al, RMSE 1.7, CV-RMSE 1.8%; minor relapse-free survival 93.7% in AAV-Sim versus 95.0% in Pendergraft et al, RMSE 1.5, CV-RMSE 1.5%) (Figure 2C).

DISCUSSION

We developed and validated the AAV-Sim, a novel state-transition microsimulation model of clinical outcomes among

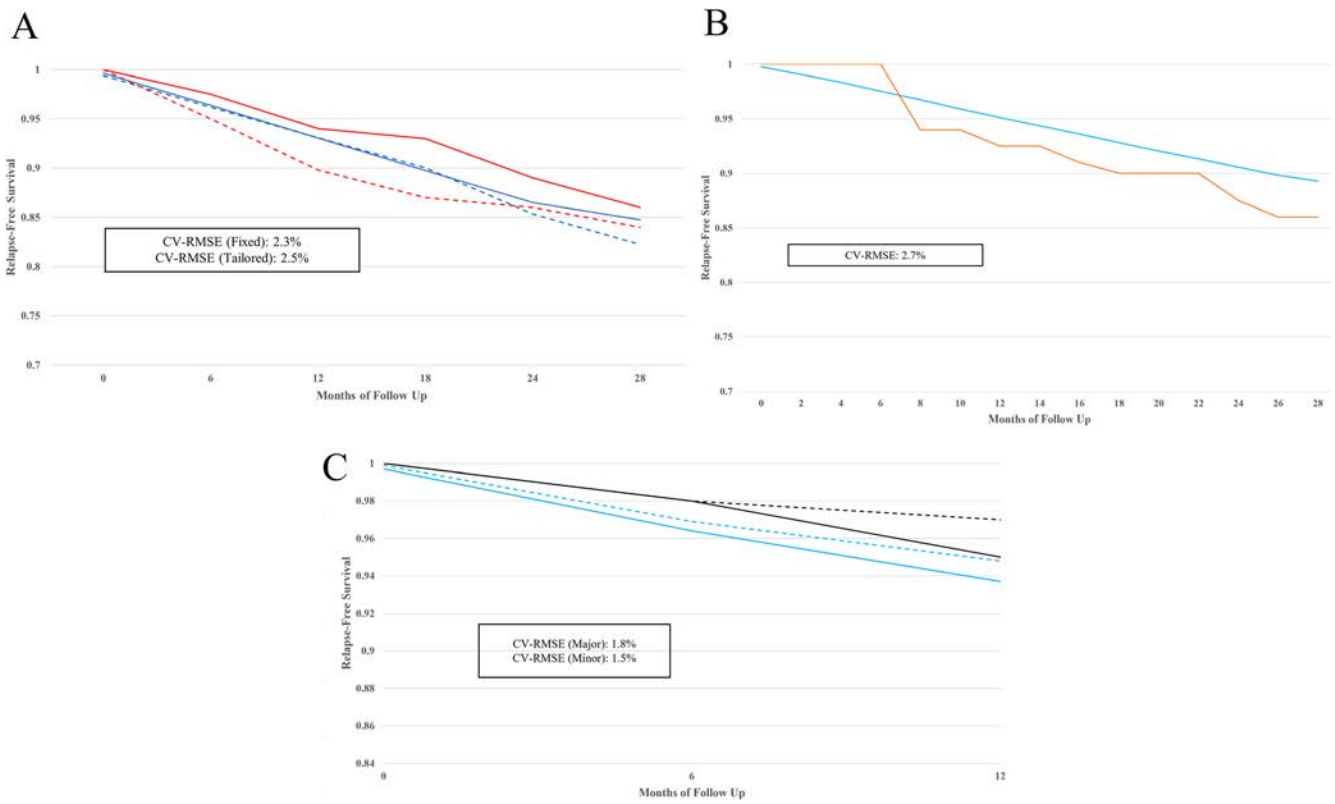

Figure 2. **A**, Internal validation of relapse-free survival: projected results from the antineutrophil cytoplasmic antibody–associated vasculitis simulation model (AAV-Sim) and observed relapse-free survival observed in the Rituximab versus Azathioprine in ANCA-Associated Vasculitis 2 (MAINRITSAN2) trial. This figure represents observed relapse-free survival in the internal validation among patients who received fixed B cell depletion treatment (solid lines) and those who received tailored treatment (broken lines) in the AAV-Sim (blue lines) and the MAINRITSAN2 trial (red lines). **B**, External validation of relapse-free survival: projected results from the AAV-Sim and observed relapse-free survival observed in the MAINRITSAN2 trial. This figure represents observed relapse-free survival in the external validation among patients who received fixed B cell depletion treatment and those who received tailored treatment in the AAV-Sim (blue line) and the MAINRITSAN1 trial (red line). **C**, External validation of minor and major relapse-free survival: projected major and minor relapse-free survival observed in the AAV-Sim and observed results from the study by Pendergraft et al (7). This figure represents observed minor (solid lines) and major (broken lines) relapse-free survival in the external validation among patients who received fixed B cell depletion treatment and those who received tailored treatment in the AAV-Sim (blue lines) and the study by Pendergraft et al (black lines). CV-RMSE = coefficient of variation of root mean square error.

patients undergoing treatment to maintain remission for AAV. This is the first validated microsimulation model of AAV and is an important first step in leveraging this methodology to inform and advance care for patients with AAV. A particular strength of this approach is its ability to incorporate data from multiple sources, including clinical trials, observational studies, and nationwide registries, as well as opinions from experts in the field for aspects of the disease course where data may be limited. Additionally, the AAV-Sim can be used to project how the preferred treatment strategies could vary when the associated outcomes are examined over longer time horizons, refined to reflect different priorities (e.g., infection or relapse), or according to disease subgroups (e.g., ANCA type, renal involvement, age).

We performed several validation steps to demonstrate the accuracy of projected outcomes of the AAV-Sim. First, clinical experts in AAV evaluated the structure of the model and assessed input parameters to confirm face validity. Second, we performed internal validation by comparing the outcomes projected in the AAV-Sim to those observed in the MAINRITSAN2 trial, which is the only study to evaluate both fixed and tailored rituximab retreatment strategies. Third, we performed external validation by comparing projected AAV-Sim outcomes to those observed in MAINRITSAN1, a large clinical trial that established the efficacy of the fixed rituximab retreatment strategy, and in an observational study reporting a single center's experience with fixed retreatment. In internal and external validation, we found that model-generated results and observed data were within a range generally accepted to indicate a good fit by MAPE and CV-RMSE (25–27). Some larger MAPE values were observed when projected and observed outcomes were close to one another, but the frequency of these outcomes was close to 0 (e.g., <10%). Projected outcomes were consistently within the confidence intervals of observed outcomes in the trials used for validation.

Table 4. External validation comparing antineutrophil cytoplasmic antibody (ANCA)–associated vasculitis simulation model (AAV-Sim) projected outcomes for the fixed strategy with those outcomes observed in the MAINRITSAN1 trial*

	AAV-Sim, %	MAINRITSAN1†	MAPE, %
Minor relapse	6.6	11.0 (4.0–21.5%)	40
Major relapse	4.0	5.0 (1.1–14.6%)	21
≥1 severe infection	19.4	19.0 (10.1–31.9%)	2
Relapse-free survival	89.2	86.0 (74.2–93.7%)	4
Overall survival	94.1	100 (93.7–100%)	6

* Values are the percentage (95% confidence interval [95% CI]) unless indicated otherwise. MAINRITSAN1 = Rituximab versus Azathioprine in ANCA-Associated Vasculitis 1 (trial); MAPE = mean average percent error.
† 95% CIs are reported for outcomes observed in the MAINRITSAN1 trial to demonstrate the accuracy of AAV-Sim projections.

In human immunodeficiency virus infection, and other chronic conditions, simulation modeling has informed management guidelines and policy (4,5). In contrast, simulation modeling has been infrequently applied in inflammatory rheumatic diseases (28–33). The AAV-Sim could have important influence because it can address highly relevant and timely concerns in AAV care (e.g., impact of delaying rituximab for vaccinations, use of alternative maintenance strategies). While our primary aim was to validate the AAV-Sim, we also confirmed in a larger simulated population that fixed versus tailored retreatment strategies yield similar rates of relapse-free survival. We also projected higher risk of death due to infection, especially during remission on fixed retreatment, an observation that could not be made in the MAINRITSAN2 trial due to lack of power.

We plan to expand the AAV-Sim to incorporate utilities and costs as a next step to enable projections that can compare the relative value of different treatments for AAV. These data can then be used to inform management guidelines, especially when resources are limited and/or costs are in a range that may be considered unacceptably high for payers. Incorporating patient preferences and costs will facilitate analyses of novel drugs entering the market to determine what society should be willing to pay beyond currently available strategies (34). Such studies are needed given the estimated price of avacopan, the newest US Food and Drug Administration–approved medication for AAV: $128,976 per year for Veterans Affairs beneficiaries (35). The AAV-Sim can be further adapted to compare other maintenance strategies and to project outcomes during remission induction.

AAV-Sim–projected deaths were more frequent than those observed in the studies used for validation for several potential reasons. First, clinical trials often exclude patients with certain comorbidities, so the observed mortality risk is lower than in the general population. For instance, the MAINRITSAN2 trial excluded patients with recent infection or malignancy, advanced heart failure, or lung disease, which increase mortality; our model inputs for background mortality were based on national vital statistics, which do not exclude patients with these comorbidities. Exclusion criteria are not detailed in MAINRITSAN1, but patients with higher mortality risk were likely excluded. Additionally, the number of patients in the MAINRITSAN1 arm treated with rituximab was small (n = 57), limiting the number of potential observed deaths. Second, MAINRITSAN1 and MAINRITSAN2 were conducted in France, where life expectancy is higher than in the US (36). Third, we assumed no loss to follow-up in the AAV-Sim, whereas 7 of 81 patients in the fixed strategy and 2 of 81 in the tailored strategy were lost to follow-up from MAINRITSAN 2. This loss to follow-up may have included patients who died.

Several factors influenced what may be perceived as lower than expected incidence of ESRD in the AAV-Sim. First, we projected outcomes in patients who had survived their initial AAV presentation without ESRD; a large portion of ESRD in AAV occurs during the initial presentation (11). Second, patients modeled in this analysis of the AAV-Sim were assumed to have preserved renal function, reflecting both the unlikely use of tailored rituximab retreatment in patients at high risk of ESRD because of AAV, as well as the patients enrolled in the trials used for validation. A strength of the AAV-Sim and this methodologic approach is that the input parameters can be varied to reflect the population of interest in future studies, and additional complexity in model structure can be developed to address other questions. Future uses of the AAV-Sim may include incorporating changes in eGFR over time in an individual and identifying an eGFR threshold over which tailored strategies may be preferable to balance risks of severe infection with relapse and ESRD. Third, we projected outcomes for 28 months, limiting the number of observed long-term outcomes.

Strengths of this study include the use of a novel methodology to project outcomes for patients with AAV, guided by clinically relevant questions, and a stepwise approach to validation that follows the recommended practices for simulation modeling (37). Despite these strengths, our study has certain limitations. First, this is a simplified model of a complex condition that required certain assumptions. Despite this, the AAV-Sim projected key outcomes with close fidelity to those observed in clinical trials. In future steps, we plan to compare alternative treatment strategies and incorporate greater complexity of dosing using more sophisticated programming methods. Second, although model outcomes included relapse, ESRD, severe infection, and mortality, additional relevant outcomes are not yet projected in the AAV-Sim. In the future, we plan to incorporate toxicities of glucocorticoids and other medications, all-cause hospitalization, quality of life, among other outcomes, especially given the focus on glucocorticoids-sparing approaches in recent and ongoing trials (38–41). Additionally, we plan to expand the model to project outcomes, particularly ESRD, in individuals who present with or develop more severe chronic kidney disease during follow-up. Third, ESRD and death occurred rarely in the studies used to

validate the AAV-Sim; we will compare model projections with observational data as they become available among patients receiving long-term B cell depletion. Last, there are no studies other than the MAINRITSAN2 trial that evaluate the impact of tailored rituximab therapy to maintain remission, so we were only able to perform an internal validation of this strategy. Reassuringly, internal validation of this arm and external validation of fixed treatment approaches closely approximated outcomes observed in other studies (3,6,7).

In conclusion, we developed and validated the AAV-Sim, a novel microsimulation model to project outcomes in individuals with AAV. This model may be applied to inform clinical guidelines and identify influential areas of uncertainty to guide the design and prioritization of future studies. With the increasing availability of multiple treatment strategies in AAV, this methodologic approach will facilitate the development of guidelines to recommend preferred treatments in all-comers with AAV and subgroups identified by demographic characteristics, ANCA type, and AAV organ involvement.

ACKNOWLEDGMENT

We acknowledge Aaron Wu for his evaluation of our model parameterization in TreeAge.

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. Wallace, Stone, Merkel, Miloslavsky, Choi, Hyle.

Acquisition of data. Wallace, Fu, Zhang.

Analysis and interpretation of data. Wallace, Fu, Zhang, Hyle.

REFERENCES

1. Wallace ZS, Yun H, Curtis JR, et al. ANCA-associated vasculitis management in the United States: data from the Rheumatology Informatics System for Effectiveness (RISE) registry. J Rheumatol 2021;48:1060–4.
2. Chung SA, Langford CA, Maz M, et al. 2021 American College of Rheumatology/Vasculitis Foundation guideline for the management of antineutrophil cytoplasmic antibody–associated vasculitis. Arthritis Care Res (Hoboken) 2021;73:1366–83.
3. Charles P, Terrier B, Perrodeau E, et al. Comparison of individually tailored versus fixed-schedule rituximab regimen to maintain ANCA-associated vasculitis remission: results of a multicentre, randomised controlled, phase III trial (MAINRITSAN2). Ann Rheum Dis 2018;77:1143–9.
4. Rutter CM, Zaslavsky AM, Feuer EJ. Dynamic microsimulation models for health outcomes: a review. Med Decis Making 2011;31:10–8.
5. Bespalov A, Barchuk A, Auvinen A, et al. Cancer screening simulation models: a state of the art review. BMC Med Inform Decis Mak 2021;21:359.
6. Guillevin L, Pagnoux C, Karras A, et al. Rituximab versus azathioprine for maintenance in ANCA-associated vasculitis. N Engl J Med 2014;371:1771–80.
7. Pendergraft WF III, Cortazar FB, Wenger J, et al. Long-term maintenance therapy using rituximab-induced continuous B-cell depletion in patients with ANCA vasculitis. Clin J Am Soc Nephrol 2014;9:736–44.
8. Miloslavsky EM, Specks U, Merkel PA, et al. Outcomes of nonseeded relapses in antineutrophil cytoplasmic antibody–associated vasculitis treated with glucocorticoids. Arthritis Rheumatol 2015;67:1629–36.
9. Iudici M, Pagnoux C, Courvoisier DS, et al. Granulomatosis with polyangiitis: study of 795 patients from the French Vasculitis Study Group registry. Semin Arthritis Rheum 2021;51:339–46.
10. Iudici M, Pagnoux C, Courvoisier DS, et al. Localized versus systemic granulomatosis with polyangiitis: data from the French Vasculitis Study Group registry. Rheumatology (Oxford) 2022;61:2464–71.
11. Lionaki S, Hogan SL, Jennette CE, et al. The clinical course of ANCA small-vessel vasculitis on chronic dialysis. Kidney Int 2009;76:644–51.
12. Smith RM, Jones RB, Specks U, et al. Rituximab as therapy to induce remission after relapse in ANCA-associated vasculitis. Ann Rheum Dis 2020;79:1243–9.
13. Wester Trejo MA, Floßmann O, Westman KW, et al. Renal relapse in antineutrophil cytoplasmic autoantibody-associated vasculitis: unpredictable, but predictive of renal outcome. Rheumatology (Oxford) 2019;58:103–9.
14. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases. United States Renal Data System. 2021 USRDS annual data report: epidemiology of kidney disease in the United States. URL: https://usrds-adr.niddk.nih.gov/2021.
15. Cook CE, Fu X, Zhang Y, et al. Validation of antineutrophil cytoplasmic antibody-associated vasculitis as the cause of end-stage renal disease in the US Renal Data System. ACR Open Rheumatol 2022;4:8–12.
16. Wallace ZS, Zhang Y, Lu N, et al. Improving mortality in end-stage renal disease due to granulomatosis with polyangiitis (Wegener's) from 1995 to 2014: data from the United States Renal Data System. Arthritis Care Res (Hoboken) 2018;70:1495–500.
17. Tan JA, Dehghan N, Chen W, et al. Mortality in ANCA-associated vasculitis: a meta-analysis of observational studies. Ann Rheum Dis. 2017;76:1566–74.
18. Agency for Healthcare Research and Quality. Healthcare Cost and Utilization Project (HCUPnet). URL: https://datatools.ahrq.gov/hcupnet.
19. Lionaki S, Blyth ER, Hogan SL, et al. Classification of antineutrophil cytoplasmic autoantibody vasculitides: the role of antineutrophil cytoplasmic autoantibody specificity for myeloperoxidase or proteinase 3 in disease recognition and prognosis. Arthritis Rheum 2012;64:3452–62.
20. Schirmer JH, Wright MN, Herrmann K, et al. Myeloperoxidase–antineutrophil cytoplasmic antibody (ANCA)–positive granulomatosis with polyangiitis (Wegener's) is a clinically distinct subset of ANCA-associated vasculitis: a retrospective analysis of 315 patients from a German vasculitis referral center. Arthritis Rheumatol 2016;68:2953–63.
21. Specks U, Merkel PA, Seo P, et al. Efficacy of remission-induction regimens for ANCA-associated vasculitis. N Engl J Med 2013;369:417–27.
22. Hogan SL, Falk RJ, Chin H, et al. Predictors of relapse and treatment resistance in antineutrophil cytoplasmic antibody-associated small-vessel vasculitis. Ann Intern Med 2005;143:621–31.

23. Charles P, Perrodeau É, Samson M, et al. Long-term rituximab use to maintain remission of antineutrophil cytoplasmic antibody-associated vasculitis: a randomized trial. Ann Intern Med 2020;173:179–87.
24. Stone JH, Merkel PA, Spiera R, et al. Rituximab versus cyclophosphamide for ANCA-associated vasculitis. N Engl J Med 2010;363:221–32.
25. Kazemian P, Wexler DJ, Fields NF, et al. Development and validation of PREDICT-DM: a new microsimulation model to project and evaluate complications and treatments of type 2 diabetes mellitus. Diabetes Technol Ther 2019;21:344–55.
26. Reddy KP, Bulteel AJ, Levy DE, et al. Novel microsimulation model of tobacco use behaviours and outcomes: calibration and validation in a US population. BMJ Open 2020;10:e032579.
27. Hyle EP, Foote JH, Shebl FM, et al. Development and validation of the age-associated dementia policy (AgeD-Pol) computer simulation model in the USA and Europe. BMJ Open 2022;12;e056546.
28. Silverstein MD, Albert DA, Hadler NM, et al. Prognosis in SLE: comparison of Markov model to life table analysis. J Clin Epidemiol 1988;41:623–33.
29. Gavan S, Bruce I, Payne K. Generating evidence to inform health technology assessment of treatments for SLE: a systematic review of decision-analytic model-based economic evaluations. Lupus Sci Med 2020;7:e000350.
30. Nguyen CM, Bounthavong M, Mendes MA, et al. Cost utility of tumour necrosis factor-α inhibitors for rheumatoid arthritis: an application of Bayesian methods for evidence synthesis in a Markov model. Pharmacoeconomics 2012;30:575–93.
31. Kamal KM, Miller LA, Kavookjian J, et al. Alternative decision analysis modeling in the economic evaluation of tumor necrosis factor inhibitors for rheumatoid arthritis. Semin Arthritis Rheum 2006;36:50–60.
32. Plumpton CO, Alfirevic A, Pirmohamed M, et al. Cost effectiveness analysis of HLA-B*58:01 genotyping prior to initiation of allopurinol for gout. Rheumatology (Oxford) 2017;56:1729–39.
33. Jutkowitz E, Dubreuil M, Lu N, et al. The cost-effectiveness of HLA-B*5801 screening to guide initial urate-lowering therapy for gout in the United States. Semin Arthritis Rheum 2017;46:594–600.
34. Neilan AM, Landovitz RJ, Le MH, et al. Cost-effectiveness of long-acting injectable HIV preexposure prophylaxis in the United States: a cost-effectiveness analysis. Ann Intern Med 2022;175:479–89.
35. Majumdar M, Nadeem D. ChemoCentryx's drug gets U.S. FDA nod for treating rare autoimmune disease. Reuters 2021. URL: https://www.reuters.com/business/healthcare-pharmaceuticals/us-fda-approves-chemocentryx-drug-treat-rare-autoimmune-disease-2021–10–08/.
36. Ho JY, Hendi AS. Recent trends in life expectancy across high income countries: retrospective observational study. BMJ 2018;362:k2562.
37. Eddy DM, Hollingworth W, Caro JJ, et al. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. Med Decis Making 2012;32:733–43.
38. Furuta S, Nakagomi D, Kobayashi Y, et al. Effect of reduced-dose vs high-dose glucocorticoids added to rituximab on remission induction in ANCA-associated vasculitis: a randomized clinical trial. JAMA 2021;325:2178–87.
39. Jayne DR, Merkel PA, Schall TJ, et al. Avacopan for the treatment of ANCA-associated vasculitis. N Engl J Med 2021;384:599–609.
40. Wallace ZS, Miloslavsky EM. Management of ANCA associated vasculitis. BMJ 2020;368:m421.
41. Walsh M, Merkel PA, Peh CA, et al. Plasma exchange and glucocorticoids in severe ANCA-associated vasculitis. N Engl J Med 2020;382:622–31.
42. Arias E, Bastian B, Xu J, et al. U.S. state life tables, 2018. Natl Center Vital Stat Rep 2021;70:1–18.

Arthritis Care & Research
Vol. 75, No. 9, September 2023, pp 1986–1995
DOI 10.1002/acr.25092

American College of Rheumatology
Empowering Rheumatology Professionals

Effectiveness and Persistence in SB4- and Reference Etanercept–Treated Rheumatoid Arthritis Patients in Ordinary Clinical Practice in Norway

Glenn Haugeberg,[1] Gunnstein Bakland,[2] Erik Rødevand,[3] Inger Johanne Widding Hansen,[4] Andreas Diamantopoulos,[5] and Are Hugo Pripp[6]

Objective. Biosimilars represent cost-effective alternatives to reference biologic disease-modifying antirheumatic drugs. Our objective was to compare drug effectiveness and drug persistence in the treatment of rheumatoid arthritis (RA), assessing the etanercept biosimilar SB4 in efficacy and safety compared with reference etanercept in a Phase III, randomized controlled trial. We applied EULAR Points to Consider for Comparative Effectiveness Research in a retrospective database study of etanercept and SB4 in patients treated in clinical practice in Norway.

Methods. Patients with RA (n = 1,455) treated with etanercept or SB4 between 2010 and 2018 at 5 centers in Norway with ≥1 year of follow-up were included. Disease outcomes (Disease Activity Score in 28 joints [DAS28] at week 52) and drug persistence were compared between unmatched etanercept (n = 575) and SB4 (n = 299) cohorts and matched analyses (n = 172, both cohorts) using propensity score (PS) matching to adjust for confounders.

Results. In unmatched analyses, the difference in change from baseline between etanercept (n = 221) and SB4 (n = 106) for DAS28 at week 52 was mean –0.02 (95% confidence interval [95% CI] –0.32, 0.27), demonstrating equivalence by the predetermined equivalence margin (±0.6). In PS-matched analyses, the difference between etanercept (n = 49) and SB4 (n = 49) was 0.03 (95% CI –0.46, 0.52), within the predefined equivalence margin. Persistence using the drug at week 52 was similar between etanercept (0.62 [95% CI 0.57, 0.65]) and SB4 (0.66 [95% CI 0.60, 0.71]) cohorts in the unmatched analysis; in PS-matched cohorts, persistence at week 52 was 0.52 (95% CI 0.44, 0.59) for etanercept and 0.68 (95% CI 0.61, 0.75) for SB4.

Conclusion. Outcomes for disease status/drug persistence at week 52 were similar between patients with RA treated with etanercept or SB4.

INTRODUCTION

The introduction of biologic disease-modifying antirheumatic drugs (bDMARDs) in 1999 led to a paradigm shift in the treatment of chronic inflammatory arthritis disorders. The tumor necrosis factor inhibitors infliximab and etanercept (ETN), licensed for the treatment of rheumatoid arthritis (RA), were the first to reach the market, receiving approvals for use in the European Union in 1999 and 2000, respectively. The costs of biologic drugs present challenges, causing restrictions to the prescribing of these drugs in several countries and subsequently contributing to inequalities of care (1–3). However, the expiration of patents for bDMARDs allowed the manufacture of biosimilars, which can be sold at lower prices. Since the first biosimilar tumor necrosis factor inhibitor, infliximab CT-P13, was approved in the European Union in 2013, additional biosimilars have become available. The ETN biosimilar SB4 was approved by the European Medicines Agency in January 2016. SB4 demonstrated similarity to reference ETN in a comprehensive biosimilarity exercise, which included a

Previously presented in part as a poster at the European Congress of Rheumatology, June 2–5, 2021, Paris. Haugeberg G, Bakland G, Rødevand E, et al. POS0612 exploring equivalence between biosimilar SB4 and reference etanercept by assessing effectiveness in rheumatoid arthritis patients treated in ordinary clinical practice.

Supported by Biogen International GmbH.

[1]Glenn Haugeberg, MD, PhD: Sørlandet Hospital, Kristiansand, and Norwegian University of Science and Technology, Trondheim, Norway; [2]Gunnstein Bakland, MD, PhD: University Hospital of Northern Norway, Tromsø, Norway; [3]Erik Rødevand, MD: St. Olavs Hospital, Trondheim, Norway; [4]Inger Johanne Widding Hansen, MD: Sørlandet Hospital, Kristiansand, Norway; [5]Andreas Diamantopoulos, MD, PhD, MPH: Martina Hansens Hospital, Sandvika, Norway; [6]Are Hugo Pripp, PhD: Oslo University Hospital and Oslo Metropolitan University, Oslo, Norway.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.25092&file=acr25092-sup-0001-Disclosureform.pdf.

Address correspondence via email to Glenn Haugeberg, MD, PhD, at glenn.haugeberg@sshf.no.

Submitted for publication March 23, 2022; accepted in revised form January 24, 2023.

SIGNIFICANCE & INNOVATIONS

- SB4 has demonstrated similarity to etanercept (ETN) in a comprehensive biosimilarity exercise, which included a 52-week, double-blind, Phase III, randomized controlled trial in patients with rheumatoid arthritis (RA). However, published real-world data on SB4 are limited.
- This study reports similar effectiveness, persistence, and safety for patients with RA who initiated treatment with ETN or SB4 for up to 2 years as part of routine clinical care at outpatient clinics in Norway. However, when accounting for differences between cohorts at baseline using propensity score matching, persistence was greater on SB4 than on ETN. Effectiveness was maintained in patients with RA who had a mandatory switch from ETN to SB4.
- These findings support outcomes from earlier biosimilarity studies and indicate that SB4 is an effective option for switching from ETN for the treatment of patients with RA.

52-week, double-blind, Phase III, randomized controlled trial (RCT) in patients with RA (4,5).

Biosimilar drugs follow a tailored approval pathway compared with reference drugs, including a Phase III RCT with high internal but low external validity. Therefore, observational studies with high external validity are important to reassure patients and physicians that no clinically meaningful differences exist between a biosimilar and its reference drug when used in routine clinical practice. Unfortunately, recent comparative effectiveness studies often do not disclose applied analytical methods in sufficient detail, with many not adjusting for confounders (6) or accounting for attrition or missing data, according to a EULAR task force systematic review (7). Compliance with these recommendations for conducting comparative effectiveness studies may contribute toward high-quality observational studies.

Although SB4 has been on the market for several years, published real-world data are limited for patients with RA (8–12). The objective of this real-world study was to compare drug effectiveness and drug persistence in ETN treatment-naive patients with RA who received treatment with ETN or the biosimilar SB4, applying the EULAR Points to Consider for Comparative Effectiveness Research. Further, we aimed to examine drug effectiveness and drug persistence in patients with RA treated with SB4 after a mandatory nonmedical switch from ETN and to explore reasons for cessation among the 3 RA treatment cohorts: ETN, SB4, and SB4 switch.

MATERIALS AND METHODS

Study design. This was a retrospective database study of ETN-naive patients who received treatment with ETN or SB4 and ETN-treated patients who switched to SB4 and had at least 1 year of follow-up data. Data extraction from the participating centers was performed between June 26 and July 1, 2019. The study followed the recommendations outlined in the EULAR Points to Consider When Analysing and Reporting Comparative Effectiveness Research with Observational Data in Rheumatology (13), as well as the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) and Good Research for Comparative Effectiveness (GRACE) guidelines (14,15).

Study population. ETN-naive patients with RA started ETN treatment between January 2010 and July 2018 at 5 centers in Norway and were followed for up to 2 years. The participating centers were University Hospital of North Norway, Tromsø; St. Olavs Hospital, Trondheim; Haukeland University Hospital, Bergen; Sørlandet Hospital, Kristiansand; and Martina Hansens Hospital, Sandvika.

Data collection. Data collection at participating centers was performed at clinical visits made as part of routine practice. Data variables collected by all centers included age, sex, body mass index (BMI), duration of disease, anti–cyclic citrullinated peptide antibodies (anti-CCP), C-reactive protein (CRP) level, erythrocyte sedimentation rate (ESR), 28 swollen and tender joint counts (SJC28 and TJC28), patient global assessment (PtGA) reported on a 0–100-mm visual analog scale, Disease Activity Score in 28 joints (DAS28), modified Health Assessment Questionnaire (MHAQ), current use of methotrexate, current conventional synthetic DMARDs (csDMARDs), including methotrexate, and order of bDMARDs. Data extraction also included reasons for drug cessation registered in the hospital clinical GoTreatIT Rheuma databases. Data were collected for ETN-naive patients who started treatment on ETN or SB4 and for patients who switched from ETN to SB4. At the time of data extraction, we examined how many patients had remained on ETN and how many patients who switched to SB4 had switched back to ETN. Data are available upon reasonable request.

Study objectives. The primary objective was to compare drug effectiveness and drug persistence of ETN and SB4 at week 52 in treatment-naive patients with RA treated in ordinary clinical practice in Norway. Secondary objectives were to further assess drug effectiveness and persistence at week 52 and week 104 in patients with RA treated with SB4 after a mandatory nonmedical switch from ETN and to explore reasons for drug cessation across the 3 RA treatment cohorts (ETN, SB4, and SB4 switch).

Study end points. Primary outcome measures were disease outcomes (DAS28 at week 52, assessed as a continuous variable) and drug persistence (measured as time to treatment discontinuation during a 52-week follow-up). The equivalence of DAS28 was determined based on a predefined equivalence

margin of ±0.6 (16). Unmatched (primary) and propensity score (PS)–matched (supportive) analyses were conducted, including a sensitivity analysis of PS-matched samples using all available data in the statistical analysis.

Secondary outcome measures included DAS28 at week 104, assessed as a continuous variable, and DAS28 at week 52 and week 104, assessed as a categorical variable based on EULAR response criteria of good, moderate, and no response to treatment (17). Other clinical outcomes assessed at week 52 and week 104 included CRP levels, ESR, SJC28, TJC28, PtGA, and MHAQ. Reasons for drug cessation were recorded. At data extraction, the number of patients who had remained on ETN without switching to SB4 was quantified, as was the number of patients who had switched back to ETN from SB4. Reasons for discontinuing treatment were also assessed; where an adverse event (AE) was given as the reason for discontinuation, specific AEs were reported if they had been recorded in the registry.

Statistical analysis. Independent samples *t*-tests and chi-square tests were used to compare baseline characteristics by treatment cohort in unmatched data. For matched data, paired samples *t*-test and McNemar's test were used to compare baseline characteristics for continuous variables and proportions, respectively.

For the primary, unmatched analysis of DAS28 outcomes between patients treated with ETN and patients treated with SB4, a conventional independent samples *t*-test was used (model 0). For the supportive, matched analyses of DAS28, PS matching was used and analyzed with a paired samples *t*-test. The primary PS model was based on clinical knowledge and adjusted for the following confounders at baseline: age, sex, DAS28, order of biologics, and concomitant csDMARDS. Additional supportive models that matched for different sets of confounders were also investigated (model 1 [M1], M2, M3, and M4): M1 adjusted for age; M2 adjusted for age and sex; M3 adjusted for age, sex, and DAS28; and M4 adjusted for age, sex, DAS28, order of biologics, and concomitant use of csDMARDs and the other clinical outcome measures (CRP level, ESR, SJC28, TJC28, PtGA, and MHAQ). The primary PS-matched model was found to be the most supportive based on clinical knowledge and data availability. A standardized difference of <0.1 indicates a good match. Drug persistence was analyzed by the Kaplan-Meier method. Kaplan-Meier estimates were calculated for week 52 and week 104 in unmatched (primary) and matched (supportive) analyses.

Secondary efficacy end points were analyzed based on the same approach as used for DAS28 and PS-matched models for supportive analyses. No imputation of missing data was performed for the yearly assessments. However, a sensitivity analysis for DAS28 that included all available data in the matched samples using regression analysis with standard errors for matched clusters was performed. For example, PS-matched pairs with only DAS28 data for 1 of the drugs at week 52 were excluded in the main matched analysis, but included in the sensitivity analysis.

Ethics approval and patient involvement. The study was approved by the regional ethical committee (Regional etisk komite Midt-Norge 2010/3078). No consent from patients was required by the committee, as all data were anonymized and collected as part of routine clinical care. Patients were not involved in the design, conduct, reporting or dissemination of this research.

RESULTS

Disposition and baseline characteristics. A total of 1,455 patients with RA from 5 participating outpatient clinics were included in this analysis (see Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25092), including 575 patients in the ETN cohort, 299 patients in the SB4 cohort, and 581 patients who had switched from ETN to SB4. Based on unmatched comparisons between the ETN and SB4 cohorts, there was a difference in DAS28 at baseline, with a mean ± SD of 4.3 ± 1.2 and 4.0 ± 1.3, respectively (Table 1). This results in a standardized difference (*d*) of 0.25. Differences were also observed at baseline between the 2 cohorts in age (*d* = 0.16), BMI (*d* = –0.13), SJC28 (*d* = 0.29), TJC28 (*d* = 0.19), and order of bDMARDs (*d* = 0.46).

After matching based on the primary PS model, there were 172 patients each in the ETN and SB4 cohorts; the mean ± SD DAS28 was 4.1 ± 1.3 and 4.1 ± 1.3, respectively (*d* = 0.00) (Table 1). Baseline characteristics showed a good overlap between the PS-matched cohorts based on $d \leq 0.1$, with the exceptions of BMI (–0.24), anti-CCP positivity (0.25), and CRP level (–0.17). For patients who switched from ETN to SB4, the mean ± SD DAS28 at baseline was 2.7 ± 1.2.

Primary outcome measure: DAS28 at week 52 (continuous). Before PS matching, the mean DAS28 at week 52 was 3.2 (95% confidence interval [95% CI] 3.0, 3.3) for the ETN cohort (n = 268) and 2.9 (95% CI 2.7, 3.1) for the SB4 cohort (n = 134) (Table 2 and Figure 1). After matching based on the primary PS model, the mean DAS28 in the baseline to week 52 period was 3.0 (95% CI 2.7, 3.3) for the ETN cohort (n = 49) and 3.2 (95% CI 2.8, 3.7) for the SB4 cohort (n = 49) (Table 3). For the switch cohort, the mean DAS28 was 2.4 (95% CI 2.3, 2.5) for the same period (n = 235). Details on the availability of patient data for the primary analysis are reported in Supplementary Figure 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25092.

The disease status in patients in the ETN cohort and the SB4 cohort at week 52 is shown in Figure 2. In the primary unmatched analysis, the mean difference in change from baseline/week 0 between ETN (n = 221) and SB4 (n = 106) cohorts was –0.02 (95% CI –0.32, 0.27) at week 52, demonstrating equivalence

Table 1. Baseline clinical data for unmatched (observed) and PS-matched ETN-naive patients with RA treated with ETN or SB4 and in patients with RA switched from ETN to SB4*

	Unmatched					PS-matched			
	ETN (n = 575)	SB4 (n = 299)	*P*	*d*	Switch SB4 (n = 581)	ETN (n = 172)	SB4 (n = 172)	*P*	*d*
Age, years	60.5 ± 15.2	58.1 ± 14.9	0.033	0.16	63.0 ± 13.1	59.0 ± 14.1	60.2 ± 14.0	0.324	−0.09
No.	474	299	–	–	581	–	–	–	–
Female, no. (%)	413 (71.8)	225 (75.3)	0.280	0.08	399 (68.7)	131 (76.2)	129 (75.0)	0.803	0.03
BMI, kg/m^2	25.8 ± 4.6	26.4 ± 5.0	0.100	−0.13	25.9 ± 4.2	25.4 ± 4.6	26.5 ± 5.0	0.039	−0.24
No.	368	278	–	–	463	143	143		
Disease duration, years	8.0 ± 9.5	8.2 ± 9.3	0.670	−0.02	14.4 ± 9.1–21.2	8.3 ± 9.7	8.6 ± 10.0	0.784	−0.03
Anti-CCP positive, no./total (%)	195/258 (75.6)	86/176 (48.9)	0.700	0.04	40/49 (81.6)	33/36 (91.7)	30/36 (83.3)	0.317	0.25
C-reactive protein, mg/liter	13.7 ± 15.8	12.3 ± 16.3	0.270	0.09	5.2 ± 8.3	11.5 ± 13.6	14.4 ± 18.6	0.065	−0.17
No.	397	282	–	–	399	169	169	–	–
ESR, mm/hour	23.4 ± 16.7	21.8 ± 16.3	0.250	0.10	16.2 ± 14.9	21.8 ± 17.0	22.4 ± 16.9	0.725	−0.04
No.	349	223	–	–	364	–	–	–	–
SJC28 (range 0–28)	4.1 ± 3.7	3.1 ± 3.1	<0.001	0.29	0.9 ± 1.7	3.6 ± 3.7	3.3 ± 3.4	0.439	0.09
No.	408	281	–	–	415	–	–	–	–
TJC28 (range 0–28)	5.6 ± 4.8	4.7 ± 4.3	0.013	0.19	1.6 ± 2.9	5.1 ± 4.5	5.0 ± 4.6	0.885	0.01
No.	408	281	–	–	415	–	–	–	–
PtGA (range 0–100 mm)	50.0 ± 22.7	51.1 ± 21.9	0.540	−0.05	32.2 ± 23.2	50.6 ± 23.1	49.4 ± 21.7	0.597	0.06
No.	414	270	–	–	406	–	–	–	–
DAS28	4.3 ± 1.2	4.0 ± 1.3	0.006	0.25	2.7 ± 1.2	4.1 ± 1.3	4.1 ± 1.3	0.965	0.00
No.	327	202	–	–	331	–	–	–	–
MHAQ (range 0–3)	0.7 ± 0.5	0.7 ± 0.4	0.950	0.01	0.4 ± 0.5	0.7 ± 0.4	0.6 ± 0.4	0.333	0.11
No.	395	272	–	–	405	167	167		
Current methotrexate, no. (%)	355 (61.7)	175 (58.5)	0.360	0.07	356 (61.3)	97 (56.4)	97 (56.4)	1.000	0.00
Current csDMARDs, no. (%)	417 (72.5)	212 (70.9)	0.610	0.04	396 (68.2)	121 (70.3)	122 (70.9)	0.903	0.01
Order of bDMARDs, no. (%)									
First	394 (68.5)	143 (47.8)	<0.001	0.46	0 (0.0)	85 (49.4)	87 (50.6)	0.872	0.02
Second	127 (22.1)	91 (30.4)	–	–	455 (78.3)	54 (31.4)	53 (30.8)	–	–
≥Third	54 (9.4)	65 (21.7)	–	–	126 (21.7)	33 (19.2)	32 (18.6)	–	–
Center, no. (%)									
University Hospital of North Norway	67 (11.7)	57 (19.1)	<0.001	0.46	46 (7.9)	26 (15.1)	31 (18.0)	<0.001	1.12
St. Olavs Hospital	156 (27.1)	113 (37.8)	–	–	107 (18.4)	69 (40.1)	50 (29.1)	–	–
Haukeland University Hospital/Helse Bergen	101 (17.6)	24 (8.0)	–	–	144 (24.8)	0 (0.0)	14 (8.1)	–	–
Sørlandet Hospital	119 (20.7)	34 (11.4)	–	–	83 (14.3)	73 (42.4)	25 (14.5)	–	–
Martina Hansens Hospital	132 (23.0)	71 (23.7)	–	–	201 (34.6)	4 (2.3)	52 (30.2)	–	–

* Values are the mean ± SD unless indicated otherwise. For the propensity score (PS)–matched population, the logistic regression model used for the PS matching adjusted for age, sex, Disease Activity Score in 28 joints (DAS28), order of biologics, and conventional synthetic disease-modifying antirheumatic drugs (csDMARDs). A *d* of <0.1 indicates a good match. anti-CCP = anti–cyclic citrullinated peptide; bDMARD = biologic DMARD; BMI = body mass index; CRP = C-reactive protein; *d* = standardized difference; ESR = erythrocyte sedimentation rate; ETN = etanercept; MHAQ = modified Health Assessment Questionnaire; no. = number of patients with available data; PtGA = patient global assessment; RA = rheumatoid arthritis; SB4 = ETN biosimilar drug; SJC28 = 28 swollen joint count; TJC28 = 28 tender joint count.

Table 2. Disease status prior to start of treatment, at baseline, and up to 104 weeks follow-up in ETN-naive patients with RA treated with ETN or SB4 in unmatched patient cohorts and in patients switched from ETN to SB4*

	52 weeks before baseline		Baseline		Baseline to 52 weeks		52–104 weeks	
Variable/treatment	No.	Mean (95% CI)	No.	Mean (95% CI)	No.	Mean (95% CI)	No.	Mean (95% CI)
DAS28								
ETN	174	3.8 (3.6, 4.0)	327	4.3 (4.2, 4.5)	268	3.2 (3.0, 3.3)	178	3.0 (2.8, 3.1)
SB4	125	3.5 (3.3, 3.8)	202	4.0 (3.8, 4.2)	134	2.9 (2.7, 3.1)	46	2.5 (2.2, 2.9)
Switch SB4	334	2.6 (2.5, 2.7)	331	2.7 (2.5, 2.8)	235	2.4 (2.3, 2.5)	200	2.4 (2.3, 2.5)
CRP, mg/liter								
ETN	216	13.2 (10.8, 15.6)	397	13.7 (12.1, 15.2)	336	7.6 (6.4, 8.9)	228	6.7 (4.7, 8.7)
SB4	189	8.4 (7.0, 9.8)	282	12.3 (10.4, 14.2)	227	6.9 (5.2, 8.5)	85	5.2 (3.2, 7.2)
Switch SB4	413	4.7 (3.9, 5.4)	399	5.2 (4.4, 6.0)	318	5.2 (4.2, 6.1)	259	4.6 (3.6, 5.7)
ESR, mm/hour								
ETN	194	21.7 (19.5, 23.9)	349	23.4 (21.6, 25.2)	286	17.1 (15.5, 18.7)	185	15.7 (13.8, 17.5)
SB4	140	20.4 (17.8, 22.9)	223	21.8 (19.6, 23.9)	155	15.7 (13.2, 18.2)	57	12.8 (8.5, 17.0)
Switch SB4	367	15.9 (14.5, 17.4)	364	16.2 (14.7, 17.7)	271	15.3 (13.4, 17.1)	231	15.6 (13.4, 17.7)
SJC28 (range 0–28)								
ETN	233	2.7 (2.3, 3.1)	408	4.1 (3.7, 4.5)	350	2.0 (1.7, 2.3)	233	1.5 (1.2, 1.9)
SB4	193	1.9 (1.5, 2.3)	281	3.1 (2.7, 3.5)	225	1.4 (1.1, 1.8)	85	0.9 (0.6, 1.3)
Switch SB4	418	0.9 (0.8, 1.1)	415	0.9 (0.7, 1.1)	316	0.6 (0.5, 0.7)	263	0.6 (0.5, 0.8)
TJC28 (range 0–28)								
ETN	233	4.1 (3.6, 4.7)	408	5.6 (5.2, 6.1)	350	3.0 (2.6, 3.4)	233	2.2 (1.8, 2.6)
SB4	193	3.5 (2.9, 4.1)	281	4.7 (4.2, 5.2)	225	2.8 (2.2, 3.3)	85	2.4 (1.6, 3.1)
Switch SB4	418	1.3 (1.1, 1.6)	415	1.6 (1.3, 1.9)	316	1.2 (0.9, 1.5)	263	0.9 (0.7, 1.2)
PtGA (0–100 mm)								
ETN	226	43.7 (40.7, 46.6)	414	50.0 (47.9, 52.2)	355	34.9 (32.4, 37.4)	237	31.6 (28.6, 34.6)
SB4	191	42.3 (38.9, 45.7)	270	51.1 (48.5, 53.7)	223	37.2 (34.0, 40.4)	90	39.1 (33.4, 44.7)
Switch SB4	420	30.0 (27.7, 32.2)	406	32.2 (29.9, 34.4)	310	30.4 (27.7, 33.1)	260	30.3 (27.5, 33.2)
MHAQ (range 0–3)								
ETN	213	0.6 (0.5, 0.6)	395	0.7 (0.6, 0.7)	351	0.5 (0.4, 0.5)	236	0.4 (0.4, 0.5)
SB4	191	0.5 (0.5, 0.6)	272	0.7 (0.6, 0.7)	224	0.5 (0.4, 0.5)	90	0.5 (0.4, 0.6)
Switch SB4	416	0.4 (0.4, 0.4)	405	0.4 (0.4, 0.5)	309	0.4 (0.4, 0.5)	258	0.4 (0.4, 0.5)

* 95% CI = 95% confidence interval; CRP = C-reactive protein; DAS28 = Disease Activity Score in 28 joints; ESR = erythrocyte sedimentation rate; ETN = etanercept; MHAQ = modified Health Assessment Questionnaire; PtGA = patient global assessment; RA = rheumatoid arthritis; SB4 = ETN biosimilar drug; SJC28 = 28 swollen joint count; TJC28 = 28 tender joint count.

based on an independent samples *t*-test. In the PS-matched analysis, the mean difference between ETN (n = 49) and SB4 (n = 49) for the primary outcome DAS28 at week 52 was 0.03 (95% CI –0.46, 0.52) using a paired samples *t*-test of matched pairs with complete data. Outcomes were consistent between the unmatched and PS-matched analyses for disease status based on DAS28 at week 52 (Figure 2). In the primary and supportive PS models for the matched analyses, 95% CIs included zero, but equivalence between the ETN and SB4 cohorts could not be determined in all models, based on 95% CIs not being entirely confined within the predefined equivalence margin of ±0.6. Equivalence was shown in all PS models, with the sensitivity analysis using all available data for PS-matched pairs.

DAS28 at week 104 (continuous). Before PS matching, the mean DAS28 in the week 52–104 period was 3.0 (95% CI 2.8, 3.1) for the ETN cohort (n = 178) and 2.5 (95% CI 2.2, 2.9) for the SB4 cohort (n = 46) (Table 2 and Figure 1). After PS matching, the mean DAS28 in the week 52–104 period was 3.4 (95% CI 2.7, 4.2) for the ETN cohort (n = 11) and 2.4 (95% CI 1.7, 3.0) for the SB4 (n = 11) cohort. In both the unmatched and PS-matched analyses, there was a reduction in disease activity as assessed by DAS28 at week 104 compared with baseline, in both the ETN- and SB4-treated patients (Figure 1).

Disease response: DAS28 at week 52 and week 104 (categorical). Disease response based on DAS28 at week 52 and week 104 was also assessed as a categorical variable defined by the EULAR response criteria. Before matching, similar proportions of patients in the ETN (n = 221) and SB4 (n = 106) cohorts achieved a good response (40.3% versus 39.6%), a moderate response (28.1% versus 24.5%), or no response (31.7% versus 35.8%) at week 52 (see Supplementary Figure 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25092). After matching based on the primary PS model, 49 patients were in each of the ETN and SB4 cohorts, of which 49.0% and 30.6%, respectively, achieved a good response at week 52 (see Supplementary Figure 3). Moderate responses at week 52 were observed in 14.3% and 34.7%, respectively, of patients in the ETN and SB4 cohorts, and no response in 36.7% and 34.7%. At week 52, the proportions of patients in the switch cohort (n = 173) achieving

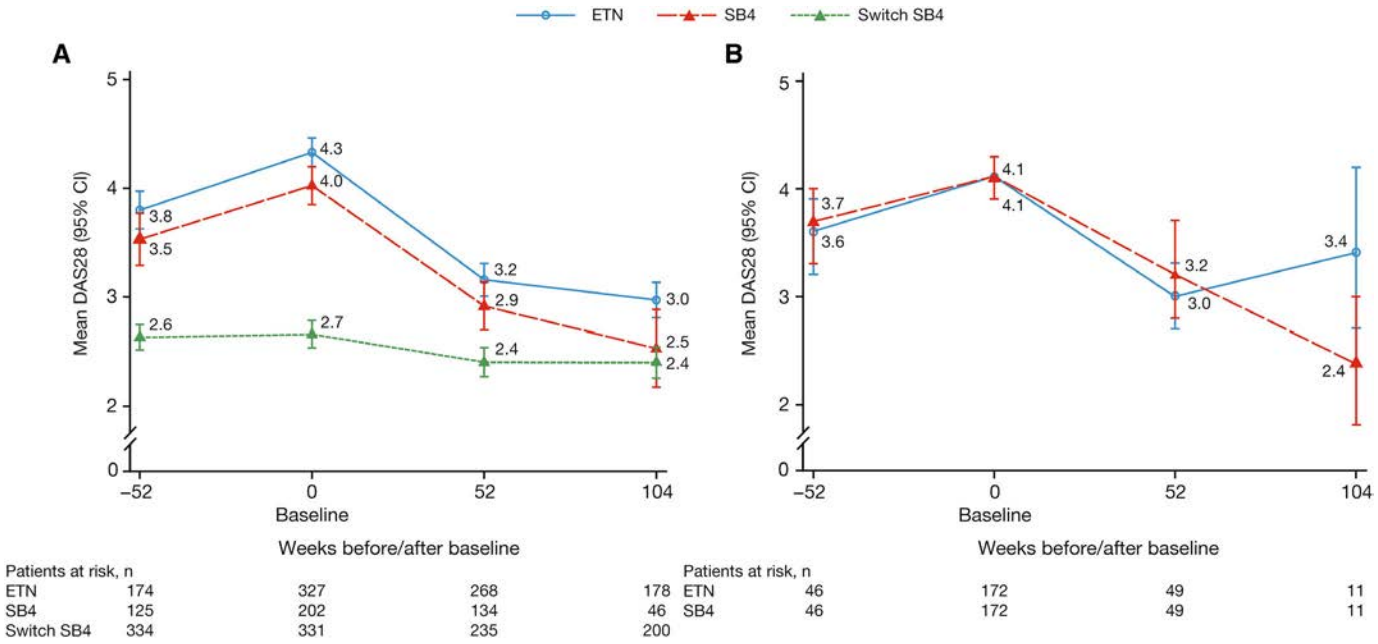

Figure 1. Disease activity expressed as Disease Activity Score in 28 joints (DAS28) over 2 years of treatment in **A**, unmatched, and **B**, propensity score–matched patients with rheumatoid arthritis. Numbers represent the numbers of patients in the unmatched and matched (primary PS model) populations. Data are shown as the mean with 95% confidence interval (95% CI). No imputation of missing data was performed. ETN = etanercept.

good or moderate responses were 9.8% and 14.5%, respectively, whereas 75.7% achieved no response.

Before matching, 37.0% versus 44.1% of patients in the ETN (n = 146) and SB4 (n = 34) cohorts, respectively, achieved a good response, 30.1% versus 14.7% achieved a moderate response, and 32.9% versus 41.2% achieved no response at week 104 (see Supplementary Figure 3, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25092). Based on the primary PS-matched analysis, 27.3% versus 54.6% of patients in the ETN (n = 11) and SB4 (n = 11) cohorts, respectively,

Table 3. Disease status before start of treatment, at baseline, and up to 104 weeks follow-up in ETN-naive patients with RA treated with ETN or SB4 in PS-matched cohorts in the primary PS model*

	52 weeks before baseline		Baseline		Baseline to 52 weeks		52–104 weeks	
Variable/treatment	No.	Mean (95% CI)	No.	Mean (95% CI)	No.	Mean (95% CI)	No.	Mean (95% CI)
DAS28								
ETN	46	3.6 (3.2, 3.9)	172	4.1 (3.9, 4.3)	49	3.0 (2.7, 3.3)	11	3.4 (2.7, 4.2)
SB4	46	3.7 (3.3, 4.0)	172	4.1 (3.9, 4.3)	49	3.2 (2.8, 3.7)	11	2.4 (1.7, 3.0)
CRP, mg/liter								
ETN	64	11.3 (7.8, 14.8)	169	11.5 (9.5, 13.6)	105	7.5 (5.0, 10.0)	18	7.6 (4.4, 10.9)
SB4	64	7.8 (5.6, 10.0)	169	14.4 (11.6, 17.2)	105	7.5 (4.7, 10.3)	18	4.2 (2.1, 6.3)
ESR, mm/hour								
ETN	51	19.6 (15.9, 23.2)	172	21.8 (19.2, 24.3)	63	15.4 (12.4, 18.4)	13	17.1 (10.0, 24.2)
SB4	51	20.9 (17.0, 24.8)	172	22.4 (19.8, 24.9)	63	18.0 (13.6, 22.4)	13	11.0 (6.8, 15.3)
SJC28 (range 0–28)								
ETN	76	2.7 (2.1, 3.4)	172	3.6 (3.1, 4.2)	108	2.0 (1.4, 2.5)	21	2.0 (0.7, 3.3)
SB4	76	1.6 (1.1, 2.1)	172	3.3 (2.8, 3.9)	108	1.7 (1.0, 2.3)	21	0.7 (0.1, 1.3)
TJC28 (range 0–28)								
ETN	76	4.4 (3.3, 5.5)	172	5.1 (4.4, 5.7)	108	3.2 (2.4, 4.0)	21	2.2 (0.7, 3.8)
SB4	76	3.4 (2.5, 4.2)	172	5.0 (4.3, 5.7)	108	2.9 (2.1, 3.7)	21	2.0 (0.5, 3.6)
PtGA (0–100 mm)								
ETN	76	47.2 (41.6, 52.9)	172	50.6 (47.2, 54.1)	111	35.7 (31.2, 40.1)	21	25.3 (14.0, 36.7)
SB4	76	40.7 (35.3, 46.2)	172	49.4 (46.1, 52.6)	111	35.6 (31.3, 39.8)	21	32.6 (22.5, 42.8)
MHAQ (range 0–3)								
ETN	75	0.6 (0.5, 0.7)	167	0.7 (0.6, 0.7)	112	0.5 (0.4, 0.6)	21	0.4 (0.2, 0.6)
SB4	75	0.5 (0.4, 0.6)	167	0.6 (0.6, 0.7)	112	0.5 (0.4, 0.5)	21	0.3 (0.2, 0.5)

* CRP = C-reactive protein; DAS28 = Disease Activity Score in 28 joints; ESR = erythrocyte sedimentation rate; ETN = etanercept; MHAQ = modified Health Assessment Questionnaire; PtGA = patient global assessment; PS = propensity score; RA = rheumatoid arthritis; SB4 = ETN biosimilar drug; SJC28 = 28 swollen joint count; TJC28 = 28 tender joint count.

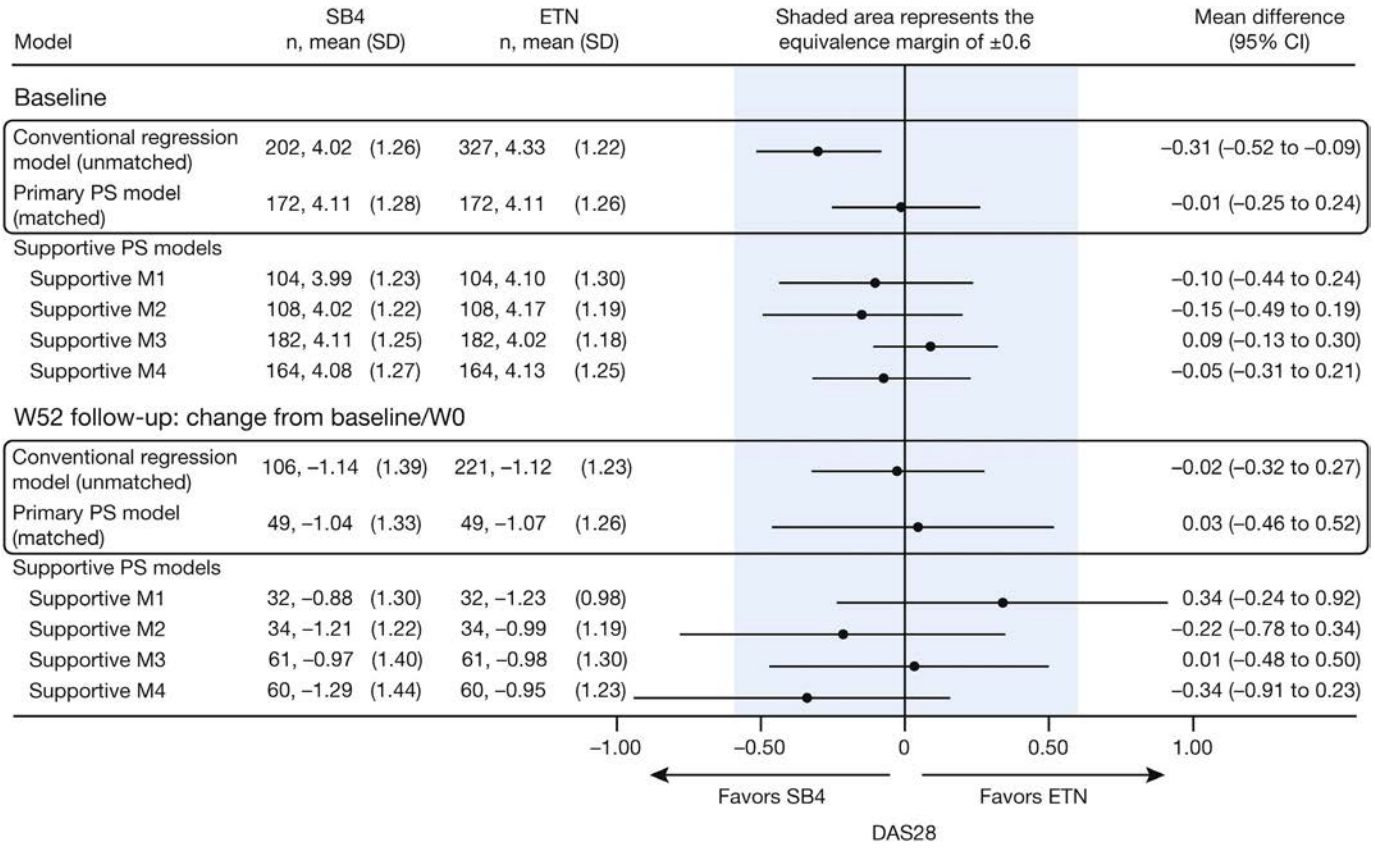

Figure 2. Comparison of effectiveness of etanercept (ETN) and the biosimilar SB4 in patients with rheumatoid arthritis on disease activity (Disease Activity Score in 28 joints [DAS28]) at baseline and week 52 (W52) follow-up for unmatched and propensity score (PS)–matched populations. Baseline shows absolute values and the differences between cohorts at baseline. Week 52 follow-up shows the change from baseline for each cohort and the differences between the cohorts at week 52. Mean differences are shown for baseline values; 1-year follow-up shows the mean difference for change from baseline. Unmatched and primary model models are highlighted. Secondary models are M1, M2, M3, and M4. 95% CI = 95% confidence interval; W0 = week 0.

achieved a good response, 27.3% versus 27.3% achieved a moderate response, and 45.5% versus 18.2% achieved no response at week 104 (see Supplementary Figure 3). At the same time point, the proportions of patients in the switch cohort (n = 148) who achieved good, moderate, or no responses were 11.5%, 15.5%, and 73.0%, respectively.

Drug persistence. In the unmatched sample, the estimated persistence at week 52 was 0.62 (95% CI 0.57, 0.65) for ETN and 0.66 (95% CI 0.60, 0.71) for SB4 (Figure 3A). The overlapping of 95% CIs for the unmatched population indicates similar persistence between ETN and SB4 cohorts. In the matched sample using the primary PS model, the estimated persistence at week 52 was 0.52 (95% CI 0.44, 0.59) for ETN and 0.68 (95% CI 0.61, 0.75) for SB4 (Figure 3B). For switched patients, the estimated persistence at week 52 was 0.80 (95% CI 0.76, 0.83).

In the unmatched sample, the estimated persistence at week 104 was 0.47 (95% CI 0.43, 0.51) for ETN and 0.56 (95% CI 0.49, 0.61) for SB4. In the matched sample using the primary PS model, the estimated persistence at week 104 was 0.37 (95% CI 0.29, 0.44) for ETN (n = 63) and 0.60 (95% CI 0.51, 0.67) for SB4 (n = 24). For nonmedical switch patients, the estimated persistence at week 104 was 0.73 (95% CI 0.69, 0.77).

A small number of patients with RA (n = 5) did not undergo the nonmedical switch from ETN to SB4, including 1 patient at Sørlandet Hospital and 4 at Martina Hansens Hospital; these patients did not contribute to this study. Similarly, a number of patients with RA (n = 48) switched back from SB4 to ETN: 3 at University Hospital of North Norway, 3 at St. Olavs Hospital, 7 at Haukeland University Hospital, 8 at Sørlandet Hospital, and 27 at Martina Hansens Hospital. Reasons for switching back to ETN were often subjective and included lack of efficacy and AEs.

Secondary effectiveness outcome measures. Secondary outcomes for the unmatched analyses at week 52 and week 104 are reported in Table 2. After PS matching based on the primary PS model, both ETN and SB4 cohorts

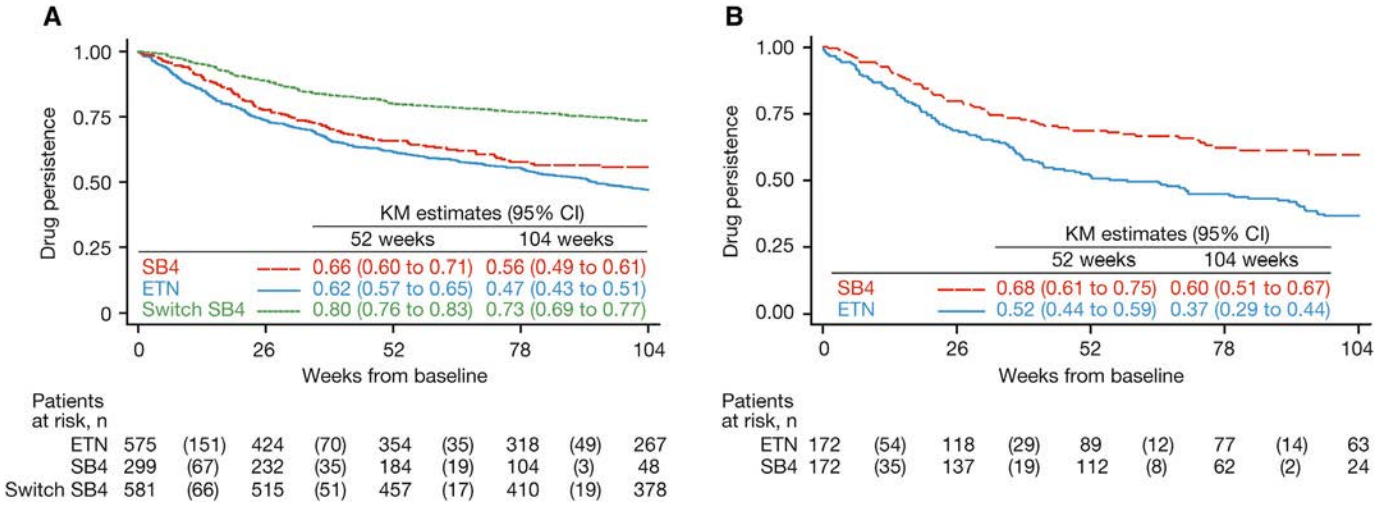

Figure 3. Kaplan-Meier (KM) plots of treatment retention rates among patients with rheumatoid arthritis, **A**, treated with etanercept (ETN) or the biosimilar SB4 or with a nonmedical switch from ETN to SB4, and **B**, treated with ETN or SB4 after propensity score (PS) matching based on the primary PS model. Listed under the graphs are the numbers of patients at risk and the numbers of patients who experienced an event and stopped treatment (shown in parentheses). 95% CI = 95% confidence interval.

experienced improvements from baseline to week 52 and week 104 in measures of disease activity (CRP level, ESR, SJC28, and TJC28) and patient-reported outcomes (PtGA and MHAQ) (Table 3).

Safety. After 104 weeks, 52.9% (n = 304) of ETN, 41.5% (n = 124) of SB4, and 26.2% (n = 152) of nonmedical switch patients had discontinued treatment (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25092). The most common reasons for drug discontinuation were AEs, occurring in 17.4%, 16.4%, and 8.1%, and lack of effect/no effect, occurring in 15.0%, 17.4%, and 9.6% of patients in the ETN, SB4, and switch cohorts, respectively. The most frequent AEs leading to discontinuation were skin involvement and infection. Reasons for stopping treatment in the PS-matched population are summarized in Supplementary Table 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25092.

DISCUSSION

This retrospective database study assessed disease activity and drug persistence in 1,455 patients with RA who were treated with ETN or the biosimilar SB4 for up to 2 years in routine clinical care at 5 outpatient clinics in Norway. Outcomes were compared between treatment cohorts using unmatched and matched analyses. For DAS28, unmatched analyses were based on independent samples *t*-tests, whereas matched analyses used PS models adjusted for confounders including age, sex, and baseline disease status. The primary outcome measure of DAS28 after 52 weeks of treatment was equivalent between cohorts of patients treated with ETN or SB4 based on independent samples *t*-tests and the applied predefined equivalence margin of ±0.6 (16). Consistent results were observed applying the primary PS model, but owing to the low number of patients with complete available disease scores at week 52 in the matched pairs, results could be uncertain. Therefore, the observed results may be limited by the nonsystematic capture of patients' disease scores.

Differences in baseline characteristics in the unmatched cohorts were observed, suggesting a selection bias in treatment initiation; hence, PS matching was investigated as a supportive analysis. The PS-matched models ensured comparability of treatment cohorts at baseline. Persistence using the drug at week 52 was similar between ETN and SB4 treatment cohorts based on the unmatched analysis, as indicated by overlapping 95% CIs. However, in PS-matched cohorts, persistence was greater for SB4 than for ETN at week 52 and week 104.

Although the frequency of drug discontinuation was higher in the ETN cohort than in the SB4 cohort (52.9% versus 41.5%), the reasons for discontinuation were consistent between cohorts and included AEs and lack of effectiveness/no effect. Further, these reasons for discontinuation occurred at similar frequencies between the 2 cohorts.

Published real-world data on SB4 are limited, particularly in patients naive to ETN. A study of the National Romanian Registry of Rheumatic Diseases followed patients with RA for 6 months and found no difference in effectiveness and safety between ETN (n = 123) and SB4 (n = 119) (8). A 2019 systematic review of SB4 real-world data found no difference in effectiveness and safety between switch or ETN-naive patients (9). Similar to

Norway, Denmark also operates a mandatory switch system. An analysis of the Danish DANBIO registry in patients with RA who switched from ETN to SB4 indicated no change in disease activity 3 months post-switch compared with the 3 months pre-switch (10). In addition, the 1-year adjusted retention rate for SB4 post-switch (0.83 [95% CI 0.79, 0.87]) was found to be somewhat lower than for a historical control group for ETN (0.90 [95% CI 0.88, 0.92]). A limitation of these analyses was that data were not reported for ETN or SB4 outcomes in treatment-naive patients who initiated treatment on ETN or SB4 (10).

Across different countries and regions, a nonmedical switch may follow a mandatory or nonmandatory switch model. Countries with mandatory switch models, including Denmark and Norway, have been shown to be more successful in using biosimilars than countries using nonmandatory models. In 2015, the infliximab biosimilar constituted as much as 90.6% of the total infliximab prescribed in Denmark 4 months after the patent expiration of the reference drug (18).

As for all observational studies, this study's limitations relate to measured and unmeasured confounding factors, attrition, and missing data. To counteract these limitations, we aimed to analyze the data and report the results in accordance with observational study recommendations, including GRACE and STROBE. The use of propensity statistics as supportive analyses mitigated the risk of selection bias, simulating a randomized study design. We analyzed the primary outcome measure with different propensity matching adjustments to explore the robustness of the results. In matched pairs analysis, missing data may have a substantial impact. Typically, a matched pair at baseline may only have data for 1 of them at week 52 and then be lost. Therefore, a sensitivity analysis was conducted using all available data, and produced outcomes consistent with those from the primary matched pairs analyses.

This study aimed to closely observe the recommendations of the EULAR Points to Consider When Analyzing and Reporting Comparative Effectiveness Research with Observational Data in Rheumatology (see Supplementary Table 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25092) (7,13). Although most recommendations were followed closely, there were some minor deviations. First, the numbers of patients who stopped and/or changed therapies over time may not have been fully captured. Second, although other analyses that are not reported here were performed, a sensitivity analysis investigating the missing data pattern was conducted by not excluding matched pairs with partially missing data. Finally, although a full statistical analysis plan had not been prepared, an outline was developed in advance of this study.

In conclusion, after 52 weeks of treatment, disease outcomes based on DAS28 were comparable between cohorts of patients treated with ETN or SB4, and equivalence for DAS28 was demonstrated based on independent sample *t*-tests. Consistent results were observed applying the primary PS model but should be interpreted with caution owing to missing patient disease scores at week 52. Persistence was similar at week 52 between the ETN and SB4 cohorts in the unmatched populations but greater for SB4 in the PS-matched analyses.

ACKNOWLEDGMENTS

The authors thank Bjørg Tilde-Fevang for providing data from Haukeland University Hospital and Ulrich Freudensprung of Biogen International GmbH for statistical guidance. They are also grateful to all doctors, nurses, secretaries, and patients at the participating outpatient clinics who have contributed in daily clinical practice to make this real-life study possible. Medical writing support for the preparation of the manuscript was provided by David Pertab, PhD, Excel Medical Affairs.

ROLE OF THE STUDY SPONSOR

Biogen International 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 Biogen International.

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. Haugeberg 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. Haugeberg, Bakland, Rødevand, Hansen, Diamantopoulos, Pripp.

Acquisition of data. Haugeberg, Bakland, Rødevand, Hansen, Diamantopoulos.

Analysis and interpretation of data. Haugeberg, Pripp.

REFERENCES

1. Bergstra SA, Branco JC, Vega-Morales D, et al. Inequity in access to bDMARD care and how it influences disease outcomes across countries worldwide: results from the METEOR-registry. Ann Rheum Dis 2018;77:1413–20.
2. Putrik P, Ramiro S, Kvien TK, et al. Inequities in access to biologic and synthetic DMARDs across 46 European countries. Ann Rheum Dis 2014;73:198–206.
3. Putrik P, Ramiro S, Kvien TK, et al. Variations in criteria regulating treatment with reimbursed biologic DMARDs across European countries: are differences related to country's wealth? Ann Rheum Dis 2014;73:2010–21.
4. Emery P, Vencovsky J, Sylwestrzak A, et al. A phase III randomised, double-blind, parallel-group study comparing SB4 with etanercept reference product in patients with active rheumatoid arthritis despite methotrexate therapy. Ann Rheum Dis 2017;76:51–7.
5. Emery P, Vencovský J, Sylwestrzak A, et al. 52-week results of the phase 3 randomized study comparing SB4 with reference etanercept in patients with active rheumatoid arthritis. Rheumatology (Oxford) 2017;56:2093–101.
6. Kay J, Schoels MM, Dörner T, et al. Consensus-based recommendations for the use of biosimilars to treat rheumatological diseases. Ann Rheum Dis 2018;77:165–74.
7. Lauper K, Kedra K, de Wit M, et al. A systematic review to inform the EULAR points to consider when analyzing and reporting comparative

effectiveness research with observational data in rheumatology. Ann Rheum Dis 2020;79 Suppl 1:2.

8. Codreanu C, Popescu CC, Mogoşan C, et al. Efficacy and safety of original and biosimilar etanercept (SB4) in active rheumatoid arthritis: a comparison in a real-world national cohort. Biologicals 2019;62: 27–32.
9. Ebbers HC, Pieper B, Issa A, et al. Real-world evidence on etanercept biosimilar SB4 in etanercept-naïve or switching patients: a systematic review. Rheumatol Ther 2019;6:317–38.
10. Glintborg B, Loft AG, Omerovic E, et al. To switch or not to switch: results of a nationwide guideline of mandatory switching from originator to biosimilar etanercept. One-year treatment outcomes in 2061 patients with inflammatory arthritis from the DANBIO registry. Ann Rheum Dis 2019;78:192–200.
11. Madenidou AV, Jeffries A, Varughese S, et al. Switching patients with inflammatory arthritis from Etanercept (Enbrel®) to the biosimilar drug, SB4 (Benepali®): a single-centre retrospective observational study in the UK and a review of the literature. Mediterr J Rheumatol 2019;30 Suppl 1:69–75.
12. Tweehuysen L, Huiskes VJ, van den Bemt BJ, et al. Open-label, non-mandatory transitioning from originator etanercept to biosimilar SB4: six-month results from a controlled cohort study. Arthritis Rheumatol 2018;70:1408–18.
13. Courvoisier D, Lauper K, Bergstra SA, et al. Points to consider when analysing and reporting comparative effectiveness research with observational data in rheumatology. Ann Rheum Dis 2020;79 Suppl 1:2.
14. Dreyer NA, Bryant A, Velentgas P. The GRACE checklist: a validated assessment tool for high quality observational studies of comparative effectiveness. J Manag Care Spec Pharm 2016;22:1107–13.
15. Von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007;370: 1453–7.
16. Fransen J, van Riel PL. The Disease Activity Score and the EULAR response criteria. Clin Exp Rheumatol 2005;23 Suppl 39:S93–9.
17. Van Gestel AM, Haagsma CJ, van Riel PL. Validation of rheumatoid arthritis improvement criteria that include simplified joint counts. Arthritis Rheum 1998;41:1845–50.
18. Jensen TB, Bartels D, Saedder EA, et al. The Danish model for the quick and safe implementation of infliximab and etanercept biosimilars. Eur J Clin Pharmacol 2020;76:35–40.

Arthritis Care & Research
Vol. 75, No. 9, September 2023, pp 1996–2010
DOI 10.1002/acr.25098

American College of Rheumatology
Empowering Rheumatology Professionals

REVIEW

Recruitment and Retention Strategies for Underrepresented Populations and Adults With Arthritis in Behavioral Interventions: A Scoping Review

Christine A. Pellegrini, Sara Wilcox, Katherine E. DeVivo, and Scott Jamieson

Objective. To identify strategies used to recruit and retain underrepresented populations and populations with arthritis or fibromyalgia (FM) into behavioral programs targeting exercise, physical activity, or chronic disease self management.

Methods. Five bibliographic databases were searched for articles published between January 2000 and May 2022. The search focused on strategies and best practices for recruiting and retaining underrepresented populations or populations with arthritis or FM into disease self-management or physical activity/exercise programs. Abstracts and full-text articles were screened for inclusion by 2 independent reviewers, and 2 reviewers extracted data from included articles.

Results. Of the 2,800 articles, a total of 43 publications (31 interventions, 8 reviews, 4 qualitative/descriptive studies) met criteria and were included. The majority of studies focused on physical activity/exercise (n = 36) and targeted African American (n = 17), Hispanic (n = 9), or arthritis populations (n = 7). Recruitment strategies that were frequently used included having race- or community-matched team members, flyers and information sessions in areas frequented by the population, targeted emails/mailings, and word of mouth referrals. Retention strategies used included having race- or community-matched team members, incentives, being flexible, and facilitating attendance. Most studies used multiple recruitment and retention strategies.

Conclusion. This scoping review highlights the importance of a multifaceted recruitment and retention plan for underrepresented populations and populations with arthritis or FM in behavioral intervention programs targeting exercise, physical activity, or chronic disease self management. Additional research is needed to better understand the individual effects of different strategies and the costs associated with the various recruitment/retention methods in underrepresented populations and populations with arthritis.

INTRODUCTION

With over 54 million adults with arthritis, arthritis is the leading cause of disability in the US (1). Participation in regular aerobic physical activity is an effective strategy for reducing pain (2–6) and fatigue (5,7–9), and improving physical function (2,3,6,10), quality of life (4), and psychological wellness (11). Despite these recommendations, only 36.2% of adults with arthritis in the US met the aerobic physical activity guidelines in 2015 (12).

The Centers for Disease Control and Prevention along with the Osteoarthritis Action Alliance have several recognized arthritis-appropriate evidence-based programs focused on physical activity and chronic disease self management (13,14). These programs are effective at increasing physical activity and improving symptoms of arthritis (15–17); however, participation in behavioral programs and public health initiatives is typically lower among underrepresented populations (18), including racial and ethnic minorities (19), rural populations (20), and those with a disability (21,22). Challenges to participation in behavioral programs or clinical trials that some of these populations face include distrust of research (23), lack of access, transportation or resources (24,25), time commitments involved with participation (26), and

The contents herein are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention.

Supported by cooperative agreement from the Centers for Disease Control and Prevention (grant U48-DP-006401).

Christine A. Pellegrini, PhD, Sara Wilcox, PhD, Katherine E. DeVivo, MPH, Scott Jamieson, MS: University of South Carolina, Columbia.

Drs. Pellegrini and Wilcox contributed equally to this work.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.25098&file=acr25098-sup-0001-Disclosureform.pdf.

Address correspondence via email to Christine A. Pellegrini, PhD, at cpellegrini@sc.edu.

Submitted for publication September 7, 2022; accepted in revised form February 2, 2023.

SIGNIFICANCE & INNOVATIONS

- Many challenges exist in recruiting underrepresented populations and populations with arthritis or fibromyalgia (FM) to participate in physical activity and disease self-management programs, as well as maintaining high levels of engagement and retention throughout the duration of the program.
- Identifying recruitment and retention strategies can help future physical activity or chronic disease self-management programs meet enrollment targets and maintain high levels of retention throughout these evidence-based programs.
- This is the first scoping review focused specifically on recruitment and retention strategies used to reach underrepresented populations and populations with arthritis or FM.
- A multifaceted plan may be necessary to recruit and retain underrepresented populations and populations with arthritis or FM in behavioral intervention programs targeting exercise, physical activity, or chronic disease self management.

lack of awareness (27). There have been several studies examining the Walk With Ease program (28,29), a physical activity program for adults with arthritis, among African American (30), Hispanic (31), and rural populations (29); However, the majority of studies include samples that were >50% White (28,29,32) and highly educated (28,32). Increasing recruitment of underrepresented populations within these behavioral programs is critical to expand generalizability and eliminate health inequalities seen in these groups.

Maintaining high levels of engagement and retention in physical activity and disease self-management programs can also be challenging among underrepresented populations (33,34). Many populations face barriers that interfere with their ability to participate, including challenges with transportation (35,36), community mobility barriers (35), and low socioeconomic status (37). To achieve the greatest benefit from these programs, retention throughout the entire program is critical.

Although reviews of recruitment and retention strategies in underrepresented populations exist (38–40), they focus on a wide range of interventions, populations, clinical disorders, and medical and behavioral outcomes. Unique barriers and enablers to participation and retention in exercise/physical activity and chronic disease self-management programs may exist, necessitating unique strategies. Thus, the purpose of this scoping review was to identify strategies that have been used to recruit and retain underrepresented populations and populations with arthritis or fibromyalgia (FM) into behavioral programs targeting exercise, physical activity, or chronic disease self management. Further, we aimed to identify the strategies that appear most effective for recruitment and retention. Identifying these strategies will help future programs recruit participants from typically underserved populations, including those with lower socioeconomic status, as well as help retain participants in evidence-based programs.

METHODS

Our protocol was guided by the Joanna Briggs Institute manual for evidence synthesis (41) and the Preferred Reporting Items for Systematic Reviews and meta-analyses guide for scoping reviews (42).

Eligibility criteria. We included studies with reported strategies and/or best practices for recruiting and retaining underrepresented populations and populations with arthritis in behavioral programs (exercise, physical activity, or chronic disease self management). Peer-reviewed journal articles were included if they were published between 2000 and 2022; were written in English; focused specifically on recruitment or retention or reported recruitment or retention strategies; were focused on individuals with arthritis, FM, or underrepresented populations (including racial and ethnic minorities, rural residence, low socioeconomic status, or individuals with arthritis or FM); and included an intervention or program for physical activity, exercise, or chronic disease self management. Reviews and qualitative studies were included if they met these criteria.

Studies that were conducted outside of the US were excluded due to potential differences in recruitment and commonly used retention strategies. Conference abstracts, dissertations, and records from ClinicalTrials.gov were excluded. While studies did not need to focus specifically on individuals with arthritis, we excluded studies if they focused entirely on a clinical diagnosis or diagnoses other than arthritis unless the study was focused on prevention (e.g., prevention of heart disease in a sample of African Americans). Because we were focused on strategies that would inform recruitment and retention of adults with arthritis in behavioral interventions, we excluded studies that focused on youth/adolescents, pregnant/postpartum women, substance use, and mental disorders.

We were interested in studies conducted outside of clinical settings (i.e., in the community) and as a result excluded studies where study recruitment was conducted entirely through a clinical setting or if the intervention was delivered entirely by clinical or hospital staff. We also excluded faith-based interventions where only churches were recruited, and the church was the sole delivery site because strategies to recruit and intervene at the organizational level differ from strategies to recruit and intervene at the individual level. Studies that only examined participant-level factors related to recruitment or retention were excluded since they would not provide examples of strategies to enhance recruitment or retention. Finally, laboratory-based or one-time studies were excluded.

Information sources. To identify potentially relevant articles, the following bibliographic databases were searched from

January 2000 to May 2022: Web of Science, Ovid Medline, CINAHL, PsycINFO, and Cochrane. The search strategies were drafted in collaboration with an experienced librarian and further refined through investigative team discussions. The final search strategy for Ovid Medline is shown in the Supplementary Materials (available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25098/abstract). Final search results were exported into EndNote and then into Covidence systematic review software (Veritas Health Innovation). Covidence removed identified duplicates, and a study investigator (SW) removed additional duplicates not detected by Covidence. The electronic database search was supplemented by scanning reference lists in articles chosen for full-text review.

Selection of sources of evidence. An initial list of inclusion and exclusion criteria was established, and 4 reviewers (CAP, SW, KED, and SJ) went through 2 rounds of screening 50 publications (100 total). After each round, the reviewers discussed challenges and disagreements and revised definitions of the inclusion and exclusion criteria. Publications that were identified in our search were evenly distributed across the 4 reviewers. Each title and abstract was reviewed for inclusion by 2 independent reviewers using Covidence software. Publications that could not be excluded based on inclusion/exclusion criteria were retained for the full-text review phase. Disagreements regarding publication inclusion were resolved by 2 reviewers via discussion until consensus was achieved.

Articles that were retained after the title and abstract reviews were uploaded to Covidence and evenly distributed across the 4 reviewers. Each full article was reviewed for inclusion by 2 independent reviewers. In the event of a discrepancy, 2 reviewers discussed until consensus was met.

Data extraction and data items. A data extraction form was developed within Covidence by 2 reviewers. The publications retained after the full-text review were evenly distributed across the 4 reviewers, and a single reviewer completed extraction. The following items were extracted from each article: study type (intervention, review, qualitative or descriptive, protocol/design), study population, study aims, inclusion and exclusion criteria, study setting and locations, type of intervention (e.g., physical activity/exercise, disease self management), intervention duration, sample size and characteristics, dates of recruitment, recruitment strategies, recruitment rates and goal, retention strategies, retention/attrition rates, and retention goals. Study investigators (CAP and SW) completed a review of data extracted to ensure accuracy.

Synthesis of results. The studies were grouped by type of program (disease self management, physical activity/exercise) and article type (intervention, descriptive/review). The programs implemented, target population, and recruitment and retention strategies were summarized.

RESULTS

Selection of sources of evidence. As shown in Figure 1, our database search identified 2,800 publications, and our review of reference lists generated an additional 30 publications. The removal of duplicates left us with 1,377 publications for title/abstract screening. The title/abstract screening excluded 1,226 publications, leaving 152 full-text articles to be assessed for eligibility. The most common reasons for exclusion of full-text articles were because they were not specifically focused on recruitment or retention (n = 46), were not conducted in the US (n = 18), were a conference abstract or protocol (n = 16), and because they recruited churches (n = 9). A total of 43 publications representing 41 studies were extracted and included in this review.

Intervention and population characteristics. Most articles were intervention papers (n = 24) (43–66) or protocol or design papers describing interventions (n = 7) (67–73). The remaining articles were qualitative or descriptive studies (n = 4) (74–77) or reviews (n = 8) (21,78–84).

Of the interventions and protocol/design papers describing interventions (Table 1), the studies were primarily focused on physical activity or exercise (n = 25) (44–48,50–55,58,59,61–66,68–73) with only 5 studies examining disease self-management programs (43,49,56,57,60,67). Of the reviews included (Table 2), 1 focused on recruitment strategies used for disease self-management programs (78), and 7 summarized recruitment and/or retention strategies for physical activity or exercise interventions (21,79–84). Of the descriptive/qualitative papers (Table 3), all focused on physical activity or exercise. One descriptive paper summarized lessons learned regarding the recruitment and retention of hard-to-reach populations for physical activity studies among individuals with arthritis (77). Of the 3 qualitative papers, 1 included interviews with African American men (76), 1 included interviews with Hispanic men (75), and 1 included focus groups with older African American adults (74). In the qualitative papers, participants were asked to discuss suggested strategies to recruit and/or retain participants from their demographic group.

Among disease self-management programs, target populations included adults with lupus (57) and those with arthritis (49,56). Three study samples were primarily African American individuals (43,57,60,67) and 2 studies included >35% Hispanic or Latino individuals (49,56). Most studies included both men and women (49,56,60); however, some only targeted men (43,67) or women (57). The intervention duration of the disease self-management programs ranged from 6 weeks (49,57) to 4 months (56), with follow-up lasting between 12 weeks (43,67) and 18 months (57).

Among physical activity or exercise interventions, a few studies targeted populations with arthritis, including

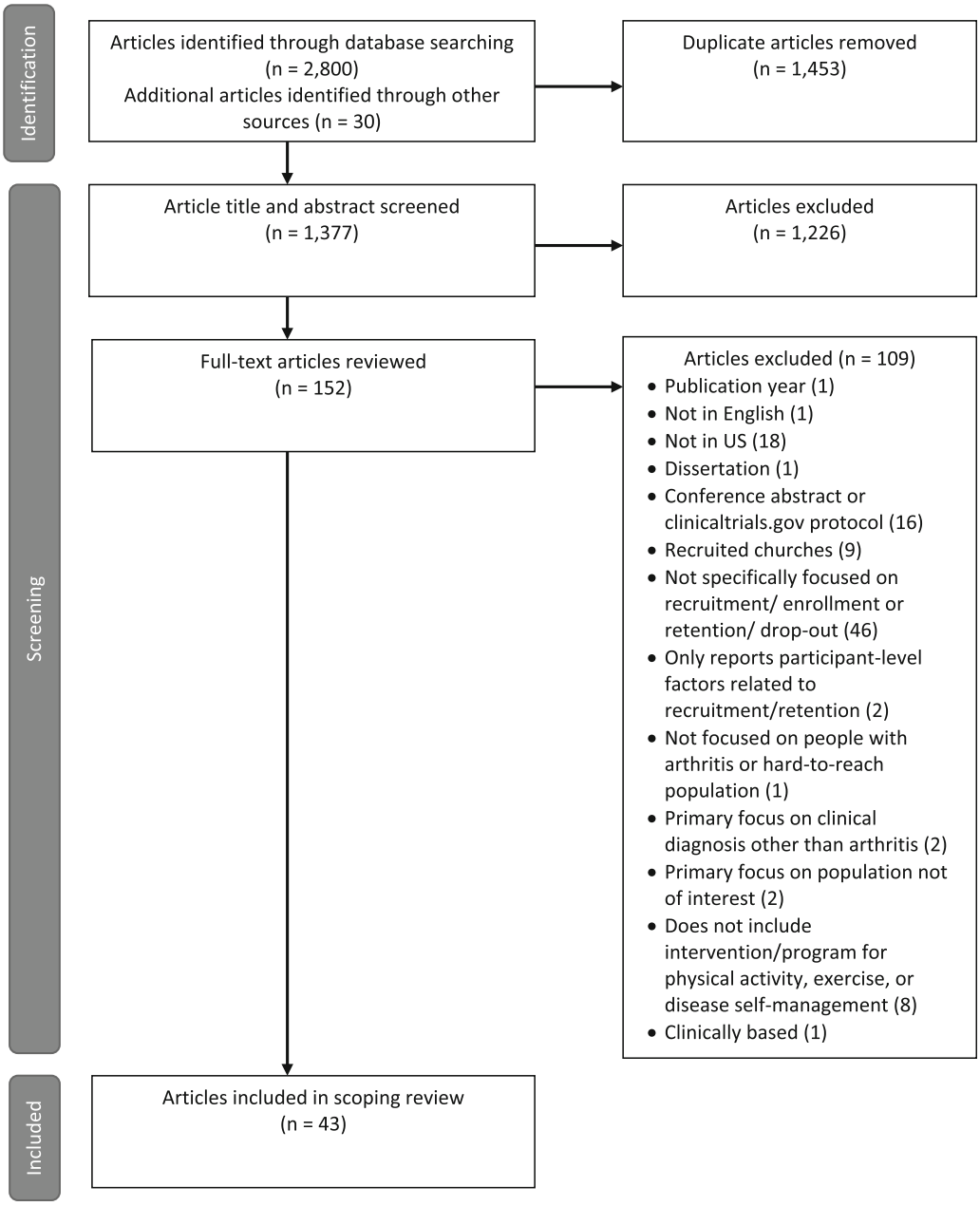

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram.

osteoarthritis (58,73), rheumatoid arthritis (73), and FM (50). The majority of studies included primarily African American (44–48,51,52,54,55,59,64–66) or Hispanic or Latino populations (53,54,62,66,69,71,72). The reviews of strategies used in physical activity/exercise interventions primarily focused on underserved or underrepresented (21,79,82–84) or African American populations (80). One study focused entirely on men (72) and 10 studies only included women (46–48,52,54,59,63–66). Of the physical activity interventions with reported intervention durations, many were ≤ 6 months in duration (46,48,50–52,55,63–65,69,72), whereas other interventions lasted 12 months (47,53,58,59,61,70,71).

Recruitment strategies and results in disease self-management programs. Recruitment strategies were described in the 1 review (78) and in all 5 studies of disease self-management programs (Tables 2 and 3), and each used multiple strategies (43,49,56,57,60,67). Three of the programs, as well as the review article, described the importance of building a relationship with the community and ensuring that those responsible for recruiting participants were race- and community-matched and trained in areas such as building trust and reaching minority individuals (43,57,60,67,78). In fact, 1 study used past participants to recruit new participants (43,67). All studies used one-on-one or face-to-face recruitment strategies, including talks,

Table 1. Intervention and population characteristics of included articles*

Author	Study location	N	Sample target population	Sample characteristics	Intervention duration	Follow-up duration
Disease self-management programs						
Drenkard et al, 2020 (57)	Atlanta, GA	699	AA women with SLE	100% AA; 100% female; mean ± SD 47.4 ± 14.0 years (usual care), mean ± SD 49.8 ± 12.3 years (WELL); 100% SLE	6 weeks	18 months
Goeppinger et al, 2009 (57)	Stanford University and UNC	921	Adults with arthritis	17% AA, 37% Hispanic, 44% White (intervention); 17% AA, 38% Hispanic, 44%; White (control); 15% male (intervention); 14% male (control); mean ± SD 54.3 ± 12.2 years old (intervention); mean ± SD 53.4 ± 12.3 years (control); 51% OA, 33% RA, 30%; FM, 14% other arthritic condition	4 months	9 months
Graham et al, 2018 (43) and Valdez et al, 2021 (67)	Springfield, MA area	245	Low/no income AA men	92% AA, 6% Latino, 2% Other; 100% male; 52% >45 years old	12 weeks	12 weeks
Ramsay et al, 2020 (60)	Detroit, MI	453	Older minority adults	86% AA, 9% multiracial, 5% White; 74% female; 30% ≥70 years old	7 weeks	NR
Reid et al, 2014 (49)	New York City, NY	201	Older adults with arthritis	32% AA, 43% Hispanic, 25% White; 82% female (adapted); 77% female (original); mean ± SD 75.5 ± 8.8 years (adapted), mean ± SD 73.3 ± 8.8 years (original)	6 weeks	24 weeks
Physical activity/ exercise programs					12 months	12 months
Adams et al, 2015 (59)	Columbia, SC and Florence, SC	530†	AA women	100% AA (for both studies); 100% female (SISTAS)	12 months	12 months
Carthron et al, 2018 (58)	Communities near UNC	173	Adults with OA	11% Black, 83% White, 6% other; 65% female; mean ± SD 65.4 ± 8.9 years old; 100% OA	NR	24 months
Chang et al, 2014 (73)	NR	340	Adults with RA or knee OA	12% AA, 72% White, 16% other (RA); 35% AA, 52% White, 13% other (knee OA); 84% female (RA); 60% female (knee OA); mean ± SD 54.8 ± 13.7 years (RA); mean ± SD 63.1 ± 12.9 years (knee OA)	24 weeks	24 weeks
Garcia et al, 2018 (72)	Tucson, AZ	50	Hispanic men	100% Hispanic; 100% male; mean ± SD 43.3 ± 11.4 years old	NR	24 months
Gilbert et al, 2018 (62)	Micropolitan community in southeast Iowa	55	Latinos	100% Latino; 64% female (RDD); 77% female (RDS); 37 years old (RDD); 34 years old (RDS)	12 months	12 months
Hammerback et al, 2012 (61)	Seattle, WA	131	Low-income, ethnically diverse	51% White; mean 70 years old	6 months	6 months
Hines-Martin et al, 2009 (55)	Three low-income inner-city communities in a southern urban city	104	Low income and primarily AA women	72% AA, 19% White, 9% American Indian (phase 1); 88% AA, 10% White, 2% American Indian (phase 2); 100% female	NR	NR
Keller et al, 2005 (54)	NR	47†	AA women	100% AA (study 1); 100% Latina (study 2); 100% female (for both studies)	NR	NR
Keller et al, 2010 (66)	NR	47†	AA and Latino/Mexican American women	100% female	12 months	12 months

(Continued)

Table 1. (*Cont'd*)

Author	Study location	N	Sample target population	Sample characteristics	Intervention duration	Follow-up duration
King et al, 2017 (41)	Santa Clara and San Mateo counties, CA	284	Older Latino adults	98% Hispanic; 79% female; mean ± SD 62.3 ± 8.4 years old	12 months	18 months
King et al, 2020 (53)	San Francisco Bay, CA area	350†	Older Latino adults	100% Latino; 100% female; mean ± SD 52 ± 9 years old	24 months	24 months
King et al, 2021 (70)	San Francisco Bay, CA area	300	Adults living in or around affordable senior public housing settings	3% AA, 17% Latino, 17% Asian, 2% >1 race; 74% female; mean ± SD 69.5 ± 10.3 years old	6 months	12 months
Nies et al, 2001 (52)	Detroit, MI area	175	AA women	66% AA; 100% female	24 weeks	48 weeks
Okhomina et al, 2020 (51)	Jackson, MS	375	AA adults	100% AA; women: mean ± SD 62.2 ± 10.2 years old; men: mean ± SD 61.8 ± 9.1 years old; 77% female	12 weeks or 24 weeks	52 weeks
Park et al, 2021 (50)	Boston, MA	226	Adults with FM	100% FM	12 weeks (phase 2); 12 months (phase 3)	12 weeks (phase 2); 12 months (phase 3)
Phillips-Caesar et al, 2015 (69)	New York, NY area	498†	AA and Hispanic adults	48% AA, 56% Hispanic; 75% female; mean ± SD 50 ± 12.6 years old		
Rubin et al, 2002 (68)	Multisite (27 sites)	3,234	Adults at risk for diabetes	45% racial or ethnic minority; 66% female; 20% ≥60 years old	NR	NR
Sharp et al, 2008 (65)	Chicago, IL	213	AA women	100% AA; 100% female; mean ± SD 46.0 ± 8.4 years old	6 months	18 months
Staffileno et al, 2017 (64)	NR	35	AA women	100% AA; 100% female; mean ± SD 33.0 ± 7.1 years old	12 weeks	12 weeks
Sullivan-Marx et al, 2011 (48)	Metropolitan northeastern city	52	Older AA women	100% AA; 100% female; completers: mean ± SD 80.4 ± 9.0 years old; non-completers: mean ± SD 77.7 ± 8.9 years old	16 weeks	NR
Wilbur et al, 2001 (63)	Chicago, IL	173	AA and White women	37% AA; 64% White; 100% female; 63% 45–49 years old; 37% ≥50 years old	24 weeks	48 weeks
Wilbur et al, 2006 (46)	Chicago, IL	281	Older adult rural women	100% AA; 100% female; mean ± SD 48.5 ± 6.0 years old	24 weeks	48 weeks
Wilbur et al, 2013 (47)	Chicago IL	288	AA women	100% AA; 100% female	48 weeks	NR
Yancey et al, 2001 (44)	NR	NR	AA women	100% AA; 100% female	NR	NR
Yancey et al, 2003 (45)	Richmond, VA	349	Low-income AA adults	87% AA; 12% White; 95% female; 50% ≥45 years old	NR	NR

* AA = African American; FM = fibromyalgia; NR = Not Reported; OA = osteoarthritis; RA = rheumatoid arthritis; RDD = random digit dialing; RDS = respondent-driven sampling; SISTAS = Sistas Inspiring Sistas Through Activity and Support; SLE = systemic lupus erythematosus; UNC = University of North Carolina; WELL = Women Empowered to Live with Lupus.
† For both studies.

Table 2. Recruitment and retention strategies of included disease self-management program reviews and physical activity/exercise program reviews*

Author	Target population	No. of articles	Recruitment strategies and recommended sites	Retention strategies
Horrell et al, 2017 (78)†	All populations	39	Local media (in 55% of studies), partner with trusted professionals in community (55%), flyers/posters/brochures in community settings (55%), advertisements in patient organizations (listservs, databases, magazines) (27%), newsletters (36%), promotion booths (27%), word of mouth (27%), presentations (27%), recruitment events (18%), internet advertisements (1%)	NR
Carroll et al, 2011 (84)‡	Underserved populations	38	Formation of partnerships between research team and community organizations; research team members ethnically and/or linguistically matched to study population; recruitment strategies and materials developed collaboratively with community organization or advisory board; targeted advertisements in newspapers, on television, radio, or posters; targeted mailings and/or telephone calls via use of phone lists from organizations; direct neighborhood canvassing (e.g., street outreach, visits, or presentations to neighborhood businesses or organizations, or door-to-door recruitment to homes); community or health fair presentations; less common but innovative approaches (e.g., social marketing via direct mailings from a magazine subscription list, pay stub advertisements, development of a DVD or audio material for lower literacy groups, use of community or youth leaders to directly appeal, timing or linking recruitment to a cultural tradition, storytelling, carefully planning culturally tailored and interactive pre-intervention/ informational sessions); social networking/word of mouth	Use of formative research and ongoing process evaluation; administrative strategies: obtaining multiple alternate contact numbers for participants, making multiple contacts with participants during study period, maintain detailed databases and tracking registries for participants, providing financial or other incentives, transportation, childcare; interpersonal strategies: persistent study team, having a positive, caring attitude toward participants, demonstrate cultural sensitivity and respect for participants' life situations, meet regularly with community advisors and study staff for problem solving, be flexible or accommodating to location and timing of study visits, having bicultural, bilingual, or ethnically matched study team members, ethnically matched individuals involved in leadership, research questions, and study design, have study staff with strong interpersonal qualities and well trained in the study procedures and protocol
Conn et al, 2013 (83)‡	Underrepresented populations§	71	Involve underrepresented members on the research team and in the design or implementation of recruitment strategies; target locations or activities frequented by target population (e.g., churches, health care facilities, schools, businesses, worksites, ESL programs, vocational training classes, beauty salons, grocery stores, worksites, ethnic civic/social/fraternal organizations; community events (e.g., health fairs); snowball sampling from friends and families of participants; targeted mass mailings, newspaper announcements, and flyers; conduct studies at specific locations to facilitate engagement (e.g., churches, community settings, fitness centers, hospitals); address cost considerations (e.g., provide childcare, choose facilities near transportation, cover transportation costs)	Include underrepresented members within study team; involve group members in retention activities; provide services to facilitate participation (e.g., transportation, childcare, alternative participation times for working subjects); employ staff who are members of the group to collect data; use measures that are culturally relevant, and if necessary, translate and back-translate measures; provide alternatives for low-literacy participants; implement efforts to find participants who may have relocated
El Masri et al, 2022 (82)‡	CALD populations	19	Recruited from various places: religious establishments, religious or cultural events, community organizations or centers, community events, stores and markets, communities with larger concentrations of CALD groups, work break sites, flyers/brochures, newspaper or online advertisements, CALD-specific radio and television,	NR

(Continued)

Table 2. (*Cont'd*)

Author	Target population	No. of articles	Recruitment strategies and recommended sites	Retention strategies
			advertisements, word of mouth; use of community health workers, incentives, forming partnerships with community, organizations, trained CALD-matched research team members	
Foster et al, 2011 (81)	All populations	47	Flyers, posters, advertisements, mail drops (31 studies); Newsletters/newspaper articles (18 studies); word-of-mouth (12 studies); internet/intranet/television/radio (11 studies); direct mail (10 studies); information sessions or health screenings (7 studies); telephone calls (6 studies); referrals (5 studies); medical campaign/press conference (4 studies)	NR
James et al, 2017 (80)‡	African American	56	Recruit from clinical setting (private medical practices, public health clinics/community health centers, hospitals, electronic medical records); flyers, newspaper and public service announcements; electronic websites, listservs, email, or postal mailings; recruit from churches or college campus; use of financial incentives	Provision of a peer advocate; texting or calling participants who did not participate; incremental incentives for each assessment phase; offering church/site financial incentive if member response rate is high; oversample in anticipation of attrition; increase human interaction
Kosma et al, 2004 (79)‡	Underrepresented populations	12	Active strategies: telephone, direct mail, face-to-face (door-to-door) recruitment, study presentations at churches, physician referrals; passive strategies: study flyers, press release, posters, television and radio announcements, newspaper advertisements, internet publicity, social networking, word of mouth; collaboration with the community and establishment of health promotion resources and programs, frequent feedback about progress and study results; staff matched to demographic characteristics of target population; overcome transportation barriers; incentives; dealing with schedule conflicts	NR
Mendoza-Vasconez et al, 2016 (21)‡	Racial and ethnic minorities, low SES, and adults with physical disabilities	NR	Recruiting racial and ethnic minorities (research staff reflective of the participants' cultural background, seek partnerships with communities, addressing safety concerns [e.g., lack of appropriate street lighting at night], use well-funded social marketing campaigns facilitating attendance [e.g., provide free transportation], multiple marketing platforms [e.g., advertisements on radio stations and newspapers, word of mouth, colorful flyers in churches]), recruiting those of low socioeconomic status (mass media, word of mouth, enhancing traditional marketing with community-based participatory approaches), recruiting individuals with physical disabilities (use more flexible inclusion criteria, multiple marketing platforms, involving community gatekeepers, maintaining regular contact)	NR

* CALD = culturally and linguistically diverse; ESL = English as a second language; NR = not reported; SES = socioeconomic status.
† Disease self-management program reviews.
‡ Physical activity/exercise program reviews.
§ Underrepresented refers to those living in the US who have any of the following backgrounds or identities: African American, Native American, Latino, Latino American, Asian, Asian American, Pacific Islander, Native Alaskan, or Native Hawaiian.

Table 3. Recruitment and retention strategies of included descriptive and qualitative studies focused on physical activity/exercise*

Author	Target population	N	Recruitment strategies and recommended sites	Retention strategies
Altpeter et al, 2011 (77)	Rural, African American, men, and lower education	1,517	Rural residents: establish diverse set of partnerships for outreach (e.g., Area Agencies on Aging, senior centers, parish nurses, cooperative extensions, etc.); word of mouth; offer transportation; African Americans: recruit African Americans to be intervention leaders or coordinators; work with community-based African American organizations (e.g., churches, sororities); men: enroll women/partners to reach men; lower educational status: partner with subsidized housing; word of mouth; work with community leaders to help with outreach	Did not differentiate between recruitment and retention strategies
Friedman et al, 2012 (76)	African American men	49	Word of mouth; mass media (radio and television with African American audiences); partner with respected organizations and businesses (e.g., Urban League, NAACP, YMCA, Salvation Army); fraternities/ lodges/formal social organizations; locations frequented by African American men (state government buildings, barbershops, gyms, senior centers, schools or universities, social groups or clubs)	NR
Garcia et al, 2017 (75)	Hispanic men	14	Family involvement (especially spouses); multimedia outlets that are tailored to Hispanic population; personalized approach; cultural, regional, socioeconomic, and linguistically tailored	NR
Walcott-McQuigg et al, 2001 (74)	African American older adults	103	Conduct needs survey (likes/dislikes) and follow up with visit or call; send flyers to surrounding neighborhood; personalize mail and telephone message when possible; bulletin boards at planned facility and nearby facility; word of mouth; train older adults to recruit others; provide transportation	Provide safe place to exercise with people of same age; start and end on time; ensure exercise classes are appropriate for older adults (intensity, choice of music; offer variety of exercises); organize walking groups; provide exercise equipment/ facilities; provide health-related feedback and record progress; serve reduced fat/calorie refreshments; promote pool playing as exercise; use competition and prizes; have a clearly identified group leader; assign partners/buddies to encourage, remind, and check on attendance; leaders and group members should get to know each other; treat participants with dignity, kindness, and respect; provide transportation

* See Table 2 for definitions.

tabling at events, and on-site sessions. All studies distributed recruitment packets and/or materials to the priority population (e.g., flyers, posters, etc.), often at community events and locations.

Traditional media such as radio, television, and magazine advertisements or appearances (43,56,67) were used in several studies and was one of the most frequent strategies identified in the review (78). Other approaches used less frequently included social media (43,67), word-of-mouth referrals (60,78), health fairs (60,78), creating a culturally tailored recruitment video (57), email and web announcements (56), professional referrals (56), and recruitment from clinics/hospitals (60).

Although 4 of the 5 disease self-management studies reported recruitment results (43,56,57,60,67), only 1 reported recruitment yields and cost by strategy (60). In this study, flyers generated the largest number of enrolled participants, whereas information sessions had the highest enrollment yield (number enrolled divided by number contacted). The average cost of recruitment per enrolled participant was $142, with word of mouth ($60), community referrals ($60), and flyers ($91) being the least costly. Health fairs ($436) and recruitment from hospital clinics ($248) were the costliest. In another study (56), it was reported that the effectiveness of the recruitment strategy depended on the population; public service announcements were most effective for Hispanics, radio talks shows and personal contact for African Americans, and advertisements in lay health magazines and announcements on arthritis websites for non-Hispanic White populations.

Retention strategies and results in disease self-management programs. Three disease self-management studies described retention strategies (43,49,57,67). Two of the studies emphasized establishing strong relationships and trust with the study participants (43,57,67). One included flexibility in data collection, frequent reminders, and monetary and nonmonetary incentives (57). The third study described adapting the program for African American and Hispanic participants (49). Three studies reported retention results (49,56,57). In 1 study, retention was higher after implementing several retention strategies (e.g., increased contact with participants, more flexibility in data collection, providing incentives) (57), and in the other, retention was higher after making adaptations to the program to better suit the priority population (49). The third study simply reported 4- and 9-month retention of participants without reporting retention strategies (56).

Recruitment strategies and results in physical activity interventions. Recruitment strategies and results were described in all but 1 (54,66) of the 24 physical activity intervention studies as well as in 7 of the review papers (see Supplementary Table 1, at http://onlinelibrary.wiley.com/doi/10.1002/acr.25098/abstract) (21,79–84). The 1 descriptive paper described successful recruitment strategies across a series of programs (77), and in the 3 qualitative papers, suggested recruitment strategies were reported from the participant perspective (74–76). Like the chronic disease self-management studies, 7 physical activity intervention studies (44,46,47,55,59,63,69), 1 descriptive paper (77), 1 qualitative paper (74), and 5 reviews (21,79,82–84) described ensuring that recruitment staff were race- and community-matched, and 4 described having a community advisory board or obtaining community input to inform recruitment activities (46,47,59,63).

A wide range of recruitment strategies were employed across the studies and reported in the reviews, and almost all studies used multiple strategies. In most studies, flyers/brochures/printed materials were distributed to recruit participants from a wide variety of sites including worksites, schools, churches, libraries, beauty and nail salons, community centers, cafeterias, grocery stores, senior housing sites, clinical centers, and other locations (21,44–47,50–53,55,61,63–65,69–72,74,75,79–84), sometimes with tailoring such as using images that reflected the study priority populations. Often studies described general recruiting at community locations where the priority population spent time such as outdoor markets, laundromats, dollar stores, childcare settings, bus depots, senior centers, community centers, and community events (45,55,61,72,74,76,77,82,83). In many studies it was reported that presentations were given and information sessions were held at community locations, such as churches and synagogues, retirement or assisted living communities, civic clubs, community agencies or organizations, and clinics (44,46,47,55,58,59,61,79,84). Community health fairs or screenings were used to recruit participants in 7 studies (44,46,47,55,59,64,68) and 3 reviews (81,83,84). One descriptive paper (77) and a qualitative paper (75) described how family involvement, particularly regarding spouses, can be helpful in recruiting men into the study.

Although we excluded studies that delivered the intervention entirely in a clinical setting or only recruited from a clinical setting, 12 studies (46,50,51,53,58,59,61,64,68,69,71,73) and 1 review (80) used clinical practices and referrals from within them as part of a broader approach to recruitment. Another common strategy used in 13 studies (44–47,50,51,55,62–65,70,75), 3 qualitative/descriptive papers (74,76,77), and mentioned in nearly all reviews of physical activity interventions (21,79,81–84) was to recruit participants via word of mouth, social networking, and/or referrals from other study participants. In some cases, this was an intentional strategy, and in other cases it was not.

Both social media (21,44,50,53,64,75,79,84) and websites (50,64,70,79–81) along with traditional media such as newspapers (44–46,50,52,53,61,68,71,79–84), radio (44,52,53,68,79,81,82,84), television or news stories (45,52,63,79,81,82,84), and magazines (56,68) were commonly used to recruit participants, including in qualitative/descriptive studies (75,76). Direct mailings were also reported in several reviews (79–81,83,84), as well as in 7 studies (44,45,51,53,68,70,71). Targeted email distribution lists were also

used in 5 studies (46,53,58,63,65). Other strategies used included using existing research infrastructure including research registries (50,64,73) and recruiting participants taking part in current or past studies/programs (48,51,58,72). Additional strategies were infrequently used across studies.

Although all but 1 intervention study reported recruitment results, results were not reported in a uniform manner. In 1 study the screening yield was reported (number screened divided by number contacted for each source), and it was found that email was most efficient (58). Screening sources were reported in 4 studies (number screened from that source divided by the total screened); with word of mouth generating the most screens in 2 studies (44,45), flyers/brochures in 1 study (65), and a university website in 1 study (64). Enrollment/randomization yield was reported in 7 studies (number enrolled or randomized divided by number contacted or screened for each source); with 2 studies finding emails (58,63), 1 newspaper advertisements (50), and 1 flyers/tabletop cards (64) as most effective. Three demonstrated no major differences across strategies (46,47,72) except that health fairs were the least efficient in 1 study (46). Finally, 9 studies reported enrollment sources (number enrolled from that source divided by the total number enrolled); with 3 demonstrating that mass mailings (53,70,71), 2 word of mouth/social networking (45,46), 1 flyers/brochures (65), and 1 using a university website (64) yielded the most enrolled participants.

In the 2 other studies, source by participant race was examined. One showed that for both Black and White participants, television and radio yielded the most enrolled participants (52). The other study found that direct mail was most effective for Black participants, direct mail and phone for Hispanic participants, screening events and medical referrals for American Indians, and print and direct mail for Asian Americans (68). The cost of recruitment strategies was reported in 7 of the 23 studies (47,50,53,64,65,68,71). Some studies reported total recruitment costs (50,53,68,71) ranging from $26,874 ($118.91/participant) (50) to $4,105,000 ($1,075/participant) (68). Sharp et al (65) reported email recruitment was least expensive ($14/participant) and Park et al (50) reported clinic referrals ($0), word of mouth ($0), and web advertisements ($9.38/participant) were the least expensive. Newspaper advertisements were the most expensive ($212/participant). Wilbur et al (47) reported costs of $74.57/participant, which included time for the telephone screening and health assessment.

Retention strategies and results in physical activity interventions. Strategies to retain participants were reported in 12 of the 23 intervention studies (44,47,52–55,58,59,63,66,69,71,72) as well as in 3 reviews (80,83,84), 1 descriptive study (77), and 1 qualitative study (74). Similar to recruitment strategies, a common retention strategy was to have peer advocates or members of the target population represented on the study team and assist with retention and data collection (74,77,80,83,84). Providing a monetary incentive for completing study visits was the most commonly reported retention strategy, cited in 6 of these studies (44,47,58,59,63,72) and 2 reviews (80,84). Providing flexibility with study visits and study activities was the next most cited strategy, reported in 4 studies (55,58,59,69) and 2 reviews (83,84). Other retention activities included visit reminders (58,59,72), providing detailed information regarding the study and/or importance of visits (52,53,69), providing childcare (54,55,66,83,84), and providing other types of nonmonetary incentives or assistance (47,52,69,74,77). Finally, several studies detailed activities conducted to build rapport, trust, and close connections with the participants and those in their family and community (54,55,58,66,72,74,84). The 6 studies that reported retention-related results only provided overall study retention or attrition and did not link findings to strategies (52,54,55,58,59,66,69). Retention rates ranged from 58% to 95%.

DISCUSSION

In this review, we identified strategies that have been used to recruit and retain underrepresented populations and populations with arthritis or FM into behavioral programs focused on exercise, physical activity, or chronic disease self management (Figure 2). A key finding is that a multifaceted recruitment and retention plan is necessary to overcome the challenges in recruiting and retaining certain populations, including those with disability, low socioeconomic status, those from rural areas, and minorities. Nearly every study reviewed incorporated multiple recruitment strategies, and although some strategies worked better than others across studies and populations, multiple approaches were typically necessary to reach enrollment targets. Retention strategies, although less frequently mentioned, were often multifaceted. No studies reviewed reported the independent effects of specific retention strategies on retention rates; however, it was clear that many retention strategies overlapped with effective recruitment strategies.

One of the common strategies reported for both recruitment and retention was the involvement of members from the priority population within the study team, either as part of the investigative team or as advisors. Many studies highlighted how this strategy helped build trust with community members, which is critical, particularly since mistrust in research is a common barrier for racial and ethnic minority populations (85). This finding is consistent with previous literature suggesting stakeholder and community involvement within clinical trials and health research can improve enrollment rates (86,87). Having race- and community-matched members of the team or advisors allows the programs to be tailored appropriately to the population, ensuring the population's beliefs, culture, and norms guide recruitment and program materials. Further, team members can help inform applicable recruitment and retention strategies, which can enhance recruitment yield and reduce recruitment costs.

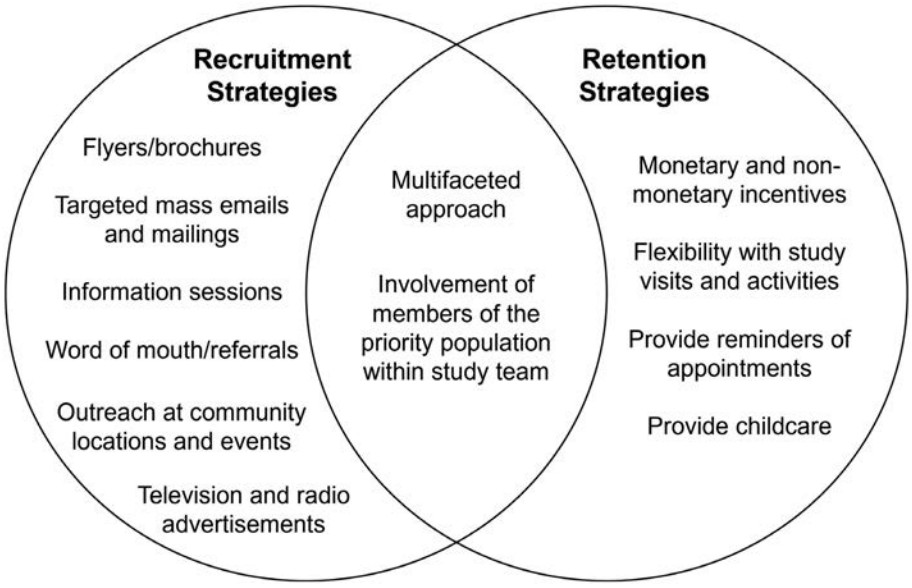

Figure 2. Summary of common recruitment and retention strategies used for underrepresented populations and adults with arthritis or fibromyalgia.

While it is difficult to directly link strategies, such as linking including members of the target population within the study team to recruitment yields, strategies that were directly tied to successful recruitment yields in underserved populations were emails, flyers/brochures, mailings, information sessions, and word of mouth/referrals. Due to the limited number of studies examining disease self-management programs and inconsistencies in reporting, it is challenging to identify the most promising strategies for each type of program. Common strategies across studies were focusing on locations the priority populations frequented, using targeted mass emails and mailings, and using passive television and radio advertisements that could reach large audiences. Some studies in this review also suggested that different recruitment strategies may be more effective in certain populations. Researchers should be encouraged to track and report enrollment rates by recruitment strategies, allowing future reviews to shed light on the most effective strategies for targeting specific underrepresented populations.

Our search also highlights the sparsity of literature focusing on strategies among individuals with arthritis or FM. Only 14% of included studies focused on adults with some form of arthritis, including lupus and FM, and only 16% of studies focused on recruitment and/or retention strategies for chronic disease self-management programs. This paucity of studies is surprising, particularly given the high prevalence of arthritis in the US (1) and known health benefits from chronic disease self-management programs (88). Since this review focused on community-based programs, research studies were omitted if recruitment was exclusively conducted through a clinical setting, thus precluding the discovery of disease self-management and physical activity programs that targeted individuals with rheumatic conditions in health care settings. Likewise, we excluded faith-based studies that solely focused on recruiting churches, although we recognize that churches are a promising setting for reaching African Americans, and increasingly, Latinos, for the promotion of physical activity (89–91).

Retention strategies were not as commonly reported as recruitment strategies and were not specifically linked to retention outcomes. Besides having members of the target population on the team and building trust, one of the most common strategies was the use of incentives, including monetary and nonmonetary. Incentives have been shown to increase retention (92) and aid in recruitment (93,94); however, the optimal ways to use them, when to use them, and what to provide are unknown.

Additional strategies frequently used to help with retention focused on flexibility and facilitating attendance. Many studies reported being flexible with scheduling study visits and activities, providing reminders of study appointments, and providing childcare. Additionally, several studies demonstrated the importance of building rapport and trust and emphasizing the importance of study visits to participants. Recently the Methods-Motivational Interviewing (MMI) approach was proposed, which has shown promise in helping with recruitment and participant engagement (95) and has been used in a Latino population (53). MMI includes an interactive prerequisite orientation session prior to the start of the program in which participant expectations and the scientific premise of the study are discussed.

While this review is novel in its focus on best practices for recruitment and retention in adults from underrepresented populations or those with arthritis or FM, this study has several

limitations. First, only studies that focused specifically on recruitment or retention were used, and strategies had to be explicitly described. Researchers might not have described all of their recruitment strategies. Further, many recruitment and retention strategies overlapped; as a result, some strategies that were used primarily for recruitment may have also influenced retention. Second, all included studies were from the US and were published in English. Finally, included studies were conducted before the COVID-19 pandemic, and only a few studies reported the use of social media. Social media may be a viable recruitment strategy (96), particularly after the pandemic resulted in shifts to more virtual and remotely delivered programs. Recruitment and retention strategies for those types of programs may differ from traditional in-person programs.

In conclusion, this review highlights the importance of a multifaceted recruitment and retention approach in underserved populations and populations with arthritis or FM in behavior intervention programs targeting exercise, physical activity, or chronic disease self management. Including race- or community-matched members within the study team and/or receiving advice regarding appropriate program tailoring and recruitment was common to aid with both recruitment and retention. Additional research is needed to better understand the individual effects of different strategies and the costs associated with the various recruitment/retention methods in underrepresented populations and populations with arthritis.

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. Pellegrini 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. Pellegrini.
Acquisition of data. Pellegrini, Wilcox, DeVivo, Jamieson.
Analysis and interpretation of data. Pellegrini, Wilcox.

REFERENCES

1. Barbour KE, Helmick CG, Boring M, et al. Vital signs: prevalence of doctor-diagnosed arthritis and arthritis-attributable activity limitation - United States, 2013–2015. MMWR Morb Mortal Wkly Rep 2017;66:246–53.
2. Kelley GA, Kelley KS, Hootman JM, et al. Effects of community-deliverable exercise on pain and physical function in adults with arthritis and other rheumatic diseases: a meta-analysis. Arthritis Care Res 2011;63:79–93.
3. Allen KD, Woolson S, Hoenig HM, et al. Stepped exercise program for patients with knee osteoarthritis: a randomized controlled trial. Ann Intern Med 2021:174:298–307.
4. Verhoeven F, Tordi N, Prati C, et al. Physical activity in patients with rheumatoid arthritis. Joint Bone Spine 2016;83:265–70.
5. Fontaine KR, Conn L, Clauw DJ. Effects of lifestyle physical activity on perceived symptoms and physical function in adults with fibromyalgia: results of a randomized trial. Arthritis Res Ther 2010;12:R55.
6. Conn VS, Hafdahl AR, Minor MA, et al. Physical activity interventions among adults with arthritis: meta-analysis of outcomes. Semin Arthritis Rheum 2008;37:307–16.
7. Kelley GA, Kelley KS, Callahan LF. Aerobic exercise and fatigue in rheumatoid arthritis participants: a meta-analysis using the minimal important difference approach. Arthritis Care Res (Hoboken) 2018;70:1735–9.
8. Mahieu MA, Ahn GE, Chmiel JS, et al. Fatigue, patient reported outcomes, and objective measurement of physical activity in systemic lupus erythematosus. Lupus 2016;25:1190–9.
9. O'Dwyer T, Durcan L, Wilson F. Exercise and physical activity in systemic lupus erythematosus: a systematic review with meta-analyses. Semin Arthritis Rheum 2017;47:204–15.
10. Busch AJ, Schachter CL, Overend TJ, et al. Exercise for fibromyalgia: a systematic review. J Rheumatol 2008:35:1130–44.
11. Kelley GA, Kelley KS, Hootman JM. Effects of exercise on depression in adults with arthritis: a systematic review with meta-analysis of randomized controlled trials. Arthritis Res Ther 2015;17:21.
12. Murphy LB, Hootman JM, Boring MA, et al. Leisure time physical activity among U.S. adults with arthritis, 2008–2015. Am J Prev Med 2017;53:345–54.
13. Centers for Disease Control and Prevention. Arthritis: lifestyle management programs. 2021. URL: https://www.cdc.gov/arthritis/interventions/index.htm.
14. Osteoarthritis Action Alliance. Arthritis-appropriate, evidence-based interventions (AAEBI). URL: https://oaaction.unc.edu/aaebi/.
15. Callahan LF, Cleveland RJ, Shreffler J, et al. Evaluation of active living every day in adults with arthritis. J Phys Act Health 2014;11:285–95.
16. Wilcox S, Dowda M, Leviton LC, et al. Active for life: final results from the translation of two physical activity programs. Am J Prev Med 2008;35:340–51.
17. Hughes SL, Seymour RB, Campbell R, et al. Impact of the fit and strong intervention on older adults with osteoarthritis. Gerontologist 2004;44:217–28.
18. Shaghaghi A, Bhopal RS, Sheikh A. Approaches to recruiting 'hard-to-reach' populations into re-search: a review of the literature. Health Promot Perspect 2011;1:86–94.
19. Wieland ML, Njeru JW, Alahdab F, et al. Community-engaged approaches for minority recruitment into clinical research: a scoping review of the literature. Mayo Clin Proc 2021;93:733–43.
20. Young L, Barnason S, Do V. Review strategies to recruit and retain rural patient participating self-management behavioral trials. Online J Rural Res Policy 2015;10:1–12.
21. Mendoza-Vasconez AS, Linke S, Munoz M, et al. Promoting physical activity among underserved populations. Curr Sports Med Rep 2016;15:290–7.
22. Murphy LB, Brady TJ, Boring MA, al. Self-management education participation among US adults with arthritis: who's attending? Arthritis Care Res (Hoboken) 2017;69:1322–30.
23. Friedman DB, Bergeron CD, Foster C, et al. What do people really know and think about clinical trials? A comparison of rural and urban communities in the South. J Community Health 2013;38:642–51.
24. Blanton S, Morris DM, Prettyman MG, et al. Lessons learned in participant recruitment and retention: the EXCITE trial. Phys Ther 2006;86:1520–33.
25. Tanner A, Kim SH, Friedman DB, et al. Barriers to medical research participation as perceived by clinical trial investigators: communicating with rural and African American communities. J Health Commun 2015;20:88–96.
26. Friedman DB, Foster C, Bergeron CD, et al A qualitative study of recruitment barriers, motivators, and community-based strategies for increasing clinical trials participation among rural and urban populations. Am J Health Promot 2015;29:332–8.

27. Winter SS, Page-Reeves JM, Page KA, et al. Inclusion of special populations in clinical research: important considerations and guidelines. J Clin Transl Res 2018;4:56–69.
28. Bruno M, Cummins S, Gaudiano L, et al. Effectiveness of two Arthritis Foundation programs: Walk With Ease, and YOU Can Break the Pain Cycle. Clin Interv Aging 2006;1:295–306.
29. Callahan LF, Shreffler JH, Altpeter M, et al. Evaluation of group and self-directed formats of the Arthritis Foundation's Walk With Ease Program. Arthritis Care Res 2011;63:1098–107.
30. Wyatt B, Mingo CA, Waterman MB, et al. Impact of the Arthritis Foundation's Walk With Ease Program on arthritis symptoms in African Americans. Prev Chronic Dis 2014;11:E199.
31. Callahan LF, Rivadeneira A, Altpeter M, et al. Evaluation of the Arthritis Foundation's Camine Con Gusto Program for Hispanic adults with arthritis. Hisp Health Care Int 2016;14:132–40.
32. Nyrop KA, Charnock BL, Martin KR, et al. Effect of a six-week walking program on work place activity limitations among adults with arthritis. Arthritis Care Res 2011;63:1773–6.
33. Conte KP, Odden MC, Linton NM, et al. Effectiveness of a scaled-up arthritis self-management program in Oregon: Walk With Ease. Am J Public Health 2016;106:2227–30.
34. Falasinnu T, Bao G, Brady TJ, et al. Factors associated with the initiation and retention of patients with lupus in the chronic disease self-management program. Arthritis Care Res (Hoboken) 2023;75:519–28.
35. Vaughan MW, Felson DT, LaValley MP, et al. Perceived community environmental factors and risk of five-year participation restriction among older adults with or at risk of knee osteoarthritis. Arthritis Care Res (Hoboken) 2017;69:952–8.
36. Wilkie R, Peat G, Thomas E, et al. Factors associated with restricted mobility outside the home in community-dwelling adults ages fifty years and older with knee pain: an example of use of the International Classification of Functioning to investigate participation restriction. Arthritis Rheum 2007;57:1381–9.
37. Theis KA, Murphy L, Hootman JM, et al. Social participation restriction among US adults with arthritis: a population-based study using the International Classification of Functioning, Disability and Health. Arthritis Care Res (Hoboken) 2013;65:1059–69.
38. Nicholson LM, Schwirian PM, Groner JA. Recruitment and retention strategies in clinical studies with low-income and minority populations: progress from 2004–2014. Contemp Clin Trials 2015;45:34–40.
39. Liu Y, Pencheon E, Hunter RM, et al. Recruitment and retention strategies in mental health trials – a systematic review. PLoS One 2018;13:e0203127.
40. UyBico SJ, Pavel S, Gross CP. Recruiting vulnerable populations into research: a systematic review of recruitment interventions. J Gen Intern Med 2007;22:852–63.
41. Peters M, Godfrey C, McInerney P, et al. Chapter 11: scoping reviews. In: JBI manual for evidence synthesis. 2020. URL: https://synthesismanual.jbi.global.
42. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169:467–73.
43. Graham LF, Scott L, Lopeyok E, et al. Outreach strategies to recruit low-income African American men to participate in health promotion programs and research: lessons from the Men of Color Health Awareness (MOCHA) Project. Am J Mens Health 2018;12:1307–16.
44. Yancey AK, Miles OL, McCarthy WJ, et al. Differential response to targeted recruitment strategies to fitness promotion research by African-American women of varying body mass index. Ethn Dis 2001;11:115–23.
45. Yancey AK, Jordan A, Bradford J, et al. Engaging high-risk populations in community-level fitness promotion: ROCK! Richmond. Health Promot Pract 2003;4:180–8.
46. Wilbur J, McDevitt J, Wang E, et al. Recruitment of African American women to a walking program: eligibility, ineligibility, and attrition during screening. Res Nurs Health 2006;29:176–89.
47. Wilbur J, Buchholz SW, Ingram DM, et al. Effectiveness, efficiency, duration, and costs of recruiting for an African American women's lifestyle physical activity program. Res Nurs Health 2013;36:487–99.
48. Sullivan-Marx EM, Mangione KK, Ackerson T, et al. Recruitment and retention strategies among older African American women enrolled in an exercise study at a PACE program. Gerontologist 2011;51 Suppl 1:S73–81.
49. Reid MC, Chen EK, Parker SJ, et al. Measuring the value of program adaptation: a comparative effectiveness study of the standard and a culturally adapted version of the Arthritis Self-Help Program. HSS J 2014;10:59–67.
50. Park M, Bannuru RR, Price LL, et al. Effective recruitment strategies in an exercise trial for patients with fibromyalgia. Trials 2021;22:557.
51. Okhomina VI, Seals SR, Marshall GD Jr. Recruitment and enrollment of African Americans into health promoting programs: the effects of health promoting programs on cardiovascular disease risk study. Ethn Health 2020;25:825–34.
52. Nies MA, Hayward DF, Chruscial H, et al. Recruitment and retention of Black women: experiences in an urban setting. Clin J Womens Health 2001;1:267–72.
53. King AC, Campero I, Sheats JL, et al. Testing the effectiveness of physical activity advice delivered via text messaging vs. human phone advisors in a Latino population: the On The Move randomized controlled trial design and methods. Contemp Clin Trials 2020;95:106084.
54. Keller CS, Gonzales A, Fleuriet KJ. Retention of minority participants in clinical research studies. West J Nurs Res 2005;27:292–306.
55. Hines-Martin V, Speck BJ, Stetson B, et al. Understanding systems and rhythms for minority recruitment in intervention research. Res Nurs Health 2009;32:657–70.
56. Goeppinger J, Lorig KR, Ritter PL, et al. Mail-delivered arthritis self-management tool kit: a randomized trial and longitudinal followup. Arthritis Rheum 2009;61:867–75.
57. Drenkard C, Easley K, Bao G, et al. Overcoming barriers to recruitment and retention of African-American women with SLE in behavioural interventions: lessons learnt from the WELL study. Lupus Sci Med 2020;7:e000391.
58. Carthron DL, Phillips A, Cuthbertson CC, et al. Four methods of recruiting couples into a longitudinal study of physical activity in people with osteoarthritis: recruitment, retention, and lessons learned. Front Public Health 2018;6:197.
59. Adams S, Heiney S, Brandt H, et al. A comparison of a centralized versus de-centralized recruitment schema in two community-based participatory research studies for cancer prevention. J Community Health 2015;40:251–9.
60. Ramsay JE, Hogan CK, Janevic MR, et al. Comparison of recruitment strategies for engaging older minority adults: results from Take Heart. J Gerontol A Biol Sci Med Sci 2020;75:922–8.
61. Hammerback K, Felias-Christensen G, Phelan EA. Evaluation of a telephone-based physical activity promotion program for disadvantaged older adults. Prev Chronic Dis 2012;9:E62.
62. Gilbert PA, Haines H, Baquero B, et al. Respondent-driven sampling to recruit Latinos in a Midwest micropolitan area: lessons learned and recommendations for translational work. J Clin Transl Sci 2018;2:245–8.

63. Wilbur J, Miller AM, Chandler P. Recruitment and cardiovascular risk characteristics of African American and Caucasian midlife women. J Cardiovasc Nurs 2001;15:88–104.
64. Staffileno BA, Zschunke J, Weber M, et al. The feasibility of using Facebook, Craigslist, and other online strategies to recruit young African American women for a web-based healthy lifestyle behavior change intervention. J Cardiovasc Nurs 2017;32:365–71.
65. Sharp LK, Fitzgibbon ML, Schiffer L. Recruitment of obese Black women into a physical activity and nutrition intervention trial. J Phys Act Health 2008;5:870–81.
66. Keller C, Fleury J, Perez A. Culturally specific strategies for retention and adherence to physical activity interventions in Hispanic women. In: Streiner DL, Sidani S, editors. When research goes off the rails: why it happens and what you can do about it. New York (NY): The Guilford Press; 2010. p.161–8.
67. Valdez LA, Gubrium AC, Markham J, et al. A culturally and gender responsive stress and chronic disease prevention intervention for low/no-income African American men: the MOCHA moving forward randomized control trial protocol. Contemp Clin Trials 2021;101: 106240.
68. Rubin RR, Fujimoto WY, Marrero DG, et al. The Diabetes Prevention Program: recruitment methods and results. Control Clin Trials 2002; 23:157–71.
69. Phillips-Caesar EG, Winston G, Peterson JC, et al. Small Changes and Lasting Effects (SCALE) trial: the formation of a weight loss behavioral intervention using EVOLVE. Contemp Clin Trials 2015;41: 118–28.
70. King AC, Campero MI, Garcia D, et al. Testing the effectiveness of community-engaged citizen science to promote physical activity, foster healthier neighborhood environments, and advance health equity in vulnerable communities: the Steps for Change randomized controlled trial design and methods. Contemp Clin Trials 2021;108: 106526.
71. King AC, Campero I, Sheats JL, et al. Testing the comparative effects of physical activity advice by humans vs. computers in underserved populations: the COMPASS trial design, methods, and baseline characteristics. Contemp Clin Trials 2017;61:115–25.
72. Garcia DO, Valdez LA, Bell ML, et al. A gender- and culturally-sensitive weight loss intervention for Hispanic males: the ANIMO randomized controlled trial pilot study protocol and recruitment methods. Contemp Clin Trials Commun 2018;9:151–63.
73. Chang RW, Semanik PA, Lee J, et al. Improving physical activity in arthritis clinical trial (IMPAACT): study design, rationale, recruitment, and baseline data. Contemp Clin Trials 2014;39:224–35.
74. Walcott-McQuigg JA, Prohaska TR. Factors influencing participation of African American elders in exercise behavior. Public Health Nurs 2001;18:194–203.
75. Garcia DO, Valdez LA, Hooker SP. Hispanic male's perspectives of health behaviors related to weight management. Am J Mens Health 2017;11:1547–59.
76. Friedman DB, Hooker SP, Wilcox S, et al. African American men's perspectives on promoting physical activity: "We're not that difficult to figure out!" J Health Commun 2012;17:1151–70.
77. Altpeter M, Houenou LO, Martin KR, et al. Recruiting and retaining hard-to-reach populations: lessons learned and targeted strategies from arthritis physical activity intervention studies. Arthritis Care Res (Hoboken) 2011;63:927–8.
78. Horrell LN, Kneipp SM. Strategies for recruiting populations to participate in the Chronic Disease Self-Management Program (CDSMP): a systematic review. Health Mark Q 2017;34:268–83.
79. Kosma M, Cardinal BJ, McCubbin JA. Recruitment techniques among understudied populations and their implications for physical activity promotion. Quest 2004;56:413–20.
80. James DC, Harville C II, Sears C, et al. Participation of African Americans in e-Health and m-Health studies: a systematic review. Telemed J E Health 2017;23:351–64.
81. Foster CE, Brennan G, Matthews A, et al. Recruiting participants to walking intervention studies: a systematic review. Int J Behav Nutr Phys Act 2011;8:137.
82. El Masri A, Kolt GS, George ES. Physical activity interventions among culturally and linguistically diverse populations: a systematic review. Ethn Health 2022;27:40–60.
83. Conn VS, Chan K, Banks J, et al. Cultural relevance of physical activity intervention research with underrepresented populations. Int Q Community Health Educ 2013;34:391–414.
84. Carroll JK, Yancey AK, Spring B, et al. What are successful recruitment and retention strategies for underserved populations? Examining physical activity interventions in primary care and community settings. Transl Behav Med 2011;1:234–51.
85. George S, Duran N, Norris K. A systematic review of barriers and facilitators to minority research participation among African Americans, Latinos, Asian Americans, and Pacific Islanders. Am J Public Health 2014;104:e16–31.
86. Crocker JC, Ricci-Cabello I, Parker A, et al. Impact of patient and public involvement on enrolment and retention in clinical trials: systematic review and meta-analysis. BMJ 2018;363:k4738.
87. Forsythe L, Heckert A, Margolis MK, et al. Methods and impact of engagement in research, from theory to practice and back again: early findings from the Patient-Centered Outcomes Research Institute. Qual Life Res 2018;27:17–31.
88. Allegrante JP, Wells MT, Peterson JC. Interventions to support behavioral self-management of chronic diseases. Annu Rev Public Health 2019;40:127–46.
89. Tristão Parra M, Porfírio GJ, Arredondo EM, et al. Physical activity interventions in faith-based organizations: a systematic review. Am J Health Promot 2018;32:677–90.
90. Dunn CG, Wilcox S, Saunders RP, et al. Healthy eating and physical activity interventions in faith-based settings: a systematic review using the reach, effectiveness/efficacy, adoption, implementation, maintenance framework. Am J Prev Med 2021;60:127–35.
91. Bopp M, Peterson JA, Webb BL. A comprehensive review of faith-based physical activity interventions. Am J Lifestyle Med 2012;6: 460–78.
92. Brueton VC, Tierney JF, Stenning S, et al. Strategies to improve retention in randomised trials: a Cochrane systematic review and meta-analysis. BMJ Open 2014;4:e003821.
93. Abdelazeem B, Abbas KS, Amin MA, et al. The effectiveness of incentives for research participation: a systematic review and meta-analysis of randomized controlled trials. PLoS One 2022;17: e0267534.
94. Caldwell PH, Hamilton S, Tan A, et al. Strategies for increasing recruitment to randomised controlled trials: systematic review. PLoS Med 2010;7:e1000368.
95. Jake-Schoffman DE, Brown SD, Baiocchi M, et al. Methods-motivational interviewing approach for enhanced retention and attendance. Am J Prev Med 2021;61:606–17.
96. Tsaltskan V, Nguyen K, Eaglin C, et al. Comparison of web-based advertising and a social media platform as recruitment tools for underserved and hard-to-reach populations in rheumatology clinical research. ACR Open Rheumatol 2022;4:623–30.

Arthritis Care & Research
Vol. 75, No. 9, September 2023, pp 2011–2021
DOI 10.1002/acr.25099

American College of Rheumatology
Empowering Rheumatology Professionals

Uptake and Spending on Biosimilar Infliximab and Etanercept After New Start and Switching Policies in Canada: An Interrupted Time Series Analysis

Alison R. McClean,[1] Lucy Cheng,[1] Nick Bansback,[1] Fiona Clement,[2] Mina Tadrous,[3] Mark Harrison,[1] and Michael R. Law[1]

Objective. Uptake of biosimilars has been suboptimal in North America. This study was undertaken to quantify the impact of various policy interventions (namely, new start and switching policies) on uptake and spending on biosimilar infliximab and etanercept in British Columbia (BC), Canada.

Methods. We used administrative claims data to identify BC residents ≥18 years of age with rheumatoid arthritis, ankylosing spondylitis, psoriatic arthritis, and/or plaque psoriasis who qualified for public drug coverage from January 2013 to November 2020. Using interrupted time series analysis, we studied the change in proportion spent on and prescriptions dispensed of biosimilar infliximab and etanercept out of the total amount per agent after new start and biosimilar switching policies were implemented.

Results. Our study included 208,984 individuals living with rheumatoid arthritis, ankylosing spondylitis, plaque psoriasis, and/or psoriatic arthritis, corresponding to 5,884 patients taking infliximab and etanercept. After the new start policy, we detected a small gradual increase in the proportion of dispensed biosimilar etanercept prescriptions of 0.65% per month (95% confidence interval [95% CI] 0.44, 0.85). The trend related to the proportion of total spending on biosimilar etanercept also increased (0.51% [95% CI 0.28, 0.73]). After the switching policy, there was a sustained increase in the proportion of dispensed biosimilar etanercept and infliximab prescriptions of 76.98% (95% CI 75.56, 78.41) and 58.43% (95% CI 52.11, 64.75), respectively. Similarly, there was a persistent increase in monthly spending on biosimilar etanercept and infliximab of 78.22% (95% CI 76.65, 79.79) and 71.23% (95% CI 66.82, 75.65), respectively.

Conclusion. We found that mandatory switching policies were much more effective than new starting policies for increasing the use of biosimilar medications.

INTRODUCTION

Worldwide spending on prescription drugs has been predicted to reach $1.6 trillion by 2026 (1). Specialty drugs, including biologics, have become a major driver of this expenditure, representing ∼30% of expenditure despite accounting for <2% of prescriptions dispensed (2,3). Two of the costliest biologics in the US, infliximab and etanercept, were responsible for $4.86 and $7.78 billion in spending, respectively, in 2019 alone (4).

Akin to generics, biosimilars can seek regulatory approval after the patent expires on the reference product. In order to gain market authorization, the Food and Drug Administration requires biosimilar manufacturers to demonstrate that their products are highly similar to the reference biologic with no clinically significant differences in efficacy, effectiveness, safety, or quality (5).

All inferences, opinions, and conclusions drawn herein are those of the author(s) and do not reflect the opinions or policies of the Data Steward(s).

Supported by the Canadian Institutes of Health Research (project grant 451221) and the 2020 Canadian Initiative for Outcomes in Rheumatology Care (grant 2020-001). Dr. McClean's work was supported by the Canadian Institutes of Health Research (Canada graduate scholarship master's award). Dr. Harrison's work was supported by the Michael Smith Foundation for Health Research (scholar award 16813). Dr. Law's work was supported by the Canada Research Chair in Access to Medicines.

[1]Alison R. McClean, PharmD, MSc, Lucy Cheng, MSc, Nick Bansback, PhD, Mark Harrison, PhD, Michael R. Law, PhD: University of British Columbia, Vancouver, British Columbia, Canada; [2]Fiona Clement, PhD: University of Calgary, Calgary, Alberta, Canada; [3]Mina Tadrous, PharmD, PhD: University of Toronto and Women's College Hospital, Toronto, Ontario, Canada.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.25099&file=acr25099-sup-0001-Disclosureform.pdf.

Address correspondence via email to Alison McClean, PharmD, MSc, at alisonmcclean@alumni.ubc.ca.

Submitted for publication June 3, 2022; accepted in revised form February 2, 2023.

SIGNIFICANCE & INNOVATIONS

- Potentially cost-saving biosimilars have been underutilized in North America; in 2023, biosimilars for the top-selling drug in the world (adalimumab) will become available in the US, raising questions regarding policies to support biosimilar change for economic stewardship.
- Although policy approaches to enhance use of biosimilars have been mostly passive, the province of British Columbia became the first jurisdiction in North America to mandate switching from reference biologics to biosimilars in order to maintain drug coverage.
- Switching policies have the potential to greatly enhance uptake of biosimilars beyond what was detected for the more passive new start initiatives, and payers seeking to increase use of biosimilars should consider switching programs.

Biosimilars offer one potential avenue to decrease spending on biologics, and projections suggest that they could reduce drug expenditures by $215 billion globally between 2022 and 2026 (1). Despite this potential, biosimilar uptake has been low in some countries, including the US and Canada (6–10).

The Centers for Medicare and Medicaid Services (CMS) have encouraged uptake of biosimilars through passive approaches, such as the introduction of unique billing codes for biosimilars (thereby facilitating unique pricing of biosimilars), necessitating biosimilar manufacturer discounts in the Part D donut hole for beneficiaries, and requiring biosimilar copayments at a rate similar to generics for those enrolled in the Low-Income Subsidy plan (11–13).

In Canada, the cost of prescription drugs may be covered by federal payers (e.g., eligible veterans), provincial or territorial governments, private commercial entities, or out-of-pocket payments. All residents are eligible for their respective provincial or territorial plan, although deductible and copayment amounts vary by income and geographic region, among other factors. Previous research has suggested that 59.5% residents of British Columbia (BC) have private insurance (14).

For the most part, Canadian policymakers have encouraged biosimilar uptake through passive new start policies, which require individuals initiating a biologic for the first time to begin treatment with a biosimilar. However, in 2019, the province of BC became the first region in North America to require individuals established on therapy to switch to a biosimilar in order to maintain provincial drug coverage. Under the first phase of these switching policies, individuals living with inflammatory arthritis and psoriasis and receiving reference etanercept and infliximab were given 6 months to switch to the relevant biosimilar. Specifically, individuals taking reference etanercept were required to switch to Brenzys (approved in 2016) or Erelzi (2017), while individuals receiving reference infliximab switched to Inflectra (2014) or Renflexis (2018) (15,16). Given the novelty of this policy in North America, we assessed changes in the uptake and spending on biosimilar infliximab and etanercept in BC following these 2 distinct policy changes.

MATERIALS AND METHODS

Study data, sample, and setting. We used linked deidentified province-level outpatient physician billings (Medical Services Plan [MSP] payment information file), hospital discharges and separations (Discharge Abstract Database [DAD]), emergency department visits (National Ambulatory Care Reporting System [NACRS]), outpatient prescription dispensation data (PharmaNet), and the Consolidation file (MSP registration) from Population Data BC from January 2013 to November 2020 (17–21). Individuals were eligible for inclusion if they 1) were ≥18 years, 2) had rheumatoid arthritis, ankylosing spondylitis, psoriatic arthritis, and/or plaque psoriasis, and 3) qualified for public drug coverage during the study period.

We identified individuals living with conditions of interest by searching for ≥1 instance of International Classification of Diseases, Ninth Revision (ICD-9)/International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Canada (ICD-10-CA) codes 714.X, M05.X, M06.X (rheumatoid arthritis); 720.X, M45.X (ankylosing spondylitis); and/or 696.X, L40.X (psoriatic arthritis/plaque psoriasis) present in the MSP, DAD, and/or NACRS. The aforementioned categories were not mutually exclusive. We used brand names in order to identify dispensations of infliximab (Remicade, Inflectra, Renflexis) or etanercept (Enbrel, Brenzys, Erelzi) in PharmaNet. As infliximab is also indicated for the treatment of Crohn's disease and ulcerative colitis, but these indications were not subjected to the policies of interest, individuals with instances of these billing codes were excluded from the infliximab cohort (Crohn's disease 555.X, K50.X; ulcerative colitis 556.X, K51.X). The number of individuals who had ever filled a prescription for infliximab and etanercept during the study period was tabulated, and those individuals were reported as taking these medicines. Groupings based on medicine use were not mutually exclusive. Demographic information was derived from the consolidation file. Age and neighborhood income quintile were defined at the first instance of an ICD-9/ICD-10-CA code of interest during the study period.

Policies of interest. In BC, access to infliximab and etanercept requires the prescriber to receive prior authorization (termed 'special authority') in order for these medicines to be covered under the provincial drug plan. On February 16, 2016 and July 18, 2017, the BC government required individuals initiating infliximab and etanercept, respectively, for certain inflammatory conditions to begin treatment with a biosimilar after receiving special authority approval. From May 27 to November 25, 2019,

individuals living with inflammatory arthritis and psoriasis receiving reference etanercept or infliximab were required to switch to a biosimilar version in order to maintain prescription drug coverage offered by the provincial government of BC. As pharmacists were not authorized to switch patients from the reference biologic to the biosimilar medicine without prescriber involvement, the government provided a 6-month period for patients and prescribers to work together to make the switch. Large commercial insurers that provide supplementary drug coverage, including Pacific Blue Cross and Green Shield Canada, also introduced biosimilar adjudication rules that mirrored the government's policies. Under the biosimilar switching policies, the provincial government continues to provide exceptional coverage for the reference product if medically needed (15).

In order to quantify biosimilar uptake prior to the biosimilar new start program, we stipulated a preintervention period of 6 months prior to the introduction of the policy (Figure 1). Observation of the new start policy ran from its introduction until the switching policy was implemented. In alignment with the 6-month phase-in period provided by the provincial government during the biosimilar switch, we incorporated a phase-in period from June to December 2019 in our model. Our postintervention period ran from the end of the phase-in period until the end of our data availability in November 2020.

Outcomes of interest. Using PharmaNet, we examined all retail pharmaceutical claims for reference and biosimilar etanercept and infliximab among our cohort members. We examined both public expenditure as well as nonpublic spending (i.e., patient and/or private insurer) on etanercept and infliximab per month. We studied the proportion of prescriptions dispensed of and the proportion of total spending on the biosimilars out of the total amount per agent (i.e., either etanercept and infliximab) during the study period.

Statistical analysis. We used interrupted time series (ITS) analysis, a rigorous quasi-experimental design, to adjust for secular trends in the study data (22). ITS utilizes repeat measures over time, before and after a program of interest is implemented, in order to estimate the effect of said policy. ITS permits the quantification of both the immediate-level change in the outcome of interest as well as the change over time (trend change). The sustained change in the proportional uptake or spending on biosimilar etanercept and infliximab was determined from the difference between the pre- and postintervention level and trend (i.e., the counterfactual).

Using segmented linear regression, we modeled the level and trend change in biosimilar etanercept and infliximab use and spending after each of the 2 policy interventions. We used generalized least squares models and included autoregressive moving average with p autoregressive and q moving-average terms based on standard diagnostic tests (23). Following best practices, we also completed a sex-stratified analysis, as well as a neighborhood income quintile–stratified analysis, for both biosimilar spending and use (24). Data were prepared using SAS, version 9.4, and analyses were conducted with R, version 4.0.5.

Ethics approval and data availability. This research was conducted in compliance with the Declaration of Helsinki, and ethics approval was obtained from the University of British Columbia's Behavioural Research Ethics Board (H20-00252). Access to data provided by the Data Steward(s) is subject to approval but can be requested for research projects through the Data Steward(s) or their designated service providers.

RESULTS

We identified 208,984 individuals, of which 104,796 (50.1%), 99,261 (47.5%), and 23,371 (11.2%) were living with rheumatoid arthritis, plaque psoriasis, or psoriatic arthritis and/or ankylosing spondylitis, respectively (these groups were not mutually exclusive) (Table 1). Our study included 4,697 (79.8%) and 1,187 (20.2%) individuals who had ever filled a prescription for etanercept and infliximab, respectively, during the study period (these

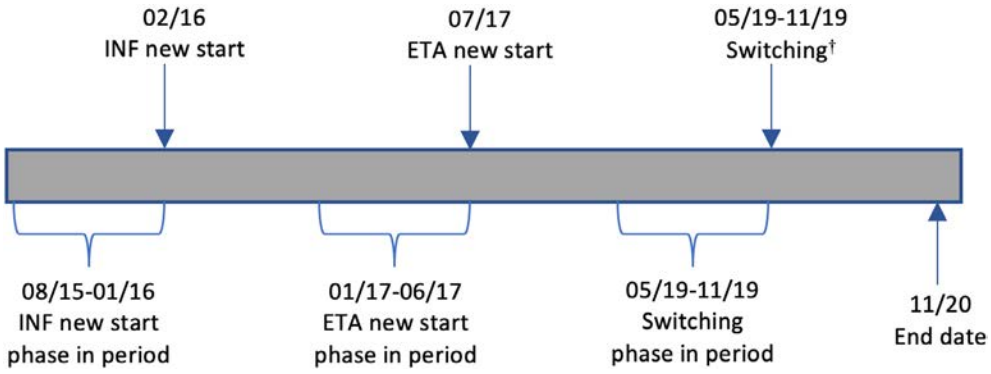

Figure 1. Study timeline. † = infliximab (INF) and etanercept (ETA) biosimilar switching among individuals with inflammatory arthritis and psoriasis.

Table 1. Sociodemographic characteristics of study participants*

Characteristic	Rheumatoid arthritis†	Psoriatic arthritis/ plaque psoriasis†	Ankylosing spondylitis†	Infliximab utilizer‡	Etanercept utilizer‡	Overall
Sex						
Female	69,823 (66.6)	52,129 (52.5)	12,268 (52.5)	708 (59.6)	2,949 (62.8)	123,475 (59.1)
Male	34,973 (33.4)	47,132 (47.5)	11,103 (47.5)	479 (40.4)	1,748 (37.2)	85,509 (40.9)
Neighborhood income quintile						
1 (lowest)	21,316 (20.3)	19,102 (19.2)	4,520 (19.3)	233 (19.6)	813 (17.3)	41,361 (19.8)
2	21,696 (20.7)	19,848 (20.0)	4,621 (19.8)	228 (19.2)	921 (19.6)	42,421 (20.3)
3	21,020 (20.0)	20,033 (20.2)	4,681 (20.0)	258 (21.7)	957 (20.4)	42,115 (20.2)
4	20,644 (19.7)	20,438 (20.6)	4,806 (20.6)	248 (20.9)	984 (20.9)	42,145 (20.2)
5 (highest)	19,376 (18.5)	19,044 (19.2)	4,512 (19.3)	210–220§	987 (21.0)	39,312 (18.8)
Missing	744 (0.7)	796 (0.8)	231 (1.0)	5–15§	35 (0.8)	1,630 (0.8)
Age group, years						
18–29	3,738 (3.6)	9,842 (9.9)	1,904 (8.1)	86 (7.2)	266 (5.6)	14,654 (7.0)
30–49	20,484 (19.5)	27,492 (27.7)	7,302 (31.2)	394 (33.2)	1,399 (29.8)	50,707 (24.3)
50–69	49,575 (47.3)	42,432 (42.7)	9,559 (40.9)	560 (47.2)	2,466 (52.5)	92,592 (44.3)
70+	30,999 (29.6)	19,495 (19.6)	4,606 (19.7)	147 (12.4)	566 (12.1)	51,031 (24.4)

* Values are the number (%) unless indicated otherwise.
† Groups based on diagnosis codes were not mutually exclusive.
‡ Groups based on biologic use were not mutually exclusive.
§ In order to protect individual privacy related to small cell counts, ranges were reported.

groups were not mutually exclusive). Overall, our study included 123,475 female subjects (59.1%), and individuals were most often age 50–69 years (n = 92,592, 44.3%).

New start policy. For etanercept, we detected a gradual monthly increase in the proportion of prescriptions dispensed that were biosimilar of 0.65% (95% confidence interval [95% CI] 0.44, 0.85), with no significant level change in use (Table 2 and Figure 2). Similar changes were quantified for the change in proportion of total spending on biosimilar etanercept post new start (Table 3). No significant changes in use or spending were detected for biosimilar infliximab after the new start policy was introduced. Comparable results were quantified among our disease-specific cohorts.

Switching policy. In terms of total spending and number of prescriptions dispensed, proportional use of biosimilar infliximab increased from 21.6% to 74.6% and 33.7% to 78.9%, respectively, over the preintervention period. Similarly, biosimilar etanercept increased from 15.8% to 71.9% and 21.6% to 79.8% in terms of proportion of total spending and prescriptions dispensed, respectively.

From our ITS, we observed significant level and trend changes among all study cohorts with respect to the proportion of biosimilar etanercept prescriptions dispensed and total spending after the switching policy was introduced (Tables 2 and 3 and Figure 2). For example, we detected a step change of 76.98% (95% CI 75.56, 78.41) in terms of the proportion of etanercept prescriptions dispensed post switch, in addition to a persistent gradual decrease of –0.95% per month (95% CI –1.04, –0.85), in our overall cohort in the post-switch period.

We also detected significant level changes in the proportion of biosimilar infliximab prescriptions dispensed and total spending among all cohorts. Among the overall cohort, we found a step change of 58.43% (95% CI 52.11, 64.75), accompanied by a gradual decrease of –0.66% per month (95% CI –1.13, –0.20) in terms of the proportion of biosimilar infliximab prescriptions dispensed after the switch policy. With respect to the proportion of total spending on biosimilar infliximab, we observed nonsignificant monthly trend changes for all groups.

Stratification by sex and neighborhood income quintile. When the proportion of biosimilar prescriptions dispensed were stratified by sex, we found similar associations across most groups (Table 4). However, a significant level change was detected post new start among male subjects receiving etanercept (but not among female subjects or overall) in terms of the proportion of biosimilar prescriptions dispensed and total spending. We also identified unique associations in terms of biosimilar spending among female subjects receiving infliximab after the new start and switch policy: there was a significant positive trend in the post new start period and a significant downward trend post switch detected among female subjects and nonsignificant changes among male subjects.

In general, our neighborhood income quintile–stratified analysis of the proportion of prescriptions dispensed of and total spending on biosimilar infliximab and etanercept demonstrated similar results to those from our overall cohort (Table 4). Notable deviations in terms of the proportion of biosimilar prescriptions dispensed include a significant level change (3.83 [95% CI 0.08, 7.59]) post infliximab new start among the fourth highest neighborhood income quintile and a nonsignificant trend post switch among lower income quintiles. There was also a positive

Table 2. Change in the proportion of prescriptions of infliximab and etanercept dispensed that were biosimilar after new start and switching policies were introduced*

	Etanercept new start policy (February 2016)		Etanercept switching policy (May 2019 to November 2019)		Infliximab new start policy (July 2017)		Infliximab switching policy (May 2019 to November 2019)	
Group	Level (95% CI)	Trend (95% CI)	Level (95% CI)	Trend (95% CI)	Level (95% CI)	Trend (95% CI)	Level (95% CI)	Trend (95% CI)
Overall	0.68 (–0.18, 1.53)	0.65 (0.44, 0.85)	76.98 (75.56, 78.41)	–0.95 (–1.04, –0.85)	1.74 (–1.61, 5.09)	0.71 (–0.14, 1.56)	58.43 (52.11, 64.75)	–0.66 (–1.13, –0.20)
RA	0.64 (–0.23, 1.52)	0.66 (0.45, 0.86)	77.95 (76.49, 79.41)	–0.99 (–1.09, –0.90)	0.63 (–2.99, 4.25)	0.74 (–0.18, 1.65)	58.97 (52.13, 65.81)	–0.74 (–1.24, –0.24)
PS	0.20 (–0.81, 1.21)	0.31 (0.07, 0.55)	81.04 (79.34, 82.74)	–0.65 (–0.76, –0.54)	2.33 (–3.15, 7.81)	0.64 (–0.76, 2.04)	53.54 (43.05, 64.02)	–0.37 (–1.15, 0.41)
AS	0.93 (–0.53, 2.39)	0.74 (0.27, 1.20)	67.88 (64.70, 71.07)	–1.24 (–1.50, –0.98)	1.60 (–3.72, 6.91)	0.64 (–0.70, 1.98)	56.13 (46.53, 65.74)	–0.81 (–1.47, –0.15)

* 95% confidence intervals (95% CIs) excluding 1 are indicative of statistical significance at $P < 0.05$. AS = ankylosing spondylitis; PS = plaque psoriasis and/or psoriatic arthritis; RA = rheumatoid arthritis.

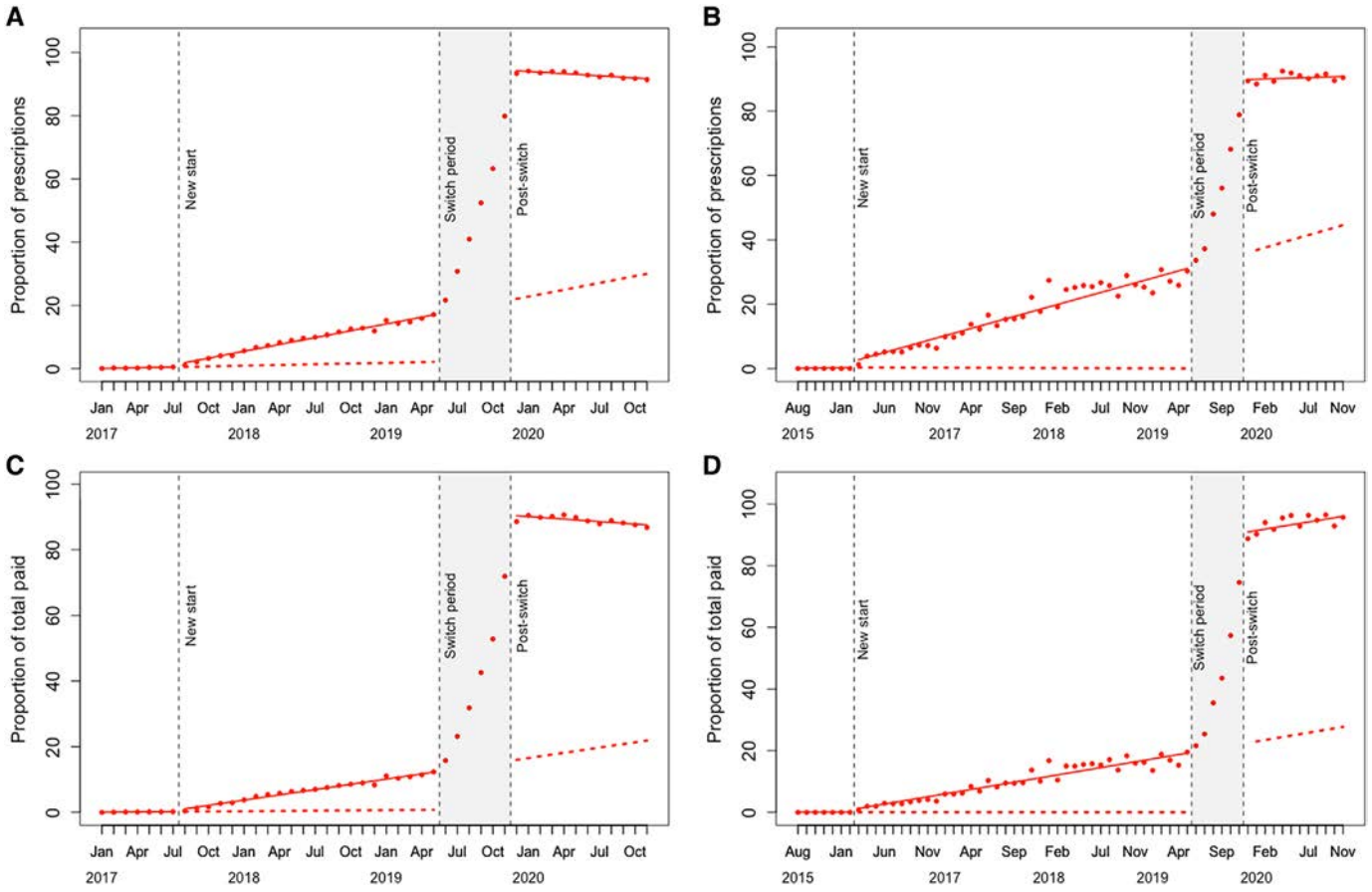

Figure 2. Interrupted time series analyses of the overall proportion of prescriptions dispensed that were biosimilar etanercept (**A**) and infliximab (**B**) and the proportion of total spending on biosimilar etanercept (**C**) and infliximab (**D**) after the new start and switching policies were introduced. Each dot represents the actual proportion of prescriptions dispensed (**A** and **B**) or the proportion of total payment (**C** and **D**) that was biosimilar in each month; the broken lines indicate the new start and switching period counterfactual.

relationship between increasing neighborhood income quintile and the magnitude of level change post infliximab switch: the lowest income quintile demonstrated an immediate increase of 49.81% (95% CI 39.62, 59.99) compared to 66.38% (95% CI 56.76, 76.00) among the highest income quintile. In terms of proportion of total spending, we detected a negative trend post infliximab switch among the highest income quintile only (0.59 [95% CI –1.02, –0.16]). No deviations were detected with respect to etanercept. Those with missing neighborhood income quintile data were excluded due to low cell counts.

DISCUSSION

Our findings suggest that mandatory biosimilar switching policies have the potential to greatly increase use beyond what was detected for new start programs. While the new start policies may have had a small impact on the monthly trend of biosimilar use, introduction of mandatory biosimilar switching resulted in an immediate increase in proportional spending and uptake, ranging from 58.43% to 78.22%, among the overall cohort.

We detected sex-based differences in the use of biosimilars. For instance, male subjects receiving etanercept demonstrated a significant level change post new start policy, whereas female subjects did not. On the other hand, a significant positive trend post infliximab new start and a significant negative trend post infliximab switch was detected among female subjects but not male subjects. Previous research has suggested that there may be differences in medication use by sex; e.g., female subjects may be less adherent to medications and may receive guideline-based care less often compared to male subjects (25). However, from this work alone, it is unclear whether there are differences in use of biosimilars by sex or whether these findings were due to random chance.

Similarly, we quantified some variations in proportional use of biosimilar infliximab based on neighborhood income quintile. We found that lower income quintiles had nonsignificant trend changes in terms of the proportion of biosimilar prescriptions dispensed post infliximab switch. In addition, there was an increase in the magnitude of the step change post switch with increasing neighborhood income quintile. A recent large systematic review

Table 3. Change in the proportion of total spending on biosimilar infliximab and etanercept after new start and switching policies were introduced*

	Etanercept new start policy (February 2016)		Etanercept switching policy (May 2019 to November 2019)		Infliximab new start policy (July 2017)		Switching policy (May 2019 to November 2019)	
Group	Level (95% CI)	Trend (95% CI)	Level (95% CI)	Trend (95% CI)	Level (95% CI)	Trend (95% CI)	Level (95% CI)	Trend (95% CI)
Overall	0.34 (–0.60, 1.28)	0.51 (0.28, 0.73)	78.22 (76.65, 79.79)	–0.79 (–0.90, –0.69)	0.77 (–1.65, 3.20)	0.47 (–0.13, 1.07)	71.23 (66.82, 75.65)	–0.01 (–0.33, 0.30)
RA	0.32 (–0.55, 1.20)	0.53 (0.33, 0.74)	80.71 (79.25, 82.17)	–0.86 (–0.96, –0.76)	0.30 (–2.64, 3.23)	0.46 (–0.27, 1.19)	71.55 (66.21, 76.89)	–0.09 (–0.48, 0.29)
PS	–0.05 (–1.21, 1.10)	0.24 (–0.03, 0.51)	76.55 (74.59, 78.50)	–0.52 (–0.65, –0.40)	1.00 (–2.79, 4.79)	0.45 (–0.49, 1.39)	67.65 (60.78, 74.52)	0.47 (–0.02, 0.95)
AS	1.33 (–0.21, 2.88)	0.76 (0.40, 1.13)	76.39 (73.81, 78.97)	–1.19 (–1.36, –1.02)	0.73 (–2.75, 4.21)	0.43 (–0.45, 1.31)	69.97 (63.50, 76.44)	0.002 (–0.47, 0.47)

* 95% confidence intervals (95% CIs) excluding 1 are indicative of statistical significance at $P < 0.05$. AS = ankylosing spondylitis; PS = plaque psoriasis and/or psoriatic arthritis; RA = rheumatoid arthritis.

Table 4. Change in the proportion of prescriptions dispensed of and total spending on biosimilar infliximab and etanercept after new start and switching policies were introduced, stratified by sex and neighborhood income quintile*

	Etanercept new start policy (February 2016)		Etanercept switching policy (May 2019 to November 2019)		Infliximab new start policy (July 2017)		Infliximab switching policy (May 2019 to November 2019)	
Group	Level (95% CI)	Trend (95% CI)	Level (95% CI)	Trend (95% CI)	Level (95% CI)	Trend (95% CI)	Level (95% CI)	Trend (95% CI)
Proportion of prescriptions dispensed by sex								
Female	0.37 (−0.66, 1.40)	0.71 (0.47, 0.96)	76.05 (74.33, 77.76)	−0.97 (−1.08, −0.85)	2.56 (−1.44, 6.55)	0.77 (−0.24, 1.77)	52.75 (45.60, 59.90)	−0.70 (−1.19, −0.22)
Male	1.56 (0.64, 2.48)	0.60 (0.36, 0.83)	78.88 (77.29, 80.47)	−0.91 (−1.00, −0.83)	0.72 (−2.94, 4.39)	0.60 (−0.32, 1.51)	65.32 (58.57, 72.08)	−0.57 (−1.06, −0.09)
Proportion of total spending by sex								
Female	−0.01 (−1.11, 1.09)	0.56 (0.30, 0.82)	77.68 (75.84, 79.51)	−0.79 (−0.91, −0.67)	0.91 (−1.27, 3.10)	0.57 (0.03, 1.11)	68.41 (64.36, 72.46)	−0.35 (−0.65, −0.05)
Male	1.07 (0.24, 1.89)	0.47 (0.27, 0.67)	79.58 (78.18, 80.97)	−0.81 (−0.90, −0.72)	0.54 (−2.10, 3.17)	0.36 (−0.32, 1.03)	75.23 (70.55, 79.91)	0.26 (−0.03, 0.55)
Proportion of prescriptions dispensed by neighborhood income quintile								
1 (lowest)	0.66 (−0.93, 2.24)	0.70 (0.30, 1.09)	79.97 (77.28, 82.66)	−0.85 (−1.01, −0.68)	0.39 (−5.41, 6.18)	0.88 (−0.57, 2.33)	49.81 (39.62, 59.99)	−0.34 (−0.99, 0.31)
2	2.31 (0.53, 4.09)	0.63 (0.20, 1.05)	73.47 (70.50, 76.43)	−0.96 (−1.16, −0.76)	1.60 (−4.45, 7.64)	0.59 (−0.99, 2.16)	54.80 (42.85, 66.75)	−0.86 (−1.76, 0.05)
3	0.25 (−1.04, 1.54)	0.87 (0.57, 1.18)	74.32 (72.17, 76.47)	−1.16 (−1.30, −1.02)	1.64 (−3.0, 6.28)	0.73 (−0.43, 1.89)	59.14 (51.00, 67.29)	−0.47 (−0.99, 0.05)
4	0.37 (−0.39, 1.14)	0.29 (0.06, 0.51)	80.43 (78.96, 81.91)	−0.78 (−0.90, −0.66)	3.83 (0.08, 7.59)	0.76 (−0.20, 1.72)	60.15 (53.47, 66.83)	−0.77 (−1.18, −0.36)
5 (highest)	−0.21 (−1.48, 1.06)	0.65 (0.28, 1.01)	75.75 (73.30, 78.20)	−0.92 (−1.12, −0.73)	0.28 (−5.20, 5.75)	0.73 (−0.64, 2.11)	66.38 (56.76, 76.00)	−0.83 (−1.45, −0.22)
Proportion of total spending by neighborhood income quintile								
1 (lowest)	0.05 (−1.76, 1.85)	0.50 (0.07, 0.92)	81.81 (78.80, 84.82)	−0.67 (−0.87, −0.47)	−0.31 (−3.77, 3.15)	0.55 (−0.32, 1.42)	72.82 (66.38, 79.26)	−0.14 (−0.63, 0.34)
2	1.56 (−0.53, 3.65)	0.51 (0.02, 1.01)	74.43 (70.95, 77.92)	−0.85 (−1.09, −0.62)	1.67 (−2.62, 5.96)	0.51 (−0.54, 1.55)	64.81 (57.26, 72.36)	0.06 (−0.45, 0.57)
3	−0.25 (−1.84, 1.35)	0.70 (0.32, 1.07)	76.30 (73.64, 78.97)	−0.90 (−1.08, −0.72)	0.00 (−2.89, 2.88)	0.48 (−0.25, 1.22)	70.44 (65.32, 75.56)	0.37 (0.05, 0.69)
4	0.88 (−0.54, 2.29)	0.37 (0.04, 0.71)	81.72 (79.36, 84.09)	−0.60 (−0.75, −0.44)	2.03 (−1.23, 5.29)	0.44 (−0.39, 1.27)	69.58 (63.79, 75.36)	−0.06 (−0.42, 0.30)
5 (highest)	−0.37 (−2.22, 1.49)	0.48 (0.02, 0.94)	77.39 (74.16, 80.62)	−0.92 (−1.15, −0.68)	0.12 (−3.73, 3.97)	0.42 (−0.55, 1.39)	80.81 (74.02, 87.58)	−0.59 (−1.02, −0.16)

* 95% confidence intervals (95% CIs) excluding 1 are indicative of statistical significance at $P < 0.05$.

of 81 studies of rheumatoid arthritis, medicine use, and socioeconomic status found variable relationships between income and medication use. For instance, lower income was associated with reduced adherence in Canada, but the reverse was true in the US (26). However, we only detected income quintile–based variation in use among infliximab switching and not with respect to etanercept. This may indicate variation according to specific factors associated with infliximab use or chance associations detected by our study.

Large increases in biosimilar use are likely in Canada as BC has continued to roll out biosimilar switching programs for patients receiving rituximab, adalimumab, enoxaparin, filgrastim, and a number of insulins, and similar policies have now been announced in a number of jurisdictions (7,15). Greater biosimilar use may make the Canadian market more appealing for biosimilar manufacturers to enter, having a feedforward effect on biologic competition.

In the US, biosimilars for the top-selling drug in the world, adalimumab, are set to be available for the first time in 2023 (13). Recent calls from the US Department of Health and Human Services for the CMS to do more to incentivize biosimilar use could consider mandatory biosimilar switching policies (11). From our findings, it appears that simply enacting biosimilar new start programs may not be sufficient to greatly enhance use of these medicines.

Notably, we detected a small but significant downward trend in the post-switch period, potentially indicating treatment failure and switches back to the reference product. Studies examining biosimilar infliximab and etanercept switching in BC found no differences in health care utilization (e.g., emergency department visits or hospitalizations) post policy (27,28). Nevertheless, concerns around the use of biosimilars remain, stemming in part from their regulatory designation as similar but not identical to the reference product in terms of efficacy and safety, their interchangeability and substitutability, and the use of clinical trial data from one disease state to substantiate claims of efficacy and safety in another (i.e., extrapolation of indication), among other factors (13,29–31). However, research examining biosimilar infliximab and etanercept switching in Denmark found no deleterious outcomes associated with the national policies, and systematic reviews have also not found evidence of significant safety or efficacy concerns with biosimilar switching (32–34).

Prior to this analysis, it was unclear to what extent a biosimilar switching policy would capture the market for a particular agent. For instance, with less expensive biosimilars (e.g., insulin glargine), one may expect that private insurers and/or individuals would be more likely to pay the difference between the biosimilar and reference in order to maintain treatment with the reference product. In addition, the BC government maintained a system by which patients would be able to receive exceptional coverage for the reference product on a case-by-case basis (15). Both provide mechanisms by which the impact of biosimilar switching policies may be decreased. Future work should examine the post-switch market structure over a longer time period, as it is unclear whether the large increases in uptake of biosimilars will be eroded as the market reaches a new equilibrium. Indeed, we detected a negative gradual trend post switch that should be monitored. Our sensitivity analysis also detected sex- and neighborhood income quintile–based differences in the use of biosimilar infliximab and etanercept, which should continue to be tracked. Our findings may be of particular interest to jurisdictions with large public payers and private insurers who are considering implementing biosimilar switching policies.

However, rapid biosimilar penetration via switching policies is just one potential policy lever. By combining national drug tendering and procurement with rapid uptake of biosimilars via both switching and new start policies, Denmark saved nearly $2 for every $3 spent on infliximab, and tendering provided similar levels of biosimilar infliximab savings in Norway (35,36). Countries including France and Belgium have combined biosimilar price caps, with prescribing quotas aimed at increasing use of biosimilars (37). The US recently approved the first interchangeable biosimilars for insulin glargine and adalimumab (which may be automatically substituted at the level of the pharmacy depending on state law), although it is too early to discern the impact of this designation (38). Interchangeable biosimilars may be particularly compatible with switching policies.

Our study has a number of limitations. We relied on administrative data to identify our cohort, which could have inaccuracies related to the diagnoses of interest. This was likely partially mitigated by the fact that etanercept is only indicated for treatment of the conditions of interest in our study and has limited off-label use. Further, any misclassification would likely have persisted over the entire study period, so it would not have modified the relative changes we observed. Our study also required only 1 instance of a diagnosis code of interest. However, by design, our study only included individuals who also received a prescription for either etanercept or infliximab and who had no previous instance of a diagnosis code for Crohn's disease or ulcerative colitis (in the case of infliximab). Therefore, we believe the potential impact of this less stringent approach on our denominators of interest (namely, total number of prescriptions dispensed or spending per month on either etanercept or infliximab) would be at least partially mitigated by design. Of note, individuals with plaque psoriasis and receiving etanercept were not required to switch until 2021. However, it was not possible to exclude individuals with plaque psoriasis from the etanercept analyses, as the relevant diagnosis codes pertain to both plaque psoriasis and psoriatic arthritis.

The NACRS does not cover all emergency department visits in BC, although this is unlikely to impact our findings. The generalizability of our results may be limited, as the magnitude of impact of biosimilar switching policies may vary by patient, physician, and payer-related factors and may also differ depending on the

medication of interest (39). Due to the nature of our study, we were unable to comment on the impact of these policies at an individual level. In addition to further exploration of the relative factors that may influence biosimilar uptake post switch, future work should examine the more long-term impacts of these policies. We were unable to comment on pricing and cost-savings related to these policies due to the presence of confidential rebates and listing agreements in Canada.

Overall, our study clearly shows that a mandatory switching policy was much more effective than a new starter policy at increasing the use of biosimilars. Thus, although new start policies may result in some small gradual increases in biosimilar use, payers can substantially influence the use of biosimilars through the implementation of mandatory biosimilar switching policies. Given the clinical similarity in their effect and potential savings, other jurisdictions and payers should seriously consider the use of these policies.

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. McClean 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. McClean, Cheng, Bansback, Clement, Tadrous, Harrison, Law.

Acquisition of data. Law.

Analysis and interpretation of data. McClean, Cheng, Bansback, Clement, Tadrous, Harrison, Law.

REFERENCES

1. IQVIA Institute for Human Data Science. The global use of medicines 2022: outlook to 2026. 2022. URL: https://www.iqvia.com/-/media/iqvia/pdfs/library/publications/the-global-use-of-medicines-2022.pdf.
2. Government of Canada. Patented Medicine Prices Review Board. Biologics in Canada. Part 1: market trends, 2018. URL: https://www.canada.ca/en/patented-medicine-prices-review/services/reports-studies/biologics-part1-market-trends.html.
3. IQVIA Institute for Human Data Science. Medicine use and spending in the US: a review of 2017 and outlook to 2022. 2018. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/medicine-use-and-spending-in-the-us-review-of-2017-outlook-to-2022.
4. IQVIA Institute for Human Data Science. Biosimilars in the United States 2020–2024: Competition, savings, and sustainability. 2020. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/biosimilars-in-the-united-states-2020-2024.
5. US Food and Drug Administration. Biosimilar development, review, and approval. 2017. URL: https://www.fda.gov/drugs/biosimilars/biosimilar-development-review-and-approval.
6. Hofmeister M, Sivakumar A, Clement F, et al. Trends in Canadian prescription drug purchasing: 2001–2020. J Pharm Policy Pract 2022;15:20.
7. McClean AR, Law MR, Harrison M, et al. Uptake of biosimilar drugs in Canada: analysis of provincial policies and usage data. CMAJ 2022;194:E556–60.
8. Chen AJ, Gascue L, Ribero R, et al. Uptake of infliximab biosimilars among the Medicare population. JAMA Intern Med 2020;180:1255–6.
9. Kim SC, Sarpatwari A, Landon JE, et al. Utilization and treatment costs of tumor necrosis factor inhibitors after the introduction of biosimilar infliximab in the United States. Arthritis Rheumatol 2020;72:1036–8.
10. Bansback N, Curtis JR, Huang J, et al. Patterns of tumor necrosis factor inhibitor (TNFi) biosimilar use across United States rheumatology practices. ACR Open Rheumatol 2020;2:79–83.
11. Murrin S, for the US Department of Health and Human Services Office of Inspector General. Medicare Part D and beneficiaries could realize significant spending reductions with increased biosimilar use. 2022. URL: https://oig.hhs.gov/oei/reports/OEI-05-20-00480.pdf.
12. Hagen T. PhRMA: federal biosimilar payment policies are having a positive effect. AJMC; 2020. URL: https://www.centerforbiosimilars.com/view/phrma-federal-biosimilar-payment-policies-are-having-a-positive-effect.
13. Yazdany J. Failure to launch: biosimilar sales continue to fall flat in the United States. Arthritis Rheumatol 2020;72:870–3.
14. Law MR, Cheng L, Kolhatkar A, et al. The consequences of patient charges for prescription drugs in Canada: a cross-sectional survey. CMAJ Open 2018;6:E63–70.
15. British Columbia Ministry of Health. Biosimilar initiative for health professionals. 2022. URL: https://www2.gov.bc.ca/gov/content/health/practitioner-professional-resources/pharmacare/prescribers/biosimilars-initiative-health-professionals.
16. Government of Canada. Drug product database online query. 2022. URL: https://health-products.canada.ca/dpd-bdpp/index-eng.jsp.
17. Population Data BC, British Columbia Ministry of Health. Medical Services Plan (MSP) payment information file. URL: https://www.popdata.bc.ca.
18. Population Data BC, Canadian Institute for Health Information. Discharge abstract database (hospital separations). URL: https://www.popdata.bc.ca.
19. Population Data BC, Canadian Institute for Health Information. National Ambulatory Care Reporting System data set. URL: https://www.popdata.bc.ca.
20. Population Data BC, British Columbia Ministry of Health. PharmaNet. Data Stewardship Committee. URL: https://www.popdata.bc.ca.
21. Population Data BC, British Columbia Ministry of Health. Consolidation File (MSP registration & premium billing). URL: https://www.popdata.bc.ca.
22. Hategeka C, Ruton H, Karamouzian M, et al. Use of interrupted time series methods in the evaluation of health system quality improvement interventions: a methodological systematic review. BMJ Glob Health 2020;5:e003567.
23. Wagner A, Soumerai S, Zhang F, et al. Segmented regression analysis of interrupted time series studies in medication use research. J Clin Pharm Ther 2002;27:299–309.
24. Government of Canada, Canadian Institutes of Health Research. Sex and gender in health research. 2021. URL: https://cihr-irsc.gc.ca/e/50833.html.
25. Manteuffel M, Williams S, Chen W, et al. Influence of patient sex and gender on medication use, adherence, and prescribing alignment with guidelines. J Womens Health (Larchmt) 2014;23:112–9.
26. Russell O, Lester S, Black RJ, et al. Socioeconomic status and medication use in rheumatoid arthritis: a scoping review. Arthritis Care Res (Hoboken) 2023;75:92–100.
27. Fisher A, Kim JD, Dormuth CR. Mandatory nonmedical switching from originator to biosimilar infliximab in patients with inflammatory arthritis and psoriasis in British Columbia: a cohort study. CMAJ 2022;10:E109–18.
28. Fisher A, Kim J, Carney G, et al. Rapid monitoring of health service use following a policy to switch patients from originator to biosimilar etanercept: a cohort study in British Columbia. BMC Rheumatol 2022;6:5.

29. Murdoch B, Caulfield T. The law and ethics of switching from biologic to biosimilar in Canada. J Can Assoc Gastroenterol 2020;3:228–33.

30. Gamez-Belmonte R, Hernandez-Chirlaque C, Arrendondo-Amador M, et al. Biosimilars: concepts and controversies. Pharmacol Res 2018; 133:251–64.

31. Teeple A, Ellis LA, Huff L, et al. Physician attitudes about non-medical switching to biosimilars: results from an online physician survey in the United States. Curr Med Res Opin 2019;35:611–7.

32. Glintborg B, Sorensen IJ, Loft AG, et al. A nationwide non-medical switch from originator infliximab to biosimilar CT-P13 in 802 patients with inflammatory arthritis: 1-year clinical outcomes from the DANBIO registry. Ann Rheum Dis 2017;76:1426–31.

33. Glintborg B, Loft AG, Omerovic E, et al. To switch or not to switch: results of a nationwide guideline of mandatory switching from originator to biosimilar etanercept. One-year treatment outcomes in 2061 patients with inflammatory arthritis from the DANBIO registry. Ann Rheum Dis 2019;78:192–200.

34. Feagan BG, Lam G, Ma C, et al. Systematic review: efficacy and safety of switching patients between reference and biosimilar infliximab. Aliment Pharmacol Ther 2019;49:31–40.

35. Jensen TB, Bartels D, Sædder EA, et al. The Danish model for the quick and safe implementation of infliximab and etanercept biosimilars. European J Clinl Phrmacol 2020;76:35–40.

36. Mestre-Ferrandiz J, Towse A, Berdud M. Biosimilars: how can payers get long-term savings? Pharmacoeconomics 2016;34:609–16.

37. Vogler S, Schneider P, Zuba M, et al. Policies to encourage the use of biosimilars in European countries and their potential impact on pharmaceutical expenditure. Front Pharmacol 2021;12:625296.

38. US Food and Drug Administration. Drugs@FDA: FDA-approved drugs. 2022. URL: https://www.accessdata.fda.gov/scripts/cder/daf/.

39. Dean EB, Johnson P, Bond AM. Physician, practice, and patient characteristics associated with biosimilar use in Medicare recipients. JAMA Netw Open 2021;4:e2034776.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 75, No. 9, September 2023, pp 2022–2031
DOI 10.1002/acr.25087

Birth Outcomes and Rehospitalizations Among Pregnant Women With Rheumatoid Arthritis and Systemic Lupus Erythematosus and Their Offspring

Namrata Singh,[1] Julianna Sabo,[1] Deborah A. Crane,[1] David R. Doody,[2] Melissa A. Schiff,[3] and Beth A. Mueller[4]

Objective. To compare obstetric/birth outcomes and rehospitalization among women with and without rheumatoid arthritis (RA) or systemic lupus erythematosus (SLE) and their infants.

Methods. This population-based retrospective cohort study identified women with RA (n = 1,223) and SLE (n = 1,354) and unexposed women with singleton births 1987–2014 in Washington State in linked vital hospital discharge records. Outcomes, including cause-specific hospitalizations <2 years postpartum, were compared by estimating adjusted relative risks (RRs) and cause-specific rehospitalization hazard ratios (HRs) with 95% confidence intervals (95% CIs).

Results. We observed increased risks of several adverse outcomes; RRs were often greatest for SLE. Women with RA/SLE more often required rehospitalization, most notably at <6 months postpartum (RA: 4% versus 2%; RR 2.22 [95% CI 1.62–3.04]; SLE: 6% versus 2%; RR 2.78 [95% CI 2.15–3.59]). Maternal postpartum rehospitalization was greatest for musculoskeletal conditions (RA: HR 19.1 [95% CI 13.6–26.8]; SLE: HR 29.8 [95% CI 22.1–40.1]). Infants of women with SLE more often had malformations (9% versus 6%; RR 1.46 [95% CI 1.21–1.75]), and increased mortality at <2 years (RR 2.11 [95% CI 1.21–3.67]). Infants of women with SLE also experienced more frequent rehospitalizations in their first year of life.

Conclusion. Women with RA or SLE and their infants experienced adverse outcomes, particularly infants of women with SLE. Maternal/infant rehospitalization was more common; most marked in the early months postpartum. Close follow-up during these time periods is crucial to minimize adverse outcomes.

INTRODUCTION

Steady increases have occurred in the number of pregnancies among women with rheumatoid arthritis (RA) or systemic lupus erythematosus (SLE) during recent decades (1,2), both conditions involving increased risks of adverse obstetric and birth outcomes (3–5). Among autoimmune conditions, SLE carries the greatest risk of poor outcome (4). These risks are well described and include intrauterine growth restriction, low birthweight, and increased caesarean delivery, among others (6,7).

Several studies have examined early outcomes among infants of mothers affected by RA or SLE, but less is known about maternal or infant longer-term postdelivery experience. In 1 study, women with SLE without a prior diagnosis of heart disease were observed to have a greater than 4-fold increased risk of new cardiac disease development requiring readmission in the postpartum period (8). A comprehensive evaluation of postdelivery rehospitalizations among women with RA or SLE or their infants has not been conducted. We aimed to compare obstetric outcomes among women with and without RA or SLE, and birth outcomes among their infants, including maternal and infant rehospitalization occurrence and reasons for rehospitalization within 2 years of delivery. Increased knowledge of the postpartum experience of this population will help identify time periods of greatest risk and help develop strategies to best support them postdelivery.

Dr. Singh's work was supported by the NIH/National Institute of Arthritis and Musculoskeletal and Skin Diseases (grant K23-AR-079588) and the Rheumatology Research Foundation. Dr. Mueller's work was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (grant R21-HD-073024).

[1]Namrata Singh, MD, MSCI, Julianna Sabo, MD, Deborah A. Crane, MD, MPH: University of Washington, Seattle; [2]David R. Doody, MS: Fred Hutchinson Cancer Research Center, Seattle, Washington; [3]Melissa A. Schiff, MD, MPH: University of New Mexico, Albuquerque, and University of Washington School of Public Health, Seattle; [4]Beth A. Mueller, DrPH: Fred Hutchinson Cancer Research Center and University of Washington School of Public Health, Seattle.

Drs. Singh and Sabo contributed equally to this work.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.25087&file=acr25087-sup-0001-Disclosureform.pdf.

Address correspondence via email to Namrata Singh, MD, MSCI, at nasingh@uw.edu.

Submitted for publication June 10, 2022; accepted in revised form January 5, 2023.

SIGNIFICANCE & INNOVATIONS

- Pregnant women with rheumatoid arthritis (RA) or systemic lupus erythematosus (SLE) have greater risks of adverse outcomes such as preeclampsia (9% RA, 15% SLE versus 6% in unexposed women) and preterm labor (10% RA, 15% SLE versus 5% in unexposed women).
- Care providers for women with these conditions should be aware of an increased rehospitalization risk in the 2 years postpartum, especially in the first 6 months.
- Infants of women with SLE have increased risks of serious outcomes, including malformations, rehospitalization, and mortality.

MATERIALS AND METHODS

Institutional Review Board approvals for this project were obtained from Washington State and the Fred Hutchinson Cancer Research Center. This retrospective cohort study evaluated the associations of RA and SLE with selected adverse pregnancy and neonatal outcomes among pregnant women in Washington State from 1987 to 2014.

Study cohort. Washington State birth-hospital discharge data (Comprehensive Hospital Abstract Reporting System) (9) contain patient-level information on all nonfederal inpatient and observation discharges in nonfederal facilities in the state. Multiple International Classification of Diseases, Ninth Revision (ICD-9) codes were available for each discharge, using Medicare-Medicaid billing standards. This file has been routinely linked to state birth records to build a linked file of birth and hospital discharge records for the delivery hospitalizations of the mother and infants and has been used extensively to examine pregnancy outcomes (10,11). Using methods previously described (12), we identified all women with live birth deliveries in Washington State during 1987–2014 in linked Washington State birth/fetal death-hospital discharge records. Briefly, we screened up to 25 hospital discharge fields in the linked birth-hospital discharge records to identify all women with ICD-9 diagnosis codes indicating RA (714.X, 725.X [n = 1,252]) or SLE (710.X [n = 1,404]) on their delivery hospital discharge record for the delivery. For each, we randomly selected from the remaining linked birth-hospital discharge records 10 women with deliveries in the same year to serve as our unexposed cohort. After excluding women with multiple gestations and/or fetal deaths, our data included 1,223 women with RA, 1,354 women with SLE, and 12,293 and 13,751 unexposed women, respectively, for analyses.

Covariate and outcome assessment. The linked birth-hospital discharge records for the delivery provided information concerning pregnancy course and maternal and infant outcomes and conditions noted at delivery and during the delivery hospitalization. To examine longer term outcomes and rehospitalizations, we further linked all subject records to subsequent hospital discharge and death records for the 2 years after delivery to assess the occurrence of, and reasons for, subsequent nonpregnancy-related rehospitalizations (ICD-9 630–679, 760–779 excluded), and to measure mortality among mothers and children.

Maternal characteristics at delivery and pregnancy information available from the birth record included the age at delivery (ranges 12–19, 20–24, 25–29, 30–34, 35–39, 40+ years), marital status, education (<12, 12, >12 years), Medicaid status, numbers of prior pregnancies and births (1, 2, 3, 4+), prior fetal deaths (0, 1+), prenatal smoking, prepregnancy body mass index (BMI; <18.5, 18.5–24.9, 25.0–29.9, 30.0+), use of fertility treatment, adequacy of gestational weight gain per American College of Gynecology recommendations (13) calculated from gestational weight gain, gestational length, and maternal prepregnancy BMI as reported on the birth record, and the Kotelchuck index of prenatal care adequacy (inadequate, intermediate, adequate, intensive) (14). Race and ethnicity (White, Black, Hispanic, Asian, American Indian/Alaska Native, and Pacific Islander, the latter 2 categories combined due to small numbers) were obtained from the birth record, largely self-reported by the mother around the time of delivery using a birth certificate worksheet with checkboxes. The 2003 standard birth certificate revision incorporated multirace checkboxes; bridging and classification of categories follows the National Center for Health Statistics guidelines (15).

Selected pregnancy and neonatal outcomes were identified by screening both the birth and linked hospital discharge records for the delivery hospitalization, as the use of both resources in combination improves the accuracy of identification of pregnancy conditions and outcomes (16). Birth records use a checkbox format and hospital discharge records contain up to 25 ICD-9 diagnosis codes to indicate outcomes. Hospital discharge records were screened to augment identification of chronic hypertension (ICD-9 401–405, 642.0–642.2, 642.7, 642.9), diabetes mellitus (ICD-9 250, 362.0, 648.01, 648.02), gestational diabetes mellitus (ICD-9 648.8, assessed only among women without existing diabetes mellitus), preeclampsia/eclampsia (ICD-9 642.4, 642.5, 642.6, 642.7), severe preeclampsia (ICD-9 642.5, 642.6), placental abruption (ICD-9 641.2, 762.1), preterm premature rupture of membranes (ICD-9 658.1, 658.2, 761.1, restricted to <37 weeks of gestation per birth record), urinary tract infection/pyelonephritis (ICD-9 599.0, 590.1, 590.8, 590.9), preterm labor (ICD-9 644.0, 644.1, 644.2), postpartum hemorrhage (ICD-9 666.X), anemia (ICD-9 280.9, 648.2), and cesarean delivery (ICD-9 diagnosis codes 669.7, 763.4 and procedure codes 74.x). Deep venous thrombosis (DVT)/pulmonary embolism (PE) (ICD-9 453.4, 415.1, 671.3, 671.4, 673.0, 673.2, V12.51, 673) was identified only by the maternal hospital discharge record.

Neonatal outcomes assessed from the birth record included gestational age at delivery <32, 32 to <37, 37+ weeks),

birthweight (<2,500, 2,500+ grams, the largest nonmissing value being 7,460 grams), small size for gestational age (<10th percentile for gestational age using birth weight nomograms derived from the population-based Washington State birth data), Apgar score <7 at 5 minutes, use of assisted ventilation for >30 minutes after delivery, neonatal intensive care unit (NICU) stay, and breastfeeding initiation (no/yes). Infant conditions assessed using the birth record and hospital discharge record in combination included fetal distress (ICD-9 656.3, 768.2–768.4) and congenital malformations (ICD-9 740–759). Maternal and infant lengths of stay for the delivery hospitalization (<3, 3–5, 6+ days) were assessed from the hospital discharge records.

Statistical analysis. Descriptive statistics characterized women with RA/SLE and unexposed women. We used multivariate Poisson regression with robust SEs to account for common outcomes to estimate relative risks (RRs) and 95% confidence intervals (95% CIs) to assess the associations of RA or SLE with selected adverse outcomes, adjusting a priori for maternal age, delivery year, and parity. Potential confounding by marital status, race/ethnicity and education was assessed, but none of these meaningfully changed the RR (>10%) so they were not retained in the final estimates. Results based on cell count sizes of <5 were suppressed. As our study period encompassed nearly 3 decades during which therapies have changed, we stratified the analyses into 2 time periods, 1987–1999 and 2000–2014, to assess possible temporal changes and tested for significant differences in risk estimates over time using a Wald test of the coefficient for the interaction of case/comparison by time period. A sensitivity analysis was also conducted by repeating analyses for nulliparous women only (RA: 516 exposed, 4,997 unexposed; SLE: 555 exposed, 5,596 unexposed).

To assess relative occurrences of rehospitalizations by ICD-9/ICD-10–based diagnosis group categories (17,18) of mothers and infants during the 2 years after delivery, Cox regressions were performed to compute hazard ratios (HRs) and 95% CI, accounting for the matching variables (birth year, sex), maternal age, and parity via baseline hazard stratification. Follow-up accrued from the delivery date through whichever came first: first rehospitalization after delivery, death, or December 31, 2014. Analyses were performed using Stata, version 15.

RESULTS

Women with RA and SLE were more likely than unexposed women to be age >35 years at delivery, slightly more likely to be married, have >12 years of education, and to have private health insurance (Table 1). They were also disproportionately more likely to be American Indian/Alaskan Native (6% RA, 4% SLE versus 2% unexposed). A lower proportion of women with RA were identified as Asian (5% versus 10% unexposed), but greater proportions of women with SLE were Asian (12% versus 8% unexposed) or Black (6% versus 4% unexposed). Prior pregnancy histories were similar between women with RA and SLE and their comparators, except women with these conditions were slightly more likely to have used fertility treatments and a greater number of women with SLE had prior fetal losses (9% versus 3% unexposed). Established diabetes mellitus levels were generally similar between women with RA and SLE and unexposed women, but chronic hypertension was more common in women with RA (5%) and SLE (10%) versus 2–3% unexposed.

Maternal outcomes. Women with RA or SLE more often required intensive levels of prenatal care (RA: 34% versus 23%; RR 1.46 [95% CI 1.33–1.60]; SLE: 55% versus 22%, RR 2.42 [95% CI 2.27–2.58]) and were more likely than women without these conditions to have less than appropriate gestational weight gain (RA: 28% versus 22%; RR 1.30 [95% CI 1.16–1.47]; SLE: 28% versus 21%; RR 1.28 [95% CI 1.12–1.45]) (Table 2). Gestational diabetes mellitus was not increased in either group, but all other conditions examined occurred more often in women with these conditions, with the exception of placental abruption and postpartum hemorrhage, which were statistically significantly increased only among women with SLE at 2% (versus 1% exposed) and 5% (versus 4%), respectively. Preeclampsia occurred more often during pregnancies of women with RA (9% versus 6%; RR 1.42 [95% CI 1.17–1.71]) or SLE (15% versus 6%; RR 2.33 [95% CI 2.01–2.70]), as did preterm rupture of membranes (RA: 6% versus 2%; RR 2.86 [95% CI 2.20–3.72]; SLE: 6% versus 2%; RR 3.28 [95% CI 2.54–4.23]). Cesarean deliveries were more common among nulliparous women in both groups (RA: 40% versus 28%; SLE: 39% versus 27%; RR 1.32 [95% CI 1.18–1.48] for both conditions). To examine whether increased cesarean delivery occurrence was due to increased levels of adverse pregnancy conditions that are indications for surgical delivery, we recalculated the RR after excluding women with preeclampsia, eclampsia, macrosomia, placenta previa, fetal distress, or malpresentation; the risk remained increased (RA: RR 1.51 [95% CI 1.25–1.83]; SLE: RR 1.46 [95% CI 1.18–1.79]). Women with RA or SLE more often had long delivery hospitalizations of 6 or more days (RA: 9% versus 2%; RR 2.43 [95% CI 1.78–3.31]; SLE: 10% versus 2%; RR 5.57 [95% CI 4.52–6.86]). Maternal deaths in both groups were too few to assess.

Infant outcomes. Infants of women with these conditions experienced all adverse outcomes examined, with the exception of fetal distress, malformations, and 5-minute Apgar score <7 among infants of women with RA. Infants of women with RA or SLE were more likely to weigh <2,500 grams (RA: 10% versus 5%; RR 2.08 [95% CI 1.72–2.52]; SLE: 21% versus 5%; RR 4.88 [95% CI 4.27–5.58]), be small for gestational age (RA: 11% versus 9%; RR 1.26 [95% CI 1.07–1.50]; SLE: 10% versus 9%; RR 2.30 [95% CI 2.04–2.59]), be delivered at <32 weeks' gestation (RA: 2% versus 1%; RR 1.83 [95% CI 1.13–2.97]; SLE: 4% versus 1%; RR 5.13 [95% CI 3.75–7.01]), or require NICU

Table 1. Characteristics of women with and without RA or SLE with singleton live birth deliveries in Washington State, 1987–2014*

Characteristic at delivery	With RA (n = 1,223)	Comparison (n = 12,293)	With SLE (n = 1,354)	Comparison (n = 13,762)
Maternal age, years				
12–19	40 (3.3)	957 (7.8)	36 (2.7)	1,157 (8.4)
20–24	173 (14.1)	2,707 (22.0)	217 (16.0)	3,113 (22.6)
25–29	297 (24.3)	3,531 (28.7)	389 (28.7)	4,055 (29.5)
30–34	416 (34.0)	3,222 (26.2)	433 (32.0)	3,425 (24.9)
35+	297 (24.3)	1,875 (15.3)	279 (20.6)	2,012 (14.6)
Marital status				
Unmarried	330 (27.0)	3,896 (31.7)	357 (26.4)	4,133 (30.1)
Married	330 (27.0)	3,896 (31.7)	996 (73.6)	9,594 (69.9)
Race and ethnicity				
White	908 (75.9)	8,664 (72.1)	931 (70.3)	9,892 (73.6)
Hispanic	91 (7.6)	1,255 (10.4)	95 (7.2)	1,465 (10.9)
Asian	63 (5.3)	1,174 (9.8)	158 (11.9)	1,079 (8.0)
Black	55 (4.6)	564 (4.7)	78 (5.9)	589 (4.4)
American Indian/Alaska Native, or Pacific Islander	79 (6.6)	356 (2.9)	63 (4.7)	412 (3.0)
Education, years†				
<12	126 (11.0)	1,993 (17.3)	105 (8.6)	2,196 (18.0)
12	240 (21.0)	2,846 (24.7)	273 (22.5)	3,121 (25.5)
13+	778 (68.0)	6,681 (58.0)	836 (68.9)	6,916 (56.5)
Health insurance				
Medicaid	442 (36.1)	5,005 (40.7)	464 (34.3)	5,519 (40.1)
Private	781 (63.9)	7,286 (59.3)	890 (65.7)	8,243 (59.9)
Prior pregnancies				
0	388 (32.5)	3,912 (32.5)	377 (28.5)	4,286 (31.8)
1	317 (26.5)	3,294 (27.3)	331 (25.0)	3,777 (28.0)
2+	490 (41.0)	4,842 (40.2)	617 (46.6)	5,426 (40.2)
Prior live births				
0	516 (42.9)	4,997 (41.3)	555 (41.8)	5,596 (41.4)
1	377 (31.4)	3,848 (31.8)	423 (31.9)	4,346 (32.1)
2+	309 (25.7)	3,241 (26.8)	349 (26.3)	3,587 (26.5)
Prior fetal deaths‡				
0	194 (97.5)	1,894 (96.9)	360 (91.4)	3,744 (96.9)
1+	5 (2.5)	61 (3.1)	34 (8.6)	118 (3.1)
Prenatal smoking				
No	1,086 (90.6)	10,893 (89.8)	1,180 (89.5)	11,967 (88.7)
Yes	113 (9.4)	1,239 (10.2)	138 (10.5)	1,518 (11.3)
Chronic hypertension				
No	1,161 (94.9)	11,989 (97.5)	1,216 (89.8)	13,469 (97.9)
Yes	63 (5.2)	305 (2.5)	139 (10.3)	294 (2.1)
Preexisting diabetes mellitus				
No	1,206 (98.6)	12,159 (98.9)	1,337 (98.7)	13,622 (99.0)
Yes	18 (1.5)	135 (1.1)	17 (1.3)	141 (1.0)
Prepregnancy BMI§				
<18.5 (low)	20 (2.3)	244 (2.8)	31 (4.3)	203 (2.8)
18.5–24.9	408 (47.5)	3,939 (46.0)	343 (47.6)	3,493 (48.2)
25.0–29.9 (overweight)	208 (24.2)	2,254 (26.3)	162 (22.5)	1,855 (25.6)
30.0+ (obese)	223 (26.0)	2,126 (24.8)	184 (25.6)	1,691 (23.3)
Fertility treatment				
No	908 (98.5)	9,198 (99.1)	768 (98.3)	7,829 (99.0)
Yes	14 (1.5)	82 (0.9)	13 (1.7)	81 (1.0)

* Values are the number (%). Numbers may not sum to totals due to missing data. BMI = body mass index; RA = rheumatoid arthritis; SLE = systemic lupus erythematosus.
† Data available for deliveries 1992 or later: 1,179 women with RA, 11,862 women without; 1,257 women with SLE, 12,751 women without.
‡ Fetal deaths at 20+ weeks among women with prior pregnancies only, data available for deliveries prior to 2003 only: 199 women with RA, 1,947 women without; 396 women with SLE, 3,868 women without.
§ Data available for deliveries 2003 or later: 935 women with RA, 9,395 women without; 792 women with SLE, 8,011 women without.

admission (RA: 12% versus 6%; RR 1.89 [95% CI 1.56–2.30]; SLE: 16% versus 6%; RR 2.71 [95% CI 2.25–3.28]) (Table 3). Infants of women with SLE were more likely to have a malformation (9% versus 6%; RR 1.46 [95% CI 1.21–1.75]) or die within 2 years (RR 2.11 [95% CI 1.21–3.67]). There were too few deaths at <2 years among infants of women with RA to assess.

Table 2. Pregnancy outcomes and complications among women with and without RA or SLE in Washington State, 1987–2014*

Outcome	RA			SLE		
	Yes (n = 1,223)	No (n = 12,293)	RR (95% CI)†	Yes (n = 1,354)	No (n = 13,762)	RR (95% CI)†
Prenatal care adequacy						
Inadequate	10.3	13.9	0.86 (0.71–1.03)	9.0	13.5	0.76 (0.63–0.92)
Intermediate	12.7	16.2	0.80 (0.68–0.84)	6.9	17.1	0.41 (0.33–0.51)
Adequate	42.8	47.3	1.00 [Ref.]	28.8	47.1	1.00 [Ref.]
Intensive	34.1	22.5	1.46 (1.33–1.60)	55.3	22.3	2.42 (2.27–2.58)
Pregnancy weight gain‡						
Appropriate	31.5	32.2	1.00 [Ref.]	34.2	31.0	1.00 [Ref.]
Less than appropriate	27.7	21.7	1.30 (1.16–1.47)	27.6	21.4	1.28 (1.12–1.45)
Excessive	40.8	46.1	0.89 (0.82–0.97)	38.2	47.5	0.81 (0.74–0.90)
During pregnancy						
Anemia	13.9	10.4	1.43 (1.23–1.67)	12.6	9.7	1.38 (1.19–1.61)
DVT/PE§	–	–	–	2.1	0.1	20.73 (10.42–41.25)
Preeclampsia/eclampsia	9.3	6.4	1.42 (1.17–1.71)	14.8	6.1	2.33 (2.01–2.70)
Severe preeclampsia	2.1	0.8	2.51 (1.62–3.90)	2.9	0.9	3.12 (2.13–4.57)
Gestational diabetes mellitus¶	9.0	6.6	1.16 (0.95–1.41)	6.8	5.6	1.08 (0.87–1.33)
Cesarean delivery#	40.3	27.7	1.32 (1.18–1.48)	38.6	26.5	1.32 (1.18–1.48)
Placental abruption	0.8	1.2	0.62 (0.31–1.22)	2.0	1.0	1.84 (1.21–2.78)
Preterm rupture of membranes	6.0	2.0	2.86 (2.20–3.72)	5.8	1.8	3.28 (2.54–4.23)
Intrauterine growth retardation	4.8	2.9	1.59 (1.21–2.09)	9.2	2.6	3.50 (2.87–4.27)
Preterm labor	10.0	5.4	1.79 (1.48–2.16)	15.1	5.1	3.07 (2.65–3.56)
Postpartum hemorrhage	4.9	4.1	1.26 (0.96–1.64)	5.3	3.9	1.34 (1.05–1.71)
Length of stay, days**						
<3	62.8	74.1	1.00 [Ref.]	51.6	74.7	1.00 [Ref.]
3–5	33.0	24.5	1.12 (1.03–1.21)	37.8	23.8	1.26 (1.17–1.36)
6+	4.3	1.5	2.43 (1.78–3.31)	10.6	1.6	5.57 (4.52–6.86)

* Values are the percentage unless indicated otherwise. 95% CI = 95% confidence interval; DVT = deep venous thrombosis; PE = pulmonary embolism; RA = rheumatoid arthritis; Ref. = reference; RR = relative risk; SLE = systemic lupus erythematosus.
† RR estimated from Poisson regression with robust SEs, adjusted for birth year, maternal age, and parity.
‡ Data available for deliveries 2003 or later: 935 women with RA, 9,395 women without; 792 women with SLE, 8,011 women without.
§ Results suppressed if cell size <5.
¶ Among women without established diabetes mellitus: 1,206 women with RA, 12,159 women without; 1,337 women with SLE, 13,622 women without.
# Among women without prior deliveries: 516 women with RA, 4,997 women without; 555 women with SLE, 5,596 women without.
** Additionally adjusted for method of delivery.

Maternal and infant rehospitalization at <2 years of delivery. Women with RA or SLE were more likely to be rehospitalized for nonpregnancy-related causes during the 2 years after delivery than were unexposed women (RA: 12% versus 10%; RR 1.33 [95% CI 1.13–1.56]; SLE: 16% versus 11%; RR 1.48 [95% CI 1.29–1.69]) (Table 4). Rehospitalization risks among women with RA (4% versus 2%; RR 2.22 [95% CI 1.62–3.04]) and SLE (6% versus 2%; RR 2.78 [95% CI 2.15–3.59]) were increased and remained increased during the first year postpartum. Infants of women with these conditions were also more likely to be rehospitalized at <2 years (RA: 10% versus 9%; RR 1.22 [95% CI 1.01–1.46]; SLE: 13% versus 9%; RR 1.59 [95% CI 1.36–1.84]). Infants of women with SLE had the greatest rehospitalization risks in the first year (RR 1.64 [95% CI 1.36–1.98] at <6 months; RR 1.63 [95% CI 1.10–2.42] at 6 to <12 months).

Maternal rehospitalization HRs were increased for women with RA or SLE for most causes, being greatest for musculoskeletal-related conditions (RA: HR 19.1 [95% CI 13.6–26.8]; SLE: HR 29.8 [95% CI 22.1–40.1]), but also increased for injury and mental health conditions (Figure 1). Among infants of women with RA, rehospitalization HRs were only increased for genitourinary conditions (HR 2.2 [95% CI 1.2–4.0]), although borderline statistically significant increases were also noted for infection- and respiratory-related conditions. Infants of women with SLE had increased rehospitalization HRs for nearly all causes, the greatest being for musculoskeletal conditions (HR 4.2 [95% CI 1.8–9.6]), with HRs >2.0 also observed for hematologic, circulatory, digestive, and genitourinary conditions, and significantly increased HRs >1.4 for infection-, endocrine-, nervous system–, and respiratory-related conditions.

Time period stratified and sensitivity analyses. Most results for mother and infant were similar when compared across the periods 1987–1999 and 2000–2014, with only modest and statistically nonsignificant changes noted from the earlier to the later time period for both RA and SLE (Table 5). The only statistically significant change noted was for SLE, with decreased preterm delivery (P = 0.02), although the RR remained markedly increased in both time periods.

Table 3. Selected outcomes among singleton infants of women with and without RA or SLE with deliveries in Washington State, 1987–2014*

Outcome	RA			SLE		
	Yes (n = 1,223)	No (n = 12,293)	RR (95% CI)†	Yes (n = 1,354)	No (n = 13,762)	RR (95% CI)†
Fetal distress	9.8	8.4	1.12 (0.93–1.34)	13.1	10.2	1.26 (1.09–1.46)
Birthweight <2,500 grams	10.1	4.9	2.08 (1.72–2.52)	20.8	4.5	4.88 (4.27–5.58)
<37 weeks' gestation	14.8	6.9	2.11 (1.81–2.46)	23.4	6.6	3.65 (3.24–4.11)
<32 weeks' gestation	1.6	1.0	1.83 (1.13–2.97)	4.4	1.0	5.13 (3.75–7.01)
Small for gestational age	11.1	9.0	1.26 (1.07–1.50)	20.3	9.0	2.30 (2.04–2.59)
Apgar score <7 at 5 minutes‡	2.6	1.9	1.26 (0.87–1.82)	4.0	1.9	2.01 (1.51–2.69)
Malformation	7.6	6.6	1.15 (0.93–1.42)	9.1	6.3	1.46 (1.21–1.75)
NICU admission§	12.0	6.4	1.89 (1.56–2.30)	15.6	5.9	2.71 (2.25–3.28)
Not breastfed§	10.7	7.7	1.60 (1.31–1.96)	12.9	7.9	1.77 (1.45–2.16)
Assisted ventilation >30 min.¶	1.7	0.8	2.21 (1.37–3.58)	2.9	0.7	4.20 (2.84–6.23)
Days hospitalized‡						
<3	69.2	81.4	1.00 [Ref.]	62.4	81.1	1.00 [Ref.]
3–5	21.6	14.4	1.19 (1.07–1.32)	22.2	15.1	1.05 (0.95–1.16)
6+	9.2	4.2	1.98 (1.61–2.43)	15.4	3.8	3.67 (3.15–4.28)
Death at <1 year#	–	–	–	1.1	0.5	2.27 (1.30–3.98)
Death at <2 years#	–	–	–	1.1	0.6	2.11 (1.21–3.67)

* Values are the percentage unless indicated otherwise. 95% CI = 95% confidence interval; NICU = neonatal intensive care unit; RA = rheumatoid arthritis; Ref. = reference; RR = relative risk; SLE = systemic lupus erythematosus.
† RR estimated from Poisson regression with robust SEs, adjusted for birth year, maternal age, and parity.
‡ 5-minute APGAR score, additionally adjusted for method of delivery.
§ Data available for deliveries 2003 or later: 935 women with RA, and 9,395 women without; 792 women with SLE, and 8,011 women without.
¶ Data available for 1,202 women with RA, and 12,087 women without with deliveries in 1989 or later.
# Results suppressed if cell size <5.

Table 4. Maternal and infant rehospitalization in first 2 years after delivery among women with and without RA and SLE*

Rehospitalization	RA			SLE		
	Yes (n = 1,223)	No (n = 12,293)	RR (95% CI)†	Yes (n = 1,354)	No (n = 13,762)	RR (95% CI)†
Maternal						
Ever	12.3	9.9	1.33 (1.13–1.56)	15.9	11.2	1.48 (1.29–1.69)
No. of rehospitalizations						
0	87.7	90.1	1.00 [Ref.]	84.1	88.8	1.00 [Ref.]
1	9.2	8.5	1.15 (0.95–1.39)	11.6	9.8	1.24 (1.06–1.45)
2+	3.2	1.4	2.43 (1.71–3.45)	4.3	1.5	3.09 (2.31–4.14)
Months to rehospitalization						
No event	87.7	90.1	1.00 [Ref.]	84.1	88.8	1.00 [Ref.]
<1	2.2	0.9	2.48 (1.63–3.77)	2.7	0.9	2.89 (1.99–4.20)
1 to <3	0.7	0.4	1.77 (0.87–3.58)	1.6	0.6	2.99 (1.85–4.85)
3 to <6	0.8	0.4	2.10 (1.08–4.08)	1.3	0.5	2.37 (1.38–4.07)
6 to <12	2.0	1.2	2.10 (1.36–3.25)	2.3	1.4	1.91 (1.31–2.79)
12 to <24	6.5	7.0	0.98 (0.79–1.23)	8.1	7.8	1.07 (0.88–1.29)
Infant						
Ever	9.6	8.6	1.22 (1.01–1.46)	12.8	8.7	1.59 (1.36–1.84)
No. of rehospitalizations						
0	90.4	91.4	1.00 [Ref.]	87.2	91.3	1.00 [Ref.]
1	7.8	7.2	1.17 (0.95–1.44)	9.4	7.2	1.41 (1.18–1.69)
2+	1.9	1.4	1.48 (0.95–2.31)	3.4	1.6	2.39 (1.74–3.28)
Months to rehospitalization						
No event	90.4	91.4	1.00 [Ref.]	87.2	91.3	1.00 [Ref.]
<1	3.4	3.4	1.05 (0.77–1.45)	4.3	3.2	1.41 (1.07–1.84)
1 to <3	1.8	1.5	1.27 (0.81–2.00)	3.0	1.7	1.97 (1.41–2.75)
3 to <6	1.3	0.9	1.62 (0.96–2.72)	1.5	0.9	1.93 (1.20–3.10)
6 to <12	1.5	1.3	1.23 (0.74–2.04)	2.1	1.4	1.63 (1.10–2.42)
12 to <24	1.6	1.4	1.30 (0.81–2.09)	1.9	1.5	1.33 (0.88–2.01)

* Values are the percentage unless indicated otherwise. 95% CI = 95% confidence interval; RA = rheumatoid arthritis; Ref. = reference; RR = relative risk; SLE = systemic lupus erythematosus.
† RR estimated from Poisson regression with robust SEs, adjusted for birth year, maternal age, and parity.

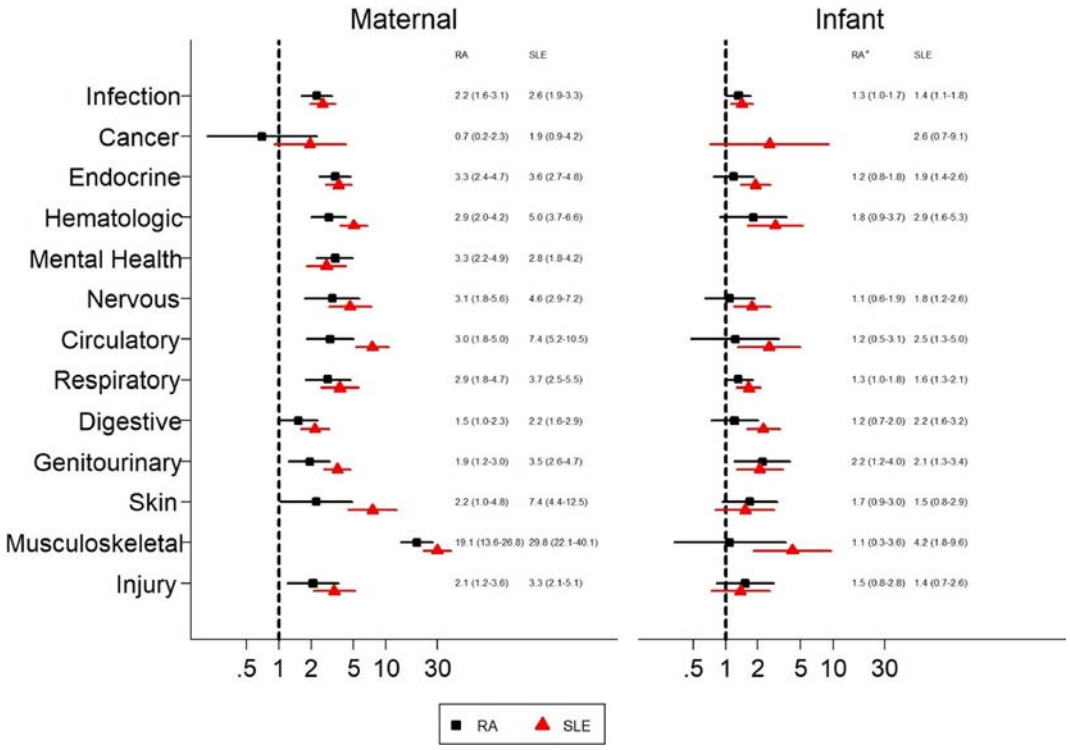

Figure 1. Hazard ratios and 95% confidence intervals for cause-related rehospitalization postdelivery among women with rheumatoid arthritis (RA) or systemic lupus erythematosus (SLE) and their infants. [a] = no cancer hospitalizations observed in infants of women with RA.

Sensitivity analyses of maternal and infant outcomes restricted to records of nulliparous women did not materially change results for either condition, with the exception that the placental abruption risk among women with SLE was markedly attenuated (odds ratio 1.07 [95% CI 0.46–2.46], results not shown).

DISCUSSION

In this population-based study, women with RA and SLE required a more intensive level of prenatal care, had inadequate weight gain, had preeclampsia, and had increased cesarean deliveries and longer hospitalizations than comparators. Infants of women with these conditions had higher rates of small for gestational age, preterm birth, and NICU admissions than unexposed women's infants. Both women with RA and SLE were more likely to be rehospitalized for nonpregnancy-related causes during the 2 years after delivery. However, reassuringly, the majority of women with RA and SLE did not have these complications, and we observed a decrease in RRs of preterm delivery over time. Our study builds upon prior literature by reporting on rehospitalization and longer-term mortality after delivery among mothers with RA or SLE and their infants.

Our results are consistent with prior studies of many outcomes, including an increased risk for intrauterine growth restriction, preeclampsia/eclampsia, preterm labor, and cesarean deliveries in women with RA or SLE (3–6) and postpartum hemorrhage in women with SLE (19). We found an increased occurrence of anemia in women with RA (14%) or SLE (13%), although an increased risk has only been previously reported in SLE (20). This complication may be explained as an association with the underlying chronic disease (in RA and SLE) and/or the increased tendency for postpartum hemorrhage among women with SLE.

Despite women with RA or SLE being more likely to receive intensive prenatal care, they were 28–30% more likely to have less than appropriate gestational weight gain (28% of women with either condition). This finding could be from the burden of their disease that prohibits adequate nutrition for the mother and fetus and may also be why we observed infants with low birthweight and size for gestational age in their mothers.

Although RA and SLE have both been identified as risk factors for the development of DVT/PE (21), the condition occurred too infrequently for us to evaluate among those with RA. We observed it among 2% of women with SLE, but with a greater than 20-fold increased risk relative to women without that condition. A synergistic effect of a baseline possibly increased DVT/PE risk during pregnancy in all women (22), and underlying SLE led to such a dramatically increased risk. The immune and coagulation systems have been postulated to share a common evolutionary origin (23), which may further explain the association between SLE and increased venous thromboembolism (VTE) risk. Additionally, we found that women with RA or SLE were more likely

Table 5. Selected outcomes and complications among women with and without RA or SLE and their infants in Washington State during 2 time periods, 1987–1999 and 2000–2014*

	RA		SLE	
Time period and outcome	1987–1999 (206/2,077)†	2000–2014 (1,017/10,216)†	1987–1999 (403/4,114)†	2000–2014 (951/9,648)†
Maternal				
Intensive prenatal care	1.20 (0.94–1.55)	1.50 (1.36–1.66)	2.26 (2.00–2.55)	2.48 (2.30–2.67)
Anemia	1.74 (1.20–2.53)	1.38 (1.17–1.62)	1.28 (0.94–1.74)	1.42 (1.19–1.69)
Preeclampsia/eclampsia	0.98 (0.57–1.70)	1.50 (1.23–1.84)	2.49 (1.91–3.25)	2.26 (1.89–2.69)
Severe preeclampsia	0.66 (0.09–5.10)	2.81 (1.79–4.43)	4.29 (2.12–8.69)	2.75 (1.74–4.34)
Gestational diabetes mellitus‡	1.37 (0.73–2.57)	1.14 (0.93–1.39)	1.53 (0.96–2.43)	0.99 (0.78–1.26)
Cesarean delivery§	1.35 (1.04–1.76)	1.33 (1.18–1.50)	1.46 (1.19–1.80)	1.26 (1.10–1.45)
Placental abruption	0.68 (0.33–1.40)	0.68 (0.33–1.40)	1.63 (0.74–3.57)	1.93 (1.18–3.14)
Intrauterine growth retardation	1.77 (0.86–3.65)	1.57 (1.17–2.11)	3.69 (2.52–5.40)	3.44 (2.73–4.34)
Preterm labor	1.64 (1.00–2.68)	1.81 (1.48–2.22)	3.27 (2.49–4.29)	2.97 (2.49–3.54)
Postpartum hemorrhage	0.91 (0.46–1.80)	1.35 (1.01–1.80)	1.15 (0.70–1.89)	1.42 (1.07–1.88)
Rehospitalized¶	1.27 (0.90–1.77)	1.34 (1.12–1.61)	1.46 (1.18–1.82)	1.48 (1.25–1.74)
Infant				
Fetal distress	0.94 (0.65–1.38)	1.18 (0.96–1.45)	1.22 (0.99–1.52)	1.28 (1.06–1.55)
Birthweight <2,500 grams	2.01 (1.24–3.26)	2.10 (1.71–2.58)	5.60 (4.40–7.14)	4.58 (3.89–5.37)
<37 weeks' gestation#	1.75 (1.14–2.69)	2.17 (1.84–2.56)	4.41 (3.56–5.47)**	3.35 (2.91–3.87)**
Small for gestational age	1.30 (0.90–1.89)	1.26 (1.04–1.53)	2.40 (1.96–2.95)	2.25 (1.93–2.61)
Malformation	1.26 (0.75–2.10)	1.13 (0.89–1.42)	1.80 (1.31–2.46)	1.32 (1.06–1.66)
Rehospitalized¶	1.43 (0.96–2.13)	1.17 (0.95–1.44)	1.80 (1.36–2.36)	1.51 (1.26–1.80)

* Values are the relative risk (RR) (95% confidence interval). RA = rheumatoid arthritis; SLE = systemic lupus erythematosus. RR was estimated from Poisson regression with robust SEs, adjusted for birth year, maternal age, and parity.
† Cases/comparison.
‡ Among women without established diabetes mellitus: 1,206 women with RA, 12,159 women without; 1,337 women with SLE, 13,622 women without.
§ Among women without prior deliveries: 516 women with RA, 4,997 women without; 555 women with SLE, 5,596 women without.
¶ Rehospitalized within 2 years of delivery discharge.
# $P = 0.01$ for difference between time periods in RRs of <37 weeks gestation for infants of women with SLE.
** Statistically significant.

to experience prolonged delivery hospitalizations, consistent with previous studies (24); prolonged hospitalization is known to lead to an increased VTE risk. Glucocorticoid use may contribute to an increased VTE risk in inflammatory states (25). Although we lacked information on medication use, at least a portion of the women with SLE were likely taking glucocorticoids at some point during pregnancy, perhaps contributing to the increased VTE risk.

Women with RA or SLE exhibited an increased risk of preeclampsia, with severe preeclampsia present in 2% of women with RA and 3% of women with SLE. This risk may have contributed to the relatively longer delivery hospitalizations we observed, possibly also due to closer monitoring by care providers even in the absence of complications. An abnormal placenta, often observed in SLE (26), holds a central role in preeclampsia and intrauterine growth restriction development. Placental abnormalities in women with RA remain a topic for future study. An increased risk of preeclampsia/eclampsia may explain the increased placental abruption risk we observed in women with SLE (27), as these complications frequently overlap (28). Our results did not indicate an increased risk of gestational diabetes mellitus in women with SLE, consistent with a previous meta-analysis (29). However, the authors of that analysis observed that glucocorticoid use among SLE patients was positively associated with an increased gestational diabetes mellitus risk.

Neonates of women with RA or SLE are often preterm and have low birthweight (30,31). Although preeclampsia (often associated with preterm delivery) was increased in women with SLE, the vast majority (73%) of preterm infants of women with SLE in our study did not have mothers with preeclampsia, suggesting that other pathways are also relevant. Our observed increased NICU admission in infants of women with RA (12%) or SLE (16%) may be partly due to preterm birth. Neonates of women with SLE experienced an increased risk of death within the first 2 years of life, possibly due to complications of prematurity/low birthweight or malformations, which were increased among infants of affected women.

Neonates born to women with RA or SLE were less likely to be breastfed, as indicated on the birth record. There are many possible reasons for this finding. Babies of mothers with RA or SLE are more often in the NICU, affecting the mothers' ability to breastfeed due to physical distance or stress affecting their milk supply. Overall, breastfeeding is decreased in babies born preterm (32). Concern over medication safety while breastfeeding is common (33), but many medications used to treat RA and SLE are safe to use while breastfeeding, some with special instruction such as delayed breastfeeding after taking the medication. Given the many benefits of breastmilk, women with

these conditions should be encouraged to breastfeed, if possible.

Increased maternal and infant rehospitalization postdelivery in women with RA or SLE has not been reported previously and yet carries significant clinical implications. Relatively large numbers of women with RA (12%) or SLE (16%) experienced a rehospitalization after delivery, with similar numbers among their infants (10% and 13%, respectively). The burden of rehospitalizations for affected families is large, especially given the challenges of self-care postpartum and of childcare in infancy and early childhood during the 2 years after delivery. Reasons for rehospitalization varied, with musculoskeletal conditions exhibiting the greatest risk of maternal rehospitalization. A flare of disease possibly contributed to these findings. Skin-, respiratory-, and genitourinary-related hospitalizations may be related to maternal immunosuppression and resultant infection. As women with SLE had an increased risk of VTE, perhaps stroke contributed to the increased occurrence that we observed of nervous system- or hematologic-related rehospitalizations. Hematologic-related rehospitalizations may also have been due to cytopenia, commonly seen in RA or SLE, and specifically anemia, which was notable in our study. Plausibly, endocrine-related rehospitalization was related to glycemic control in the setting of glucocorticoid use.

Preterm infants experience increased rehospitalization rates due to respiratory- and infection-related causes (34). Infants of women with SLE may have a greater likelihood of infection (35), suggesting that this increased risk may be due to the women's impaired ability to provide antibodies to their offspring in the setting of immunosuppression. Maternal immunosuppressive medication has not been shown to cause increased infant infection risk or significant immune system dysfunction beyond slight alteration to cell counts within the first year of life (36).

This study has several limitations. Our data included no information about disease activity, SSA/SSB positivity, antiphospholipid antibody positivity, or medications, all of which would have been helpful to understand disease severity and reasons for rehospitalization. Another limitation is that the state hospital discharge data do not include federal (military) hospitals. This absence is unlikely to have affected results, given the fact that few deliveries (3–5%) in Washington State occurred in military hospitals, and that pregnant women with conditions requiring intensive prenatal care who might otherwise use military hospitals were likely referred to high-level obstetric care in community hospitals for delivery.

The use of ICD codes at the time of delivery to identify exposed women and the lack of outpatient information were limitations. Inclusion of women with RA or SLE in the unexposed group would have biased results toward null. Conversely, women with preeclampsia or other pregnancy complications may have been more likely to have RA or SLE identified at delivery hospitalization (biasing results away from null). If the exposed groups contained largely women with more severe disease, this selection bias would render results relevant to only a segment of this population. Based on earlier work with these data, the positive predictive value (PPV) for identification of RA based on ICD codes in these records is 100% (37), indicating that our exposed women truly had this disease; 40% sensitivity suggests that some misclassification of women with RA into the unexposed group may have occurred. Identification of RA in administrative data is improved when medication, laboratory, and specialist care data are available, and with access to >1 patient contact record (38). The PPV for SLE diagnosis (versus medical records review) was 93% using health plan and birth certificate records (during 1 year prior to pregnancy through delivery) including >1 inpatient visit with a relevant diagnosis code or >2 outpatient visits >30 days apart (39). PPV ranges from 60% to 98% using other algorithms, including hospital discharge records (40–42).

Our results are based on a single state. Results from other regions may differ, due to underlying differences in populations (e.g., Washington State has lower prenatal smoking and obesity levels than some other states) or health care delivery. Finally, although our data are population-based, they are based on the largely White population of our state. Future analyses with similar data from more diverse regions are needed.

Our study also has several strengths. It is a large population-based cohort study spanning multiple years with evaluation of selected less-studied outcomes in pregnancy among women with RA or SLE. We also present results on rehospitalizations within 2 years of delivery among women with RA or SLE and their infants.

Women with RA or SLE should be monitored closely in the year following delivery to minimize rehospitalization, particularly during the first 6 months postpartum. Infants of women with SLE should be monitored closely following delivery, and further study should be conducted to elucidate reasons for poor outcomes. These novel findings can improve the health of these groups as well as minimize the burden on the medical system. Hospitalizations are burdensome for the patients, their caregivers, and the health care system. Further research is needed to understand their cause and to develop strategies for preventing rehospitalizations postpartum.

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. Singh 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. Singh, Sabo, Crane, Schiff, Mueller.
Acquisition of data. Singh, Sabo, Crane, Doody, Schiff, Mueller.
Analysis and interpretation of data. Doody, Mueller.

REFERENCES

1. Aljary H, Czuzoj-Shulman N, Spence AR, et al. Pregnancy outcomes in women with rheumatoid arthritis: a retrospective population-based cohort study. J Matern Fetal Neonatal Med 2020;33:618–24.

2. Mehta B, Luo Y, Xu J, et al. Trends in maternal and fetal outcomes among pregnant women with systemic lupus erythematosus in the United States: a cross-sectional analysis. Ann Intern Med 2019;171: 164–71.
3. Norgaard M, Larsson H, Pedersen L, et al. Rheumatoid arthritis and birth outcomes: a Danish and Swedish nationwide prevalence study. J Intern Med 2010;268:329–37.
4. Strouse J, Donovan BM, Fatima M, et al. Impact of autoimmune rheumatic diseases on birth outcomes: a population-based study. RMD Open 2019;5:e000878.
5. Wallenius M, Salvesen KA, Daltveit AK, et al. Systemic lupus erythematosus and outcomes in first and subsequent births based on data from a national birth registry. Arthritis Care Res (Hoboken) 2014;66:1718–24.
6. Bundhun PK, Soogund MZ, Huang F. Impact of systemic lupus erythematosus on maternal and fetal outcomes following pregnancy: a meta-analysis of studies published between years 2001–2016. J Autoimmun 2017;79:17–27.
7. He WR, Wei H. Maternal and fetal complications associated with systemic lupus erythematosus: an updated meta-analysis of the most recent studies (2017–2019). Medicine (Baltimore) 2020;99:e19797.
8. Syeda SK, Wen T, Wright JD, et al. Postpartum cardiac readmissions among women without a cardiac diagnosis at delivery. J Matern Fetal Neonatal Med 2022;35:4768–74.
9. Washington State Department of Health. Comprehensive Hospital Abstract Reporting System (CHARS) hospital discharge system documentation. URL: https://doh.wa.gov/data-statistical-reports/healthcare-washington/hospital-and-patient-data/hospital-discharge-data-chars.
10. Mueller BA, Crane D, Doody DR, et al. Pregnancy course, infant outcomes, rehospitalization, and mortality among women with intellectual disability. Disabil Health J 2019;12:452–9.
11. Schiff MA, Doody DR, Crane DA, et al. Pregnancy outcomes among visually impaired women in Washington State, 1987–2014. Disabil Health J 2021;14:101057.
12. Crane DA, Doody DR, Schiff MA, et al. Pregnancy outcomes in women with spinal cord injuries: a population-based study. PM R 2019;11:795–806.
13. Siega-Riz AM, Viswanathan M, Moos MK, et al. A systematic review of outcomes of maternal weight gain according to the Institute of Medicine recommendations: birthweight, fetal growth, and postpartum weight retention. Am J Obstet Gynecol 2009;201:339.
14. Kotelchuck M. An evaluation of the Kessner Adequacy of Prenatal Care Index and a proposed Adequacy of Prenatal Care Utilization Index. Am J Public Health 1994;84:1414–20.
15. Centers for Disease Control and Prevention, National Center for Health Statistics. National Vital Statistics System instruction manuals. URL: https://www.cdc.gov/nchs/nvss/instruction_manuals.htm.
16. Lydon-Rochelle MT, Holt VL, Cardenas V, et al. The reporting of preexisting maternal medical conditions and complications of pregnancy on birth certificates and in hospital discharge data. Am J Obstet Gynecol 2005;193:125–34.
17. Chow EJ, Cushing-Haugen KL, Cheng GS, et al. Morbidity and mortality differences between hematopoietic cell transplantation survivors and other cancer survivors. J Clin Oncol 2017;35:306–13.
18. Mueller BA, Doody DR, Weiss NS, et al. Hospitalization and mortality among pediatric cancer survivors: a population-based study. Cancer Causes Control 2018;29:1047–57.
19. Yasmeen S, Wilkins EE, Field NT, et al. Pregnancy outcomes in women with systemic lupus erythematosus. J Matern Fetal Med 2001;10:91–6.
20. Clowse ME, Jamison M, Myers E, et al. A national study of the complications of lupus in pregnancy. Am J Obstet Gynecol 2008;199:127.
21. Lee JJ, Pope JE. A meta-analysis of the risk of venous thromboembolism in inflammatory rheumatic diseases. Arthritis Res Ther 2014; 16:435.
22. Devis P, Knuttinen MG. Deep venous thrombosis in pregnancy: incidence, pathogenesis and endovascular management. Cardiovasc Diagn Ther 2017;7 Suppl 3:S309–19.
23. Dahlback B. Coagulation and inflammation: close allies in health and disease. Semin Immunopathol 2012;34:1–3.
24. Chakravarty EF, Nelson L, Krishnan E. Obstetric hospitalizations in the United States for women with systemic lupus erythematosus and rheumatoid arthritis. Arthritis Rheum 2006;54:899–907.
25. Majoor CJ, Sneeboer MM, de Kievit A, et al. The influence of corticosteroids on hemostasis in healthy subjects. J Thromb Haemost 2016;14:716–23.
26. Jiang M, Wang Y, Fu Q, et al. Preeclampsia risk prediction model for Chinese pregnant patients with systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2020;72:1602–10.
27. Tikkanen M. Placental abruption: epidemiology, risk factors and consequences. Acta Obstet Gynecol Scand 2011;90:140–9.
28. Vinet E, Genest G, Scott S, et al. Causes of stillbirths in women with systemic lupus erythematosus. Arthritis Rheumatol 2016;68: 2487–91.
29. Dong Y, Dai Z, Wang Z, et al. Risk of gestational diabetes mellitus in systemic lupus erythematosus pregnancy: a systematic review and meta-analysis. BMC Pregnancy Childbirth 2019;19:179.
30. Abdwani R, Al Shaqsi L, Al-Zakwani I. Neonatal and obstetrical outcomes of pregnancies in systemic lupus erythematosus. Oman Med J 2018;33:15–21.
31. Smith CJ, Forger F, Bandoli G, et al. Factors associated with preterm delivery among women with rheumatoid arthritis and women with juvenile idiopathic arthritis. Arthritis Care Res (Hoboken) 2019;71:1019–27.
32. Lau C. Breastfeeding challenges and the preterm mother-infant dyad: a conceptual model. Breastfeed Med 2018;13:8–17.
33. Ikram N, Eudy A, Clowse ME. Breastfeeding in women with rheumatic diseases. Lupus Sci Med 2021;8:e000491.
34. Underwood MA, Danielsen B, Gilbert WM. Cost, causes and rates of rehospitalization of preterm infants. J Perinatol 2007;27:614–9.
35. Bender Ignacio RA, Madison AT, Moshiri A, et al. A population-based study of perinatal infection risk in women with and without systemic lupus erythematosus and their infants. Paediatr Perinat Epidemiol 2018;32:81–9.
36. Motta M, Ciardelli L, Marconi M, et al. Immune system development in infants born to mothers with autoimmune disease, exposed in utero to immunosuppressive agents. Am J Perinatol 2007;24:441–7.
37. Schwartz NR, Crane DA, Doody DR, et al. Assessment of the accuracy of identification of selected disabilities and conditions in hospital discharge data for pregnant women. Epidemiology 2020;31:687–91.
38. Chung CP, Rohan P, Krishnaswami S, et al. A systematic review of validated methods for identifying patients with rheumatoid arthritis using administrative or claims data. Vaccine 2013;31 Suppl 10: K41–61.
39. Andrade SE, Scott PE, Davis RL, et al. Validity of health plan and birth certificate data for pregnancy research. Pharmacoepidemiol Drug Saf 2013;22:7–15.
40. Bernatsky S, Linehan T, Hanly JG. The accuracy of administrative data diagnoses of systemic autoimmune rheumatic diseases. J Rheumatol 2011;38:1612–6.
41. Hanly JG, Thompson K, Skedgel C. Identification of patients with systemic lupus erythematosus in administrative healthcare databases. Lupus 2014;23:1377–82.
42. Moores KG, Sathe NA. A systematic review of validated methods for identifying systemic lupus erythematosus (SLE) using administrative or claims data. Vaccine 2013;31 Suppl 10:K62–73.