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 guide-lines, health care economics, health care policy, educational, social, and public health issues, and future trends in rheumatology practice.

Volume 71, No. 11, November 2019

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 71 • November 2019 • NO. 11

Cover image: The figure on the cover (from Zhao et al, page 1430) depicts the infrared thermal image and coronal STIR magnetic resonance image (MRI) of a child with an active chronic nonbacterial osteomyelitis lesion. Note the increased temperature in the distal tibia/fibula area and the MRI hyperintensity in the right distal tibia and surrounding tissue.

Arthritis Care & Research
Vol. 71, No. 11, November 2019, pp 1401–1403
DOI 10.1002/acr.23913

American College of Rheumatology
Empowering Rheumatology Professionals

EDITORIAL

Antinuclear Antibodies and Lupus: Label Versus Meaning

David I. Daikh

The myriad of organ-specific and systemic autoimmune diseases presents many clinical challenges. Systemic lupus erythematosus (SLE) is the prototype of autoimmune disease because of the systemic nature of the disease, its complex and diverse pathogenesis, and the prominent presence of autoantibodies. While most patients with lupus have measurable autoantibodies, some do not. This can present additional diagnostic challenges and raises questions about the underlying pathogenesis of disease. Correspondingly there is ongoing interest in antinuclear antibody (ANA)–negative SLE.

In this issue, Tarazi and colleagues examine this question using an analysis of cases of lupus skin disease and utilizing a university-based dermatology clinic database of cases of cutaneous lupus erythematosus (CLE) (1). This database was queried for patients who had a negative ANA and, if so, whether or not they met classification criteria for SLE. The database includes patients diagnosed with various subtypes of CLE. For the purpose of this analysis, only cases for which there was confirmation of a diagnosis of CLE by skin biopsy and a known ANA titer were included. Cases of chilblain lupus and lupus tumidus were also excluded. Among the more than 450 cases in this clinical database, 309 met these criteria and had available or adequate clinical data. Of these 309, 111 had a negative ANA (36.9%). Among 81 cases with multiple ANAs reported, 27 (33.3%) had a result that fluctuated between positivity and negativity. These results emphasize that the absence of an ANA is common among patients with CLE.

Admirably, the analysis also focused on the method of ANA determination. This is important because of the variety and lack of standardization among methods used to detect ANA in clinical laboratories and because some of these methods are less accurate than others. For example, while the use of automated enzyme-linked immunosorbent assay (ELISA) ANA detection has become increasingly common, this method can be associated with a significant false-negative rate. Thus, a task force of the College of American Pathologists and the American College of Rheumatology (ACR) has recommended that indirect immunofluorescence on standardized HEp-2 cells remains the gold standard for ANA testing (2), and variability in currently used methods has highlighted the need to develop worldwide standards for this test (3). A related and perhaps more important issue is the appropriate cutoff for ANA positivity. A higher threshold for determining ANA positivity will favor specificity, while a lower threshold favors sensitivity in identifying most patients with SLE. However, when defined in relation to a control group, in a given laboratory this threshold will be influenced by the method, individual laboratory technique, and the normative population. Thus, assuming that the laboratory utilizes a high-quality method that has been validated in relation to HEp-2 indirect immunofluorescence, it is appropriate to give consideration to the threshold for negativity established by that laboratory, as well as to the absolute value in deciding whether a patient is ANA negative or positive. This was the approach that was generally used to categorize ANA status in this study, although 3 subjects with ANA titers of 1:160 reported by the laboratory as negative were counted as positive in the analysis due to the presence of an elevated titer. The vast majority of ANA in this study group was detected by indirect immunofluorescence, but 4 of the ANA-negative subjects had a negative ANA determined by ELISA.

Focusing on the negative or fluctuating ANA cases, the authors selected those that appeared to have SLE, based on chart review and applying ACR and Systemic Lupus International Collaborating Clinics (SLICC) classification criteria (4). Of the 111 patients who were ANA negative, 20 met either the ACR and/or SLICC SLE criteria (18.0%). A total of 12 cases met ACR criteria, and 8 met both ACR and SLICC criteria. Of the 27 patients with a fluctuating ANA, 12 met one or both criteria for SLE (44.4%).

These results emphasize that patients with SLE can be ANA negative. However, the prevalence of negative ANA described in this database is much higher than in many other descriptions of SLE. For example, in a composite analysis of a large number of clinical and epidemiologic studies of SLE, the sensitivity of a positive ANA titer at 1:80 for SLE was 97.8% (5). This difference may reflect that the patients in the current analysis were evaluated in a tertiary care center and perhaps that some of the cases

David I. Daikh, MD, PhD: San Francisco VA Medical Center and University of California, San Francisco.

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

Address correspondence to David I. Daikh, MD, PhD, San Francisco VA Medical Center, 4150 Clement Street, San Francisco, CA 94121. E-mail: David.Daikh@ucsf.edu.

Submitted for publication April 24, 2019; accepted in revised form April 30, 2019.

were referred for dermatologic evaluation specifically because lupus was suspected in an ANA-negative patient. The absence of detectable ANA can occur for a number of reasons, including measurement techniques that may produce false-negative or fluctuating results, and patient factors such as low levels of serum immunoglobulin (6), duration of disease, and response to therapy, particularly corticosteroids (7). Some patients with negative ANA nevertheless have measurable levels of anti-DNA or anti-Sm antibodies, such as the 3 case examples of ANA-negative SLE provided in the current study. Interestingly, a small number of ANA-negative patients have autoantibodies to non-nuclear antigens (8). As noted, estimates of the frequency of ANA-negative SLE depend on the cutoff chosen. For example, in an analysis of the initial serum samples from 1,137 patients classified as having SLE by SLICC criteria, 7.7% (88 patients) had an ANA titer less than 1:160 (9). This group had lower disease activity and was more commonly treated with steroids than the overall group. Interestingly, 24 of these ANA-negative patients had measurable autoantibodies to specific nuclear antigens (dsDNA and Ro), and 17 had isolated cytoplasmic staining on indirect immunofluorescence. Thus, even using the more stringent cutoff of 1:160 for ANA, only 4.1% of this large cohort did not have evidence of autoantibody production. Many would consider 1:80 an appropriate threshold for ANA positivity in considering a diagnosis of SLE, particularly with compatible clinical features. Additional autoantigen-specific assays might identify a few more autoantigen-positive, ANA-negative patients in such a cohort. Such data indicates that active SLE without the presence of autoantibodies is likely a rare event, if it occurs at all. Nevertheless, ANA-negative SLE definitely is a clinical entity, and this description of CLE patients nicely demonstrates why ANA positivity should not be used as an absolute requirement for the diagnosis of SLE.

The emphasis on ANA in this study by Tarazi et al also underscores the need to distinguish the practice of diagnosing a complex autoimmune disease in an individual patient in the clinic from classifying a patient as having the disease according to set classification criteria developed for the purpose of identifying patients to be grouped together for study and analysis in a clinical trial. The needs of clinical care favor specificity, which helps ensure that the many people with low levels of ANA but no autoimmune disease are not misdiagnosed as having SLE, but they also demand flexibility so as not to miss diagnosing the unusual patient, e.g, the ANA-negative SLE patient. Thus, the seemingly discordant argument made by Tarazi et al (that on one hand, the threshold for ANA titers in diagnosing SLE should be 1:160 rather than 1:80, while on the other, that a positive ANA should not be required for the classification of SLE) makes sense in the context of using the test results for clinical diagnosis. In contrast, classification criteria are developed with an eye to maximizing both sensitivity and specificity in order to capture the broadest but still appropriate group of patients for study and to help ensure the validity of the results. These considerations have much overlap, which is why classification criteria can be helpful diagnostic tools, but classification criteria are not designed for the purpose of diagnosis. The argument is also sometimes made, as in this study, that overly restrictive classification criteria prevent patients from access to trials. While it is valuable to be able to offer patients access to studies of new therapies, trials are conducted to establish the effectiveness of therapy, and this must be the paramount goal. This is particularly true in the case of a complex disease like SLE, where there have been many challenges to designing and conducting successful clinical trials (10).

The subject of ANA-negative lupus is again generating significant discussion in light of new classification criteria, proposed by the ACR and the European League Against Rheumatism, that utilize ANA positivity as an entry criterion (11), and the subject remains highly relevant to clinical study design. For example, experience in early-phase studies of belimumab, currently the unique example of a successful pivotal therapeutic trial in SLE, showed the importance of carefully selecting entry criteria. In that case, recognition that there was no benefit to ANA-negative or anti-dsDNA–negative patients in phase II studies positively informed the development of entry criteria for the ultimately successful phase III trials (12). Although the selection of entry criteria for such trials is complex and must balance multiple considerations, this example underscores the value of classification criteria for clinical trials and the importance of not equating them with diagnosis in a trial setting. This analysis of patients with CLE similarly underscores the importance of not limiting the diagnosis of SLE to a set of classification criteria. Without more specific and reliable tests, that diagnosis continues to require a comprehensive evaluation of an individual patient by an experienced clinician.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Tarazi M, Gaffney GR, Kushner CJ, Chakka S, Werth VP. Cutaneous lupus erythematosus patients with a negative antinuclear antibody meeting the American College of Rheumatology and/or Systemic Lupus International Collaborating Clinics criteria for systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2019;71:1404–9.
2. Kavanaugh A, Tomar R, Reveille J, Solomon DH, Homburger HA. Guidelines for clinical use of the antinuclear antibody test and tests for specific autoantibodies to nuclear antigens. American College of Pathologists. Arch Pathol Lab Med 2000;124:71–81.
3. Pisetsky DS, Spencer DM, Lipsky PE, Rovin BH. Assay variation in the detection of antinuclear antibodies in the sera of patients with established SLE. Ann Rheum Dis 2018;77:911–3.
4. Petri M, Orbai AM, Alarcón GS, Gordon C, Merrill JT, Fortin PR, et al. Derivation and validation of the Systemic Lupus International

Collaborating Clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum 2012;64:2677–86.

5. Leuchten N, Hoyer A, Brinks R, Schoels M, Schneider M, Smolen J, et al. Performance of antinuclear antibodies for classifying systemic lupus erythematosus: a systematic literature review and meta-regression of diagnostic data. Arthritis Care Res (Hoboken) 2018;70:428–38.
6. White NJ, Winearls CG, Ledingham JG. Systemic lupus erythematosus and nephritis: severe relapse with disappearance of antinuclear antibodies. Br Med J 1980;281:194–5.
7. Barnett EV, North AF, Condemi JJ. Antinuclear factors in systemic lupus erythematosus and rheumatoid arthritis. Ann Intern Med 1965;63:100–8.
8. Provost TT, Ahmed AR, Maddison PJ, Reichlin M. Antibodies to cytoplasmic antigens in lupus erythematosus: serologic marker for systemic disease. Arthritis Rheum 1977;20:1457–63.
9. Choi MY, Clarke AE, St. Pierre Y, Hanly JG, Urowitz MB, Romero-Diaz J, et al. Antinuclear antibody–negative systemic lupus erythematosus in an international inception cohort. Arthritis Care Res (Hoboken) 2019;71:893–902.
10. Dall'Era M, Bruce IN, Gordon C, Manzi S, McCaffrey J, Lipsky PE. Current challenges in the development of new treatments for lupus. Ann Rheum Dis 2019;78:729–35.
11. Tedeschi SK, Johnson SR, Boumpas D, Daikh D, Dörner T, Jayne D, et al. Developing and refining new candidate criteria for systemic lupus erythematosus classification: an international collaboration. Arthritis Care Res (Hoboken) 2018;70:571–81.
12. Wallace DJ, Stohl W, Furie RA, Lisse JR, McKay JD, Merrill JT, et al. A phase II, randomized, double-blind, placebo-controlled, dose-ranging study of belimumab in patients with active systemic lupus erythematosus. Arthritis Rheum 2009;61:1168–78.

Arthritis Care & Research
Vol. 71, No. 11, November 2019, pp 1404–1409
DOI 10.1002/acr.23916

American College of Rheumatology
Empowering Rheumatology Professionals

Cutaneous Lupus Erythematosus Patients With a Negative Antinuclear Antibody Meeting the American College of Rheumatology and/or Systemic Lupus International Collaborating Clinics Criteria for Systemic Lupus Erythematosus

Meera Tarazi, Rebecca G. Gaffney, Carolyn J. Kushner, Srita Chakka, and Victoria P. Werth

Objective. Systemic lupus erythematosus (SLE) is a disorder that is heterogeneous and can be difficult to diagnose. One hallmark of the disease is the presence of antinuclear antibodies (ANAs), a feature that has been incorporated into multiple classification criteria over the years. In this study, we used a database of patients with cutaneous lupus erythematosus (CLE) to determine how many had a negative ANA and met criteria for SLE using the American College of Rheumatology (ACR) and/or Systemic Lupus International Collaborating Clinics (SLICC) criteria.

Methods. We used a database of 301 biopsy-proven CLE patients that contained information including ANA status and the presence of features of SLE. The database was searched for patients who had a negative ANA result and whether or not they met SLE criteria using the ACR and/or SLICC criteria.

Results. Of the 301 patients with biopsy-proven CLE and a known ANA, 111 had a negative ANA test (36.9%) and 27 had an ANA test that fluctuated (33.3%). In all, 20 ANA-negative patients met SLE criteria (18.0%), and 12 patients with a fluctuating ANA test met SLE criteria (44.4%). Of all patients who had either a negative or fluctuating ANA result and who met criteria for SLE (n = 32), 27 patients had involvement of ≥1 organ system other than skin (84.4%), and 13 patients had involvement of ≥2 organ systems other than skin (40.6%).

Conclusion. Our results show that an ANA is not always present in patients with systemic disease. This fact should be taken into consideration when devising SLE classification criteria to be used for clinical trials.

INTRODUCTION

Systemic lupus erythematosus (SLE) is an autoimmune disorder that affects multiple organ systems and is characterized by a variety of autoantibodies. Diagnosis can be difficult because of its heterogeneous presentation. In an attempt to help guide providers in the diagnosis of this complex disease, Cohen and Canoso (1) devised the first classification criteria for SLE in 1971. These guidelines required the presence of 4 of 14 criteria for the classification of SLE (1). In 1982, the American College of Rheumatology (ACR) created a classification system, later revised in 1997, that required 4 of 11 clinical criteria to be met in order for a patient to be classified as having SLE (2,3). These criteria incorporated serologic tests that had not been a part of the preliminary criteria. The 1982 version showed improvement of the original criteria in both sensitivity (96%) and specificity (96%) when tested using data from patients with SLE and control patients from 18 different clinics (2).

Although the ACR criteria were a step in the right direction in diagnosing patients with SLE, they were criticized by many for placing too much weight on cutaneous criteria (4,5). The ACR criteria assign malar rash, discoid rash, photosensitivity, and oral ulcers as separate criteria, which meant patients with cutaneous lupus erythematosus (CLE) could then be diagnosed as having SLE based on their skin disease alone. Several publications proposed re-evaluation of SLE criteria with more input from dermatologists (4,5).

In 2012, the Systemic Lupus International Collaborating Clinics (SLICC) group developed new SLE criteria to address the shortcomings of the ACR criteria. SLICC required that patients

Supported by the Veterans Health Administration, Office of Research and Development, Biomedical Laboratory Research and Development. Dr. Werth's work was supported by the NIH (grant K24-AR-02207).

Meera Tarazi, BA, Rebecca G. Gaffney, BA, Carolyn J. Kushner, BA, Srita Chakka, BA, Victoria P. Werth, MD: Corporal Michael J. Crescenz VA Medical Center and Perelman School of Medicine, University of Pennsylvania, Philadelphia.

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

Address correspondence to Victoria P. Werth, MD, Department of Dermatology, Perelman Center for Advanced Medicine, Suite 1-330A, 3400 Civic Center Boulevard, Philadelphia, PA 19104. E-mail: werth@pennmedicine.upenn.edu.

Submitted for publication March 20, 2018; accepted in revised form April 30, 2019.

SIGNIFICANCE & INNOVATIONS

- As seen in a database of patients with cutaneous lupus erythematosus, approximately 18.0% of patients with a negative antinuclear antibody (ANA) still meet American College of Rheumatology (ACR) and Systemic Lupus International Collaborating Clinics (SLICC) criteria for systemic lupus erythematosus (SLE).
- Our database also showed that approximately 33.3% of our patients who were tested for multiple ANAs had a result that fluctuated between positivity and negativity over time, of which 44.4% qualified as having SLE using the current ACR and/or SLICC criteria.
- These results should be taken into consideration when devising SLE classification criteria to be used in clinical trials because requiring a positive ANA for diagnosis of SLE may exclude a number of patients from receiving appropriate treatment.

meet ≥1 immunologic criterion in addition to the clinical criteria to be classified as having SLE. Additionally, a patient could qualify as having SLE if they had biopsy-proven lupus nephritis in the presence of positive antinuclear antibodies (ANAs) or anti-dsDNA antibodies (6). This revision increased sensitivity to 97% in the validation set but showed a decrease in specificity as compared to the ACR criteria, reaching 84% in the validation set (6).

Since the development of ACR and SLICC criteria, they have been the predominant systems used in diagnosing SLE. Although these criteria have led to improvement in the diagnosis of SLE, modifications have been proposed in an effort to maintain the sensitivity seen in the SLICC criteria while improving specificity (7). Maximizing sensitivity and specificity in classification criteria for SLE is crucial for the inclusion of patients with significant disease in clinical trials. To do so, the appropriate clinical and immunologic items must be included. In light of potential new SLE criteria being developed, we used a database of CLE patients seen at the dermatology clinic of the University of Pennsylvania to investigate the number of CLE patients who present with negative ANA results and meet the ACR and/or SLICC criteria for SLE.

PATIENTS AND METHODS

This was a retrospective study using a database of CLE patients seen at the autoimmune dermatology clinic of the University of Pennsylvania. Subjects included in the database were patients who met criteria for having CLE and were ages ≥18 years. All patients recruited into the database were diagnosed based on clinical presentation (lesion morphology and symptoms suggestive of systemic lupus), serologic findings, and pathologic findings. At a minimum, patients must have been diagnosed with 1 subtype of CLE to be recruited into the database. Although a biopsy was performed for the majority of patients (especially if the clinical presentation was insufficient to warrant a diagnosis of CLE), in cases where there was overwhelming clinical evidence suggestive of lupus, a biopsy may not have been performed. For the purposes of this analysis, patients who did not have biopsy-proven CLE were excluded. There were no inclusion or exclusion criteria based on race, ethnic origin, or sex. The database consisted of 454 patients with CLE who had enrolled from 2007 to 2017 and included information regarding patient demographics, subtype of CLE, features of SLE, and laboratory markers of lupus, including ANA status. This information was collected during regularly scheduled patient visits at the dermatology clinic of the University of Pennsylvania. The database was searched for all CLE patients who had a negative ANA test, and results were subsequently confirmed through a review of these patients' medical records. Additionally, the laboratory technique used to detect ANAs was noted (if available), as well as the ANA titer in cases where an indirect immunofluorescence (IIF) assay was performed. ANA status was determined from the laboratory's interpretation of the test, not from the reported titer. Information from the database was also used to determine whether or not patients with negative ANA results met SLE criteria using either the ACR and/or SLICC criteria. Features of SLE reported in the database were once again verified using patient medical records. Medical records were also used to ensure that ANA status was checked after patients developed features of SLE, not prior. Finally, patients with chilblain lupus and tumid lupus were excluded from this analysis.

RESULTS

Our database contained 454 subjects with a diagnosis of CLE. Of these 454 patients, 20 had withdrawn from the study and therefore could not be used for our analysis, and 28 had an unknown ANA (6 were missing medical records, 8 were nonadherent in having an ANA checked, 9 were being followed by

Table 1. CLE subtypes in ANA-negative and ANA-fluctuating patients*

ANA-negative	ANA-fluctuating
DLE (n = 60)	DLE (n = 15)
SCLE (n = 38)	SCLE (n = 8)
ACLE (n = 3)	DLE/hypertrophic lupus overlap (n = 2)
DLE/hypertrophic lupus overlap (n = 3)	ACLE (n = 1)
DLE/lupus panniculitis overlap (n = 2)	Hypertrophic lupus (n = 1)
DLE/SCLE overlap (n = 2)	–
Hypertrophic lupus (n = 2)	–
Lupus panniculitis (n = 1)	–

* CLE = cutaneous lupus erythematosus; ANA = antinuclear antibody; DLE = discoid lupus erythematosus; SCLE = subacute CLE; ACLE = acute CLE.

Table 2. IIF assay titers of ANA negative patients (n = 91)*

Titer	No. patients
<1:40	38
<1:160	21
1:80	20
<1:80	7
1:40	5

* IIF = indirect immunofluorescence; ANA = antinuclear antibody.

outside providers who did not send records of an ANA, and 5 were missing an ANA for unknown reasons). Of the remaining 406 patients, 105 either had not undergone a biopsy or had the chilblain lupus or tumid lupus subtypes of CLE and were therefore excluded. Our final population was made up of 301 biopsy-proven CLE patients with a known ANA result. A total of 111 had a negative ANA result (36.9%) and 27 of 81 patients who had multiple ANAs checked had a result that fluctuated between positivity and negativity (33.3%).

CLE subtypes and laboratory technique for ANA detection. Information on CLE subtypes of the ANA-negative and ANA-fluctuating patients is shown in Table 1. In sum, 91 of the 111 ANA-negative patients were detected using an IIF assay. These tests were performed at different laboratories, each using different titer cutoffs for determining a result as negative. ANA titers for the 91 ANA-negative patients are shown in Table 2. Of the 111 ANA-negative patients, 4 had ANAs detected by enzyme-linked immunosorbent assay (ELISA). Finally, 16 ANA-negative patients did not have information available regarding the laboratory method used (the ANA status was reported in physician notes, but the original laboratory records were not found in the chart).

Of the 27 ANA-fluctuating patients, 24 had ANAs checked using an IIF assay. Original laboratory results for 3 patients were missing (the ANA status was reported in physician notes, but the patient charts did not include the original laboratory results), and therefore the laboratory techniques used could not be verified for these patients. Additionally, 10 of the 24 fluctuating-IIF patients had their ANA checked at the same laboratory.

CLE patients meeting SLE criteria. Of the 111 patients who were ANA negative, 20 met either the ACR and/or SLICC SLE criteria (18.0%). A total of 12 patients met ACR criteria for SLE, and 8 patients met both ACR and SLICC criteria (Figure 1). Two patients met the SLICC criteria on the basis of biopsy-proven lupus nephritis with a positive anti-dsDNA antibody. Of the 27 patients with a fluctuating ANA, 12 met SLE criteria (44.4%), with 2 patients meeting ACR criteria, 3 meeting SLICC criteria, and 7 meeting both ACR and SLICC criteria (Figure 1). Twenty-four fluctuating-ANA patients had their ANA checked using an IIF assay, and 13 of these patients went from positive to negative, with 5 meeting SLE criteria (38.5%). Two of the 7 who went from negative to positive met SLE criteria (28.6%). Three of the 4 patients who had multiple fluctuations met SLE criteria (75.0%).

Of the 29 patients who were ANA negative or had a fluctuating ANA and met ACR criteria, the 5 most frequent criteria that were met were photosensitivity, discoid rash, arthritis, malar rash, and oral ulcers, with 26, 23, 22, 17, and 15 patients meeting these criteria, respectively (Table 3). Of the 19 patients who were

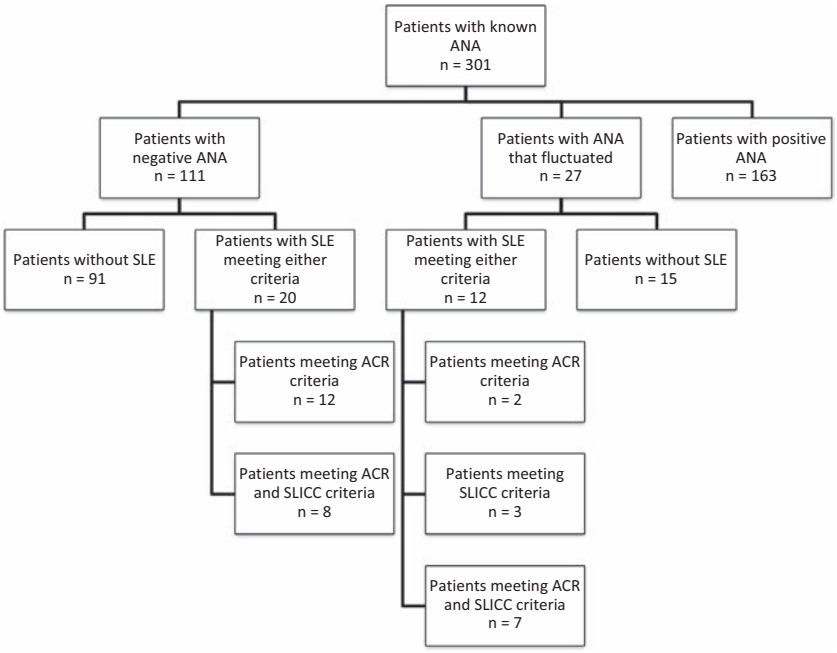

Figure 1. Diagram demonstrating antinuclear antibody (ANA) status of patients with cutaneous lupus erythematosus and whether they met American College of Rheumatology (ACR) and/or Systemic Lupus International Collaborating Clinics (SLICC) criteria for systemic lupus erythematosus (SLE).

Table 3. ANA-negative/ANA-fluctuating CLE patients meeting ACR and/or SLICC criteria for SLE*

ACR criteria (n = 29)	SLICC criteria (n = 18)
Photosensitivity (n = 26)	Chronic cutaneous lupus (n = 16)
Discoid rash (n = 23)	Acute cutaneous lupus (n = 15)
Arthritis (n = 22)	Arthritis (n = 11)
Malar rash (n = 17)	Low complement (n = 9)
Oral ulcers (n = 15)	Anti-dsDNA antibody (n = 9)
Leukopenia (n = 12)	Leukopenia (n = 7)
Anti-dsDNA antibody (n = 9)	Renal involvement (n = 7)
Renal involvement (n = 8)	Oral ulcers (n = 5)
Anti-Sm antibody (n = 4)	Nonscarring alopecia (n = 5)
Anticardiolipin antibody (n = 2)	Anti-Sm antibody (n = 4)
Serositis (n = 1)	Anticardiolipin antibody (n = 2)
Neurologic involvement (n = 1)	Serositis (n = 1)
Thrombocytopenia (n = 1)	Neurologic involvement (n = 1)
–	Thrombocytopenia (n = 1)

* Values in parentheses are the number. ANA = antinuclear antibody; CLE = cutaneous lupus erythematosus; ACR = American College of Rheumatology; SLICC = Systemic Lupus International Collaborating Clinics; SLE = systemic lupus erythematosus.

ANA negative or had a fluctuating ANA and met SLICC criteria, the 5 most frequent criteria that were met were chronic cutaneous lupus, acute cutaneous lupus, arthritis, low complement, and a positive anti-dsDNA, with 16, 15, 11, 9, and 9 patients meeting these criteria, respectively (Table 3). Additionally, of the 32 patients who met either ACR or SLICC criteria for SLE, 27 patients had involvement of ≥1 other nonmucocutaneous organ system (84.4%), with 22 patients having arthritis, 12 with leukopenia, 8 with renal involvement, 1 with serositis, 1 with neurologic involvement, and 1 having thrombocytopenia with a count below 100×10^3 cell/μl. We also found that 13 of the 32 patients meeting SLE criteria had involvement of ≥2 other nonmucocutaneous organ systems (40.6%) (see Supplementary Appendix 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23916/abstract).

Patients with childhood-onset CLE. There were 21 patients with childhood-onset lupus in our database (diagnosed at age ≤18 years). Of these patients, 10 (47.6%) were in the ANA-negative group. One ANA-negative patient met both SLICC and ACR criteria, having a discoid rash, arthritis, a malar rash, oral ulcers, photosensitivity, renal involvement, a positive anti-dsDNA, and low complement. Two ANA-negative patients met only the ACR criteria: 1 had arthritis, oral ulcers, photosensitivity, and a discoid rash, and 1 had arthritis, leukopenia, photosensitivity, a discoid rash, and malar rash.

DISCUSSION

As was demonstrated through results from our CLE database, approximately 36.9% of patients with CLE seen at the autoimmune dermatology clinic at the University of Pennsylvania presented with a negative ANA result, and approximately 18.0% of patients with a negative ANA result met a diagnosis of SLE using either the ACR and/or the SLICC criteria. To highlight the concern for excluding patients with a negative ANA result from an SLE diagnosis, we present 3 cases below of patients with flagrant SLE who lack this serology result.

Case 1 (patient 254), a 41-year-old female, presented in 2002 with joint pain, fatigue, oral and nasal ulcers, a fever, discoid skin lesions, and a malar rash. Laboratory workup was significant for proteinuria and leukopenia, and a skin biopsy result supported a diagnosis of CLE. She was initially prescribed hydroxychloroquine and prednisone 10 mg daily for 2 years before her dose was tapered. Five years after initial presentation, she developed seizures, which her physician attributed to her SLE. In 2015, serology results revealed a positive anti-dsDNA, but her ANA titer of 1:80 was reported as negative. Additionally, anti-SSA and anti-SSB tests were negative. Based on this presentation, the patient qualified as having SLE based on both the ACR and SLICC criteria, despite having a reported negative ANA result.

Case 2 (patient 358) was a 56-year-old female who was originally diagnosed with discoid lupus in 1993 and who presented to our clinic in 2014 with discoid lesions, photosensitivity, fatigue, fevers of up to 101°F, and arthritis, with laboratory results significant for leukopenia. She was found to have positive anticardiolipin and anti-Sm antibodies but had a negative ANA result (titer <1:40). Based on her presentation, she met SLE criteria using both the ACR and SLICC criteria.

Case 3 (patient 452) was a 23-year-old male who presented to his dermatologist because of scalp alopecia, which was subsequently diagnosed as discoid lupus from a skin biopsy result. One year after his original presentation, the patient presented to the emergency department with a rash, 10-pound weight loss, lower-extremity edema, fatigue, and anorexia and was found to have an elevated creatinine level (6.7), pancytopenia, and a purpuric rash. A kidney biopsy result showed membranous glomerulonephritis, consistent with lupus nephritis class IV-S and class V. He was treated with intravenous solumedrol 1,000 mg daily with a transition to prednisone 60 mg daily. He received a serologic workup a month later, which showed an ANA result with a titer of 1:80 that was reported as negative (test performed at the University of Pennsylvania using an IIF assay). While he had normal complement, and negative anti-Sm, anti-SSA, and anti-SSB antibodies, he had a positive anti-dsDNA antibody. A repeat ANA test was checked 5 months later, which was also reported as negative. Once again, this patient met the ACR criteria of discoid rash, leukopenia, renal involvement, and a positive anti-dsDNA antibody, as well as meeting SLICC criteria based on his biopsy-proven lupus nephritis and positive anti-dsDNA antibody.

As seen in the aforementioned patient examples, SLE is a disease that is heterogeneous, with varying presentations. The spectrum of disease severity is also wide, with some patients

presenting with predominantly mucocutaneous symptoms, while others present with severe systemic disease affecting vital organs. Although a positive ANA result is a sensitive marker in identifying patients with SLE, our results and highlighted cases demonstrate that an ANA is not always present in patients with significant systemic disease. This fact should be considered when creating classification criteria to be used for entry into clinical trials.

Our findings also demonstrated that immunologic titers can change over time within a single individual, with 33.3% of our patients who were tested for multiple ANAs fluctuating from a positive to negative ANA result, or vice versa. This result may be an underrepresentation of the number of patients who truly fluctuate because our analysis was limited to patients who had repeat ANA serology results checked. Therefore, requiring ANA positivity for a diagnosis of SLE would effectively exclude some of the patients in this database, depending on when in their disease course their ANA was checked.

Additional problems arise when requiring ANA positivity for a diagnosis of SLE, including the lack of laboratory technique standardization. Currently, the gold standard for ANA testing is using an IIF assay on Hep-2 cells (8,9). However, this method is highly labor intensive, requiring technical expertise to perform serial dilutions of positive sera and to visually determine the staining patterns. As demand for ANA testing has increased, alternate techniques have been developed and employed by large commercial laboratories for ANA testing, including ELISA and multiple assays as cost-saving methods. These methods have shown up to a 35% increase in false negatives, which led the ACR to create an antinuclear antibody task force to recommend that an IIF assay should remain the gold standard for ANA testing (8). This recommendation led to an international initiative to create recommendations for ANA testing by different methods, led by 2 groups of experts: the European Autoimmunity Standardization Initiative and the International Union of Immunologic Societies/World Health Organization/Arthritis Foundation/Centers for Disease Control and Prevention Autoantibody Standardizing Committee (9). These organizations put forth 25 recommendations, recognizing that a proper ANA-IIF assay is dependent on reagents, equipment, and other local factors, and therefore the screening dilution should be defined locally. The organizations stated that an abnormal ANA result should be the titer above the 95th percentile of a healthy control population, which generally is a 1:160 dilution using HEp-2(000) substrates (9). These groups also recognized that using an IIF assay has its limitations, including the time and skill required to perform the test, and therefore included alterative assays in its recommendations as long as the platform being used is specified in the laboratory report (9). Agmon-Levin et al (9) also made the point that an IIF assay is not perfect and that a negative ANA result at 1:160 dilution in the setting of clinical suspicion should not exclude disease diagnosis.

Despite these recommendations, a study by Damoiseaux et al (10) showed that there is still significant variability in laboratory techniques being used to test for ANAs. They found that although 87% of laboratories performed ANA testing by the IIF assay technique, multiple laboratories did not have IIF testing available, most notably in Norway (50% performed the IIF assay), Portugal (64% performed the IIF assay), and Ukraine (40% performed the IIF assay). Moreover, in 35.6% of laboratories, reading of the ANA slides was only done by 1 observer as opposed to 2. Finally, Damoiseaux et al found that overall, a 1:80 dilution was employed most often (in 60.5% of laboratories), compared to a 1:40 (in 15.6% of laboratories) and a 1:160 (in 15% of laboratories) (10). Results from this study show that despite the recommendations, laboratory techniques still vary significantly and are imperfect. This variability was seen in our patient population as well, which could account for some of the fluctuating ANA results. This variability was also seen with the different cutoff titers used at different laboratories, with some subjects having an ANA titer of 1:80 or higher reported as negative, while others in the database had the same titer or lower being reported as positive. Of note, 3 of our patients had a titer of 1:160 reported as negative, which we chose to include in the ANA-positive group due to elevated titer. While one could make the argument that a titer of 1:80 or above should be considered positive (making some of the reported ANA-negative patients actually positive), 50 of the 91 patients who received an IIF assay to check their ANA result (54.9%) still had a titer of <1:80. Therefore, we still believe that an ANA result of 1:80 should not be used as an entry criterion for clinical trials. Additionally, in our experience, laboratories often do not report the titer after it is determined as negative, making it difficult to use a single titer (1:80, for instance) as a universal cutoff for negativity.

Limitations of this study include its retrospective nature. Although most laboratories reported using the IIF assay technique to detect ANAs, 4 of our 111 ANA-negative patients had ANAs measured using the ELISA technique. Additionally, we were unable to confirm which laboratory technique was used to test ANAs for 16 of our 111 ANA-negative patients. Another limitation of this study was the use of multiple laboratories to detect ANAs, each using a different titer cutoff to indicate a negative result. This laboratory variability could also account for some of the ANA fluctuations seen within the same patient. Finally, this study excluded patients ages <18 years, although patients who developed childhood-onset lupus were included as long as their current age was ≥18 years.

Given the heterogeneous and complex nature of SLE, creating a classification system is important to help guide physicians in the diagnosis of this potentially devastating disease to avoid delay in treatment. In light of new systems being proposed for classification of SLE, we reviewed a database of CLE patients seen at the University of Pennsylvania to identify the number of patients with a negative ANA result who qualified as having SLE based on the present classification criteria (SLICC and/or ACR). We found that approximately 36.9% of our CLE patients had a negative ANA result, and that of those patients, 18.0% qualified as having SLE using either the ACR or SLICC criteria. We also found that approx-

imately 33.3% of our patients had an ANA status that changed over time, and that of those patients, 44.4% met SLE criteria. Additionally, we found that 84.4% of our ANA-negative or ANA-fluctuating patients who met SLE criteria had involvement of at least 1 other nonmucocutaneous organ system. Finally, we point out the absence of standardized laboratory techniques in testing for ANAs, which may result in some false negatives that would effectively exclude patients with systemic disease from clinical trials and potential treatments. Given these results, we caution against the use of ANA positivity as a requirement for the diagnosis of SLE.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. Werth 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. Tarazi, Werth.
Acquisition of data. Tarazi, Gaffney, Kushner, Chakka.
Analysis and interpretation of data. Tarazi, Gaffney, Kushner, Werth.

REFERENCES

1. Cohen AS, Canoso JJ. Criteria for the classification of systemic lupus erythematosus: status 1972 [editorial]. Arthritis Rheum 1972;15:540–3.
2. Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1982;25:1271–7.
3. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus [letter]. Arthritis Rheum 1997;40:1725.
4. Albrecht J, Berlin JA, Braverman IM, Callen JP, Costner MI, Dutz J, et al. Dermatology position paper on the revision of the 1982 ACR criteria for systemic lupus erythematosus. Lupus 2004;13:839–49.
5. Petri M, Magder L. Classification criteria for systemic lupus erythematosus: a review. Lupus 2004;13:829–37.
6. Petri M, Orbai AM, Alarcón GS, Gordon C, Merrill JT, Fortin PR, et al. Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum 2012;64:2677–86.
7. Aringer M, on behalf of SLE Classification Criteria Steering Committee. Classification of SLE: challenges and potential solutions [abstract]. Ann Rheum Dis 2017;76:4.
8. Meroni PL, Schur PH. ANA screening: an old test with new recommendations. Ann Rheum Dis 2010;69:1420–2.
9. Agmon-Levin N, Damoiseaux J, Kallenberg C, Sack U, Witte T, Herold M, et al. International recommendations for the assessment of autoantibodies to cellular antigens referred to as anti-nuclear antibodies. Ann Rheum Dis 2014;73:17–23.
10. Damoiseaux J, Agmon-Levin N, Van Blerk M, Chopyak V, Eriksson C, Heijnen I, et al. From ANA-screening to antigen-specificity: an EASI-survey on the daily practice in European countries. Clin Exp Rheumatol 2014;32:539–46.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 71, No. 11, November 2019, pp 1410–1418
DOI 10.1002/acr.23777

Persistently Frequent Emergency Department Utilization Among Persons With Systemic Lupus Erythematosus

Jiha Lee,[1] Judith Lin,[2] Lisa Gale Suter,[3] and Liana Fraenkel[3]

Objective. In order to identify opportunities to improve outpatient care, we evaluated patients with systemic lupus erythematosus (SLE) who persistently frequent the emergency department (ED).

Methods. We conducted a retrospective study of patients with SLE who frequented the ED for ≥3 visits in a calendar year, from 2013 to 2016. Persistent users were those who met criteria for persistent use for at least 2 of the 4 years, and limited users for 1 of the 4 years. Each ED encounter was categorized as SLE-related, infection-related, pain-related, or other. We compared ED use between persistent and limited users, and analyzed factors associated with pain-related encounters among persistent users through multivariate logistic regression.

Results. We identified 77 participants who had 1,143 encounters as persistent users, and 52 participants who had 335 encounters as limited users. Persistent users accounted for 77% of ED use by patients with SLE who frequented the ED. Pain-related ED visits were more common among persistent users (32%) than limited users (18%). Among persistent users, most pain-related encounters were discharged from the ED (69%) or within 48 hours of admission (20%). Persistent users with pain-related encounters accounting for >10% of ED use were more likely to be obese, have fewer comorbid conditions, and be on long-term opioid therapy.

Conclusion. Pain is a major cause of ED use. Patients with SLE who persistently utilize the ED for pain are likely to be noncritically ill, as evidenced by frequent discharges from the ED and short-stay admissions. Patients with SLE who persistently frequent the ED for pain represent a viable target for interventions to improve outpatient quality of care.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a chronic inflammatory autoimmune disorder associated with substantial socioeconomic burden and health care resource utilization. Inpatient care accounts for the largest component of direct costs (1–4). However, patients with SLE have more ED visits than hospital admissions, with 40–70% having at least 1 ED visit in a year (1–6). In addition, hospitalizations are increasingly originating from the ED (7), and approximately 20% of admissions of patients with SLE are avoidable (8). For these reasons, understanding ED utilization among persons with SLE may provide insight into drivers of both health care resource utilization and poor quality of care for SLE in the outpatient setting.

As in the general population, SLE patients with low sociodemographic status, lower education level, and poor adherence more frequently utilize the ED, and account for the majority of all ED visits (9). The definition of frequent ED use is variable, but frequent ED users generally account for 4.5–8% of all ED patients and 21–28% of all ED visits (10). Frequent ED use has been generally thought to arise from difficulty in access to primary or specialty care (11,12). However, studies show that most frequent ED users have insurance coverage and are more likely to utilize all existing forms of health care resources including outpatient care (9,10,13–15). In addition, there is evidence to suggest that the use of ED, for most people, is an affirmative choice over other sources of health care rather than a last resort (16).

It is increasingly recognized that frequent ED users are not a homogenous population (10,17,18). In the general population, studies have demonstrated that most individuals cease to qualify as frequent ED users within a year (17,19). This brief period of frequent ED use may be due to an acute event requiring multiple ED visits, pregnancy-related complications, or flare of a chronic dis-

Dr. Suter was supported by the Centers for Medicare & Medicaid Services. Dr. Fraenkel was supported by the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases grant AR-060231).

[1]Jiha Lee, MD, MHS: Yale University School of Medicine, New Haven, Connecticut, and University of Michigan, Ann Arbor; [2]Judith Lin, MD: Yale University School of Medicine, New Haven, Connecticut; [3]Lisa Gale Suter, MD, Liana Fraenkel, MD, MPH: Yale University School of Medicine, New Haven, Connecticut, and VA Connecticut Healthcare System, West Haven, Connecticut.

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

Address correspondence to Jiha Lee, MD, MHS, University of Michigan, Department of Internal Medicine, Division of Rheumatology, 300 North Ingalls Building, Room 7C27, Ann Arbor, MI 48109-5422. E-mail: jihalee@med.umich.edu.

Submitted for publication August 23, 2017; accepted in revised form October 2, 2018.

SIGNIFICANCE & INNOVATIONS

- Systemic lupus erythematosus (SLE) is associated with substantial socioeconomic burden and health care resource utilization. SLE patients with low socioeconomic status, irrespective of their access to care, frequent the emergency department (ED). This pattern of ED use suggests a gap in the care of SLE.
- Increasingly, it is recognized that frequent ED use is not a stable phenomenon. Most patients who utilize the ED at a high rate only experience a brief period of frequent ED use (<12 months); however, a subgroup continue to frequent the ED for years.
- Understanding persistently frequent ED use in SLE can help provide insight into opportunities to reduce health care resource utilization and improve quality of care.

ease. In contrast, a small but consistent percentage persistently frequent the ED for years (17,19). Causes and, therefore, interventions for this subgroup of patients are likely to be different than for those with a brief period of frequent ED use. Understanding the factors underlying persistently frequent ED use may help inform interventions to improve chronic disease management and care coordination in the outpatient setting.

In this study, we sought to identify patients with SLE who persistently frequented the ED throughout a 4-year period. We examined the characteristics and patterns of ED utilization at the individual patient- and encounter-levels. Our research aimed to answer the following questions: What are the demographic and disease characteristics of patients with SLE who persistently frequent the ED? How do persistently frequent users compare to those with limited frequent ED use? Is persistently frequent ED use associated with certain comorbidities, in particular, chronic pain?

PATIENTS AND METHODS

Subjects. We performed an electronic health record (EHR)–based query in Epic software for a cohort for which International Statistical Classification of Diseases and Related Health Problems, Tenth Revision code of M32 for SLE was entered at least once in either the problem list, encounter diagnosis, or as a billing code during the study period between January 1, 2013 and December 1, 2016, and who met criteria for persistently frequent ED use. Persistently frequent ED use was defined as having >3 ED visits during the 12-month calendar year, similar to previous studies on ED utilization among persons with SLE (9), for at least 2 of the 4 years during the study period, consecutive or nonconsecutive, between 2013 and 2016, at a large urban tertiary medical center.

We then verified the diagnosis of SLE through in-depth EHR review. Only those who met American College of Rheumatology criteria for SLE (20), or had SLE documented by a rheumatologist, nephrologist, or dermatologist, or were on active immunosuppressive therapy treatment for no other medical condition, were determined to have a verified diagnosis of SLE and were included in the study.

For those diagnosed with SLE during the study period, we reviewed and censored ED encounters preceding the diagnosis unless diagnosis of SLE was probable at the time of visit based on physician documentation and/or serologic evaluation. We then reevaluated the number of ED encounters for these newly diagnosed patients with SLE to ensure that they still met criteria for persistently frequent ED use after removal of censored visits. In instances of patient death prior to close of the study period, we reviewed the number of ED visits from study inception to time of death to ensure fulfillment of criteria for persistently frequent ED use.

To understand the comparative magnitude and pattern of ED utilization among persistently frequent users, we performed a second EHR-based query and applied the same criteria to verify diagnosis of SLE and number of ED encounters, in order to identify SLE patients who had limited frequent ED use. Limited use was defined as meeting criteria for frequent ED use for 1 out of the 4 years during the study period.

The study protocol was approved by the Human Research Protection Program at our institution. We collected patient- and encounter-level data through retrospective in-depth physician review of the EHR, using a standardized data abstraction template.

Patient-level measures. We collected demographic information including age, sex, and race/ethnicity. Zip code information was collected in order to calculate the Area of Deprivation Index (ADI) (21). The ADI is a geographic area–based measure of socioeconomic deprivation (22). It combines 17 different indicators of socioeconomic status, including level of education, income, employment, value of assets, and poverty level derived from decennial census data. Higher ADI values represent greater deprivation. We also queried the EHR for primary insurance coverage at time of enrollment, and categorized type of insurance as Medicaid, Medicare, or private/commercial.

We collected information on SLE history, including manifestations, disease duration, and organ involvement prior to the index encounter through in-depth retrospective EHR review. For those with lupus nephritis, we reviewed treatment history, and/or active renal replacement therapy through either hemodialysis or peritoneal dialysis, and/or transplant status. Information related to lupus disease activity at time of ED encounter was not consistently available in the EHR. We also collected medication history, including exposure to glucocorticoids, hydroxychloroquine (HCQ), and/or additional disease-modifying antirheumatic drugs (DMARDs), such as azathioprine, methotrexate, and mycophenolate mofetil. We categorized long-term opioid therapy as having prescription for daily or near-daily use of opioids for at least 90 days, or total days of opioid supply >120 days (23,24). We also collected information on relevant medical comorbidities including depression.

Encounter-level measures. We classified disposition of each encounter as discharged from the ED or admitted to the hospital. For encounters resulting in ED-initiated admission, we obtained information on initial admission floor status (i.e., observation, medical/surgical floor, step down unit [SDU], intensive care unit [ICU]) and length of stay in the hospital (number of days). We categorized ED-initiated admissions without a claims code for ED critical care, not admitted to the SDU/ICU, and discharged within 48 hours as potentially avoidable short-stay (PASS) admissions.

We categorized each ED encounter as SLE-related, infection-related, pain-related, or "other." This categorization was applied after discharge, either from the ED or after ED-initiated admission. We classified encounters into 1 of these 4 groups based on the principal discharge diagnosis, which was supported by physician documentation and diagnostic evaluation results. An encounter was classified as being SLE-related if a patient presented with an SLE flare or SLE-related disease activity, and/or was prescribed glucocorticoids, HCQ, or other DMARDs during the encounter by a rheumatologist, nephrologist, or dermatologist. An encounter was classified as infection-related if a patient had positive culture, or imaging diagnostic of infection, and received antibiotics in either the ED or on discharge. An encounter was classified as pain-related if the primary discharge diagnosis was for pain not attributable to SLE, trauma, or without a specific etiology or organic cause based on unremarkable diagnostic evaluation (e.g., no changes in electrocardiogram, no elevation in troponin, no abnormal imaging), and without indication for invasive or surgical intervention. By study definition, categories of SLE- and pain-related encounters were mutually exclusive; however, an encounter could be infection-related and SLE- or pain-related. For those few cases (n = 8), the encounter was classified according to the principal discharge diagnosis. Encounters that were not related to SLE, infection, or pain were classified as "other" (see Supplementary Tables 1 and 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23777/abstract).

Analyses. Demographic and disease characteristics were described using mean ± SDs and proportions, as appropriate. We compared the distribution of encounters by category group at discharge from either the ED or after ED-initiated admission. In addition, for ED encounters that led to admission, we analyzed the length of stay and initial admission floor status to identify PASS admissions.

We also compared sociodemographic and disease characteristics between patients with SLE who persistently frequented the ED during the study period to those who had limited frequent ED use, using *t*-test for continuous measures and either the chi-square test or Fisher's exact test for categorical measures. Variables with *P* value < 0.1 or with clinical significance were then included in a multivariate logistic regression model. The same analytic approach was conducted to assess factors related to higher propensity to utilize the ED for pain-related encounters among patients with SLE who persistently frequent the ED. We compared a group of persistent users who had pain-related encounters accounting for >10% of their total ED use to those for whom pain-related encounters constituted ≤10% of ED use. Data were analyzed using Stata, version 14.2.

RESULTS

Subject characteristics. We initially identified 187 participants with possible SLE who met criteria for persistently frequent ED use, and 132 who met criteria for limited frequent ED use from 2013 to 2016, through EHR query. After in-depth retrospective EHR review to verify diagnosis of SLE, and to censor ED encounters for dates of SLE diagnosis and death, 77 participants with SLE met all inclusion criteria for persistently frequent ED use and 52 met all inclusion criteria for limited frequent ED use during the study period (see Supplementary Appendix 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23777/abstract).

Overall (n = 129), most of the participants were young African American women (n = 77 [59.7%]) with a mean ± SD age of 41.5 ± 15.6 years. All had some form of insurance, with most having Medicaid or Medicare as their primary coverage (n = 106 [82.2%]). ADI score was higher compared to the region (mean ± SD 87.3 ± 26.7), reflecting higher neighborhood socioeconomic deprivation. Most were being treated with glucocorticoids (74.4%) and/or some form of DMARD (89.1%) during the study period.

Characteristics for persistent and limited users are presented in Table 1. Approximately 1 in 3 persistent users (31.2%) and 1 in 5 limited users (19.2%) had diagnosis of depression. Long-term opioid therapy was nearly 3 times more prevalent among persistent users (37.7%) than limited users (13.5%). More persistent users had renal involvement on dialyses (19.5%) compared to limited users (5.8%).

In multivariate analysis, patients with SLE who persistently frequented the ED were more likely to be African American, have Medicare as their primary insurance coverage, be treated with dialysis, and be receiving long-term opioid therapy, compared to those with limited frequent ED use (Table 2).

ED encounters in persistently frequent versus limited frequent users. The 77 patients with SLE who persistently frequented the ED had 1,143 ED encounters and the 52 patients with limited frequent ED use had 335 ED encounters. Persistent users had more than twice the average number of ED encounters (mean 14.8 ± SD 8.8) compared to limited users (mean ± SD 6.4 ± 2.0) during the study period ($P < 0.001$). Patients with SLE who persistently frequented the ED had more encounters that led to ED-initiated admission (48.6%) than limited users (39.7%) ($P = 0.004$). More encounters were pain-related among those who persistently used the ED (32.4%) compared to those with limited use (18.2%) ($P < 0.001$). On average, persistent users had a

Table 1. Demographics and disease characteristics of SLE patients who were limited and persistently frequent ED users, 2013–2016*

Variables	Persistently frequent ED use (n = 77)	Limited frequent ED use (n = 52)	*P*
Demographics			
Age, mean ± SD years	42.3 ± 15.4	40.3 ± 15.9	0.482
Women	70 (90.9)	46 (88.5)	0.205
Race			0.026
White	11 (14.3)	16 (30.8)	
African American	53 (68.8)	24 (46.1)	
Hispanic/Latino	13 (16.9)	12 (23.1)	
Insurance			< 0.01
Medicaid	38 (49.3)	37 (71.1)	
Medicare	28 (36.4)	3 (5.8)	
Private/commercial	11 (14.3)	12 (23.1)	
ADI score, mean ± SD†	105.6 ± 11.7	102.6 ± 10.6	0.143
Comorbidities			
Psychiatric diagnosis	27 (35.1)	17 (32.7)	0.077
Depression	24 (88.9)	10 (58.8)	0.131
Hypertension	48 (62.3)	30 (57.7)	0.597
Hyperlipidemia	14 (18.2)	18 (34.6)	0.034
Diabetes	18 (23.4)	10 (19.2)	0.575
Coronary artery disease	8 (10.4)	5 (9.6)	0.886
Cardiovascular accident	7 (9.1)	7 (13.5)	0.434
Congestive heart failure	11 (14.3)	5 (9.6)	0.430
Asthma	15 (19.5)	9 (17.3)	0.756
COPD	2 (2.6)	3 (5.8)	0.360
No. of comorbidities, mean ± SD	1.6 ± 1.5	1.7 ± 1.3	0.614
BMI, mean ± SD kg/m^2	30.4 ± 9.8	29.5 ± 7.4	0.584
Long-term opioid therapy	29 (37.7)	7 (13.5)	0.003
SLE characteristics			
Disease duration ≥10 yrs	29 (38.7)	18 (34.6)	0.642
Renal involvement	34 (44.2)	18 (34.6)	0.279
LN on dialyses	15 (44.1)	3 (16.7)	0.027
LN with transplant	7 (20.5)	3 (16.7)	0.489
Lung involvement	15 (19.5)	16 (30.8)	0.141
Pericarditis	14 (18.2)	13 (25.0)	0.350
Medication use			
None	6 (7.8)	1 (1.9)	0.149
Glucocorticoids	55 (71.4)	41 (78.8)	0.344
HCQ	59 (76.6)	45 (86.5)	0.162
Other DMARD†	39 (50.6)	34 (65.4)	0.098
AZA	15 (38.5)	16 (47.0)	0.141
MTX	8 (20.5)	10 (29.4)	0.155
MMF	24 (61.5)	15 (44.1)	0.778

* Values are the number (%) unless indicated otherwise. SLE = systemic lupus erythematosus; ED = emergency department; ADI = Area of Deprivation Index; COPD = chronic obstructive pulmonary disease; BMI = body mass index; LN = lupus nephritis.
† Sum of number of participants being treated with azathioprine (AZA), methotrexate (MTX), and myocophenolate mofetil (MMF) exceed number of participants being treated with other disease-modifying antirheumatic drugs (DMARDs; other than hydroxychloroquine [HCQ]), as some were concomitantly on >1 DMARD.

mean ± SD 4.8 ± 6.1 pain-related encounters and limited users had a mean ± SD 1.2 ± 1.4 pain-related encounters during the study period ($P < 0.001$). One in 4 persistent users (26%) had >5 pain-related encounters, whereas 1 single limited user (1.9%) had >5 pain-related encounters between 2013 and 2016 ($P < 0.001$). Infection-related (12.9%) and SLE-related (6.7%) encounters were less common among persistently frequent users compared to limited users (15.5% and 10.5%, respectively). Encounters categorized as other accounted for the majority of ED use for persistent (48.0%) and limited (56%) users.

ED utilization among SLE patients who persistently frequent the ED. The 77 patients who persistently frequented the ED accounted for 77% of all ED use by patients with SLE who had ≥3 ED visits in a calendar year between 2013 and 2016. Of the 1,143 encounters incurred by patients with SLE who persistently frequented the ED, 588 (51.4%) resulted in discharge from the ED and 555 (48.6%) led to ED-initiated admissions. A substantial portion of encounters resulting in discharge from the ED were pain-related (43.7%), some were infection-related (10.4%), and few were SLE-related (1.4%) (Figure 1). The 8 encounters

Table 2. Factors associated with persistently frequent ED use compared to limited frequent ED use*

Variable	OR (95% CI)	*P*
Age	0.99 (0.96–1.03)	0.966
Women	1.67 (0.40–7.03)	0.482
Race		
White	Ref.	–
African American	5.24 (1.63–16.84)	0.005†
Hispanic/Latino	2.12 (0.52–8.68)	0.295
Insurance		
Medicaid	Ref.	–
Medicare	15.77 (3.8–73.65)	<0.001†
Private/commercial	1.71 (0.57–5.15)	0.342
No. of comorbidities	0.70 (0.48–1.01)	0.061
Depression	1.97 (0.66–5.82)	0.222
Long-term opioid therapy	3.09 (1.02–9.38)	0.046†
Renal involvement on dialysis	5.03 (1.06–23.84)	0.042†
Other DMARD‡	0.44 (0.18–1.08)	0.075

* ED = emergency department; OR = odds ratio; 95% CI = 95% confidence interval; Ref. = reference; DMARD = disease-modifying antirheumatic drug.
† Significant at $P < 0.05$.
‡ DMARDs other than hydroxychloroquine.

categorized as SLE-related on discharge from the ED involved evaluation either by a rheumatologist or a nephrologist during the ED course, and had documentation to support findings of SLE-related activity/complications in the EHR. Among encounters resulting in ED-initiated admission, 20.4% were pain-related, 15.5% were SLE-related, and 12.4% were infection-related.

Among encounters that led to ED-initiated admission, the majority of pain-related encounters (65.5%) resulted in admission with discharge within 48 hours and were significantly more likely than any other encounter category group to meet criteria for PASS admissions ($P < 0.001$). Infection-related encounters were least likely to lead to admission with discharge within 48 hours (19.8%), and were more often initially admitted to the SDU/ICU (12.8%). In comparison, 43.5% of SLE-related encounters resulted in admissions with discharge within 48 hours. Among the 56.5% of SLE-related encounters resulting in ED-initiated admissions with a length of stay longer than 48 hours, 10.3% were initially to the SDU/ICU.

The number of participants having at least 1 ED visit related to each encounter category group varied. Thirty-two participants (41.6%) had ≥1 SLE-related encounters, 55 participants (71.4%) had at least 1 infection-related encounter, and 61 participants (79.2%) had at least 1 pain-related encounter. All patients had at least 1 ED encounter classified as "other."

Patient characteristics associated with pain-related encounters among SLE patients who persistently frequent the ED. We observed a high burden of pain among patients with SLE who persistently frequented the ED, with 50.7% of encounters coding pain as the chief concern at initiation of the ED encounter. Pain was the presenting symptom for 51 (66.2%) of SLE-related encounters, 38 (25.8%) of infection-related encounters, and 171 (31.1%) of "other" encounters. Of the 580 encounters with pain symptoms reported at presentation, 320 (55.2%) were categorized as pain-related encounters upon discharge. These pain-related encounters, as aforementioned, accounted for one-third of ED use by patients with SLE who persistently frequented the ED, representing 61 participants (79.2%). We observed a wide range in the frequency of pain-related encounters among participants with at least 1 pain-related encounter. One participant had a single pain-related encounter, whereas another had 31 pain-related encounters during the study period.

In order to understand factors associated with higher propensity to utilize the ED for pain, we compared characteristics of participants who had pain-related encounters accounting for >10% of their total ED use to those with pain-related encounters accounting for ≤10% of their total ED use (Table 3). Participants with higher propensity to persistently frequent the ED for pain-related encoun-

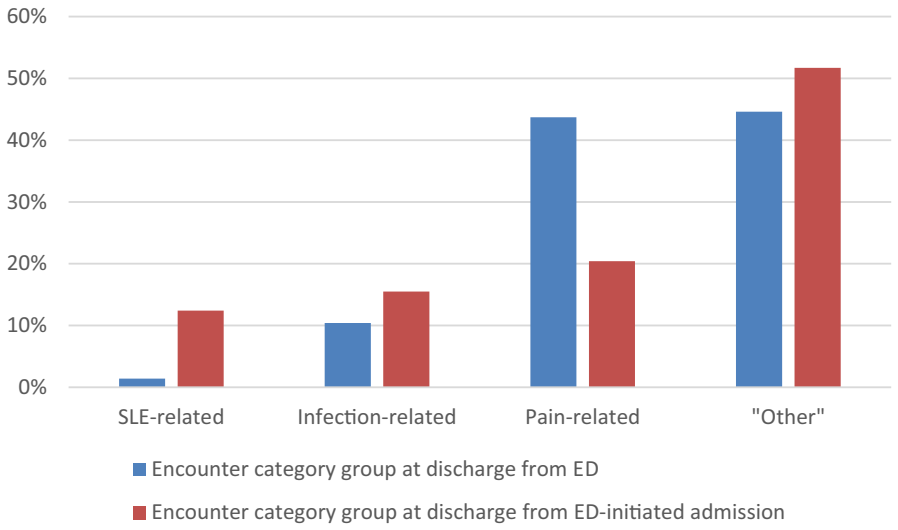

Figure 1. Proportion of emergency department (ED) encounters in each encounter category group at discharge from either the ED or after ED-initiated admission among patients with systemic lupus erythematosus (SLE) who persistently frequented the ED from 2013 to 2016.

Table 3. Comparison of patient characteristics with various degrees of pain-related encounters among SLE patients who persistently frequented the ED*

Variables	Pain-related ED encounters ≤10%	Pain-related ED encounters >10%	*P*
Participants	24 (31.2)	53 (68.8)	–
Demographics			
Age, mean ± SD years	48.0 ± 18.4	39.7 ± 13.2	0.028
Women	22 (91.7)	48 (90.6)	0.623
Race			0.001
White	5 (20.8)	6 (11.3)	
African American	10 (41.7)	43 (81.1)	
Hispanic/Latino	9 (37.5)	4 (7.6)	
Insurance			0.892
Medicaid	12 (50.0)	26 (49.1)	
Medicare	8 (33.3)	20 (37.7)	
Private/commercial	4 (16.7)	7 (13.2)	
ADI score, mean ± SD	100.8 ± 10.7	107.7 ± 11.6	0.016
Comorbidities			
Hypertension	17 (70.8)	31 (58.5)	0.218
Hyperlipidemia	8 (33.3)	6 (11.3)	0.028
Diabetes mellitus	9 (37.5)	9 (17.0)	0.049
Coronary artery disease	4 (16.7)	4 (7.5)	0.205
Congestive heart failure	8 (33.3)	3 (5.7)	0.003
Cerebrovascular accident	4 (16.7)	3 (5.7)	
COPD	2 (8.3)	0 (0.0)	0.094
Asthma	4 (16.7)	11 (20.7)	0.467
No. of comorbidities, mean ± SD	2.3 ± 2.0	1.3 ± 1.1	0.019
Depression	10 (41.7)	14 (26.4)	0.181
Long-term opioid therapy	5 (20.8)	24 (45.3)	0.040
BMI, mean ± SD kg/m^2	27.3 ± 6.9	31.8 ± 10.6	0.058
SLE disease characteristics			
Disease duration ≥10 years	7 (50.0)	23 (54.8)	0.757
Renal involvement	13 (54.2)	21 (39.6)	0.234
LN on dialyses	9 (69.2)	6 (28.6)	0.024
LN with transplant	1 (7.7)	5 (23.8)	0.237
Medication use			
Glucocorticoids	16 (66.7)	39 (73.6)	0.534
Hydroxychloroquine	16 (66.7)	43 (81.1)	0.165
Other DMARD†	8 (33.3)	31 (58.5)	0.041

* Values are the number (%) unless indicated otherwise. SLE = systemic lupus erythematosus; ED = emergency department; ADI = Area of Deprivation Index; COPD = chronic obstructive pulmonary disease; BMI = body mass index; LN = lupus nephritis; DMARD = disease-modifying antirheumatic drug.
† DMARDs other than hydroxychloroquine.

ters were younger (*P* = 0.028), more likely to be African American (*P* = 0.001), and came from more socioeconomically deprived neighborhoods (*P* = 0.016). No difference in the prevalence of depression was observed, although, long-term opioid therapy was more common in this group (*P* = 0.040). In addition, participants with >10% pain-related encounters had fewer comorbid conditions (*P* = 0.019) and were more likely to be treated with DMARDs rather than HCQ (*P* = 0.041). In multivariate analysis, African Americans, fewer comorbid conditions, long-term opioid therapy, and higher BMI were associated with higher propensity to utilize the ED for pain (Table 4).

Characteristics of SLE patients who persistently frequent the ED with pain-related PASS admissions. One in 5 hospitalized encounters were pain-related upon discharge from ED-initiated admission among patients with SLE who persistently frequented the ED, of which 74 encounters (65.6%) met criteria for PASS admissions. Table 5 shows the 25 participants (32.5%) who accounted for the 74 pain-related PASS admissions. All of the 25 participants were female except one. The mean ± SD age was 38.4 ± 13.8 years, 18 (72%) were African American, 5 (20%) were white, and 2 (8%) were Hispanic. All participants had some form of insurance; 24 (96%) had public insurance (either Medicaid or Medicare) and only 1 (4%) had private/commercial insurance as their primary insurance. Of the 25 persistent users with pain-related PASS admissions, 13 (52.0%) were on long-term opioid therapy. Even within this subgroup of participants, heterogeneity in the frequency of pain-related PASS admissions was observed. Fourteen participants (56%) had ≤2 pain related PASS admissions, whereas 1 participant accounted for 10 (13.5%) of these encounters. Overall, the 25 persistently frequent ED users with pain-related PASS admissions constituted

Table 4. Patient characteristics associated with higher propensity to utilize the ED for pain among SLE patients who persistently frequented the ED*

Variables	OR (95% CI)	*P*
Age	1.00 (0.95–1.06)	0.870
Race		
African American	Ref.	–
White	0.25 (0.02–3.09)	0.283
Hispanic	0.02 (0.00–0.17)	<0.001
Area of Deprivation Index	1.05 (0.98–1.12)	0.201
No. of comorbidities	0.54 (0.33–0.89)	0.015
Long-term opioid therapy	7.50 (1.19–47.43)	0.032
BMI	1.12 (1.01–1.24)	0.034
Other DMARD use†	2.55 (0.50–12.97)	0.258

* ED = emergency department; SLE = systemic lupus erythematosus; OR = odds ratio, 95% CI = 95% confidence interval; BMI = body mass index; DMARD = disease modifying antirheumatic drug.
† DMARDs other than hydroxychloroquine.

one-third of the study participants and accounted for 43.8% of all ED encounters.

DISCUSSION

To our knowledge, this is the first study to characterize persistently frequent ED use among patients with SLE. In this study, patients with SLE who frequented the ED were mostly young African American females, all of whom had some form of insurance. Persistent users were more likely to have Medicare as their primary insurance and be on long-term opioid therapy compared to limited users. Medicare was associated with persistent use when adjusted for age, and may be confounded by dialysis status and other factors unaccounted for in this study that relate to permanent disability or disability benefit status, which are eligibility criteria for Medicare coverage. Long-term opioid therapy and depression were each observed in 1 in 3 patients with SLE who persistently frequented the ED.

In this study, persistent users disproportionately utilized the ED compared to limited users, and mostly for non-lupus–related pain reasons. Chronic pain, a symptom frequently experienced by patients with SLE (25,26), was a major cause of ED utilization and ED-initiated admissions among patients with SLE who persistently frequented the ED. These patients were more likely to be non-critically ill, as evidenced by frequent discharge from the ED and PASS admissions. And thus, SLE patients who persistently frequent the ED for chronic pain represent a viable and high-impact target for early intervention and education to improve chronic care management and coordination.

Lessons on how to improve the delivery of care to patients with SLE may be learned from other chronic diseases, such as sickle cell anemia. Sickle cell and SLE share certain character-

Table 5. Characteristics of 25 persistently frequent ED users with pain-related PASS admissions and pattern of ED utilization during the study period*

Patient ID	Age	Sex	Race/ Ethnicity	Insurance	ADI	LTOT	No. pain PASS	Total ED visit	No. SLE-related	No. infection-related	No. pain-related	No. other †
P01	22	F	AA	Medicaid	108.73	Yes	1	31	11	1	5	16
P02	21	F	AA	Medicaid	107.47	Yes	9	41	1	1	26	13
P03	46	F	AA	Medicaid	115.60	Yes	1	17	2	2	12	1
P08	33	F	AA	Medicare	114.64	Yes	2	11	1	0	3	7
P09	36	M	AA	Medicare	115.60	Yes	5	31	7	0	23	1
P10	28	F	AA	Medicaid	109.07	Yes	4	37	2	1	15	19
P15	51	F	AA	Medicaid	108.24	No	5	12	0	1	9	2
P16	22	F	AA	Medicare	126.32	No	1	18	1	4	3	10
P21	36	F	AA	Medicaid	111.65	No	3	11	2	1	5	3
P23	44	F	Hispanic	Medicaid	96.64	No	1	10	1	2	1	6
P26	36	F	White	Medicaid	89.39	Yes	10	42	0	1	32	9
P27	22	F	AA	Medicaid	103.25	No	1	13	2	1	9	1
P29	27	F	AA	Medicaid	86.08	No	3	21	2	6	12	1
P32	39	F	AA	Medicaid	116.35	No	1	10	0	4	3	3
P34	67	F	AA	Medicare	102.61	No	2	11	0	0	9	2
P36	37	F	AA	Medicare	126.82	No	2	10	0	3	5	2
P43	48	F	White	Medicaid	108.86	Yes	4	19	2	3	11	3
P44	39	F	AA	Medicare	107.10	Yes	6	22	2	1	14	5
P49	60	F	White	Medicare	81.76	Yes	1	11	0	3	4	4
P54	72	F	AA	Medicare	116.71	Yes	1	7	0	0	4	3
P58	47	F	AA	Priv./Comm.	97.76	No	1	10	0	1	3	6
P62	32	F	AA	Medicaid	112.22	Yes	1	19	2	0	10	7
P65	30	F	Hispanic	Medicaid	115.60	No	1	9	0	0	6	3
P69	24	F	White	Medicare	109.09	Yes	4	49	0	6	8	35
P73	40	F	White	Medicare	101.70	No	4	29	5	3	17	4

* ED = emergency department; PASS = pain-related potentially avoidable short-stay admissions; ADI = Area of Deprivation Index; LTOT = long-term opioid therapy; SLE = systemic lupus erythematosus; AA = African American; Priv./Comm. = private/commercial.
† Encounters that were not related to SLE, infection, or pain.

istics in that both are complex chronic diseases, with periods of exacerbation, which disproportionately affect young African Americans and are frequently associated with chronic pain. Outpatient pain has been shown to be predictive of ED utilization among patients with sickle cell disease and intensive ambulatory management with frequent outpatient visits has been successful in reducing health care resource utilization (27–29). However, despite projected therapeutic efficacy and cost-effectiveness of ambulatory chronic pain management, compliance with and sustained improvement of health care resource utilization through nonpharmacologic pain management may be challenging. Studies have identified poor social support and communication with providers, limitation of financial and transportation resources, reliance on opioids, and lack of belief in and inadequacy of pain control as barriers to multimodality pain management (30–33). For these reasons and because of findings showing that regardless of access to care, some patients continue to preferentially utilize the ED for ambulatory care–sensitive conditions (such as chronic pain [16,34,35]), ED-based interventions for chronic pain management (such as case management, use of chronic pain protocols, and pain specialist consultation in the ED) should be developed to complement outpatient services. Some studies have explored the use of individualized home pain management programs and community health workers who provide social support, navigation of health systems and resources, and counseling, for the management of chronic pain in sickle cell disease (36,37). Web-based nonpharmacologic interventions may also be a viable option for chronic pain management in young patients with SLE, who have ready access to and familiarity with technology, but often limited access to outpatient specialty pain clinics (38–41).

This study has several limitations. Findings are based on a small number of participants at a single tertiary medical center. The cohort of patients with SLE who persistently frequented the ED, however, is expected to be small, as frequent ED users typically consist of 4.5–8% of all ED patients, and persistently frequent ED users are a smaller subgroup of this population (10,17). In addition, the criteria to confirm diagnosis of SLE was designed to have high specificity for this study, further limiting the size of the cohort. In future studies, utilization of validated EHR-based search algorithms with high positive predictive value to identify SLE patients would increase both generalizability and reproducibility. Although based on a small cohort, this study included a comparison cohort, incorporated a substantial number of unique ED encounters, and detailed information on patient- and encounter-level variables for each visit that were obtained through retrospective in-depth physician chart review. However, because data on lupus-related disease activity (either through validated or laboratory measures) at time of each ED encounter were not consistently available, we were unable to assess the relationship between SLE disease activity, pain, and ED utilization. Findings from this study would be strengthened by conducting key informant qualitative interviews. Patients with SLE who persistently frequent the ED can be engaged to elicit their perception of, and barriers to, ambulatory care coordination and chronic pain management (as relevant). Clinical impression at time of care transition from the ED and factors influencing physician decision for admissions can inform understanding of ED-initiated admissions. In this study, ED encounters were categorized using a priori criteria based on the principal discharge diagnosis. Further delineation of "other" encounters, particularly those that led to ED-initiated admission and were more likely to have greater complexity and discharge diagnosis codes, may provide further insight into the burden of pain not attributable to lupus and persistently frequent ED utilization. In addition, information on health care resource utilization during admission, especially during the first 48 hours, would allow for factors associated with PASS admissions to be ascertained, and should be included in future studies to inform opportunities to reduce ED-initiated admission of noncritically ill patients with SLE and improve outpatient chronic disease management.

In conclusion, patients with SLE who persistently frequented the ED were young African American females, who were living in more economically deprived areas, and had a high burden of depression and long-term opioid therapy. Pain was a major cause of both ED utilization and ED-initiated admissions, most of which were PASS admissions. Patients with SLE who persistently frequent the ED, particularly for pain, would benefit from targeted early interventions, in both the ED and outpatient settings, to improve chronic disease management and care coordination.

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. Lee 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. Lee, Suter, Fraenkel.
Acquisition of data. Lee, Lin, Fraenkel.
Analysis and interpretation of data. Lee, Suter, Fraenkel.

REFERENCES

1. Nichol MB, Shi S, Knight TK, Wallace DJ, Weisman MH. Eligibility, utilization, and costs in a California Medicaid lupus population. Arthritis Rheum 2004;51:996–1003.
2. Garris C, Jhingran P, Bass D, Engel-Nitz NM, Riedel A, Dennis G. Healthcare utilization and cost of systemic lupus erythematosus in a US managed care health plan. J Med Econ 2013;16:667–77.
3. Garris C, Shah M, Farrelly E. The prevalence and burden of systemic lupus erythematosus in a medicare population: retrospective analysis of medicare claims. Cost Eff Resour Alloc 2015;13:9.
4. Furst DE, Clarke A, Fernandes AW, Bancroft T, Gajria K, Greth W, et al. Resource utilization and direct medical costs in adult systemic lupus erythematosus patients from a commercially insured population. Lupus 2013;22:268–78.
5. Zhu TY, Tam LS, Li EK. Cost-of-illness studies in systemic lupus erythematosus: a systematic review. Arthritis Care Res (Hoboken) 2011;63:751–60.
6. Li T, Carls GS, Panopalis P, Wang S, Gibson TB, Goetzel RZ. Long-term medical costs and resource utilization in systemic lupus erythe-

matosus and lupus nephritis: a five-year analysis of a large Medicaid population. Arthritis Rheum 2009;61:755–63.
7. Schuur JD, Venkatesh AK. The growing role of emergency departments in hospital admissions. N Engl J Med 2012;367:391–3.
8. Ward MM. Avoidable hospitalizations in patients with systemic lupus erythematosus. Arthritis Rheum 2008;59:162–8.
9. Panopalis P, Gillis JZ, Yazdany J, Trupin L, Hersh A, Julian L, et al. Frequent use of the emergency department among persons with systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2010;62:401–8.
10. LaCalle E, Rabin E. Frequent users of emergency departments: the myths, the data, and the policy implications. Ann Emerg Med 2010;56:42–8.
11. Richardson LD, Hwang U. Access to care: a review of the emergency medicine literature. Acad Emerg Med 2001;8:1030–6.
12. Milbrett P, Halm M. Characteristics and predictors of frequent utilization of emergency services. J Emerg Nurs 2009;35:191–8.
13. Hunt KA, Weber EJ, Showstack JA, Colby DC, Callaham ML. Characteristics of frequent users of emergency departments. Ann Emerg Med 2006;48:1–8.
14. Zuckerman S, Shen YC. Characteristics of occasional and frequent emergency department users: do insurance coverage and access to care matter? Med Care 2004;42:176–82.
15. Block L, Ma S, Emerson M, Langley A, de la Torre D, Noronha G. Does access to comprehensive outpatient care alter patterns of emergency department utilization among uninsured patients in East Baltimore? J Prim Care Community Health 2013;4:143–7.
16. Ragin DF, Hwang U, Cydulka RK, Holson D, Haley LL Jr, Richards CF, et al. Reasons for using the emergency department: results of the EMPATH Study. Acad Emerg Med 2005;12:1158–66.
17. Johnson TL, Rinehart DJ, Durfee J, Brewer D, Batal H, Blum J, et al. For many patients who use large amounts of health care services, the need is intense yet temporary. Health Aff (Millwood) 2015;34:1312–9.
18. Harris LJ, Graetz I, Podila PS, Wan J, Waters TM, Bailey JE. Characteristics of hospital and emergency care super-utilizers with multiple chronic conditions. J Emerg Med 2016;50:e203–14.
19. Mandelberg JH, Kuhn RE, Kohn MA. Epidemiologic analysis of an urban, public emergency department's frequent users. Acad Emerg Med 2000;7:637–46.
20. Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1982;25:1271–7.
21. Health Innovation Program: 2000 Area Deprivation Index. 2014; URL: https://www.hipxchange.org/ADI.
22. Singh GK. Area deprivation and widening inequalities in US mortality, 1969–1998. Am J Public Health 2003;93:1137–43.
23. Chou R, Fanciullo GJ, Fine PG, Adler JA, Ballantyne JC, Davies P, et al. Clinical guidelines for the use of chronic opioid therapy in chronic noncancer pain. J Pain 2009;10:113–30.
24. Von Korff M, Saunders K, Thomas Ray G, Boudreau D, Campbell C, Merrill J, et al. De facto long-term opioid therapy for noncancer pain. Clin J Pain 2008;24:521–7.
25. Atzeni F, Cazzola M, Benucci M, Di Franco M, Salaffi F, Sarzi-Puttini P. Chronic widespread pain in the spectrum of rheumatological diseases. Best Pract Res Clin Rheumatol 2011;25:165–71.
26. Di Franco M, Bazzichi L, Casale R, Sarzi-Puttini P, Atzeni F. Pain in systemic connective tissue diseases. Best Pract Res Clin Rheumatol 2015;29:53–62.
27. Koch KL, Karafin MS, Simpson P, Field JJ. Intensive management of high-utilizing adults with sickle cell disease lowers admissions. Am J Hematol 2015;90:215–9.
28. Jonassaint CR, Beach MC, Haythornthwaite JA, Bediako SM, Diener-West M, Strouse JJ, et al. The association between educational attainment and patterns of emergency department utilization among adults with sickle cell disease. Int J Behav Med 2016;23:300–9.
29. Ezenwa MO, Molokie RE, Wang ZJ, Yao Y, Suarez ML, Angulo V, et al. Outpatient pain predicts subsequent one-year acute health care utilization among adults with sickle cell disease. J Pain Symptom Manage 2014;48:65–74.
30. Bair MJ, Matthias MS, Nyland KA, Huffman MA, Stubbs DL, Kroenke K, et al. Barriers and facilitators to chronic pain self-management: a qualitative study of primary care patients with comorbid musculoskeletal pain and depression. Pain Med 2009;10:1280–90.
31. Park J, Hirz CE, Manotas K, Hooyman N. Nonpharmacological pain management by ethnically diverse older adults with chronic pain: barriers and facilitators. J Gerontol Soc Work 2013;56:487–508.
32. Jerant AF, von Friederichs-Fitzwater MM, Moore M. Patients' perceived barriers to active self-management of chronic conditions. Patient Educ Couns 2005;57:300–7.
33. Simmonds MJ, Finley EP, Vale S, Pugh MJ, Turner BJ. A qualitative study of veterans on long-term opioid analgesics: barriers and facilitators to multimodality pain management. Pain Med 2015;16:726–32.
34. Brown LE, Burton R, Hixon B, Kakade M, Bhagalia P, Vick C, et al. Factors influencing emergency department preference for access to healthcare. West J Emerg Med 2012;13:410–5.
35. DeLia D, Cantor JC, Brownlee S, Nova J, Gaboda D. Patient preference for emergency care: can and should it be changed? Med Care Res Rev 2012;69:277–93.
36. Hsu LL, Green NS, Donnell Ivy E, Neunert CE, Smaldone A, Johnson S, et al. Community health workers as support for sickle cell care. Am J Prev Med 2016;51 Suppl 1:S87–98.
37. Crosby LE, Simmons K, Kaiser P, Davis B, Boyd P, Eichhorn T, et al. Using quality improvement methods to implement an electronic medical record (EMR) supported individualized home pain management plan for children with sickle cell disease. J Clin Outcomes Manag 2014;21:210–7.
38. Nevedal DC, Wang C, Oberleitner L, Schwartz S, Williams AM. Effects of an individually tailored web-based chronic pain management program on pain severity, psychological health, and functioning. J Med Internet Res 2013;15:e201.
39. Macea DD, Gajos K, Daglia Calil YA, Fregni F. The efficacy of web-based cognitive behavioral interventions for chronic pain: a systematic review and meta-analysis. J Pain 2010;11:917–29.
40. Stinson J, White M, Isaac L, Campbell F, Brown S, Ruskin D, et al. Understanding the information and service needs of young adults with chronic pain: perspectives of young adults and their providers. Clin J Pain 2013;29:600–12.
41. Williams DA. Web-based behavioral interventions for the management of chronic pain. Curr Rheumatol Rep 2011;13:543–9.

Arthritis Care & Research
Vol. 71, No. 11, November 2019, pp 1419–1424
DOI 10.1002/acr.23792

American College of Rheumatology
Empowering Rheumatology Professionals

BRIEF REPORT

Azathioprine and Mycophenolate Mofetil Adherence Patterns and Predictors Among Medicaid Beneficiaries With Systemic Lupus Erythematosus

Candace H. Feldman,[1] Jamie Collins,[1] Zhi Zhang,[1] Chang Xu,[1] S. V. Subramanian,[2] Ichiro Kawachi,[2] Daniel H. Solomon,[1] and Karen H. Costenbader[1]

Objective. Azathioprine (AZA) and mycophenolate mofetil (MMF) are immunosuppressants frequently used in the treatment of moderate-to-severe systemic lupus erythematosus (SLE). We studied longitudinal patterns and predictors of adherence to AZA and MMF in a nationwide US SLE cohort.

Methods. In the Medicaid Analytic eXtract (2000–2010) database, we identified patients with SLE who initiated AZA or MMF (no use in the prior 6 months) with ≥12 months of continuous follow-up. We dichotomized adherence at 80%, with ≥24 of 30 days per month considered adherent. We used group-based trajectory models to estimate monthly adherence patterns and multivariable multinomial logistic regression to determine the association between demographic, SLE and utilization-related predictors, and the odds ratios (OR) of belonging to a nonadherent versus the adherent trajectory, separately for AZA and MMF.

Results. We identified 2,309 AZA initiators and 2,070 MMF initiators with SLE. Four-group trajectory models classified 17% of AZA and 21% of MMF initiators as adherent. AZA and MMF nonadherers followed similar trajectory patterns. African American race (OR 1.67 [95% confidence interval (95% CI) 1.20–2.31]) and Hispanic ethnicity (OR 1.58 [95% CI 1.06–2.35]) increased odds of AZA nonadherence; there were no significant associations between race/ethnicity and MMF nonadherence. Male sex and polypharmacy were associated with lower odds of nonadherence to both medications; lupus nephritis was associated with lower odds of nonadherence to MMF (OR 0.74 [95% CI 0.55–0.99]).

Conclusion. Adherence to AZA or MMF over the first year of use was rare. Race, sex, and lupus nephritis were modestly associated with adherence, but the magnitude, direction, and significance of predictors differed by medication, suggesting the complexity of predicting adherence behavior.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a complex autoimmune disease with a range of organ system manifestations. Patients with moderate-to-severe disease often receive immunosuppressants, either azathioprine (AZA) or mycophenolate mofetil (MMF), often interchangeably, to control lupus nephritis, serositis, hematologic abnormalities, arthritis, and cutaneous disease (1,2). Adherence to medications for SLE varies from 20% to 80% depending on the population studied, the medication, and the method used to measure adherence (e.g., self-reported surveys, blood levels, prescription refill data) (3–6). Higher rates of nonadherence have been observed among younger age groups, African American and Hispanic patients, and individuals with less education (4,7,8). Studies have varied as to whether polypharmacy and disease severity affect the risk for nonadherence.

In Medicaid, the largest public insurance in the US, primarily for low-income individuals, our prior work demonstrated that

Supported by the Rheumatology Research Foundation (Investigator Award to Dr. Feldman) and the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases grant K23-AR-071500 to Dr. Feldman, grants R01057327 and K24-AR-066109 to Dr. Costenbader, grant K24-AR-055989 to Dr. Solomon, and grant P30-AR-072577 to Drs. Solomon and Costenbader).

[1]Candace H. Feldman, MD, ScD, Jamie Collins, PhD, Zhi Zhang, MS, Chang Xu, Daniel H. Solomon, MD, MPH, Karen H. Costenbader, MD, MPH: Brigham and Women's Hospital, Boston, Massachusetts; [2]S. V. Subramanian, PhD, Ichiro Kawachi, MBChB, PhD: Harvard T. H. Chan School of Public Health, Boston, Massachusetts.

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

Address correspondence to Candace H. Feldman, MD, ScD, 60 Fenwood Road, Office 6016P, Boston, MA 02115. E-mail: cfeldman@bwh.harvard.edu.

Submitted for publication August 7, 2018; accepted in revised form October 16, 2018.

SIGNIFICANCE & INNOVATIONS

- Adherence to the 2 most frequently used immunosuppressive medications for systemic lupus erythematosus (SLE), azathioprine (AZA) and mycophenolate mofetil (MMF), is overall extremely poor during the first year of use among a national cohort of Medicaid beneficiaries with SLE.
- Adherence overall was slightly better to MMF compared to AZA; however, adherence to both medications declined significantly for nearly 80% of patients over the first year of use.
- While demographic factors including African American race, Hispanic ethnicity, and younger age were associated with higher odds of nonadherence among AZA initiators, they were significantly less strongly associated with nonadherence among MMF initiators. This suggests that a single set of patient characteristics does not consistently predict nonadherence patterns across medications.

fewer than 20% of patients were adherent to hydroxychloroquine (HCQ), defined as ≥80% of days covered with prescription refills during the first year of use (8). There were higher odds of nonadherence among patients with SLE who were young, African American, taking fewer medications, and with less severe disease. We also observed that HCQ adherence was dynamic and for most declined over the first year of use. In this study, we aimed to assess patterns of adherence over the first year of AZA or MMF use. We hypothesized that like HCQ, adherence would decline over time and that, while predictors of nonadherence would be similar, patterns would suggest better adherence to MMF because patients' illness may be slightly more severe and therefore more invested in continuing their medication.

PATIENTS AND METHODS

Patient cohort. We used the Medicaid Analytic eXtract (MAX) database with demographic data, billing claims, healthcare utilization, and drug dispensing data from 2000–2010 for the 29 most populated US states (86% of Medicaid beneficiaries nationwide). We excluded all claims from Ohio because detailed medication dispensing data were not available, and we additionally excluded all individuals without drug dispensing data, including those who were hospitalized for the entire follow-up period. We identified 2 cohorts of patients with SLE (≥2 International Classification of Diseases, Ninth Edition [ICD-9] codes for SLE [710.0] for discharge diagnoses or physician claims ≥30 days apart) with either AZA or MMF dispensing within 365 days of an SLE code (8). We required ≥6 months of continuous enrollment without use of AZA or MMF prior to the date of initiation (index date). We allowed AZA initiators to previously receive MMF and vice versa, and we required ≥365 days of continuous enrollment following the index date to assess adherence.

Measures of adherence. We used prescription refill data to measure adherence, which has previously been validated in claims data (9). We calculated the proportion of days covered (PDC) beginning at the index date for 365 days (number of days covered divided by 365 × 100, subtracting hospitalized days from the numerator and denominator) and classified individuals with PDC ≥80% as adherent (10). We also measured adherence to MMF and to AZA monthly over the 12-month period, and each month was classified either as adherent (1) or nonadherent (0) depending on whether ≥24 of 30 days (80%) were covered. The majority (>85%) of our cohort received a 1-month supply of AZA or MMF in accordance with Medicaid policies in most states.

Potential correlates of adherence. We measured potential predictors during the 6 months prior to and including the index date in the AZA and MMF cohorts. Demographic factors included age and state of residence at the index date, sex, race/ethnicity, and region from MAX, and zip code-level median household income from American Community Survey data (11). We included diabetes mellitus, smoking, lupus nephritis, use of antidepressant medications, SLE-related laboratory tests and medications (HCQ, immunosuppressants, and corticosteroids), number of medications at the index date, days' supply of first AZA/MMF prescription fill, vaccinations, health-care utilization, and the SLE risk-adjustment index (see Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23792/abstract) (8,12,13). We also examined models that included comorbidities (thromboembolism, pulmonary, hepatic, cardiovascular and cerebrovascular disease, substance abuse, obesity, and malignancy) but did not include them in our final models as they did not contribute significantly and have not been shown in prior studies to be strongly associated with adherence.

Statistical analyses. We compared baseline characteristics and PDC for AZA and MMF using descriptive statistics. We used our binary indicators of monthly adherence to construct group-based trajectory models (GBTMs) to classify patients by adherence separately for AZA and MMF. GBTMs are used to identify latent patterns in longitudinal data with repeated measures and have been previously applied to prescription refill data to uncover adherence patterns over the first year of use (8,14). We evaluated AZA and MMF GBTMs ranging from 3 to 6 trajectory groups and based our model choice on a combination of Bayesian information criteria with lower values considered preferable, reasonable distribution across groups, posterior probabilities ≥80% for each group, and

explanatory potential (15). We then used multinomial logistic regression models for both AZA and MMF to determine the odds of belonging to a nonadherent trajectory compared to the persistently adherent trajectory for demographic, utilization, and SLE-related predictors.

We conducted sensitivity analyses, censoring at potential conditions that may have resulted in physician-recommended discontinuation rather than nonadherence. We censored at the beginning of the nearest preceding refill for first discharge diagnosis code for serious infection and for any code for neutropenia or transaminitis for both cohorts, and additionally for pregnancy or colitis for the MMF cohort (see Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23792/abstract). We conducted all analyses using SAS software, version 9.4 and used the Proc Traj add-on package for GBTMs. Data were obtained from the Centers for Medicare and Medicaid Services through a data use agreement. Data are presented in accordance with their policies. The Partners Healthcare Institutional Review Board approved this study.

RESULTS

We identified 2,309 AZA initiators and 2,070 MMF initiators. AZA initiators were slightly older with a mean ± SD age of 36.1 ± 11.8 years compared to 33.4 ± 11.6 years for MMF (Table 1). The percentage of females was slightly higher among AZA initiators as was the percentage of African Americans. On average, AZA initiators at baseline had less severe SLE with a lower mean SLE risk adjustment index, lower prevalence of lupus nephritis, fewer overall medications, and less immunosuppressant use. Corticosteroid use was comparable between initiators of the 2 drugs. The mean ± SD PDC for AZA initiators beginning at the index date of new use was 40% ± 29%, with 15% classified as adherent (PDC ≥80%) compared to 44% ± 30% for MMF, with 18% classified as adherent ($P < 0.001$). We did not observe significant changes in either AZA or MMF adherence by index date year (see Supplementary Figure 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23792/abstract). When we varied the adherence threshold, 7.3% of AZA initiators and 8.6% of MMF initiators had PDCs ≥90%, and 21.9% of AZA initiators and 25.7% of MMF initiators had PDCs ≥70%.

We examined 3- to 6-group GBTMs for both AZA and MMF initiators. We aimed to balance model fit with explanatory power and to compare similar numbers of trajectories for both medications. A 4-group model provided an adequate fit for both drugs. The mean posterior probabilities for each trajectory were >80%, and each trajectory had a reasonably balanced distribution of individuals. The Bayesian information criteria for the 3-group models were slightly smaller than for the 4-group models, but we chose the 4-group models because we felt that the explanatory potential was greater, and the other model fit criteria were met (15).

Table 1. Baseline characteristics of azathioprine and mycophenolate initiators with SLE in Medicaid, 2000–2010*

Characteristics	Azathioprine (n = 2,309)	Mycophenolate mofetil (n = 2,070)
Age, mean ± SD years	36.1 ± 11.8	33.4 ± 11.6
Female	2,138 (92.6)	1,857 (89.7)
Race/ethnicity		
African American	1,089 (47.2)	925 (44.7)
White	579 (25.1)	506 (24.4)
Hispanic	476 (20.6)	453 (21.9)
Asian	93 (4.0)	114 (5.5)
American Indian/Alaska Native	22 (1.0)	31 (1.5)
Region		
Northeast	523 (22.7)	577 (27.9)
South	837 (36.3)	680 (32.9)
Midwest	378 (16.4)	368 (17.8)
West	571 (24.7)	445 (21.5)
Median (25th, 75th percentile) household income, $†	41,643 (33,659–51,948)	42,557 (33,995–55,565)
Diabetes mellitus	297 (12.9)	242 (11.7)
SLE risk adjustment index, mean ± SD	1.4 ± 2.2	2.1 ± 2.5
Lupus nephritis	503 (21.8)	1,162 (56.1)
No. drugs, mean ± SD	5.0 ± 3.6	5.5 ± 3.8
Antidepressant use	439 (19.0)	209 (10.1)
Corticosteroid use	1,898 (82.2)	1,704 (82.3)
Hydroxychloroquine use	1,315 (57.0)	938 (45.3)
Immunosuppressant use‡	374 (16.2)	526 (25.4)
No. of SLE-related laboratory tests, mean ± SD§	3.3 ± 4.2	3.7 ± 5.0

* Values are the number (%) unless indicated otherwise. SLE = systemic lupus erythematosus.
† By zip code.
‡ Methotrexate, leflunomide, tacrolimus, sulfasalazine, cyclosporine, cyclophosphamide, azathioprine (for mycophenolate mofetil initiators), and mycophenolate mofetil (for azathioprine initiators).
§ Blood urea nitrogen, creatinine, complement C3 and C4, erythrocyte sedimentation rate, C-reactive protein level, anti–double-stranded DNA, and urinalysis.

Overall, we observed similar patterns for the 4-group trajectory model (Figure 1). In the persistently adherent trajectory (group 4), there were 384 (17%) AZA initiators and 441 (21%) MMF initiators. In the persistently nonadherent trajectory (group 1), there were 1,030 (45%) AZA initiators and 883 (43%) MMF initiators. Among AZA initiators, 2 groups with more dynamic nonadherent patterns (group 2 and group 3) steadily declined until between 6–7 months, at which point adherence for group 3 plateaued at ~45% of days/month covered, while group 2 continued to decline. Among MMF initiators, of those with more dynamic nonadherent patterns, group 2 precipitously declined initially and then plateaued at ~30% of days/month covered. Group 3 declined more slowly over the course of use, remaining just below the adherent range of ≥80% of days covered until between months 4 and 5.

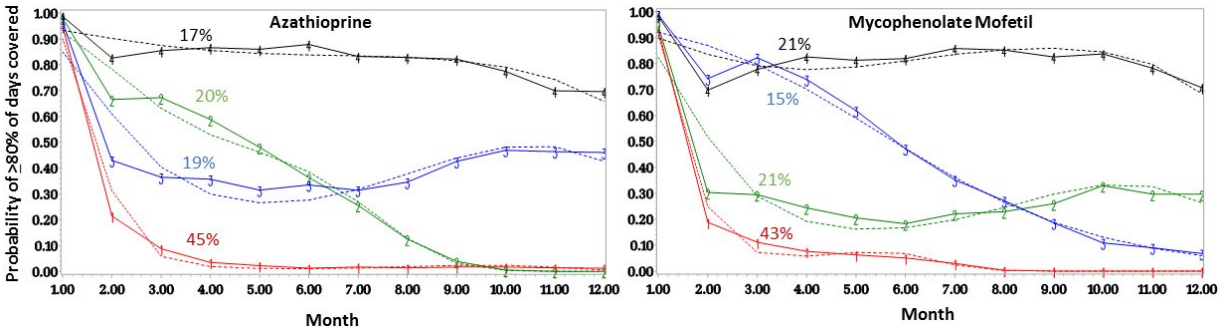

Figure 1. Group-based trajectory models demonstrating monthly adherence patterns for patients with systemic lupus erythematosus enrolled in Medicaid over the first year of azathioprine and mycophenolate mofetil use, with group 4 (**black**) as the persistently adherent trajectory and group 1 (**red**) as the persistently nonadherent trajectory. Groups 2 (**green**) and 3 (**blue**) show more dynamic nonadherent patterns.

We examined multinomial logistic regression models comparing the odds of belonging to the nonadherent trajectories (groups 1–3) versus the persistently adherent trajectory (group 4) for both AZA and MMF initiators (Table 2). Among AZA initiators, we observed increased odds of belonging to the persistently nonadherent trajectory (group 1) versus the persistently adherent trajectory (group 4) among patients with SLE who were African American (odds ratio [OR] 1.67 [95% confidence interval (95% CI) 1.20–2.31]) or Hispanic (OR 1.58 [95% CI 1.06–2.35]) compared to white and in the age 18–35 year group (OR 1.60 [95% CI 1.10–2.34]) compared to the oldest (age 51–65 years). We found >2 times higher odds of belonging to the declining and then plateauing nonadherent trajectory (group 3) versus the persistently adherent trajectory (group 4) for African Americans and Hispanics, and reduced odds of belonging to group 3 versus 4 among individuals living in areas with less than or equal to the median of the zip code median household income compared to areas above the median (OR 0.63 [95% CI 0.47–0.86]). Male sex and increased number of medications were associated with reduced odds of belonging to nearly all nonadherent trajectories compared to the most adherent (group 4).

Among MMF initiators, we did not observe statistically significant associations by race/ethnicity comparing any of the nonadherent trajectories (groups 1–3) to the persistently adherent trajectory (group 4). Similar to AZA, we observed reduced odds of belonging to the persistently nonadherent trajectory (group 1) versus the persistently adherent (group 4) among males compared to females (OR 0.67 [95% CI 0.45–0.99]) and for increased medication use (OR 0.90 [95% CI 0.87–0.94]). Specific to MMF initiators, we observed reduced odds of belonging to the persistently nonadherent versus the persistently adherent group among patients with SLE with lupus nephritis (OR 0.74 [95% CI 0.55–0.99]) and increased odds among those living in areas below or equal to the median household income compared to above (OR 1.33 [95% CI 1.02–1.72]). Increased number of emergency department visits were also associated with increased odds of belonging to the least adherent groups (groups 1 and 2) versus the most adherent group (group 4).

In sensitivity analyses, censoring at indications that may have resulted in physician-recommended discontinuation of AZA or MMF, we observed only modestly increased adherence estimates. Among AZA initiators, censoring at ≥1 ICD-9 code for serious infection, transaminitis or neutropenia, the mean ± SD PDC was 45% ± 30%, with 18.8% categorized as adherent (PDC ≥80%). Among MMF initiators, censoring at serious infection, transaminitis, neutropenia, colitis, or pregnancy, the mean ± SD PDC was 49% ± 29%, with 22.7% categorized as adherent.

DISCUSSION

Overall, we observed profoundly poor adherence among Medicaid beneficiaries with SLE who initiated AZA or MMF over the first year of use; less than one-quarter of patients refilled their medications 80% of the time or more. These rates of nonadherence were similar among HCQ initiators also enrolled in Medicaid (8). Our findings were also in line with a prior study that utilized trajectory models to describe adherence patterns among statin initiators; only 23.4% were persistently adherent over the first 15 months of use (14).

In our cohorts, adherence was slightly better among MMF initiators compared to AZA initiators, and groups of MMF nonadherers appeared to remain at least partially adherent for longer. The populations of AZA and MMF initiators were somewhat different; AZA initiators were slightly older, included more females, likely because this medication is compatible with pregnancy whereas MMF is not, and had less severe SLE, with a lower prevalence of lupus nephritis and prior immunosuppressant use. Predictors of nonadherence differed as well. We observed primarily demographic associations with AZA nonadherence; African American race, Hispanic ethnicity, female sex, and younger age were associated with increased odds of nonadherence. However, among MMF initiators, although we did see an association between female sex, younger age, and nonadherence,

Table 2. Multivariable multinomial regression comparing the odds of belonging to a nonadherent trajectory (groups 1–3) to the most adherent trajectory (group 4, reference) for azathioprine and mycophenolate mofetil initiators with SLE*

Predictors	Azathioprine (n = 2,309)			Mycophenolate mofetil (n = 2,070)		
	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
Total (%)	1,030 (44.6)	459 (19.9)	436 (18.9)	883 (42.7)	441 (21.3)	305 (14.7)
Age, years						
18–34	1.60 (1.10–2.34)†	1.75 (1.12–2.73)†	1.48 (0.96-2.29)	1.14 (0.76–1.71)	1.15 (0.71–1.86)	1.95 (1.12–3.42)†
35–50	1.36 (0.95–1.96)	1.67 (1.09–2.57)†	1.17 (0.76–1.79)	0.83 (0.55–1.26)	0.96 (0.59–1.56)	1.44 (0.82–2.54)
51–65 (reference)						
Sex	0.59 (0.39–0.91)†	0.68 (0.42–1.11)	0.54 (0.32–0.92)†	0.67 (0.45–0.99)†	0.85 (0.56–1.31)	0.83 (0.51–1.34)
Male						
Female (reference)						
Race/ethnicity						
African American	1.67 (1.20–2.31)†	1.62 (1.11–2.35)†	2.05 (1.39–3.02)†	1.33 (0.95–1.86)	1.20 (0.82–1.76)	0.72 (0.47–1.09)
Hispanic	1.58 (1.06–2.35)†	1.37 (0.87–2.15)	2.00 (1.26–3.19)†	0.94 (0.64–1.39)	0.85 (0.55–1.32)	0.88 (0.56–1.40)
Asian	1.02 (0.52–1.99)	0.85 (0.40–1.84)	1.52 (0.72–3.18)	0.64 (0.36–1.13)	0.62 (0.33–1.16)	0.59 (0.30–1.17)
American Indian/ Alaska native	0.78 (0.24–2.51)	0.61 (0.15–2.52)	NR	1.88 (0.57–6.25)	1.49 (0.39–5.65)	0.77 (0.16–3.83)
White (reference)						
SLE risk adjustment index	0.98 (0.91–1.05)	0.97 (0.90–1.05)	0.98 (0.90–1.06)	0.99 (0.92–1.05)	0.97 (0.90–1.04)	1.00 (0.93–1.08)
Lupus nephritis	1.06 (0.75–1.49)	1.22 (0.83–1.79)	1.31 (0.89–1.93)	0.74 (0.55–0.99)†	0.98 (0.70–1.36)	0.95 (0.66–1.37)
Diabetes mellitus	1.11 (0.74–1.67)	1.74 (1.11–2.71)†	1.30 (0.81–2.08)	1.05 (0.70–1.59)	0.95 (0.59–1.53)	1.02 (0.61–1.68)
Household income	0.80 (0.61–1.04)	0.99 (0.73–1.34)	0.63 (0.47–0.86)†	1.33 (1.02–1.72)†	1.28 (0.95–1.72)	1.26 (0.92–1.75)
≤median						
>median (reference)						
No. of medications	0.90 (0.86–0.93)†	0.94 (0.90–0.98)†	0.92 (0.88–0.96)†	0.90 (0.87–0.94)†	0.90 (0.86–0.94)†	0.99 (0.95–1.03)
Outpatient visits	1.01 (0.98–1.03)	0.99 (0.97–1.02)	0.99 (0.9–1.01)	0.99 (0.96–1.01)	0.99 (0.96–1.01)	0.98 (0.95–1.00)
Hospitalizations	1.08 (0.96–1.21)	1.05 (0.93–1.20)	1.09 (0.96–1.24)	0.98 (0.89–1.08)	0.96 (0.86–1.08)	0.98 (0.87–1.10)
ED visits	1.05 (0.99–1.12)	1.05 (0.98–1.13)	1.01 (0.94–1.09)	1.16 (1.08–1.25)†	1.14 (1.05–1.23)†	1.06 (0.97–1.15)

* Values are odds ratios (95% confidence interval) unless indicated otherwise. Models were additionally adjusted for calendar year of index date, state of residence at index date, days' supply of medication at first dispensing, SLE-related medication use (hydroxychloroquine, immunosuppressants, corticosteroids), number of SLE-related laboratory tests, antidepressant use, smoking, obesity, and influenza and pneumococcal vaccinations. SLE = systemic lupus erythematosus; ED = emergency department.
† Significant.

the findings were less consistent across nonadherent trajectories, and we did not observe significant associations by race/ethnicity. Among MMF initiators, we observed reduced odds of persistent nonadherence associated with lupus nephritis and increased medication use and increased odds associated with more emergency room visits. The median zip-code level household income was slightly higher for MMF initiators compared to AZA, and interestingly, while we observed increased odds of nonadherence among MMF initiators living in lower versus higher median income zip codes, the trend was in the opposite direction among AZA initiators for unclear reasons.

In agreement with prior studies, we did not find 1 dominant factor to be consistently associated with all nonadherence patterns across different SLE-related medications. While many factors likely contribute to adherence behavior, their relationship is not necessarily constant across medications, populations, or over time of use. Differences between SLE-related medications, including affordability, tolerability, regimen complexity, interactions with other medications, or lifestyle factors, likely contribute to varying patterns of adherence. In addition, aspects that cannot be measured in a study like ours, which relied on claims data, including the doctor-patient relationship or beliefs about SLE and medication safety, play a role as well. This complexity suggests that a simple variable-based algorithm to predict a person's adherence pattern over time to multiple medications is likely unrealistic.

This study has limitations. We utilized prescription refill data to determine adherence, which may not always represent use; however, prior studies have shown this to be a valid method (9,16). We did not have data regarding initial AZA or MMF prescribing and therefore could not capture primary nonadherence. We used a cutoff of ≥80% to indicate adherence, and although this is accepted in the chronic disease literature, it is unclear whether it correlates with the physiologic levels needed to have a clinically meaningful effect. In addition, we could not distinguish between medically indicated discontinuation and nonadherence. However, we excluded all hospitalized time during which adherence could not be readily measured and conducted sensitivity analyses censoring at potential indications for physician-recommended discontinuation, which resulted in adherence estimates similar to our primary analyses. We lacked qualitative measures of potential predictors of nonadherence, as well as actual laboratory results and medical records to understand fluctuations in disease activity. We also lacked

information on socioeconomic status, such as individual income or education, and we cannot exclude the possibility that racial/ethnic differences may be markers for differences in socioeconomic circumstances. We lacked more recent data later than 2010; however, between 2000 and 2010, we did not appreciate significant fluctuations in either AZA or MMF adherence.

In our study, we leveraged 2 large cohorts of AZA and MMF initiators to examine adherence patterns over time. Although we found relatively similar patterns between the 2 medications, MMF initiators seemed to stay at least partially adherent for longer, possibly due to an understanding of the need to treat more severe manifestations such as lupus nephritis. Adherence overall, however, was very poor, and in this vulnerable, low-income SLE population, more needs to be done to facilitate persistent adherence to efficacious, standard of care medications to ultimately reduce disparities in outcomes.

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. Feldman 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. Feldman, Collins, Zhang, Xu, Subramanian, Kawachi, Solomon, Costenbader.

Acquisition of data. Feldman, Costenbader.

Analysis and interpretation of data. Feldman, Collins, Zhang, Xu, Subramanian, Kawachi, Solomon, Costenbader.

REFERENCES

1. Hahn BH, McMahon MA, Wilkinson A, Wallace WD, Daikh DI, Fitzgerald JD, et al. American College of Rheumatology guidelines for screening, treatment, and management of lupus nephritis. Arthritis Care Res (Hoboken) 2012;64:797–808.
2. Tunnicliffe DJ, Singh-Grewal D, Kim S, Craig JC, Tong A. Diagnosis, monitoring, and treatment of systemic lupus erythematosus: a systematic review of clinical practice guidelines. Arthritis Care Res (Hoboken) 2015;67:1440–52.
3. Julian LJ, Yelin E, Yazdany J, Panopalis P, Trupin L, Criswell LA, et al. Depression, medication adherence, and service utilization in systemic lupus erythematosus. Arthritis Rheum 2009;61:240–6.
4. Durcan L, Clarke WA, Magder LS, Petri M. Hydroxychloroquine blood levels in systemic lupus erythematosus: clarifying dosing controversies and improving adherence. J Rheumatol 2015;42:2092–7.
5. Feldman CH, Yazdany J, Guan H, Solomon DH, Costenbader KH. Medication nonadherence is associated with increased subsequent acute care utilization among Medicaid beneficiaries with systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2015;67:1712–21.
6. Costedoat-Chalumeau N, Houssiau F, Izmirly P, Guern VL, Navarra S, Jolly M, et al. A prospective international study on adherence to treatment in 305 patients with flaring SLE: assessment by drug levels and self-administered questionnaires. Clin Pharmacol Ther 2018;103:1074–82.
7. Garcia-Gonzalez A, Richardson M, Garcia Popa-Lisseanu M, Cox V, Kallen MA, Janssen N, et al. Treatment adherence in patients with rheumatoid arthritis and systemic lupus erythematosus. Clin Rheumatol 2008;27:883–9.
8. Feldman CH, Collins J, Zhang Z, Subramanian SV, Solomon DH, Kawachi I, et al. Dynamic patterns and predictors of hydroxychloroquine nonadherence among Medicaid beneficiaries with systemic lupus erythematosus. Semin Arthritis Rheum 2018;48:205–13.
9. Steiner JF, Prochazka AV. The assessment of refill compliance using pharmacy records: methods, validity, and applications. J Clin Epidemiol 1997;50:105–16.
10. Yeaw J, Benner JS, Walt JG, Sian S, Smith DB. Comparing adherence and persistence across 6 chronic medication classes. J Manag Care Pharm 2009;15:728–40.
11. Integrated Public Use Microdata Series. National Historical Geographic Information System database, version 12.0. University of Minnesota. URL: https://www.nhgis.org/.
12. Ward MM. Development and testing of a systemic lupus-specific risk adjustment index for in-hospital mortality. J Rheumatol 2000;27:1408–13.
13. Chibnik LB, Massarotti EM, Costenbader KH. Identification and validation of lupus nephritis cases using administrative data. Lupus 2010;19:741–3.
14. Franklin JM, Shrank WH, Pakes J, Sanfelix-Gimeno G, Matlin OS, Brennan TA, et al. Group-based trajectory models: a new approach to classifying and predicting long-term medication adherence. Med Care 2013;51:789–96.
15. Nagin DS. Group-based modeling of development. Cambridge (MA): Harvard University Press; 2005.
16. Steiner JF, Koepsell TD, Fihn SD, Inui TS. A general method of compliance assessment using centralized pharmacy records: description and validation. Med Care 1988;26:814–23.

Arthritis Care & Research
Vol. 71, No. 11, November 2019, pp 1425–1429
DOI 10.1002/acr.23802

American College of Rheumatology
Empowering Rheumatology Professionals

BRIEF REPORT

Mock Recruitment for the Study of Antimalarials in an Incomplete Lupus Erythematosus Trial

David R. Karp,[1] Benjamin F. Chong,[1] Judith A. James,[2] Cristina Arriens,[2] Mariko Ishimori,[3] Daniel J. Wallace,[3] Duanping Liao,[4] and Nancy J. Olsen[4]

Objective. Recruitment to randomized clinical trials is expensive and often falls short of goals, limiting achievement of measurable outcomes. To prepare for a trial in patients with incomplete forms of lupus, a mock recruitment protocol was carried out at 4 proposed study sites. The objective was to determine levels of interest in patients and to uncover potential barriers to enrollment.

Methods. After obtaining institutional review board approval, study coordinators approached individuals who generally fit proposed criteria for the trial. A standardized script was followed in a structured interview. Levels of interest were determined and any reasons for concerns were collected with an open-ended format.

Results. A total of 45 subjects were interviewed, of which 73% expressed an interest in the trial, and 64% said they were likely to enroll. Concerns of those who were not interested included risk of hydroxychloroquine, desire not to receive placebo, and lack of time for participation.

Conclusion. The mock recruitment suggests that the trial will be attractive to suitable patients. The concerns raised support other data indicating that provision of information is crucial to achieving enrollment goals. Mock recruitment of potential investigators should be considered also to address referral concerns.

INTRODUCTION

Systemic lupus erythematosus (SLE) is defined by a combination of characteristic autoantibodies along with clinical and laboratory evidence of immune-mediated pathology. While there are no diagnostic standards for SLE, criteria were developed first in 1982 (1) by the American College of Rheumatology, revised in 1997 (2), and revised again in 2012 by the Systemic Lupus International Collaborating Clinics (SLICC) (3). These classification criteria were designed to ensure that homogeneous cohorts were used for observational or interventional clinical research. In each case, a person previously diagnosed with an autoimmune disorder is classified as having SLE based on the accumulation of clinical and laboratory features common to expertly adjudicated cases.

The laboratory and clinical features that lead to lupus classification typically evolve over time (4). Autoantibodies such as antinuclear antibody (ANA) and anti-Ro are often detectable 5 years or more before the first clinical symptom. Retrospective studies have documented that ANA-positive individuals have evidence of altered innate and adaptive immunity, including elevated levels of interferon-driven cytokines and memory B cells (5). Clinically, the features of SLE can be present over several years before formal classification is possible. This has given rise to the term incomplete lupus erythematosus (ILE), which may be used to refer to individuals with clinical or serologic features of lupus who are progressing to incident SLE, as well as individuals who never develop the requisite number of criteria for formal classification (6,7).

The diagnostic concept of ILE is problematic for several reasons. While a percentage of patients with ILE slowly develop

ClinicalTrials.gov identifier: NCT03030118.

The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

Supported by the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases grants U34-AR-067392 and P30-AR-073750, National Institute of General Medical Sciences grant U54-GM-104938, and National Center for Advancing Translational Sciences grant UL1-TR-0001105).

[1]David R. Karp, MD, PhD, Benjamin F. Chong, MD, MSCS: University of Texas Southwestern Medical Center, Dallas; [2]Judith A. James, MD, PhD, Cristina Arriens, MD, MSCS: Oklahoma Medical Research Foundation and University of Oklahoma Health Sciences Center, Oklahoma City; [3]Mariko Ishimori, MD, Daniel J. Wallace, MD: Cedars-Sinai Medical Center, Los Angeles, California; [4]Duanping Liao, MD, PhD, Nancy J. Olsen, MD: Penn State College of Medicine, Hershey, Pennsylvania.

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

Address correspondence to David R. Karp, MD, PhD, University of Texas Southwestern Medical Center, Dallas, TX 75390. E-mail: david.karp@utsouthwestern.edu.

Submitted for publication April 26, 2018; accepted in revised form October 23, 2018.

SIGNIFICANCE & INNOVATIONS

- The majority of patients with incomplete lupus erythematosus were interested in a clinical trial comparing untested standard of care to placebo.
- However, only 64% said they would be likely to enroll in such a trial.
- Successful recruitment for a trial such as this, which tests a treatment available as part of standard of care, will require additional education of potential subjects and effort to minimize the impact of the study on participants' lives.
- Study success will likely require that twice as many potential participants are screened as need to enroll for completion of the trial.

SLE, it is not clear whether that subset can be identified and given therapy to prevent progression of their disease. Retrospective analyses of several large cohorts of such patients reveal that two-thirds or more of them are treated with antimalarials such as hydroxychloroquine (HCQ) despite the lack of a Food and Drug Administration indication for this condition (8,9). This drug has been thoroughly tested in cases of SLE, where it has been shown to prevent disease flares, decrease damage, improve survival, and delay onset of disease (10). However, HCQ has not been rigorously tested in patients with ILE. While a recent study showed that ILE patients taking HCQ had lower expression levels of type I interferon-inducible genes than those not taking the drug, the long-term clinical benefit of this medication is unknown (7).

The Study of Anti-Malarials in Incomplete Lupus Erythematosus (SMILE; NCT03030118) is a multicenter, randomized, placebo-controlled, clinical trial of HCQ in patients with ILE. The primary outcome of this trial is the change in the rate of accumulation of SLICC criteria for the classification of lupus. Approximately 200 patients are to be followed for 2 years while receiving the study medication. In order to plan for this trial, it was necessary to learn whether patients with ILE would be willing to participate. While HCQ is not formally approved for this condition, it is commercially available, and most physicians feel comfortable using it in patients who are not diagnosed with SLE. Therefore, the prospect of being randomized to receive a placebo may make it difficult to recruit a sufficient number of participants. To study this and judge the feasibility of the clinical trial, a mock recruitment was conducted for SMILE. This recruitment has given important insights into the willingness of patients to participate and informed our recruitment strategies.

SUBJECTS AND METHODS

The current study was performed at the University of Texas Southwestern Medical Center, the Milton S. Hershey Medical Center of Penn State University, the Oklahoma Medical Research Foundation, and the Cedars-Sinai Medical Center. It was approved by an institutional review board at each site.

Study participants were identified during routine clinical care in the rheumatology and/or dermatology practices of the participating institutions. They met the provisional inclusion criteria for SMILE: 1) ages 18–45 years; 2) ANA ≥1:160 by indirect immunofluorescence testing; 3) presence of 1 or 2 additional SLICC criteria for the classification of lupus; 4) either sex; 5) be able to provide informed consent.

Participants who met the appropriate criteria were read a script (see Supplementary Appendix 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23802/abstract) that described the rationale and conduct of the SMILE trial. The risks and potential benefits of HCQ and placebo were explained, as was the fact that HCQ was already available from their provider. The physician or study coordinator then obtained informed consent from the patient for their participation. Basic demographic data and information on personal or family history of autoimmune diseases were collected from the patients who consented. Participants were then asked the following 4 questions:

1) I have told you about a proposed clinical trial or study of HCQ to treat early or mild lupus-like illness similar to the condition your doctor is treating you for. Do you have any questions about the proposed clinical trial? Can I explain it better? (Participants could answer yes or no).
2) If the proposed clinical trial of HCQ to treat early or mild lupus-like illness was enrolling patients today, how much interest do you have in learning more about the trial (for example, look at consent form, talk to family, talk to other doctors)? (Participants could answer 0% [no interest], 25%, 50%, 75%, or 100% [very interested]).
3) If the proposed clinical trial of HCQ to treat early or mild lupus-like illness was enrolling patients today, how likely would you be to enroll in the trial, considering all the information you have today? (Participants could answer 0% [definitely not enroll], 25%, 50%, 75%, or 100% [definitely would enroll]).
4) If you answered 0%, 25%, or 50% to either Question 1 or 2, can you tell us what your concerns are? You may pick more than one. (Participants were offered the following prompts: Risks from HCQ appear too great; don't want to risk getting placebo; no time to participate; already tried HCQ in the past; want to try something else; other [free text]).

Additional data, including the presence of ANA, other serology, immunologic features of SLE, presence of cytopenias, serositis, musculoskeletal symptoms, cutaneous symptoms, and constitutional symptoms (e.g., fatigue, fever), were obtained from the clinical chart. All data were entered anonymously into a Research Electronic Data Capture (REDCap) database hosted

Table 1. Interest and likelihood of enrolling in a trial of hydroxychloroquine for incomplete lupus erythematosus*

Feature	No.	Interest in learning more		Likelihood of enrolling	
		0–50%	75–100%	0–50%	75–100%
Ethnicity/race†					
Non-Hispanic White	31	7	24	11	20
African American	6	1	5	2	4
Hispanic	5	2	3	3	2
Previous diagnosis of autoimmune disease	8	1	7	3	5
Family history of autoimmune disease	19	4	15	3	5
Serology other than ANA	14	2	12	3	11
Cytopenias	4	1	3	2	2
MSK symptoms	27	7	20	8	19
Cutaneous symptoms	26	6	20	9	17
Serositis	1	0	1	1	0
Constitutional symptoms	20	3	17	5	15

* All comparisons of low versus high interest and low versus high likelihood of enrollment were not significant by Fisher's exact test. ANA = antinuclear antibody; MSK = musculoskeletal.
† Two participants were Asian and 1 declined to identify their race.

by the University of Texas Southwestern Center for Translational Medicine (11).

For the purposes of analysis, participants were categorized as likely to enroll in the clinical trial if they answered Question 3 with either 75% or 100%. A Fisher's exact test was performed to test the significance of differences in the proportion of such participants based on clinical and demographic characteristics.

RESULTS

A total of 45 subjects were included. The median age was 35 years (range: 16–57 years) and the median duration of symptoms prompting evaluation by a rheumatologist or dermatologist was 3 years (range: 1–21 years). Most participants were female (n = 43; 96%) and non-Hispanic white (n = 31; 69%). A minority of participants were African American (n = 6; 13%) or Hispanic (n = 5; 11%). Some reported that they themselves had previously been given a diagnosis of an autoimmune disease (n = 8; 18%), while a greater number indicated that a family member had been diagnosed with an autoimmune disease (n = 19; 42%). In some of these patients (n = 14; 31%), autoantibodies other than an ANA were known to be present at the time of the interview. Musculoskeletal symptoms (arthralgia/arthritis or myalgia) and cutaneous symptoms were the most common clinical features (n = 27; 60% and n = 26; 58% of participants, respectively), and constitutional symptoms were also relatively common (n = 20; 44%); these were primarily fatigue or low-grade fever. No participant had known contraindications for the use of HCQ.

Most of the participants (n = 42; 93%) felt that the description of the proposed clinical trial was adequate. One subject requested more information on the constituents of the placebo capsule. Overall, 33 (73%) of the interviewees expressed interest in the trial, and a majority (n = 29; 64%) said they were likely to enroll. None of the clinical or demographic features listed in Table 1 were significantly associated with the likelihood of the responses "interested in the trial" or "likely to enroll in the trial."

The 16 interviewees who were not interested or were not likely to participate in the clinical trial were asked for their reasons. Some (n = 4; 25%) were concerned that the stated risks of HCQ appeared too great, and there was also concern about the risk of receiving the placebo (n = 3; 19%). Eight of those who were not interested (50%) felt that they did not have time to participate in a trial. Free text answers also included concerns about the number of medications that the respondents were already taking and the logistics of participation (Table 2).

DISCUSSION

Recruitment and enrollment in randomized clinical trials is a major challenge of clinical investigation. Clinical trials in human subjects are expensive and recruitment costs are a significant proportion of the expense. Failure to reach recruitment goals has been estimated to occur in at least 20% of trials and often results in premature study closure and decreased reliability of the data generated. As a consequence, translation of research to clinical practice is impaired, and implementation of advancements in disease treatment is slowed. Recognition of these problems has led to reexamination of approaches to recruitment, which has generally been an empiric process that is addressed only after all other elements of the study are in place. The significant and unmet need for development of innovative tools for recruitment, education, and outreach has been shown by the fact that the National Institutes of Health has supported initiatives in this area (www.trialinnovationnetwork.org).

Table 2. Respondent concerns about participation in the proposed clinical trial

Participant statement
Fear of aggravating celiac disease
Depends on job schedule, difficult to make appointments
Uncertain about starting a new medication for the patient, personally
Just concerned about starting a new study
Symptoms are not that significant to want to take another medication
Give it more thought
What about comparing 2 doses of Plaquenil?
Wonder about side effects
On too many other medications
Want to do more research about the drug
Would like to remain off medication and would not want placebo if I needed medication

Systematic reviews of recruitment failure have shown some common themes. Most often, investigators overestimate the number of eligible participants who meet the inclusion criteria (12). This is a possibility with the SMILE trial as well. While the presence of ANA in the general population is common, potential participants in SMILE must meet several other inclusion criteria, including age and the presence of 1 or 2 objective SLICC criteria, as well as not having taken HCQ previously. In another review of recruitment strategies, Caldwell and colleagues determined that various strategies to increase the potential participants' knowledge of the health problem under study through computer-aided instruction, videos, and seminars were the most helpful (13). Others have emphasized the greater effectiveness of recruitment to clinical trials that enroll African Americans and Latinos when preconsent education was provided (14). This was anticipated for the SMILE trial with the development of culturally appropriate recruitment materials and Spanish language trial information.

The proposed study in patients with ILE presents some unique challenges. One is the condition itself. Eligible patients will be those who have certain inclusion criteria but who do not satisfy the standard SLE classification criteria. The concept that this is a lupus-related trial but that it does not include patients with SLE has required outreach education with colleagues who are the major referral sources for eligible patients. Another challenge is the double-blind, placebo-controlled design. Approximately 1 in 5 interviewees who thought they would not enroll expressed concern about placebo treatment as the reason. However, given the inherent variability in the proposed study population, placebo and active arms were deemed essential to evenly distribute the many unknown variables. The long-term trial objective of prevention of SLE also presents challenges. A major barrier to enrollment in a previous study on prevention of cardiovascular disease in SLE patients was found to be the poor understanding of the importance of prevention on the part of both patients and the physicians who were sources of referrals (15). That previous experience differs from the SMILE trial in that symptomatic improvement was not a likely outcome given that the target was prevention of atherosclerosis, and the interventions proposed were unlikely to alter the patient's perception of well-being. In contrast, individuals who receive active treatment in the SMILE trial have a likelihood of improved symptoms, which will be useful for retention as well.

It is also important to consider the baseline prevalence of persons eligible for the SMILE trial. A formal count of individuals meeting all inclusion and exclusion criteria was not possible. Preliminary estimates at the trial sites indicated that 12–25% of all referrals were for ANA positivity (not shown). At one site (University of Texas Southwestern), an automated process monitors all patients who receive ANA testing and are in the appropriate age range for SMILE. Typically, there are ~75 such people per month, with 1–2 per week meeting the other entry criteria for SMILE. This suggests that while recruitment for SMILE is numerically feasible, robust strategies such as tools that search the electronic health record and repeated outreach to referring physicians and community rheumatologists are necessary for success.

One relevant question is whether patients and providers would support enrollment in a trial of an agent that is readily available for off-label use, as is the case for HCQ. Erkan and colleagues recently reported the experience of an international multicenter trial designed to explore efficacy of HCQ thrombosis prevention in patients with antiphospholipid antibody syndrome (16). The trial was designed by a consortium of investigators who determined the urgent need for such a study. The design was not a blinded study, with HCQ as the active agent, compared to a control group that received no treatment. The sample size estimate was for ~1,000 patients, but after 2 years of recruitment only 2% of this goal was reached and the study was closed without achieving interpretable results. The study encountered many problems, including cost and availability of the active drug, as well as a reluctance on the part of some of the enrollment sites to administer what was perceived as prophylactic therapy to relatively healthy individuals. Patients also expressed an unwillingness to take a medication for what they saw as a low risk of events. The enrollment goal was very high; by contrast, the power calculations for the SMILE trial suggest only 192 completed patients will be required. Whether any of the issues in this previous trial might have been uncovered and mitigated by a mock recruitment exercise or by an alternative design is not known. However, it did not appear that off-label use of HCQ was a major reason for failure of recruitment.

A common theme in clinical trial strategy development is that engagement and education of both potential patients and referring providers in the health issues that are being addressed by the intervention trial are essential elements of successful recruitment. The reason for the trial appears to have greater impact than elements of the trial process itself. In the SMILE mock recruitment exercise, one-quarter of those who were not interested or not likely to participate expressed concerns about the risks of HCQ, which are in fact low and which might be addressed with appropri-

ate educational materials. A mock trial simulation in patients with lupus at 2 centers reported similar findings in that patients recommended community engagement to provide information about the disease and the impact of the proposed study on patients (17). These investigators also suggested that study teams be sensitive to the needs of patients, including concerns about scheduling and time constraints, which were issues that were expressed by half of those who were not interested in the SMILE mock recruitment. This consideration was taken into account when planning the visit schedule and length of visits in the final protocol.

The attitudes of colleagues who would be sources of patient referrals for enrollment were not surveyed in the mock exercise and likely would be of value to include in such prestudy activities. Appropriate educational materials might be developed in response to concerns about randomization, placebo treatment, and drug risk. The cited experience of the multinational study of HCQ and thrombosis shows that even when the study question is considered by experts in the field to be of high importance, individual providers may be reluctant to participate. In the case of SMILE, patients and providers will be assured that participants will be observed more closely than would be the case in routine care. If any individual progresses to an SLE classification, study exit is mandated so that appropriate therapy can be initiated.

ACKNOWLEDGMENTS

We appreciate the support of the study coordinators and data entry assistants Danielle Feger, Jamie Carter, Carl McAloose, Virginia Roberts, and Azza Mutwally Badr.

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. Karp 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. Karp, Chong, James, Arriens, Ishimori, Wallace, Liao, Olsen.

Acquisition of data. Karp, Chong, James, Arriens, Ishimori, Wallace, Olsen.

Analysis and interpretation of data. Karp, Arriens, Liao, Olsen.

REFERENCES

1. Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1982;25:1271–7.
2. 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.
3. Petri M, Orbai AM, Alarcón GS, Gordon C, Merrill JT, Fortin PR, et al. Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum 2012;64:2677–86.
4. Heinlen LD, McClain MT, Merrill J, Akbarali YW, Edgerton CC, Harley JB, et al. Clinical criteria for systemic lupus erythematosus precede diagnosis, and associated autoantibodies are present before clinical symptoms. Arthritis Rheum 2007;56:2344–51.
5. Slight-Webb S, Lu R, Ritterhouse LL, Munroe ME, Maecker HT, Fathman CG, et al. Autoantibody-positive healthy individuals display unique immune profiles that may regulate autoimmunity. Arthritis Rheumatol 2016;68:2492–502.
6. Aberle T, Bourn RL, Munroe ME, Chen H, Roberts VC, Guthridge JM, et al. Clinical and serologic features in patients with incomplete lupus classification versus systemic lupus erythematosus patients and controls. Arthritis Care Res (Hoboken) 2017;69:1780–8.
7. Olsen NJ, McAloose C, Carter J, Han BK, Raman I, Li QZ, et al. Clinical and immunologic profiles in incomplete lupus erythematosus and improvement with hydroxychloroquine treatment. Autoimmune Dis 2016;2016:8791629.
8. Al Daabil M, Massarotti EM, Fine A, Tsao H, Ho P, Schur PH, et al. Development of SLE among "potential SLE" patients seen in consultation: long-term follow-up. Int J Clin Pract 2014;68:1508–13.
9. Wieczorek IT, Propert KJ, Okawa J, Werth VP. Systemic symptoms in the progression of cutaneous to systemic lupus erythematosus. JAMA Dermatol 2014;150:291–6.
10. Olsen NJ, Schleich MA, Karp DR. Multifaceted effects of hydroxychloroquine in human disease. Semin Arthritis Rheum 2013;43:264–72.
11. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. 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.
12. Briel M, Olu KK, von Elm E, Kasenda B, Alturki R, Agarwal A, et al. A systematic review of discontinued trials suggested that most reasons for recruitment failure were preventable. J Clin Epidemiol 2016;80:8–15.
13. Caldwell PH, Hamilton S, Tan A, Craig JC. Strategies for increasing recruitment to randomised controlled trials: systematic review. PLoS Med 2010;7:e1000368.
14. Dunlop AL, Leroy ZC, Logue KM, Glanz K, Dunlop BW. Preconsent education about research processes improved African Americans' willingness to participate in clinical research. J Clin Epidemiol 2011;64:872–7.
15. Costenbader KH, Brome D, Blanch D, Gall V, Karlson E, Liang MH. Factors determining participation in prevention trials among systemic lupus erythematosus patients: a qualitative study. Arthritis Rheum 2007;57:49–55.
16. Erkan D, Unlu O, Sciascia S, Belmont HM, Branch DW, Cuadrado MJ, et al. Hydroxychloroquine in the primary thrombosis prophylaxis of antiphospholipid antibody positive patients without systemic autoimmune disease. Lupus 2017;961203317724219.
17. Lim SS, Kivitz AJ, McKinnell D, Pierson ME, O'Brien FS. Simulating clinical trial visits yields patient insights into study design and recruitment. Patient Prefer Adherence 2017;11:1295–307.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 71, No. 11, November 2019, pp 1430–1435
DOI 10.1002/acr.23804

BRIEF REPORT

A Pilot Study of Infrared Thermal Imaging to Detect Active Bone Lesions in Children With Chronic Nonbacterial Osteomyelitis

Yongdong Zhao,[1] Ramesh S. Iyer,[1] Lucas Reichley,[1] Assaf P. Oron,[2] Nancy E. Gove,[2] Averi E. Kitsch,[3] Debosmita Biswas,[3] Seth Friedman,[1] Savannah C. Partridge,[3] and Carol A. Wallace[1]

Objective. Chronic nonbacterial osteomyelitis (CNO) is an autoinflammatory bone disease. An inexpensive and rapid imaging tool, infrared thermal imaging, was evaluated for its utility to detect active bone lesions in extremities of children with CNO.

Methods. Children with suspected active CNO and healthy controls were enrolled. All subjects underwent infrared thermal imaging of the lower extremities. Patients in the CNO group also received a magnetic resonance imaging (MRI) examination. Hyperintensity within bone marrow on a fluid-sensitive T2-weighted MRI sequence was considered confirmatory for inflammation. Infrared thermal data were analyzed using custom software by dividing the leg below the knee into 3 equal segments longitudinally and adding the distal femur segment as an equal length above the knee. Median and 95th percentile temperatures were recorded for each leg segment. Temperature differences between inflamed and uninflamed segments in all subjects (both intersubject and intrasubject) were evaluated using a linear mixed-effects model.

Results. Thirty children in the suspected/known CNO group and 31 healthy children were enrolled. In the healthy control group, males had significantly higher temperature in their lower extremities than females ($P < 0.05$). There was no difference in temperature detected between inflamed leg segments of patients with CNO versus uninflamed leg segments of the healthy control group. However, within the CNO group, significantly higher temperatures were detected for inflamed versus uninflamed distal tibia/fibula segments ($P < 0.01$).

Conclusion. Children with active CNO lesions in the distal tibia/fibula exhibited higher regional temperatures on average than healthy extremities. Larger studies are warranted to further evaluate the clinical utility of infrared thermal imaging for CNO detection.

INTRODUCTION

Chronic nonbacterial osteomyelitis (CNO) is an autoinflammatory bone disease, which may evolve to bony destruction and deformity. The diagnosis of CNO is based on clinical bone pain, lytic and/or sclerotic bone lesion by radiography, and/or edema on magnetic resonance imaging (MRI) once malignancy and infection have been excluded (1,2). The disease is most common in long bones (1,3,4). Radiographs are relatively insensitive in detecting CNO, with a rate of 13% in 1 study when compared to MRI-confirmed sites of disease (5). MRI has become the gold standard for monitoring CNO due to its combination of superior sensitivity and lack of ionizing radiation (6).

A CNO lesion on MRI is defined as a focal region of signal hyperintensity within bone marrow on a short-tau inversion recovery (STIR) sequence with corresponding reduced T1-weighted signal (6–9). However, MRI examinations are expensive and require sedation in young children and therefore are used sparingly. A

Dr. Zhao's work was supported by the Seattle Children's Research Institute (Academic Enrichment Fund) and the Childhood Arthritis and Rheumatology Research Alliance.

[1]Yongdong Zhao, MD, PhD, Ramesh S. Iyer, MD, Lucas Reichley, BS, Seth Friedman, PhD, Carol A. Wallace, MD: Seattle Children's Hospital and University of Washington, Seattle; [2]Assaf P. Oron, PhD (current address: Institute for Disease Modeling, Bellevue, Washington), Nancy E. Gove, PhD: Seattle Children's Research Institute, Seattle, Washington; [3]Averi E. Kitsch, BS, Debosmita Biswas, MS, Savannah C. Partridge, PhD: University of Washington and Seattle Cancer Care Alliance, Seattle.

Dr. Zhao has received research support from Bristol-Myers Squibb. No other disclosures relevant to this article were reported.

Address correspondence to Yongdong Zhao, MD, PhD, Seattle Children's Hospital, MA 7.110, 4800 Sand Point Way NE, Seattle, WA 98105. E-mail: yongdong.zhao@seattlechildrens.org.

Submitted for publication May 6, 2018; accepted in revised form October 30, 2018.

SIGNIFICANCE & INNOVATIONS

- To our knowledge, this is the first study using infrared thermal imaging to determine active inflammation in bone.
- In children with chronic nonbacterial osteomyelitis, median and 95th percentile temperatures of inflamed long bones were significantly higher than uninflamed bones in distal tibia/fibula.
- Within healthy controls, males had significantly higher temperature than females for all lower extremity regions evaluated ($P < 0.05$).

quick and noninvasive tool to distinguish potential patients with CNO needing further workup with MRI from the large population of children with benign bone pain could dramatically shorten time to diagnosis and treatment.

Current thermal imaging technology enables real-time 2-dimensional temperature mapping with high spatial and temporal resolution. Infrared thermal imaging has been used as a screening tool in various conditions, including skin cancer, deep venous thrombosis, and rheumatoid arthritis (10–12), but has not yet been investigated in CNO. We hypothesize that the inflammation caused by CNO can be detected by thermal imaging based on elevated temperature in the affected skeletal site. The feasibility of this application was investigated in our proof-of-concept study.

MATERIALS AND METHODS

Institutional review board approval was obtained from the authors' tertiary-care, multidisciplinary pediatric hospital prior to the study. Two groups, including children with possible or established CNO and healthy children, were enrolled after consenting. Inclusion criteria of the CNO group were: previously established patients with CNO with suspected active lesions within long bones of the lower extremities (focal warmth and/or swelling and/or persistent pain) or new patients with focal uptake of tracer in bone scan or typical radiographic findings of CNO; age between 2 and 18 years; and clinical indication of MRI of an extremity warranted. Exclusion criteria of the CNO group were: skin infection in imaged area that could interfere with thermal imaging results; additional sedation time required for the research portion of the MRI; or inability to cooperate with the acquisition of thermal imaging.

Inclusion criteria of the healthy control group were: age between 2 and 18 years and normal skeletal health. Exclusion criteria of the control group were: current medication use, including nonsteroidal antiinflammatory drugs (NSAIDs); skin infection in the imaged area that could have interfered with thermal imaging results; or inability to cooperate with the acquisition of thermal imaging. The healthy group did not undergo MRI.

Sample size calculation. For 20 patients with CNO who have an MRI-confirmed active lesion, a temperature difference of +1°C should be detectable with a 99% power, assuming a per person standard deviation of 1°C as reported in active arthritis (11), in a 2-sided paired *t*-test. We aimed to enroll 30 subjects to account for missing data. The sample size of the healthy control group was set as 30–35 patients to match the CNO group.

Image acquisition. For the CNO group, non-contrast MRI images (including a STIR sequence) of the affected extremity with 2 planes (typically coronal and axial) were obtained based on clinical indication, and a research MRI examination of the contralateral extremity was obtained during the same scanning session. Twelve patients in the CNO group received an MRI scan of the entire femur and tibia. The entire tibia was imaged in 8 patients, while the entire femur was imaged in 3 patients. Five patients received partial images of either femur or tibia/fibula, or both due to clinician's preference.

All subjects from the CNO group also received infrared thermal imaging of the MRI-imaged extremities from 4 views (anterior, posterior, medial, and lateral) either prior to or following the MRI examination within a week. Thermal imaging was performed using a Fluke TiR32 thermal imager with 76,800 pixels (320 × 240) (detection range –20 to 150°C, sensitivity ≤0.04°C) by trained staff to ensure sharp focus, consistent camera leveling, and stabilization. The entire imaging session took <5 minutes. Subjects exposed their feet and entire legs to room air and rested for at least 10 minutes prior to imaging to allow stabilization and equilibration of skin temperature. Ambient temperature was set at 22.2°C (72°F). Subjects posed in standardized positions to ensure consistency of image acquisition. Imaging was performed with subjects standing on a carpet to avoid influence from the cold floor on body temperature and away from potentially interfering items such as metal panels, door knobs, computer screens, and other people. A subset of patients and controls were imaged 3 times consecutively for reproducibility analysis.

Analysis of thermal images. The spatial and temperature data from infrared thermal images were exported from SmartView software (Fluke, Inc.). Data were then analyzed using customized semiautomated software developed in Matlab (MathWorks). The lower legs were divided equally into 3 segments (proximal, mid, and distal) longitudinally by placing crosshairs at the medial and lateral sides of the knees and ankles from each view (see Supplementary Figure 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23804/abstract), and the distal femur segment was defined as the same length as the proximal tibia/fibula. Minimum, median, 95th percentile, and maximum temperatures were recorded for each leg segment. Analysis time was ~1 minute per image. Operators were blinded to subject diagnosis.

Demographic, clinical and laboratory data collection. Demographic information included sex, age, ethnicity, and race. Clinical data included body height, weight, oral temperature, presence or absence of bone swelling, pain or warmth, physician global assessment (range 0–10), Childhood Health Assessment Questionnaire score (C-HAQ) (range 0–3), patient self-reported severity of pain (range 0–10), parent assessment of overall health (range 0–10), and current medications. Laboratory data, including complete blood cell count, C-reactive protein level, and erythrocyte sedimentation rate, were collected if available.

Grading of CNO lesions from MRI. MRI images were scored by an experienced pediatric radiologist (RSI), who was blinded to both clinical and thermal imaging results. Briefly, long bones including femur, tibia, and fibula were divided into 3 equal segments longitudinally. The presence of hyperintensity in bone marrow (presumed inflammation) within these segments and/or surrounding soft tissue on the STIR sequence was reported by the radiologist. Grading of severity of bone marrow hyperintensity was based on the extent of bone or soft tissue affected similarly, as previously reported (13). Mild was defined as <25% of the area affected, moderate as 25–50%, and severe as >50%.

Statistical analysis. Children with non-analyzable thermal images were excluded from analysis. Within the CNO group, thermal data from areas without MRI confirmation were also excluded. All bone segments (distal femur, proximal tibia/fibula, mid tibia/fibula, distal tibia/fibula) were divided into 3 groups: the healthy control group; the CNO group without inflammation/bone marrow hyperintensity (MRI-negative) group; and the CNO group with

Table 1. Patient characteristics for suspected active chronic nonbacterial osteomyelitis group and healthy control group*

Variables	CNO group (n = 27)	Healthy controls (n = 31)	*P*
Age at enrollment, years	10.2 (3.2–16.8)	9.7 (3.3–16.2)	0.11
Height, cm	139.8 (97.8–160.3)	134.0 (96.6–161.4)	0.11
Weight, kg	36.3 (18.9–69.1)	29.8 (14.4–61.1)	0.04
BMI	17.8 (12.9–28.2)	16.0 (14.1–24.7)	0.04
Oral temperature, °C	36.6 (36.0–37.1)†	36.9 (36.3–37.6)‡	0.004
Female, no. (%)	14 (52)	18 (58)	0.83
Race, no. (%)			
White	22 (81)	22 (71)	0.12
Asian	1 (4)	7 (23)	
Other	2 (7)	2 (6)	
Clinical symptoms, no. (%)			
Focal swelling	8 (30)		
Focal tenderness	18 (67)		
Focal warmth	13 (48)		
Associated psoriasis	2 (7)		
Other rash	1 (4)		
Laboratory findings			
WBC, $\times 10^3/mm^3$	7.5 (3.0–13.5)		
Platelet count, $\times 10^3/mm^3$	329.5 (194.0–411.0)		
Hct, %	37.7 (31.7–45.5)		
CRP level, mg/dl	0.8 (0.8–3.3)		
ESR, mm/hour (normal 0–20)	13.0 (5.0–34.0)		
Parent/patient measures			
Patient's pain (range 0–10)	3.0 (0.0–7.0)		
Parent's global assessment of overall health (range 0–10)	1.0 (0.0–8.0)		
C-HAQ	0.2 (0.0–1.8)		
Physician global assessment of disease activity (range 0–10)	2.0 (0.0–5.0)		
Treatment at study entry, no. (%)§			
No therapy	9 (33)		
NSAIDs	15 (65)		
Methotrexate	5 (19)		
Biologic	2 (7)		
Pamidronate	0 (0)		

* Values are median (range) unless indicated otherwise. CNO = chronic nonbacterial osteomyelitis; BMI = body mass index; WBC = white blood cell; Hct = hematocrit; CRP = C-reactive protein level; ESR = erythrocyte sedimentation rate; C-HAQ = Childhood Health Assessment Questionnaire; NSAIDs = nonsteroidal antiinflammatory drugs.
† In 19 subjects.
‡ In 20 subjects.
§ Some patients were receiving multiple medications.

inflammation/bone marrow hyperintensity (MRI-positive) group. Categorical and continuous demographic variables were summarized and separated by group. A linear mixed-effects model with the patient as a random factor was used to inspect the association between the median, 95th percentile highest, and maximum temperatures of each region of interest (ROI) and lesion diagnosis by MRI in the CNO group with the assumption of normal in the healthy control group while adjusting for age and sex. The mid-tibia/fibula segment was excluded from this analysis due to low incidence of lesions occurring in this segment in the study cohort.

We accounted for multiple testing using the false discovery rate (FDR), with FDR < 0.2 considered significant, appropriate for a small-sample pilot study. The reliability of repeated imaging was analyzed using intraclass correlation, and analysis was performed with R software, version 3.4 (R Foundation).

RESULTS

A total of 30 children from the CNO group and 31 from the healthy group were enrolled. There were 27 children within the CNO group who completed MRI examinations and had analyzable thermal imaging. Patient characteristics from each group are summarized and compared in Table 1. The weight and body mass index (BMI) were significantly greater in the CNO group than in the healthy control group ($P < 0.05$). The healthy control group had a higher overall body temperature (36.9°C versus 36.6°C; $P = 0.004$). Within the CNO group, 30% of subjects had focal bone swelling, whereas 67% had focal tenderness, and 48% exhibited focal warmth on physical examination. Laboratory values were normal for the majority of subjects in the CNO group. Patient self-reported pain, parental global assessments of overall health, C-HAQ, and physician global assessment were in the mild or moderate ranges. At study entry, 65% of patients with CNO were taking NSAIDs, while 33% were not receiving therapy.

Anterior views of 3 replicate thermal images from a total of 20 patients with CNO and healthy controls were analyzed by the same operator, and the intraclass correlation coefficients of thermal measurements ranged from 0.936 to 0.981 in all 4 segments, demonstrating excellent repeatability. A total of 19 children within the CNO group had MRI-confirmed active bone inflammation (hyperintensity in bone marrow) in their lower extremities. Overall, 26 distal, 2 mid (excluded from analysis), 18 proximal tibia/fibula, and 12 distal femur lesions were detected on MRI. The severity of bone edema within these lesions is shown in Figure 1. Images and thermal analysis results from a representative case are provided in Supplementary Figure 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23804/abstract.

Within healthy control subjects, there was a significant difference in sex in regard to extremity infrared temperatures ($P < 0.05$ for medial, lateral, and posterior views; FDR < 0.1). Males had an average of 0.69–0.98°C higher temperatures than females in all views. Age was inversely correlated with temperatures, with a slope of 0.03–0.12°C per year in all views.

Using a mixed model with covariates including age, sex, health/inflammation status (control, CNO with inflammation, CNO without inflammation), and bone segment to determine the difference in temperatures, there were no significant differences in median, 95th percentile, or maximum temperatures between inflamed bone segments from children with CNO and healthy controls in any view (see Supple-

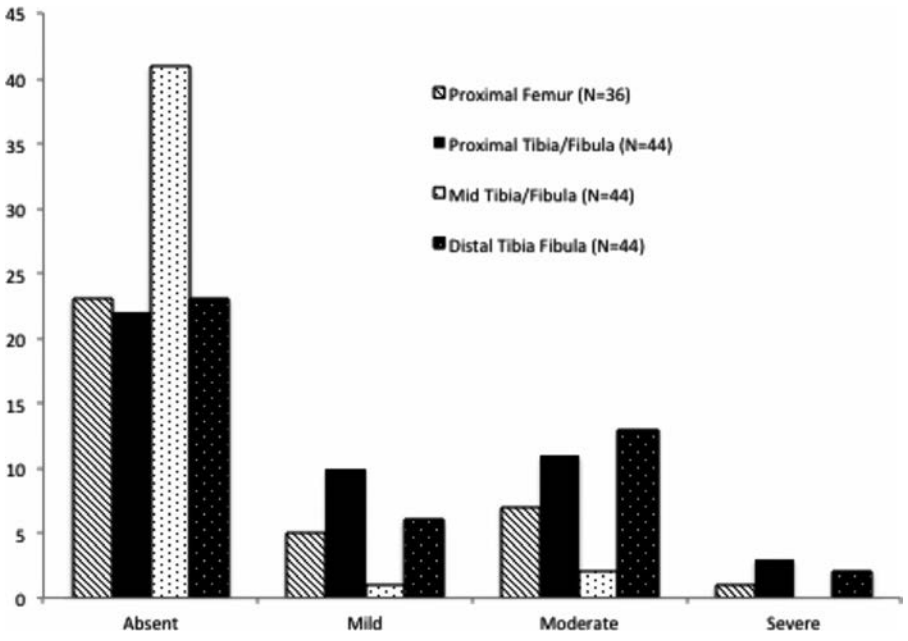

Figure 1. Number of extremities with hyperintensity within bone marrow in each studied leg segment from magnetic resonance imaging within the chronic nonbacterial osteomyelitis group.

mentary Figure 3, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23804/abstract). When comparing bone segments within the CNO group, the median, 95th percentile, and maximum temperatures from inflamed bone segments were higher than those from all uninflamed counterpart bone segments (in controls and patients with CNO) from anterior ($P < 0.05$, FDR < 0.05), medial ($P < 0.01$, FDR 0.02), and lateral ($P < 0.05$, FDR < 0.05) views. In the posterior view, only the maximum temperature was significantly higher in inflamed bone segments compared to the uninflamed counterparts ($P = 0.02$, FDR 0.03).

Inflamed distal tibia/fibula segments had significantly higher median and 95th percentile temperatures than uninflamed counterparts ($P < 0.01$ in all views). The mean difference between the 2 groups ranged from 0.7°C to 1.7°C, with the greatest difference noted in the medial view (Figure 2). Exploratory subanalysis showed no significant difference between inflamed and uninflamed distal tibia/fibula segments within the same individual (n = 7). Distal femur and proximal tibia/fibula temperatures did not differ between legs that were inflamed and those that were uninflamed from any view ($P > 0.05$).

DISCUSSION

Our study demonstrates potential for applying thermal imaging to identify inflamed bones within patients with CNO. There was a significant difference in the median, 95th percentile, and maximum temperatures between inflamed and uninflamed distal tibia/fibula within the CNO group in our study. The difference was most pronounced from the medial view, which corresponds to the location of the distal tibia likely due to less interference from overlying tissue. Subanalysis did not show temperature difference within the same subjects because of more limited sample size. Alternatively, there were no differences in temperature between inflamed extremities in the CNO group and healthy extremities in healthy controls. This was likely related to differences in patient characteristics between the CNO and healthy control groups. Body temperature was higher in healthy controls than in patients with CNO. Furthermore, patients with CNO were on average heavier with greater BMIs and the thickness of tissue overlying bones may reduce the heat emission for detection.

Our study also showed that standardization of imaging acquisition and analysis is crucial when evaluating thermal imaging as a diagnostic or disease-monitoring tool. For quantitative study of a specific extremity, clear definition of body position and view of the subject is required. Normal temperature ranges from healthy children may vary among different ages and sexes, so it will be important to perform a larger scale study to establish a reference temperature library of each segment for each age and sex. Furthermore, an internal reference within each patient may be identified to enhance the sensitivity of detection through repeated measures at different disease states.

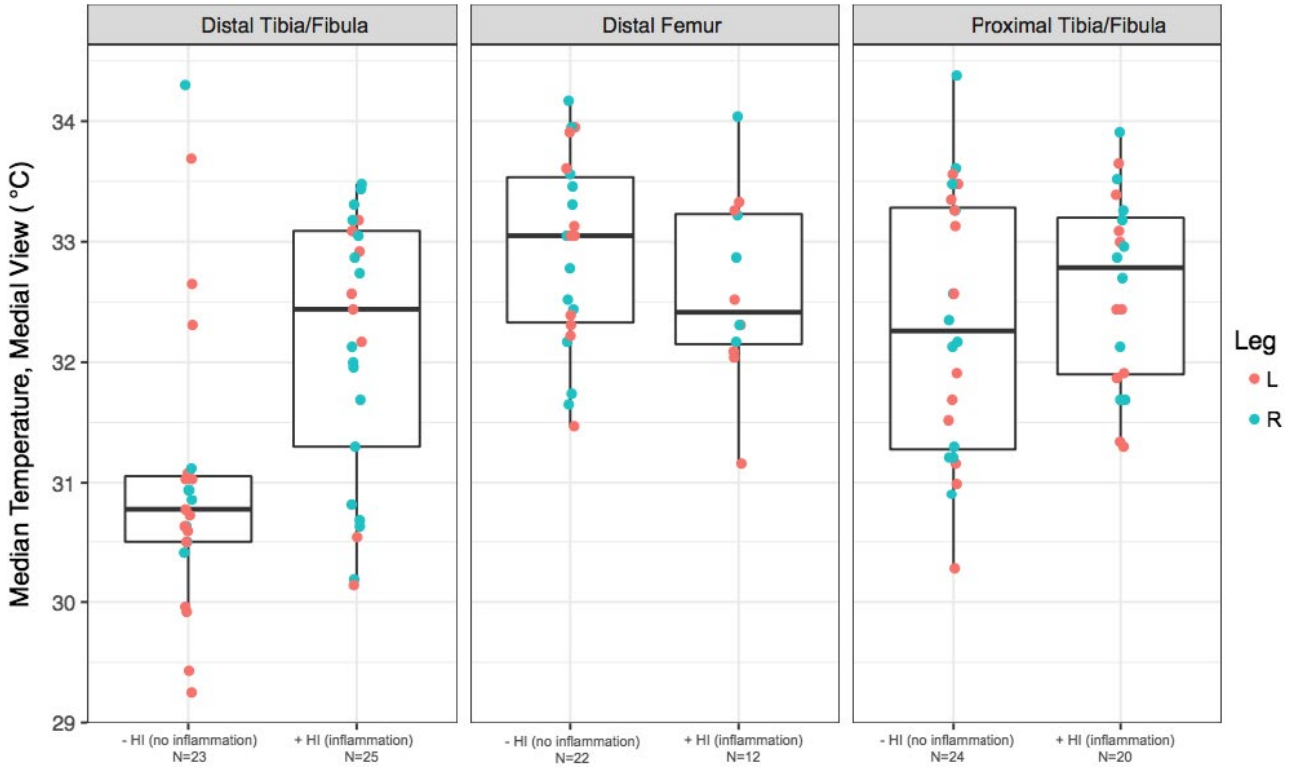

Figure 2. Mean difference between median and 95th percentile temperatures of patients with chronic nonbacterial osteomyelitis, medial view. The 95th percentile temperature increased in inflamed distal tibia/fibula compared to that from uninflamed counterparts. HI = hyperintensity.

Customized software allowed automation of image analysis to potentially reduce operator error and variation. Using easily identifiable body landmarks also ensured standardization of ROI selection for reproducible data analysis. Instead of using an arbitrary ROI drawn by the operator, the border of the ROI was defined based on sharp changes of temperature between the environment and body surface. The heat map can be useful to locate a hot spot within an area of interest for more focused analysis; however, this approach is subject to false-positive error and operator bias. Therefore, we instead used a less subjective segment-based ROI definition method. This was a proof-of-concept study and focused on lower extremities due to the high prevalence of CNO lesions within lower extremity long bones.

Only 70% of our CNO patient population had active lesions in their lower extremities based on MRI assessment, which decreased the power of our study. However, all patients with CNO were clinically suspected to have active lesions based on current symptoms and physical findings. This again demonstrates the poor predictability of actual disease status solely based on history and physical exam. This discrepancy underscores the importance of obtaining objective imaging studies to accurately assess disease activity. The majority of active CNO lesions in patients in this study were located in the distal femur, proximal and distal tibia, and distal fibula, as reported by others (1,4). Many lesions were graded as mild based on the affected size. It is unknown to what degree the size of the lesion from MRI may be directly associated with temperature.

Our study had several limitations. Our sample size was small due to the rarity of this disease. In particular, this limited our sensitivity to detect intrasubject differences between inflamed and uninflamed leg segments in patients with CNO, with only 7 patients meeting criteria for this subanalysis. Healthy controls were not strictly matched to the CNO group for age, sex, and BMI. The reproducibility of the temperature measurements over several days was not assessed due to difficulty in retaining subjects for repeat evaluations. Finally, not all MRI examinations included the entire femur and tibia.

Using infrared thermal imaging, children with active CNO lesions in the distal tibia/fibula exhibited higher regional temperatures versus healthy extremities in the same cohort. A larger and longitudinal study is needed to further evaluate this technique as a convenient, easy, and cost-effective tool to screen for patients suspected of having CNO and needing additional evaluation by MRI.

ACKNOWLEDGMENTS

The authors thank the participants of the study. They also thank Mr. Christopher Budech and Ms. Ching Hung for their help with infrared thermal imaging, Alena Beloborody for thermal imaging analysis, Dr. Antoinette Lindberg for her kind referrals of participants, and Dr. Jeffrey Ojemann for graciously allowing the study team to use the infrared thermal camera. Drs. Eric Allenspach and Matthew Basiaga critically reviewed the manuscript.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. Zhao 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. Zhao, Iyer, Reichley, Oron, Partridge, Wallace.

Acquisition of data. Zhao, Iyer, Reichley, Partridge, Wallace.

Analysis and interpretation of data. Zhao, Oron, Gove, Kitsch, Biswas, Friedman, Partridge.

REFERENCES

1. Jansson A, Renner ED, Ramser J, Mayer A, Haban M, Meindl A, et al. Classification of non-bacterial osteitis: retrospective study of clinical, immunological and genetic aspects in 89 patients. Rheumatology (Oxford) 2007;46:154–60.
2. Roderick MR, Shah R, Rogers V, Finn A, Ramanan AV. Chronic recurrent multifocal osteomyelitis (CRMO): advancing the diagnosis. Pediatr Rheumatol Online J. 2016;14:47.
3. Borzutzky A, Stern S, Reiff A, Zurakowski D, Steinberg EA, Dedeoglu F, et al. Pediatric chronic nonbacterial osteomyelitis. Pediatrics 2012;130:e1190–7.
4. Wipff J, Costantino F, Lemelle I, Pajot C, Duquesne A, Lorrot M, et al. A large national cohort of French patients with chronic recurrent multifocal osteitis. Arthritis Rheumatol 2015;67:1128–37.
5. Fritz J, Claussen CD, Carrino JA, Horger MS. Chronic recurrent multifocal osteomyelitis: comparison of whole-body MR imaging with radiography and correlation with clinical and laboratory data. Radiology 2009;252:842–51.
6. Ferguson PJ, Sandu M. Current understanding of the pathogenesis and management of chronic recurrent multifocal osteomyelitis. Curr Rheumatol Rep 2012;14:130–41.
7. Morbach H, Schneider P, Schwarz T, Hofmann C, Raab P, Neubauer H, et al. Comparison of magnetic resonance imaging and Technetium-labelled methylene diphosphonate bone scintigraphy in the initial assessment of chronic non-bacterial osteomyelitis of childhood and adolescents. Clin Exp Rheumatol 2012;30:578–82.
8. Khanna G, Sato TS, Ferguson P. Imaging of chronic recurrent multifocal. Radiographics 2009;29:1159–77.
9. Iyer RS, Thapa MM, Chew FS. Chronic recurrent multifocal osteomyelitis: review. AJR Am J Roentgenol 2011;196:S87–91.
10. Ring EF, Ammer K. Infrared thermal imaging in medicine. Physiol Meas 2012;33:R33–46.
11. Spalding SJ, Kwoh CK, Boudreau R, Enama J, Lunich J, Huber D, et al. Three-dimensional and thermal surface imaging produces reliable measures of joint shape and temperature: a potential tool for quantifying arthritis. Arthritis Res Ther 2008;10:R10.
12. Deng F, Tang Q, Zeng G, Wu H, Zhang N, Zhong N. Effectiveness of digital infrared thermal imaging in detecting lower extremity deep venous thrombosis. Med Phys 2015;42:2242–8.
13. Zhao Y, Chauvin NA, Jaramillo D, Burnham JM. Aggressive therapy reduces disease activity without skeletal damage progression in chronic nonbacterial osteomyelitis. J Rheumatol 2015;42:1245–51.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 71, No. 11, November 2019, pp 1436–1443
DOI 10.1002/acr.23783

Prospective Determination of the Incidence and Risk Factors of New-Onset Uveitis in Juvenile Idiopathic Arthritis: The Research in Arthritis in Canadian Children Emphasizing Outcomes Cohort

Jennifer J. Y. Lee,[1] Ciarán M. Duffy,[1] Jaime Guzman,[2] Kiem Oen,[3] Nick Barrowman,[1] Alan M. Rosenberg,[4] Natalie J. Shiff,[5] Gilles Boire,[6] Elizabeth Stringer,[7] Lynn Spiegel,[8] Kimberly A. Morishita,[2] Bianca Lang,[7] Deepti Reddy,[1] Adam M. Huber,[7] David A. Cabral,[2] Brian M. Feldman,[8] Rae S. M. Yeung,[8] Lori B. Tucker,[2] and Karen Watanabe Duffy,[1] for the ReACCh-Out Investigators

Objective. Identification of the incidence of juvenile idiopathic arthritis (JIA)–associated uveitis and its risk factors is essential to optimize early detection. Data from the Research in Arthritis in Canadian Children Emphasizing Outcomes inception cohort were used to estimate the annual incidence of new-onset uveitis following JIA diagnosis and to identify associated risk factors.

Methods. Data were reported every 6 months for 2 years, then yearly to 5 years. Incidence was determined by Kaplan-Meier estimators with time of JIA diagnosis as the reference point. Univariate log-rank analysis identified risk factors and Cox regression determined independent predictors.

Results. In total, 1,183 patients who enrolled within 6 months of JIA diagnosis met inclusion criteria, median age at diagnosis of 9.0 years (interquartile range [IQR] 3.8–12.9), median follow-up of 35.2 months (IQR 22.7–48.3). Of these patients, 87 developed uveitis after enrollment. The incidence of new-onset uveitis was 2.8% per year (95% confidence interval [95% CI] 2.0–3.5) in the first 5 years. The annual incidence decreased during follow-up but remained at 2.1% (95% CI 0–4.5) in the fifth year, although confidence intervals overlapped. Uveitis was associated with young age (<7 years) at JIA diagnosis (hazard ratio [HR] 8.29, $P < 0.001$), positive antinuclear antibody (ANA) test (HR 3.20, $P < 0.001$), oligoarthritis (HR 2.45, $P = 0.002$), polyarthritis rheumatoid factor negative (HR 1.65, $P = 0.002$), and female sex (HR 1.80, $P = 0.02$). In multivariable analysis, only young age at JIA diagnosis and ANA positivity were independent predictors of uveitis.

Conclusion. Vigilant uveitis screening should continue for at least 5 years after JIA diagnosis, and priority for screening should be placed on young age (<7 years) at JIA diagnosis and a positive ANA test.

INTRODUCTION

Juvenile idiopathic arthritis (JIA) is the most common pediatric rheumatic disease, with an estimated incidence of 11.9 per 100,000 person-years and a prevalence of 1–4 per 1,000 children (1–3). Uveitis is the most common extraarticular manifestation of JIA, and JIA-associated uveitis is the most common form of uveitis in childhood, with potential complications compromising eyesight (4–6). It is often asymptomatic, highlighting the importance of vigilant monitoring to ensure prompt detection.

Supported by the Canadian Institutes of Health Research (grant QNT 69301) and the Fast Foundation, Toronto, Canada.

[1]Jennifer J. Y. Lee, MD, Ciarán M. Duffy, MB, BCh, MSc, Nick Barrowman, PhD, Deepti Reddy, MSc, Karen Watanabe Duffy, MD: Children's Hospital of Eastern Ontario and University of Ottawa, Ottawa, Ontario, Canada; [2]Jaime Guzman, MD, MSc, Kimberly A. Morishita, MD, MHSc, David A. Cabral, MD, Lori B. Tucker, MD: British Columbia Children's Hospital and the University of British Columbia, Vancouver, British Columbia, Canada; [3]Kiem Oen, MD: University of Manitoba, Winnipeg, Manitoba, Canada; [4]Alan M. Rosenberg, MD: Royal University Hospital and University of Saskatchewan, Saskatoon, Saskatchewan, Canada; [5]Natalie J. Shiff, MD, MHSc: University of Florida Health Shands Children's Hospital and University of Florida, Gainesville; [6]Gilles Boire, MD, MSc: Centre Hospitalier Universitaire de Sherbrooke and Université de Sherbrooke, Sherbrooke, Quebec, Canada; [7]Elizabeth Stringer, MD, Bianca Lang, MD, Adam M. Huber, MD, MSc: IWK Health Centre and Dalhousie University, Halifax, Nova Scotia, Canada; [8]Lynn Spiegel, MD, Brian M. Feldman, MD, MSc, Rae S. M. Yeung, MD, PhD: Hospital for Sick Children and University of Toronto, Toronto, Ontario, Canada.

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

Address correspondence to Karen Watanabe Duffy, MD, Children's Hospital of Eastern Ontario, Division of Rheumatology, Department of Pediatrics, 401 Smyth Road, Ottawa, ON K1H 8L1, Canada. E-mail: kwatanabeduffy@cheo.on.ca.

Submitted for publication June 8, 2018; accepted in revised form October 9, 2018.

SIGNIFICANCE & INNOVATIONS

- In a large cohort of children with newly diagnosed juvenile idiopathic arthritis (JIA) followed prospectively for up to 5 years, the incidence of new-onset uveitis was 2.8% per year. Although there was a decrease in the annual incidence from the first to the fifth year after diagnosis, the confidence intervals overlapped and the annual incidence was still 2.1% in the fifth year.
- Independent predictors for JIA-related uveitis were young age (<7 years) at JIA diagnosis and a positive antinuclear antibody (ANA) test, while JIA subtype and female sex were not.
- Vigilant surveillance for uveitis should continue for at least the first 5 years after JIA diagnosis and priority for screening placed on children age <7 years at JIA diagnosis with a positive ANA test.

The frequency of uveitis in JIA cohorts has been most often reported in the literature as the prevalence of uveitis (proportion of cases at a given time) or cumulative incidence (rate of occurrence of new cases over the total period of study). These estimates are highly variable, ranging from 1.4% to 38%, with a recent study of a large cohort describing a point prevalence of 13% in 2002 and a period prevalence of 11.6% in 2013 (7–13). Differences in study design may largely account for the variability among reported estimates, with the majority of the studies being retrospective case series from tertiary pediatric rheumatology centers using different juvenile arthritis classification criteria and variable durations of follow-up (9). Furthermore, preliminary information from a study by Consolaro et al (14) reported geographic variability in uveitis prevalence rates.

Although previous studies have shown that uveitis develops most frequently early in the disease (8,15,16), there are very limited data about its precise risk during the course of JIA. The annual incidence (rate of occurrence of new cases in a given year) after JIA diagnosis is more informative about the risk of uveitis than frequency estimates over a reference period of time. The annual incidence of uveitis in JIA can only be determined by a systematic prospective study of newly diagnosed JIA patients.

Routine screening for JIA-associated uveitis is crucial to optimize good visual outcomes since most JIA patients who develop uveitis will have no ocular symptoms. Current guidelines for screening frequency are based on previously reported risk factors, including JIA subtype, younger age at JIA onset, antinuclear antibody (ANA) positivity, and shorter disease duration (8,17,18). Studies continue to investigate risk factors associated with the development of uveitis (16,18,19), and a large retrospective study by Calandra et al (20) challenged the importance of the JIA oligoarthritis subtype and female sex as independent risk factors. Higher inflammatory activity as reflected by an elevated erythrocyte sedimentation rate (ESR) at the time of JIA diagnosis also has been implicated as a predictor of uveitis, although after evaluation of a large national cohort, Heiligenhaus et al (8) concluded that ESR had no significant influence. In their multiethnic JIA cohort, Saurenmann et al (21) found no effect of ethnicity on the rate of occurrence of JIA-associated uveitis.

Both the assessment of risk factors and estimates of uveitis incidence over time from JIA diagnosis are necessary to formulate evidence-based screening recommendations. The Research in Arthritis in Canadian Children Emphasizing Outcomes (ReACCh-Out) cohort is a large prospective inception cohort with 1,497 newly diagnosed JIA patients followed for up to 5 years (22,23). It provides demographic, clinical, and laboratory data, including the presence of uveitis. We used this data set to estimate the annual incidence of new-onset uveitis following JIA diagnosis and the independent risk factors associated with its development.

PATIENTS AND METHODS

The methods of the ReACCh-Out study have been described previously (22,23). To summarize, consecutive newly diagnosed patients with JIA defined by the International League of Associations for Rheumatology (ILAR) criteria (24) were recruited from January 2005 to December 2010 at all 16 Canadian pediatric rheumatology referral centers. Study visits were scheduled every 6 months for the first 2 years and then annually to 5 years. Demographic, clinical, and medication data were collected at each study visit using a standardized data collection form. Laboratory data were collected as clinically indicated. ANAs as detected by immunofluorescence were considered positive at a titer ≥1:80. At every study visit, the rheumatologist was asked to report the presence of uveitis at any time since the last study visit, whether there was active uveitis at the time of the visit, the date of the last eye examination, and the type of uveitis presentation (asymptomatic [without ocular symptoms] or symptomatic). Data were not collected on the specific symptoms reported or severity of uveitis. At clinic visits occurring in between study visits, the rheumatologist was asked to report all current medications, including ophthalmic medications. The ILAR classification for JIA subtype was first assigned by the attending pediatric rheumatologist and then confirmed by 4 investigators based on submitted study data (CMD, KO, RSMY, LBT) (22,23).

In this study, we included ReACCh-Out patients with a diagnosis of JIA made within 6 months of enrollment who had ≥1 follow-up visit and documentation on the presence or absence of uveitis. Data for study visits completed up to June 30, 2011 were analyzed. The study was approved by the research ethics boards at each institution and conducted in conformity with the Declaration of Helsinki. Informed consent was obtained from the parent/guardian and assent from participants as appropriate.

Table 1. Demographic characteristics of the juvenile idiopathic arthritis (JIA) uveitis cohort*

Characteristic	Full cohort (n = 1,183)	Primary, excluding prevalent cases (n = 87)	Secondary, including prevalent cases (n = 100)†
Female, no. (%)	762 (64.4)	67 (77.0)	79 (79.0)
Age at JIA diagnosis, years	9.0 (3.8–12.9)	3.0 (2.0–5.1)	3.4 (2.1–5.5)
Diagnosis to enrollment duration, months	0.5 (0.0–1.8)	0.7 (0.0–2.1)	0.7 (0.0–2.1)
Follow-up since enrollment, months	35.2 (22.7–48.3)	42.1 (28.3–57.9)	41.2 (25.7–57.8)
JIA subtype distribution, no. (%)			
Oligoarthritis	469 (40)	49 (56)	55 (55)
Polyarthritis RF negative	239 (20)	18 (21)	22 (22)
Enthesitis-related	165 (14)	1 (1)	1 (1)
Undifferentiated	119 (10)	11 (13)	11 (11)
Systemic	73 (6)	2 (2)	3 (3)
Psoriatic	71 (6)	4 (5)	6 (6)
Polyarthritis RF positive	47 (4)	2 (2)	2 (2)
Potential risk factors, no. (%)‡			
ANA positive§	539 (49.8)	65 (76.5)	73 (76.0)
RF positive¶	82 (7.9)	5 (6.0)	5 (5.3)
HLA-B27 positive	126 (21.7)	4 (15.4)	6 (18.2)
Age <7 years at diagnosis	467 (39.5)	74 (85.1)	81 (81.0)
Medications, no. (%)#			
Methotrexate	584 (49.4)	64 (73.6)	73 (73.0)
Other DMARDs**	223 (20.6)	18 (20.7)	18 (18.0)
Biologics	116 (10.7)	12 (13.8)	15 (15.0)

* Values are the median (interquartile range) unless indicated otherwise. RF = rheumatoid factor; ANA = antinuclear antibody; DMARDs = disease-modifying antirheumatic drugs.
† Uveitis cohort including patients with uveitis at enrollment.
‡ Not all patients had autoantibody screening performed; denominator is based on the number of tests performed.
§ ANA was considered positive at a titer ≥1:80. ANA status was missing for 100 patients (8.4%).
¶ RF tested positive at least once, different from the diagnosis of polyarthritis–RF-positive JIA, which requires polyarthritis and 2 positive RF tests done at least 3 months apart.
# Includes all medications that patients received throughout the study.
** Other DMARDs included primarily cyclosporine, hydroxychloroquine, leflunomide, and sulfasalazine.

Statistical analysis. Analyses were performed with R software, version 3.4.3 (25,26). Descriptive statistics were used to describe the demographics of the patient population. Age at JIA diagnosis, the time from diagnosis to enrollment, and the study follow-up time were reported as median values with the interquartile range (IQR 25th-75th percentile). The date of new-onset uveitis was defined as the date of the first visit when the patient was reported to have developed uveitis or to have started ophthalmic corticosteroids. Interval censoring analyses using the parametric Weibull approach were used to confirm that there was no substantial bias as a result of this methodology (see supplementary material, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23783/abstract).

The proportion of patients who had not yet experienced uveitis at each time point was estimated using the Kaplan-Meier method. The reference point for time was the date of JIA diagnosis. Patients who were lost to follow-up before a first occurrence of uveitis were treated as observations censored at the time of the last follow-up. The annual incidence of new-onset uveitis for each of the first to fifth years after JIA diagnosis was computed as the negative logarithm of the Kaplan-Meier estimate (the cumulative hazard) at these time points. The incidence during the first 5 years after the diagnosis of JIA (cumulative incidence) was computed as 1 minus the Kaplan-Meier estimate. Patients with uveitis at enrollment were considered prevalent cases and excluded from incidence calculations in the primary analysis because the objective of this study was to determine the incidence of new-onset uveitis during follow-up. In a secondary analysis including all patients with uveitis, these patients were entered as having uveitis in the first year.

Potential risk factors for new-onset uveitis identified through a review of the literature (7–9,16,19) were examined using Kaplan-Meier analyses. These risk factors were female sex, a positive ANA test, age at diagnosis of JIA, JIA subtype, HLA-B27 status, rheumatoid factor (RF) status, and active joint count. To have sufficient patients in each category for the analyses, JIA subtypes were grouped into 3 categories: oligoarthritis (persistent and extended-course), polyarthritis RF negative, and other (remaining JIA subtypes). Age at diagnosis was evaluated as a dichotomous variable, age <7 or age ≥7 years. The threshold for age of 7 years was identified based on findings from Saurenmann et al (19), who reported a different uveitis risk in patients age <7 years at the time of JIA diagnosis, particularly in girls. Statistical analyses were performed to confirm that the age cutoff of 7 years was appropriate

Table 2. Incidence estimates for juvenile idiopathic arthritis (JIA)–associated uveitis*

Year after JIA diagnosis	Primary, excluding prevalent cases (n = 87)†	Secondary, including prevalent cases (n = 100)‡
Year 1	3.4 (2.3–4.5)	4.5 (3.3–5.8)
Year 2	3.0 (1.8–4.2)	3.0 (1.8–4.2)
Year 3	2.8 (1.3–4.2)	2.8 (1.3–4.2)
Year 4	2.5 (0.6–4.5)	2.5 (0.6–4.5)
Year 5	2.1 (0–4.5)	2.1 (0–4.5)
Years 1–5		
Average annual incidence	2.8 (2.0–3.5)	3.0 (2.2–3.7)
Cumulative 5-year incidence	12.9 (9.6–16.1)	13.9 (10.5–17.1)

* Values are the percentage (95% confidence interval). Estimates are derived from Kaplan-Meier survival analyses.
† Excludes 13 patients with uveitis at study enrollment.
‡ Includes 13 patients with uveitis at study enrollment, considered as having onset of uveitis during the first year.

and included Cox regression using a spline analysis (see supplementary material, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23783/abstract). Log-rank univariate analysis was used to assess the statistical significance of each risk factor.

The Cox proportional hazards regression model was then used to determine the independent contribution of each univariable risk factor. Since active joint count can vary from 1 visit to the next, this measurement was analyzed as a time-varying covariable (26). Risk factors with a *P* value less than 0.1 in the univariable analyses were included in the multivariable analysis. A *P* value of less than 0.05 was considered statistically significant in the multivariable analysis.

RESULTS

Patient demographics. The ReACCh-Out study recruited 1,497 patients with JIA between January 2005 and December 2010; 1,183 patients fulfilled inclusion criteria for this study and their characteristics are reported in Table 1. The median age at diagnosis was 9.0 years (IQR 3.8–12.9) with a median time from JIA diagnosis to enrollment of 0.5 (IQR 0–1.8) months, confirming that these patients were newly diagnosed with JIA. Patients were followed prospectively for a median of 35.2 months (IQR 22.7–48.3) after enrollment. A total of 467 patients (39.5%) of the total JIA cohort were age <7 years at the time of JIA diagnosis. A total of 539 of 1,083 patients (49.8%) with data were ANA positive, 82 of 1,042 patients (7.9%) were RF positive, and 126 of 581 patients (21.7%) were HLA-B27 positive.

Uveitis cohort. A total of 100 patients were identified as having developed uveitis, of whom 13 patients had uveitis at the time of study enrollment and 87 developed new-onset uveitis during follow-up. Of the 100 patients with uveitis, 54 (54%) had asymptomatic uveitis, 17 (17%) had symptomatic uveitis, and 29 (29%) had uveitis whose nature at presentation was not available. Characteristics of the uveitis cohort (including and excluding

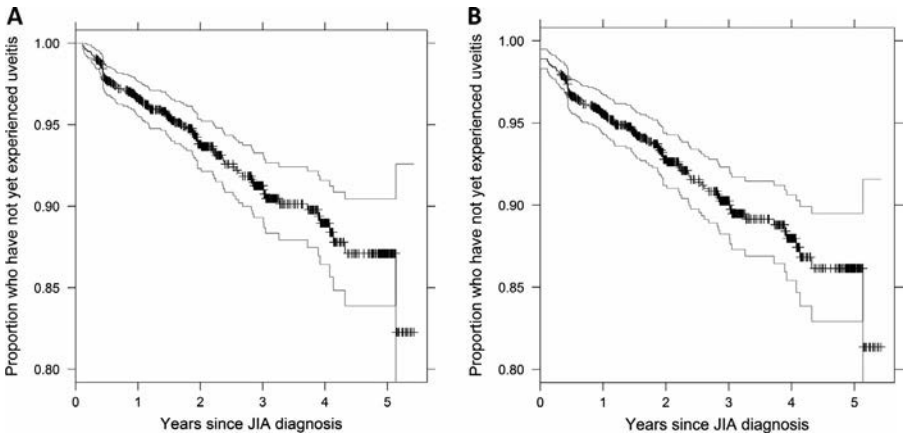

Figure 1. Kaplan-Meier analysis of time to new-onset uveitis in juvenile idiopathic arthritis (JIA) patients. Censored observations are indicated by vertical tick marks. Point-wise 95% confidence intervals are shown by curves above and below the Kaplan-Meier curve. Patients were enrolled at a median of 0.5 months (interquartile range 0–1.9) from diagnosis. **A**, Primary uveitis cohort (patients with uveitis present at enrollment were excluded from this analysis). **B**, Secondary uveitis cohort (13 patients with uveitis present at enrollment were included in this analysis and considered to have uveitis onset during the first year).

Table 3. Potential risk factors for new-onset uveitis in juvenile idiopathic arthritis (JIA) patients*

Potential risk factor	Unadjusted HR (95% CI)	*P*	Adjusted HR (95% CI)	*P*
Female (ref. male)	1.80 (1.09–2.96)	0.02	1.20 (0.72–1.99)	0.48
ANA positive (ref. negative)	3.20 (1.94–5.29)	<0.001	2.23 (1.32–3.78)	0.002
Age at diagnosis <7 years (ref. ≥7 years)	8.29 (4.60–14.96)	<0.001	6.57 (3.58–12.06)	<0.001
JIA subtype (ref. other)		0.00		0.5
Oligoarthritis	2.45 (1.45–4.12)	–	1.17 (0.68–2.02)	–
Polyarthritis RF negative	1.65 (0.87–3.12)	–	0.86 (0.44–1.69)	–
HLA-B27 positive (ref. negative)	0.68 (0.24–1.98)	0.46	–	–
RF positive (ref. negative)	0.75 (0.31–1.85)	0.51	–	–
Active joint count	0.98 (0.92–1.04)	0.42	–	–

* Unadjusted hazard ratios (HRs) were obtained by univariate log rank tests. Adjusted HRs were obtained by Cox proportional hazards regression entering variables with $P < 0.10$ in univariate tests. Prevalent uveitis cases at enrollment were excluded from these analyses. 95% CI = 95% confidence interval; ref. = reference; ANA = antinuclear antibody; RF = rheumatoid factor.

prevalent cases) are shown in Table 1. The majority of patients with uveitis had oligoarthritis (55%) or polyarthritis RF negative (22%). The next most frequent JIA subtype was undifferentiated arthritis (11%), 3% had systemic JIA, and the least common subtype was enthesitis-related arthritis (ERA) (1%). Most patients with uveitis were ANA positive (76%). While the median age at JIA diagnosis for the uveitis cohort was 3.4 years (IQR 2.1–5.5), the oldest patient was 18.4 years at uveitis onset. This patient had been diagnosed 2.8 years earlier with polyarthritis–RF-negative JIA and was ANA positive.

Incidence of new-onset uveitis. The estimated incidence rates are shown in Table 2. The incidence of new-onset uveitis during the first 5 years was 2.8% per year (95% confidence interval [95% CI] 2.0–3.5). The annual incidence was highest in the first year after JIA diagnosis at 3.4% (95% CI 2.3–4.5) and lowest in the fifth year at 2.1% (95% CI 0–4.5), although the 95% CIs overlapped. The cumulative incidence of new-onset uveitis during the first 5 years after JIA diagnosis was 12.9% (95% CI 9.6–16.1). In secondary analyses, which included the prevalent cases (13 patients with uveitis at the time of enrollment), the incidence for the first year increased from 3.4% to 4.5% (95% CI 3.3–5.8) and the overall incidence increased from 2.8% to 3.0% per year (95% CI 2.2–3.7), with a 5-year cumulative incidence of 13.9% (95% CI 10.5–17.1). Kaplan-Meier plots illustrate uveitis-free survival from JIA diagnosis to 5 years for the JIA cohort with and without patients with uveitis at enrollment (Figure 1).

Risk factors associated with uveitis development. The risk factors significantly associated with the development of uveitis in the univariable analysis were female sex (hazard ratio [HR] 1.80, $P = 0.02$), oligoarthritis subtype (HR 2.45, $P = 0.002$), and polyarthritis RF-negative subtype (HR 1.65, $P = 0.002$) compared to other subtypes, positive ANA (HR 3.20, $P < 0.001$), and young age (<7 years) at JIA diagnosis (HR 8.29, $P < 0.001$) (Table 3 and Figure 2). In the multivariable analysis, the independent predictors identified were age <7 years at JIA diagnosis (HR 6.57, $P < 0.001$) and a positive ANA (HR 2.23, $P = 0.002$) (Table 3).

DISCUSSION

This is one of the few large prospective studies to assess new-onset uveitis and its risk factors in a multicenter national inception cohort of JIA patients. To our knowledge, it is the first to estimate annual uveitis incidence in a cohort after JIA diagnosis. The annual incidence of new-onset uveitis was 3.4% in the first year after JIA diagnosis and 2.1% in the fifth year, with a cumulative incidence during 5 years of 12.9%. When the 13 patients with uveitis at the time of study enrollment are considered as

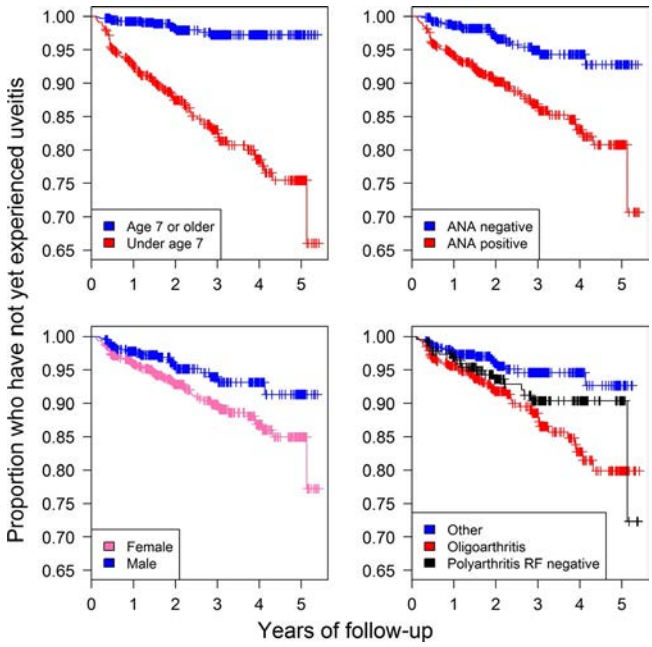

Figure 2. Kaplan-Meier analysis of potential risk factors for juvenile idiopathic arthritis–associated uveitis. Analyses excluded patients with uveitis at enrollment. Censored observations are indicated by vertical tick marks. ANA = antinuclear antibody; RF = rheumatoid factor.

cases of new-onset uveitis in the first year of follow-up, the incidence for the first year increases from 3.4% to 4.5%, and the cumulative incidence during 5 years becomes 13.9%.

Previous studies show that in 75–80% of cases of JIA-associated uveitis, uveitis develops within the first 2 to 3 years and in 90–93% within 4 years after the onset of arthritis (7,8,16). Our study provides a more precise indication of the magnitude of uveitis risk in patients with JIA for each of the first 5 years after JIA diagnosis. Although there was a decrease in incidence from 3.4% (95% CI 2.3–4.5) in the first year to 2.1% (95% CI 0–4.5) in the fifth year, confidence intervals overlapped. Fewer patients were followed up to the 5-year time point, resulting in less precise estimates for the fifth year. Due to the lack of a definitive decrease in annual incidence, our findings support continued rigorous surveillance for at least the first 5 years after JIA diagnosis.

As expected, new-onset uveitis was observed most frequently in the oligoarthritis and polyarthritis RF-negative JIA subtypes. We chose to combine persistent and extended-course oligoarthritis subtypes, a strategy supported by previous authors who have reported no difference in the occurrence of uveitis between these subtypes as well as the development of uveitis prior to extension (16,18). Only 1 of 171 patients with ERA developed uveitis. This finding is in contrast to other studies that have shown a larger proportion of ERA patients with uveitis (12,18). Since prolonged disease duration and HLA-B27 positivity are reported risk factors for acute uveitis in spondyloarthritis (27), the low frequency of uveitis observed in our study may be due to the earlier disease course of this cohort and to the relatively lower frequency of HLA-B27 in patients with ERA in our cohort (44%).

We identified uveitis in 3 systemic JIA patients, whose diagnosis of systemic JIA was confirmed by the primary rheumatologist. These patients were all ANA negative and developed uveitis within 3 years of diagnosis. None had documented complications of uveitis. Uveitis is uncommon in systemic JIA, and in contrast to our findings, others have reported that it is more likely to occur after the first 3 years of disease (28). Saurenmann et al (16) reported 1 case in 157 systemic JIA patients (0.6%), and Tappeiner et al (12) documented uveitis in 2% of 227 systemic JIA patients.

Our study confirmed risk factors associated with new-onset uveitis, including female sex, ANA positivity, young age (<7 years) at JIA diagnosis, and JIA subtype (16,19,29,30). Active joint count was not statistically significant in the univariable analysis. We identified young age (<7 years) at diagnosis of JIA and ANA positivity as independent predictors, while female sex and JIA subtype were not, confirming the findings in the retrospective analysis of 1,189 Italian children with JIA by Calandra et al (20). Our results suggest that the risk factors influencing JIA-associated uveitis are actually young age at JIA diagnosis and positive ANA test, which are typical of oligoarthritis, as opposed to the JIA subtype itself. The association of ANA positivity with uveitis development has been well described (16,18,19,29). In addition, Ravelli et al (30) proposed that ANA-positive patients constitute a more homogenous group, compared with classification into different JIA subtype categories. The threshold for age of 7 years at JIA diagnosis was selected based on previous studies (19), and we performed further statistical analyses with locally weighted scatterplot smoothing to confirm that the age cutoff of 7 years was appropriate.

There is increasing interest in the potential influence of treatment on the development of new-onset uveitis in JIA. From the national German database, Tappeiner et al (12) concluded that disease modifying antirheumatic drugs (DMARDs) have significantly reduced the risk for uveitis. However, they acknowledged their inability to identify in their cohort whether DMARDs were initiated for uveitis or whether uveitis developed following treatment. It was not within the scope of our study to evaluate the influence of treatment on the incidence of uveitis.

In the recently published European consensus-based recommendations for the management of JIA-associated uveitis, the authors concluded that evidence for previously identified risk factors is suboptimal (31). By confirming risk factors for JIA-associated uveitis, and by identifying these independent risk factors and the estimated annual incidence of new-onset uveitis after JIA diagnosis, our study provides data for the development of an evidence-based approach to the formulation of screening guidelines.

The greatest strengths of our study are in its prospective inception cohort design, which provided the systematic data collection necessary to calculate the incidence of new-onset uveitis and evaluate risk factors, using accurately assigned JIA subtype classifications and the designation of time of JIA diagnosis as the point of reference in the analyses. Our cohort provides valuable information about uveitis incidence during the first 5 years after JIA diagnosis, when patients are considered most at risk. However, we were limited by the decreased proportion of patients followed to 5 years, which may have resulted in less precise estimates in the fifth year and by the lack of data beyond 5 years.

Given that data were collected by protocol for study visits at 6 month intervals in the first 2 years and then annually to 5 years, uveitis diagnosed in the months before a study visit may not have been reported. However, those cases would have been captured by the report of ophthalmic medications at visits documented between study visits. Since the date of uveitis diagnosis was taken as the date of the visit when it was first reported, this method may have had a potential impact on the calculation of the annual incidence. However, we used interval-censoring analyses with the parametric Weibull approach to confirm that there was no substantial difference as a result of this methodology.

In this large prospective inception cohort, the incidence of new-onset uveitis in newly diagnosed JIA patients was 2.8% per year for the first 5 years after JIA diagnosis. The cumulative incidence during the first 5 years of disease was 12.9%. Young age

(<7 years) at JIA diagnosis and a positive ANA test were independent predictors of new-onset uveitis. During each of the 5 years of follow-up, there was continued development of new-onset uveitis without a definitive decrease in risk.

These findings support continued vigilant surveillance for uveitis for at least the first 5 years after JIA diagnosis and support the idea that priority for screening should be placed on young age (<7 years) at JIA diagnosis and a positive ANA test. Further studies are required to determine the ongoing annual incidence and risk of new-onset uveitis during the subsequent years of JIA follow-up and the impact of contemporary treatment on the development of uveitis.

ACKNOWLEDGMENTS

The authors thank all of the families for their generous participation and the following Canadian colleagues who are ReACCh-Out study investigators: Roxana Bolaria, Katherine Gross, Kristin Houghton, Ross E. Petty, Stuart E. Turvey (British Columbia Children's Hospital and University of British Columbia, Vancouver); Janet Ellsworth (Stollery Children's Hospital and University of Alberta, Edmonton); Susanne Benseler, Nicole Johnson, Paivi Miettunen, Heinrike Schmeling (Alberta Children's Hospital and University of Calgary, Calgary); Roberta A. Berard (Children's Hospital, London Health Science Centre and Western University, London); Maggie Larché (McMaster University, Hamilton); Bonnie Cameron, Ronald M. Laxer, Deborah M. Levy, Rayfel Schneider, Earl Silverman, Shirley M. L. Tse (Hospital for Sick Children and University of Toronto, Toronto); Michele Gibbon (national ReACCh-Out study coordinator), Roman Jurencak, Johannes Roth (Children's Hospital of Eastern Ontario and University of Ottawa, Ottawa); Anne-Laure Chetaille, Jean Dorval (Centre Hospitalier Universitaire de Laval and Université Laval, Quebec); Alessandra Bruns (Centre Hospitalier Universitaire de Sherbrooke and Université de Sherbrooke, Sherbrooke); Sarah Campillo, Gaëlle Chédeville, Claire LeBlanc, Rosie Scuccimarri (McGill University Health Centre and McGill University, Montreal); Debbie Feldman (Université de Montréal, Montréal); Elie Haddad, Claire St. Cyr (CHU Ste. Justine and Université de Montréal, Montréal); Suzanne E. Ramsey (IWK Health Centre and Dalhousie University, Halifax); and Paul Dancey (Janeway Children's Health and Rehabilitation Centre and Memorial University, Saint John's).

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. Watanabe Duffy 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. Lee, Duffy, Guzman, Oen, Watanabe Duffy.

Acquisition of data. Duffy, Guzman, Oen, Rosenberg, Stringer, Spiegel, Lang, Huber, Cabral, Feldman, Yeung, Tucker, Watanabe Duffy.

Analysis and interpretation of data. Lee, Duffy, Guzman, Oen, Barrowman, Rosenberg, Shiff, Boire, Stringer, Spiegel, Morishita, Lang, Reddy, Huber, Feldman, Tucker, Watanabe Duffy.

REFERENCES

1. Harrold LR, Salman C, Shoor S, Curtis JR, Asgari MM, Gelfand JM, et al. Incidence and prevalence of juvenile idiopathic arthritis among children in a managed care population, 1996–2009. J Rheumatol 2013;40:1218–25.
2. Prakken B, Albani S, Martini A. Juvenile idiopathic arthritis. Lancet 2011;377:2138–49.
3. Manners PJ, Bower C. Worldwide prevalence of juvenile arthritis: why does it vary so much? J Rheumatol 2002;29:1520–30.
4. Angeles-Han ST, Yeh S, Vogler LB. Updates on the risk markers and outcomes of severe juvenile idiopathic arthritis-associated uveitis. Int J Clin Rheumatol 2013;8.
5. Berk AT, Koçak N, Unsal E. Uveitis in juvenile arthritis. Ocul Immunol Inflamm 2001;9:243–51.
6. Angeles-Han ST, Griffin KW, Harrison MJ, Lehman TJ, Leong T, Robb RR, et al. Development of a vision-related quality of life instrument for children ages 8–18 years for use in juvenile idiopathic arthritis–associated uveitis. Arthritis Care Res (Hoboken) 2011;63:1254–61.
7. Grassi A, Corona F, Casellato A, Carnelli V, Bardare M. Prevalence and outcome of juvenile idiopathic arthritis-associated uveitis and relation to articular disease. J Rheumatol 2007;34:1139–45.
8. Heiligenhaus A, Niewerth M, Ganser G, Heinz C, Minden K. Prevalence and complications of uveitis in juvenile idiopathic arthritis in a population-based nation-wide study in Germany: suggested modification of the current screening guidelines. Rheumatology (Oxford) 2007;46:1015–9.
9. Carvounis PE, Herman DC, Cha S, Burke JP. Incidence and outcomes of uveitis in juvenile rheumatoid arthritis, a synthesis of the literature. Graefes Arch Clin Exp Ophthalmol 2006;244:281–90.
10. Kunnamo I, Kallio P, Pelkonen P. Incidence of arthritis in urban Finnish children: a prospective study. Arthritis Rheum 1986;29:1232–8.
11. Tappeiner C, Klotsche J, Schenck S, Niewerth M, Minden K, Heiligenhaus A. Temporal change in prevalence and complications of uveitis associated with juvenile idiopathic arthritis: data from a cross-sectional analysis of a prospective nationwide study. Clin Exp Rheumatol 2015;33:936–44.
12. Tappeiner C, Schenck S, Niewerth M, Heiligenhaus A, Minden K, Klotsche J. Impact of antiinflammatory treatment on the onset of uveitis in juvenile idiopathic arthritis: longitudinal analysis from a nationwide pediatric rheumatology database. Arthritis Care Res (Hoboken) 2016;68:46–54.
13. Oren B, Sehgal A, Simon JW, Lee J, Blocker RJ, Biglan AW, et al. The prevalence of uveitis in juvenile rheumatoid arthritis. J AAPOS 2001;5:2–4.
14. Consolaro A, Ruperto N, Filocamo G, Lanni S, Bracciolini G, Garrone M, et al, for the PRINTO. Seeking insights into the epidemiology, treatment and outcome of childhood arthritis through a multinational collaborative effort: introduction of the EPOCA study. Pediatr Rheumatol Online J 2012;10:39.
15. Angeles-Han ST, McCracken C, Yeh S, Jang SR, Jenkins K, Cope S, et al. HLA associations in a cohort of children with juvenile idiopathic arthritis with and without uveitis. Invest Ophthalmol Vis Sci 2015;56:6043–8.
16. Saurenmann RK, Levin AV, Feldman BM, Rose JB, Laxer RM, Schneider R, et al. Prevalence, risk factors, and outcome of uveitis in juvenile idiopathic arthritis: a long-term followup study. Arthritis Rheum 2007;56:647–57.

17. Cassidy J, Kivlin J, Lindsley C, Nocton J, Section on Rheumatology, Section on Ophthalmology. Ophthalmologic examinations in children with juvenile rheumatoid arthritis. Pediatrics 2006;117:1843–5.
18. Bolt IB, Cannizzaro E, Seger R, Saurenmann RK. Risk factors and longterm outcome of juvenile idiopathic arthritis-associated uveitis in Switzerland. J Rheumatol 2008;35:703–6.
19. Saurenmann RK, Levin AV, Feldman BM, Laxer RM, Schneider R, Silverman ED. Risk factors for development of uveitis differ between girls and boys with juvenile idiopathic arthritis. Arthritis Rheum 2010;62:1824–8.
20. Calandra S, Gallo MC, Consolaro A, Pistorio A, Lattanzi B, Bovis F, et al. Female sex and oligoarthritis category are not risk factors for uveitis in Italian children with juvenile idiopathic arthritis. J Rheumatol 2014;41:1416–25.
21. Saurenmann RK, Rose JB, Tyrrell P, Feldman BM, Laxer RM, Schneider R, et al. Epidemiology of juvenile idiopathic arthritis in a multiethnic cohort: ethnicity as a risk factor. Arthritis Rheum 2007;56:1974–84.
22. Guzman J, Oen K, Tucker LB, Huber AM, Shiff N, Boire G, et al, for the ReACCh-Out investigators. The outcomes of juvenile idiopathic arthritis in children managed with contemporary treatments: results from the ReACCh-Out cohort. Ann Rheum Dis 2015;74:1854–60.
23. Oen K, Tucker L, Huber AM, Miettunen P, Scuccimarri R, Campillo S, et al. Predictors of early inactive disease in a juvenile idiopathic arthritis cohort: results of a Canadian multicenter, prospective inception cohort study. Arthritis Rheum 2009;61:1077–86.
24. Petty RE, Southwood TR, Manners P, Baum J, Glass DN, Goldenberg J, et al, and International League of Associations for Rheumatology. International League of Associations for Rheumatology classification of juvenile idiopathic arthritis: second revision, Edmonton, 2001. J Rheumatol 2004;31:390–2.
25. Gentleman R, Ihaka R. The R project for statistical computing. 2017. URL: https://www.R-project.org/.
26. Therneau T, Lumley T. Survival: survival analysis version 2.44-1.1. 2019. URL: https://CRAN.R-project.org/package=survival.
27. Canouï-Poitrine F, Kemta Lekpa F, Farrenq V, Boissinot V, Hacquard-Bouder C, Comet D, et al. Prevalence and factors associated with uveitis in spondylarthritis patients in France: results from an observational survey. Arthritis Care Res (Hoboken) 2012;64:919–24.
28. BenEzra D, Cohen E, Behar-Cohen F. Uveitis and juvenile idiopathic arthritis: a cohort study. Clin Ophthalmol 2007;1:513–8.
29. Woreta F, Thorne JE, Jabs DA, Kedhar SR, Dunn JP. Risk factors for ocular complications and poor visual acuity at presentation among patients with uveitis associated with juvenile idiopathic arthritis. Am J Ophthalmol 2007;143:647–55.
30. Ravelli A, Felici E, Magni-Manzoni S, Pistorio A, Novarini C, Bozzola E, et al. Patients with antinuclear antibody–positive juvenile idiopathic arthritis constitute a homogeneous subgroup irrespective of the course of joint disease. Arthritis Rheum 2005;52:826–32.
31. Constantin T, Foeldvari I, Anton J, de Boer J, Czitrom-Guillaume S, Edelsten C, et al. Consensus-based recommendations for the management of uveitis associated with juvenile idiopathic arthritis: the SHARE initiative. Ann Rheum Dis 2018;77:1107–17.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 71, No. 11, November 2019, pp 1444–1449
DOI 10.1002/acr.23778

BRIEF REPORT

Relationship Between Esophageal Abnormalities on Fluoroscopic Esophagram and Pulmonary Function Testing in Juvenile Systemic Sclerosis

Lusine Ambartsumyan,[1] Hengqi B. Zheng,[1] Ramesh S. Iyer,[1] Jennifer Soares,[1] Gretchen Henstorf,[2] and Anne M. Stevens[3]

Objective. Juvenile systemic sclerosis (SSc) is a disabling autoimmune condition that affects multiple organs in addition to skin, notably the gastrointestinal and pulmonary systems. The relationship between esophageal abnormalities and pulmonary disease in juvenile SSc is not well understood. We describe associations between radiologic esophageal abnormalities and pulmonary function.

Methods. Clinical and radiographic data of children ages >18 years who fulfilled the 2007 Pediatric Rheumatology Provisional Classification Criteria for juvenile SSc between 1994 and 2016 were reviewed. Fluoroscopic upper gastrointestinal (UGI) studies, high-resolution computed tomography (HRCT), and pulmonary function tests (PFTs) within 12 months of presentation to Seattle Children's Hospital were extracted.

Results. Twenty-one children with juvenile SSc (67% female, ages 8–17 years) were studied. Esophageal abnormalities, defined as abnormal esophageal peristalsis and/or bolus clearance, were found in 12 patients. Abnormal esophagus on UGI tests was not associated with gastrointestinal or pulmonary symptoms, disease duration, use of medications (proton pump inhibitor or immunosuppressant), or specific autoantibodies. Compared with patients with a normal esophagus on UGI tests, children with an abnormal esophagus had decreased PFTs: mean forced expiratory volume in 1 second 96% versus 78% ($P = 0.03$), forced vital capacity 94% versus 76% ($P = 0.02$), and vital capacity 95% versus 76% ($P = 0.02$). Children with an abnormal esophagus on UGI tests had a larger mean esophageal diameter on HRCT (14.6 mm compared to 8.5 mm; $P < 0.01$).

Conclusion. There was an association between esophageal and pulmonary disease in children with juvenile SSc. Esophageal findings on UGI tests or HRCT, despite lack of symptoms, should raise concern for esophageal dysfunction and prompt heightened surveillance for concurrent lung disease.

INTRODUCTION

Juvenile systemic sclerosis (scleroderma; SSc) is a rare connective tissue disease with an estimated incidence of <1 per million children that carries significant morbidity and mortality (1). Progressive multisystem organ involvement can occur in juvenile SSc, including respiratory and gastrointestinal fibrosis leading to >10% mortality within the first 5 years of diagnosis (2).

Recently, the Childhood Arthritis and Rheumatology Research Alliance (CARRA) described pulmonary involvement in 34% of juvenile SSc patients enrolled in a registry (3). With the use of more sensitive high-resolution computed tomography (HRCT), the prevalence of interstitial lung disease (ILD) has been reported in up to 91% of children (4). Distinct findings on HRCT and a restrictive pattern on plethysmography pulmonary function tests (PFTs), as evidenced by decline in forced vital capacity (FVC) and diffusion capacity of the lung for carbon monoxide (DL_{CO}), are used to screen for ILD (5). However, despite vigorous screening and treatment, there is frequent development of severe ILD in patients with SSc.

The gastrointestinal tract is affected in 25–92% of children with SSc (2,3). The CARRA group recently reported gastrointestinal

[1]Lusine Ambartsumyan, MD, Hengqi B. Zheng, MD, Ramesh S. Iyer, MD, Jennifer Soares, MD: Seattle Children's Hospital, University of Washington, Seattle; [2]Gretchen Henstorf, MD: Seattle Children's Research Institute, Seattle, Washington; [3]Anne M. Stevens, MD, PhD: Seattle Children's Hospital, University of Washington, and Seattle Children's Research Institute, Seattle, Washington.

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

Address correspondence to Lusine Ambartsumyan, MD, Director, Gastrointestinal Motility Program, Division of Gastroenterology and Hepatology, Seattle Children's Hospital, 4800 Sand Point Way NE, Seattle, WA 98105. E-mail: lusine.ambartsumyan@seattlechildrens.org.

Submitted for publication May 9, 2018; accepted in revised form September 25, 2018.

SIGNIFICANCE & INNOVATIONS

- We describe the relationship between esophageal abnormalities on imaging and pulmonary function tests in children with juvenile systemic sclerosis (SSc).
- Abnormal esophageal peristalsis and/or bolus clearance on upper gastrointestinal (UGI) tests were associated with signs of decreased pulmonary function (decreased forced vital capacity, forced expiratory volume in 1 second, and vital capacity) and increased esophageal diameter on high-resolution computed tomography (HRCT).
- Abnormal esophageal findings were not associated with disease duration, frequency of symptoms, or autoantibodies.
- Neither gastrointestinal nor pulmonary symptoms predicted organ involvement as identified by UGI tests or HRCT. In juvenile SSc, symptoms alone should not guide screening for esophageal or pulmonary abnormalities.

involvement in 42% of their pediatric cohort marked by dysmotility (20%), gastroesophageal reflux (19%), dysphagia (17%), esophagitis (3%), and malabsorption (2%). Notably, children with gastrointestinal disease had worse quality of life, specifically, poor patient-reported quality of life and physician-reported functional disability (3). The esophagus is the most commonly affected gastrointestinal organ in patients with SSc. Esophageal abnormalities are associated with morbidity related to pulmonary disease (2,6).

The aim of this study was to describe the relationship between esophageal and pulmonary disease in juvenile SSc. We hypothesized that there is an association between esophageal findings on imaging (fluoroscopic upper gastrointestinal [UGI] studies and HRCT and abnormal spirometry and PFTs).

PATIENTS AND METHODS

Patient selection. The Seattle Children's Hospital Scleroderma Registry from March 1994 to February 2016 was used for this retrospective cross-sectional study with approval from the Seattle Children's Institutional Review Board. Patients ages <18 years who fulfilled the 2007 Pediatric Rheumatology Provisional Classification Criteria for juvenile SSc (7) were included in the study. Medical records were reviewed to obtain demographic, clinical, and laboratory data. UGI tests, HRCTs of the chest, and PFTs were obtained within 12 months of initial presentation to the clinic.

Clinical and laboratory data. Baseline demographics and clinical data were obtained from rheumatology clinic notes at the time of UGI tests. Disease duration was defined as the time between development of Raynaud's phenomenon or non-Raynaud's phenomenon symptoms attributed to juvenile SSc and the time of UGI tests. Pulmonary symptoms (cough, shortness of breath) and gastrointestinal symptoms (dysphagia, gastroesophageal reflux, weight loss) at the time of UGI tests were recorded. Laboratory data, including C-reactive protein (CRP) level, erythrocyte sedimentation rate (ESR), and hemoglobin at the time of UGI tests and autoantibody status at time of diagnosis (antinuclear, anti-Ro/SSA, anti-La/SSB, anti–double-stranded DNA, anti-Sm, anti–cyclic citrullinated peptide, anti–Scl-70, and anticentromere), were obtained. The use of proton pump inhibitor or immunosuppression drugs (corticosteroids, rituximab, methotrexate, mycophenolate, and cyclophosphamide) was assessed from the clinic visit within 1 month before or after the UGI tests.

Radiographic studies (UGI and HRCT). The UGI parameters that were evaluated included abnormalities in esophageal peristalsis, bolus clearance, and the presence of spontaneous or provoked gastroesophageal reflux. Abnormal esophagus on UGI tests was defined by abnormal peristalsis or abnormal bolus clearance. All HRCTs were reviewed by the study pediatric radiologist (RSI), who was blinded to the clinical status of the patient. The HRCT parameters evaluated included esophageal appearance (patulous, dilated, fluid-filled), esophageal diameter (maximum luminal diameter in mm), and the presence of lung disease. Lung disease was further characterized into interstitial pulmonary changes due to juvenile SSc and nonspecific interstitial changes that did not fit diagnostic criteria for classic pulmonary findings of juvenile SSc, which were classified as indeterminate.

Pulmonary function tests. PFTs (spirometry and plethysmography) performed within 12 months of the UGI tests were identified. The following PFT parameters were used for analysis: forced expiratory volume in 1 second percent predicted (FEV_1pp), forced vital capacity percent predicted (FVCpp), total lung capacity percent predicted (TLCpp), vital capacity percent predicted (VCpp), residual volume/total lung capacity (RV/TLC), residual volume percent predicted (RVpp), forced expiratory flow 25–75% percent predicted (FEFpp), forced expiratory volume at 1 second to forced expiration percent predicted (FEV_1/FVCpp), and diffusion capacity over alveolar volume adjusted for hemoglobin percent predicted (DLco/V_A adj pp). Results were expressed as percentages of normal predicted values for age, race, sex, height, and hemoglobin values for PFT parameters as recommended by the American Thoracic Society (8).

Statistical analysis. Data are reported as mean ± SEM. Numeric comparisons were made using Student's *t*-test or analysis of variance when indicated. Nonparametric testing was used when indicated. Proportions were compared using the chi-square test. *P* values less than 0.05 were considered significant. All analyses were performed using SPSS software, version 24.

RESULTS

Clinical characteristics. Twenty-one patients with juvenile SSc were included in the study (7 male, 14 female; mean age 12.8

Table 1. Clinical characteristics of patients at time of UGI tests*

Characteristic	Normal UGI test (n = 9)	Abnormal UGI test (n = 12)
Age, years	13.1 ± 1.5 (8.1, 17.8)	12.5 ± 2.3 (9.8, 16.1)
Female, no. (%)	7 (77.8)	7 (58.3)
Disease subtype, no. (%)		
Limited	8 (89)	4 (33)
Diffuse	1 (11)	8 (67)
Weight, kg	46.6 ± 6.0 (20.0, 79.9)	39.6 ± 2.0 (24.7, 53.8)
Height, cm	151.2 ± 6.0 (112.3, 173.0)	150.6 ± 3.0 (130.3, 162.0)
Body mass index, kg/m^2	19.7 ± 4.5 (15.7, 27.6)	17.5 ± 3.3 (13.6, 21.8)
Body mass index Z-score	0.1 ± 0.4 (−1.3, 1.9)	−0.6 ± 0.3 (−2.74, 0.94)
Disease duration, months of symptoms	37.7 ± 34 (7, 95)	15.5 ± 3.6 (1, 39)

* Values are the mean ± SE (range) unless indicated otherwise. There were no significant differences between the 2 groups (independent samples *t*-test, $P > 0.05$). UGI = upper gastrointestinal.

years; range 8.1–17.8 years). The main clinical characteristics are shown in Table 1. All patients had UGI tests, noncontrast HRCT of the chest, and PFTs completed within 12 months of the first clinical rheumatology evaluation. Of the 21 patients, 9 (2 male, 7 female; mean age 13.0 years) had a normal esophagus on UGI tests and 12 (5 male, 7 female; mean age 12.5 years) had an abnormal esophagus on UGI tests (peristalsis or bolus clearance). There were no significant differences in age, sex, or body mass index between the 2 groups ($P > 0.05$).

Thirteen of 21 children (62%) reported gastrointestinal symptoms at the time of UGI tests, including dysphagia (n = 4), gastroesophageal reflux (n = 6), and weight loss (n = 4) (Table 2). Symptoms did not predict esophageal findings on UGI tests: 5 of the 9 with a normal esophagus on UGI tests compared to 8 of the 12 with an abnormal esophagus reported gastrointestinal symptoms ($P = 0.60$).

Six of 21 children (29%) had pulmonary symptoms, including shortness of breath (n = 4) and/or cough (n = 3) (Table 2). Patients with an abnormal esophagus on UGI tests tended to have increased respiratory symptoms, although the association was not significant in this small cohort. More children with abnormal esophageal findings on UGI tests had pulmonary symptoms (5 of 12 versus 1 of 9; $P = 0.18$).

Disease duration at the time of UGI tests was longer in children with a normal esophagus, although the difference was not significant (37.7 ± 11.3 versus 15.5 ± 3.4 months; $P = 0.09$). There were no significant differences in laboratory values (hemoglobin, ESR, CRP level) or frequency of autoantibodies at the time of diagnosis between the 2 groups. Likewise, the use of a proton pump inhibitor, immunosuppression, or glucocorticoids at the time of UGI tests did not differ between groups (Table 2). A significant association was found between abnormal esophageal findings on UGI tests and subtype of disease: 89% of patients with diffuse juvenile SSc and 33% with limited juvenile SSc had abnormal esophageal findings on the UGI tests ($P = 0.02$).

Radiographic and pulmonary function characteristics. Three of the 12 patients with an abnormal esophagus on UGI tests had a dilated esophagus, and 1 had an esophageal stricture. Eleven patients (92%) had abnormal esophageal peristalsis and 10 (83.3%) were found to have abnormal esophageal bolus clearance. Pulmonary function decreased overall in children with an abnormal esophagus on UGI tests, including FEV_1pp (mean ± SEM 96 ± 9% versus 78 ± 5%; $P = 0.03$), FVCpp (mean ± SEM 94 ± 8% versus 76 ± 4%; $P = 0.02$), and VCpp (mean ± SEM 95 ± 8% versus 76 ± 4%; $P = 0.02$) (Figure 1). Although DL_{CO}/V_A adj pp declined in patients with an abnormal esophagus, the difference was not significant (mean ± SEM 89 ± 3% versus 81 ± 6%; $P = 0.21$). FEV_1/FVC was similar in children with a normal and an abnormal

Table 2. Comparison of laboratory values, medications, and symptoms between patients with normal and abnormal esophageal findings on UGI tests*

	Normal UGI test (n = 9)	Abnormal UGI test (n = 12)
Laboratory values		
CRP (normal <0.8 mg/liter)	0.7 ± 0.3 (0, 0.8)	0.8 ± 0.3 (0, 1.2)
ESR (normal <15 mm/hour)	7.3 ± 3.6 (0, 10)	9.5 ± 11.4 (1, 38)
Hemoglobin (gm/dl)	13.3 ± 1.5 (10.3, 15)	12.9 ± 1.5 (11, 15.9)
Proton pump inhibitor no. (%)	6 (66.7)	9 (75)
Immunosuppression no. (%)		
Any immunosuppression	4 (44.4)	6 (50)
Cyclophosphamide	0 (0)	1 (8)
MTX	1 (11.1)	2 (16.7)
MMF	0 (0)	1 (8)
Rituximab	0 (0)	3 (25)
Glucocorticoids	4 (44.4)	5 (41.7)
Symptoms at presentation, no. (%)		
Gastrointestinal	5 (55.6)	8 (66.7)
Dysphagia	3 (33.3)	1 (8)
Reflux	4 (44.4)	2 (16.7)
Weight loss	1 (11.1)	3 (25)
Pulmonary	1 (11.1)	5 (41.7)
Shortness of breath	1 (11.1)	3 (25)
Cough	0 (0)	3 (25)
Pulmonary hypertension	0 (0)	1 (8)

*Values are the mean ± SE (range) unless indicated otherwise. There were no significant differences between the 2 groups. Numerical comparisons were made using independent samples *t*-tests and proportions were compared using chi-square tests, $P > 0.05$. UGI = upper gastrointestinal; CRP = C-reactive protein; ESR = erythrocyte sedimentation rate; MTX = methotrexate; MMF = mycophenolate mofetil.

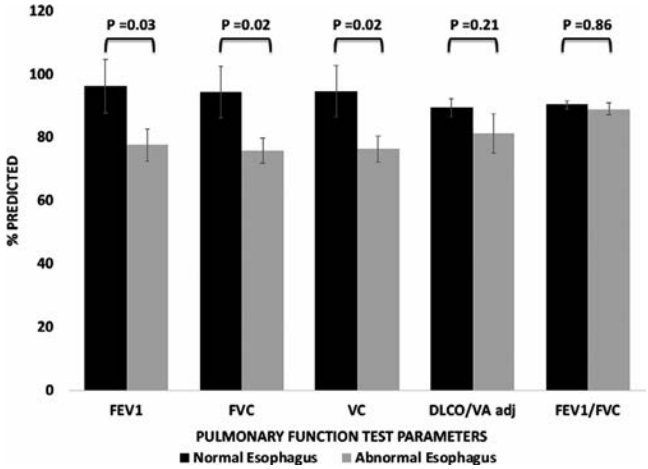

Figure 1. Pulmonary function in patients with normal versus abnormal esophageal findings on fluoroscopic upper gastrointestinal studies (mean ± SEM). Abnormal esophagus is defined by either abnormal bolus clearance or abnormal peristalsis by UGI tests. FEV_1 = forced expiratory volume in 1 second; FVC = forced vital capacity; VC = vital capacity; DLco/VA = diffusing capacity for carbon monoxide corrected for alveolar volume.

esophagus on UGI tests (mean ± SEM 90 ± 1 versus 89 ± 2; $P = 0.86$).

Spontaneous or provoked gastroesophageal reflux was evaluated on UGI tests in 19 patients and detected in 8. Reflux was noted in 5 of 10 children (50%) with abnormal esophageal findings compared to 3 of 9 (33.3%) with normal esophageal findings ($P > 0.05$). The presence of reflux on UGI tests was not associated with any of the mean PFT values ($P > 0.05$).

To assess the relationship between esophageal abnormalities on UGI tests and HRCT of the chest, the esophageal findings on HRCT were compared between children with a normal and an abnormal esophagus on UGI tests. Of the 21 HRCTs, 16 were available and reviewed by the study pediatric radiologist (RSI) to evaluate the esophageal diameter (mm), esophageal appearance, and lung findings of fibrosis/alveolitis. In 5 patients, the HRCT was performed at an outside institution and images were not available for review. Five of 16 children (31.3%) with an abnormal esophagus on UGI tests also had abnormal esophageal findings on HRCT, described as patulous, dilated, and/or fluid-filled. None of the children with abnormal esophageal findings on HRCT had a normal esophagus on UGI tests. Children with an abnormal esophagus on UGI tests had significantly larger esophageal diameters on HRCT (normal esophagus mean ± SD 8.5 ± 1.2 mm versus 14.6 ± 4.3 mm for abnormal esophagus; $P < 0.01$). In a subanalysis, abnormal bolus clearance on UGI tests was associated with abnormal esophageal findings ($P = 0.02$) and larger esophageal diameter on HRCT ($P < 0.01$).

We next assessed the relationship between esophageal abnormalities on UGI tests and ILD on HRCT. We found that 6 of 16 children (37.5%) had alveolitis or fibrosis on HRCT consistent with juvenile SSc, and 3 of 16 (18.8%) had abnormal lung findings on HRCT, characterized by a few scattered nodules 4 mm or less, but which were nonspecific findings and not necessarily suggestive for juvenile SSc. In our small cohort, we did not find an association between esophageal findings on UGI tests and lung findings on HRCT ($P = 0.8$).

DISCUSSION

To the best of our knowledge, this is the first study to directly show the relationship between the gastrointestinal and pulmonary systems in children with juvenile SSc. In the Seattle cohort presented here, abnormalities in esophageal peristalsis or bolus clearance visualized on fluoroscopic UGI tests were associated with signs of decreased pulmonary function. There were no differences in disease duration, frequency of gastroesophageal or pulmonary symptoms, or presence of autoantibodies between the children with normal and abnormal esophageal findings.

Esophageal abnormalities in ILD have been reported in adults with conflicting results. Marie et al reported significantly decreased DLco and increased frequency of ILD on HRCT in an adult cohort with severe esophageal motor disturbances on esophageal manometry compared to those with normal or moderate esophageal motor dysfunction (9). Kimmel et al (6) recently reported that patients with absent esophageal contractility on esophageal manometry had lower DLco and FVC on PFTs compared to those with normal esophageal manometry findings. In contrast, in a prospective study of 105 adults with SSc, Gilson et al (10) did not demonstrate an association between esophageal manometric findings and either FVC or DLco. We demonstrated a restrictive pattern of abnormalities on PFTs in children with juvenile SSc who had abnormal esophageal peristalsis or bolus clearance on UGI tests. We found a significant decrease in FVC, FEV_1, and VC but did not see any differences in DLco.

There are very limited data describing esophageal function in juvenile SSc. In a small cohort of 14 patients, Weber et al (11) demonstrated that 64% had an increased reflux index and 86% had an elevated number of reflux episodes on 24-hour intraesophageal pH monitoring. Symptoms of reflux were only present in 21% and thus not reliable to predict pathologic gastroesophageal reflux (11). In our cohort, 29% of children had symptoms of gastroesophageal reflux and 42% had spontaneous or provoked reflux on UGI tests. Our study did not reveal an association between pulmonary function and spontaneous reflux on UGI tests. This result may be secondary to insufficient assessment of gastroesophageal reflux when using UGI tests.

Symptoms did not predict organ involvement as identified by UGI tests or HRCT, in agreement with previous studies. In our patient cohort, 62% of patients had gastrointestinal symptoms. We did not find an association between pulmonary or gastrointestinal symptoms and abnormal esophageal findings on UGI tests. Weber et al (11) also reported that only 3 of the 14 patients in their pediatric cohort had symptoms of reflux and none had symptoms

of dysphagia, despite increased frequency and duration of reflux episodes on 24-hour intraesophageal pH monitoring. These findings indicate that symptoms alone, especially in the pediatric population, should not be used to guide further screening for esophageal or pulmonary abnormalities.

We demonstrated that children with an abnormal UGI test had larger esophageal diameters detected by HRCT. Richardson et al (12) reported that adult patients with SSc who had a larger esophageal diameter on HRCT had an increase in severity of imaging evidence of ILD (fibrosis, ground glass opacities) and a decrease in FVCpp and DLcopp. Farrokh et al (13) reported that up to 70% of patients were found to have esophageal dilation on HRCT and demonstrated an association between esophageal dilation and severe lung fibrosis on HRCT. The finding of wide esophageal diameter on routine HRCT and its relationship with worsening PFTs in SSc suggests that structural abnormalities of the esophagus on HRCT should not be overlooked; increased surveillance of pulmonary function may be warranted.

There are several limitations in our study. We conducted a retrospective study with a small sample size. Given the sample size, it was difficult to divide the patients into prepubertal- and pubertal-onset juvenile SSc, previously shown to have different outcomes (2). In addition, we used fluoroscopic UGI tests to evaluate abnormalities in esophageal function as depicted by peristalsis and/or bolus clearance. Although abnormalities of peristalsis may be seen on UGI tests, the gold standard and the standard of care for the evaluation of esophageal motility is esophageal manometry (14). There is no current literature standard on what constitutes a patulous esophagus on imaging or abnormal peristalsis on UGI tests, and therefore subjectivity exists in the radiology findings. In addition, the UGI test is not a recommended test to detect gastroesophageal reflux due to reduced sensitivity (15), and thus the true prevalence of reflux in our study population may have been underestimated. Another potential limitation is the ability to distinguish between SSc-related ILD (SSc-ILD) from another lung pathology on HRCT. We were only able to analyze 16 of 21 patients and included findings that were more typically associated with fibrosis, but these changes are not specific to SSc-ILD and may be seen in other ILDs.

In conclusion, this is the first study to show the relationship between gastrointestinal and pulmonary findings in children with juvenile SSc. Abnormal esophageal findings on UGI tests were associated with decreased FVC, FEV_1, and VC and increased esophageal diameter on HRCT. Abnormal esophageal findings were not associated with disease duration, frequency of symptoms, or the presence of autoantibodies. Thus, imaging studies remain the gold standard for screening and monitoring patients. We speculate that in children with juvenile SSc, abnormalities in esophageal function (abnormal peristalsis and bolus clearance) and compromised esophageal integrity (patulous or dilated esophagus) allow for stasis of esophageal contents, chronic microaspiration, and progression of ILD. Future prospective longitudinal studies are needed to further delineate the causative role of esophageal dysmotility and abnormal esophageal bolus clearance in development and progression of lung disease in children with juvenile SSc.

AUTHOR CONTRIBUTIONS

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

Acquisition of data. Ambartsumyan, Zheng, Iyer, Henstorf, Stevens.

Analysis and interpretation of data. Ambartsumyan, Zheng, Iyer, Stevens.

REFERENCES

1. Foeldvari I. Update on juvenile systemic sclerosis. Curr Rheumatol Rep 2015;17:1–5.
2. Martini G, Vittadello F, Kasapcopur O, Magni Manzoni S, Corona F, Duarte-Salazar C, et al. Factors affecting survival in juvenile systemic sclerosis. Rheumatology (Oxford) 2009;48:119–22.
3. Stevens BE, Torok KS, Li SC, Hershey N, Curran M, Higgins GC, et al. Clinical characteristics and factors associated with disability and impaired quality of life in children with juvenile systemic sclerosis: results from the Childhood Arthritis and Rheumatology Research Alliance Legacy Registry. Arthritis Care Res (Hoboken) 2018;70:1806–13.
4. Seely JM, Jones LT, Wallace C, Sherry D, Effmann EL. Systemic sclerosis: using high-resolution CT to detect lung disease in children. AJR Am J Roentgenol 1998;170:691–7.
5. Solomon JJ, Olson AL, Fischer A, Bull T, Brown KK, Raghu G. Scleroderma lung disease. Eur Respir Rev 2013;22:6–19.
6. Kimmel JN, Carlson DA, Hinchcliff M, Carns MA, Aren KA, Lee J, et al. The association between systemic sclerosis disease manifestations and esophageal high-resolution manometry parameters. Neurogastroenterol Motil 2016;28:1157–65.
7. Zulian F, Woo P, Athreya BH, Laxer RM, Medsger TA Jr, Lehman TJ, et al. The Pediatric Rheumatology European Society/American College of Rheumatology/European League against Rheumatism provisional classification criteria for juvenile systemic sclerosis. Arthritis Rheum 2007;57:203–12.
8. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al. Standardisation of spirometry. Eur Respir J 2005;26:319–38.
9. Marie I, Dominique S, Levesque H, Ducrotté P, Denis P, Hellot MF, et al. Esophageal involvement and pulmonary manifestations in systemic sclerosis. Arthritis Rheum 2001;45:346–54.
10. Gilson M, Zerkak D, Wipff J, Dusser D, Dinh-Xuan AT, Abitbol V, et al. Prognostic factors for lung function in systemic sclerosis: prospective study of 105 cases. Eur Respir J 2010;35:112–7.
11. Weber P, Ganser G, Frosch M, Roth J, Hulskamp G, Zimmer KP. Twenty-four hour intraesophageal pH monitoring in children and adolescents with scleroderma and mixed connective tissue disease. J Rheumatol 2000;27:2692–5.
12. Richardson C, Agrawal R, Lee J, Almagor O, Nelson R, Varga J, et al. Esophageal dilatation and interstitial lung disease in systemic sclerosis: a cross-sectional study. Semin Arthritis Rheum 2016;46:109–14.

13. Farrokh D, Abbasi B, Fallah-Rastegar Y, Mirfeizi Z. The extrapulmonary manifestations of systemic sclerosis on chest high resolution computed tomography. Tanaffos 2015;14:193–200.
14. Kahrilas PJ, Bredenoord AJ, Fox M, Gyawali CP, Roman S, Smout AJ, et al. The Chicago Classification of esophageal motility disorders, v3.0. Neurogastroenterol Motil 2015;27:160–74.
15. Vandenplas Y, Rudolph CD, Di Lorenzo C, Hassall E, Liptak G, Mazur L, et al. Pediatric gastroesophageal reflux clinical practice guidelines: joint recommendations of the North American Society for Pediatric Gastroenterology, Hepatology, and Nutrition (NASPGHAN) and the European Society for Pediatric Gastroenterology, Hepatology, and Nutrition (ESPGHAN). J Pediatr Gastroenterol Nutr 2009;49:498–547.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 71, No. 11, November 2019, pp 1450–1458
DOI 10.1002/acr.23782

Early Targeted Combination Treatment With Conventional Synthetic Disease-Modifying Antirheumatic Drugs and Long-Term Outcomes in Rheumatoid Arthritis: Ten-Year Follow-Up Results of a Randomized Clinical Trial

Vappu Rantalaiho,[1] Tia Sandström,[2] Juhani Koski,[3] Pekka Hannonen,[4] Timo Möttönen,[5] Oili Kaipiainen-Seppänen,[6] Timo Yli-Kerttula,[7] Markku J. Kauppi,[8] Toini Uutela,[9] Timo Malmi,[10] Heikki Julkunen,[11] Leena Laasonen,[11] Hannu Kautiainen,[12] and Marjatta Leirisalo-Repo,[2] for the NEO-RACo Study Group

Objective. The short-term outcomes of remission-targeted treatments of rheumatoid arthritis (RA) are well-established, but the long-term success of such strategies is speculative, as is the role of early add-on biologics. We assessed the 10-year outcomes of patients with early RA treated with initial remission-targeted triple combination of conventional synthetic disease-modifying antirheumatic drugs (csDMARDs), 7.5-mg prednisolone, and additional infliximab (IFX) or placebo infusions.

Methods. Ninety-nine patients with early, DMARD-naive RA were treated with a triple combination of csDMARDs and prednisolone and randomized to double-blind receipt of infusions of either IFX (the Finnish Rheumatoid Arthritis Combination Therapy Trial [FIN-RACo] + IFX) or placebo (FIN-RACo + placebo) during the first 6 months. After 2 years, the treatment strategies became unrestricted, but the treatment goal was strict remission in the TNF-Blocking Therapy in Combination With Disease-Modifying Antirheumatic Drugs in Early Rheumatoid Arthritis (NEO-RACo) study. At 10 years, the clinical and radiographic outcomes and the drug treatments used between 5 and 10 years were assessed.

Results. Ninety patients (91%) were followed after 2 years, 43 in the FIN-RACo + IFX and 47 in the FIN-RACo + placebo group. At 10 years, the respective proportions of patients in strict NEO-RACo remission and in Disease Activity Score using 28 joints remission in the FIN-RACo + IFX and FIN-RACo + placebo groups were 46% and 38% ($P = 0.46$) and 82% and 72% ($P = 0.29$), respectively. The mean total Sharp/van der Heijde score was 9.8 in the FIN-RACo + IFX and 7.3 in the FIN-RACo + placebo group ($P = 0.34$). During the 10-year follow-up, 26% of the FIN-RACo + IFX group and 30% of the FIN-RACo + placebo group had received biologics ($P = 0.74$).

Conclusion. In early RA, excellent results can be maintained up until 10 years in most patients treated with initial combination csDMARDs and remission-targeted strategy, regardless of initial IFX/placebo infusions.

ClinicalTrials.gov identifier: NCT00908089.

Dr. Rantalaiho's work was supported by the Competitive State Research Financing of the Expert Responsibility Area of Tampere University Hospital (grants 9R022 and 9T053) and Tampereen Reumayhdistys. Dr. Sandström's work was supported by Suomen Reumatologinen Yhdistys.

[1]Vappu Rantalaiho, MD, PhD: Tampere University and Tampere University Hospital, Tampere, Finland; [2]Tia Sandström, MD, Marjatta Leirisalo-Repo, MD: University of Helsinki and Helsinki University Hospital, Helsinki, Finland; [3]Juhani Koski, MD, PhD: Mikkeli Central Hospital, Mikkeli, Finland; [4]Pekka Hannonen, MD: Jyväskylä Central Hospital, Jyväskylä, Finland; [5]Timo Möttönen, MD: Turku University Central Hospital and University of Turku, Turku, Finland; [6]Oili Kaipiainen-Seppänen, MD, PhD: Kuopio University Hospital, Kuopio, Finland; [7]Timo Yli-Kerttula, MD, PhD: Rauma Regional Hospital, Rauma, Finland; [8]Markku J. Kauppi, MD: Päijät-Häme Central Hospital, Lahti, and Tampere University, Tampere, Finland; [9]Toini Uutela, MD, PhD: Lapland Central Hospital, Rovaniemi, Finland; [10]Timo Malmi, MD: Seinäjoki Central Hospital, Seinäjoki, Finland; [11]Heikki Julkunen, MD, PhD, Leena Laasonen, MD, PhD: Helsinki University Central Hospital, Helsinki, Finland; [12]Hannu Kautiainen, BA: Kuopio University Hospital, Kuopio, and Folkhälsan Research Center, Helsinki, Finland.

Dr. Rantalaiho has received research support from Pfizer and Celgene (less than $10,000 each). Dr. Sandström has received speaking fees, consulting fees, and/or honoraria from Bristol-Myers Squibb, UCB Pharma, and Roche (less than $10,000 each). Dr. Hannonen has received consulting fees, speaking fees, and/or honoraria from AbbVie, Bristol-Myers Squibb, MSD, Orion, Pfizer, and Roche (less than $10,000 each). Dr. Möttönen has received speaking fees, consulting fees, and/or honoraria from Pfizer, Roche, Amgen, and MSD (less than $10,000 each). Dr. Kaipiainen-Seppänen has received speaking fees, consulting fees, and/or honoraria from Boehringer Ingelheim (less than $10,000) and research support from Medac, MSD, AbbVie, Roche, and Pfizer. Dr. Yli-Kerttula has received speaking fees, consulting fees, and/or honoraria from AbbVie, Pfizer, UCB, MSD, Bristol-Myers Squibb, and Roche (less than $10,000 each). Dr. Malmi has received research support from Pfizer, Roche, and UCB. Dr. Leirisalo-Repo has received consulting fees from Pfizer, Eli Lilly, Boehringer Ingelheim, and Roche (less than $10,000 each) and research support from Schering-Plough. No other disclosures relevant to this article were reported.

Address correspondence to Vappu Rantalaiho, MD, PhD, Department of Internal Medicine, Centre for Rheumatic Diseases, Tampere University Hospital, P.O. Box 2000, Tampere FI-33521, Finland. E-mail: vappu.rantalaiho@uta.fi.

Submitted for publication March 13, 2018; accepted in revised form October 2, 2018.

SIGNIFICANCE & INNOVATIONS

- In a 10-year follow-up, a majority of rheumatoid arthritis patients remains in remission or in very low disease activity, with well-preserved functional ability and minimal radiographic progression when initially treated actively with a triple combination of conventional synthetic disease-modifying antirheumatic drugs (csDMARDs) and low-dose glucocorticoids.
- To maintain remission, one-third of the patients need continued combination csDMARD and low-dose glucocorticoid treatment and one-third need escalation to biologic DMARDs; in one-third of the patients the treatments can be tapered.

INTRODUCTION

Early and sustained remission is the current indisputable paradigm in the treatment of rheumatoid arthritis (RA) (1), and because of the modern treatment options, it has become reality to an increasing number of patients (2). However, because this chronic disease still cannot be cured, the answer to the question for how long the remission can be sustained, and by what means, remains unclear. There appears to be a very early window of opportunity, before any structural joint damage emerges, during which the initiation of treatment with disease-modifying antirheumatic drugs (DMARDs) results in an increased rate of remissions (3), but how long this early effect lasts is of interest. Further, because the definitions of remission vary, depending on their strictness, the pace of long-term structural damage progression as well as the functional capacity within each remission category that is reached may vary correspondingly (4).

There are few trials using the modern treat-to-target approach with truly long-term follow-ups (at least 10 years), or comprehensive follow-up coverage (5–7). Our previous analyses of the study TNF-Blocking Therapy in Combination With Disease-Modifying Antirheumatic Drugs in Early Rheumatoid Arthritis (NEO-RACo) have shown that in early RA, an intensified initial combination treatment strategy (Finnish Rheumatoid Arthritis Combination Therapy Trial [FIN-RACo]) with methotrexate (MTX), sulfasalazine, hydroxychloroquine, and low-dose 7.5-mg prednisolone for 2 years, and free, active, remission-targeted DMARD treatment thereafter, resulted in very low disease activity in most patients at 2 and 5 years. This treatment also resulted in minimal to no radiographic joint damage progression in most patients, regardless of double-blind induction therapy with infliximab (IFX) or placebo for the first 6 months (8,9). In the current study we report the 10-year outcomes of these patients.

PATIENTS AND METHODS

Study design and patients. The NEO-RACo trial was a multicenter, investigator-initiated study that recruited 99 patients with early, active RA fulfilling the American College of Rheumatology (ACR) 1987 criteria (10). The patients were treated with an intensified FIN-RACo regimen for 2 years, as previously described, and in

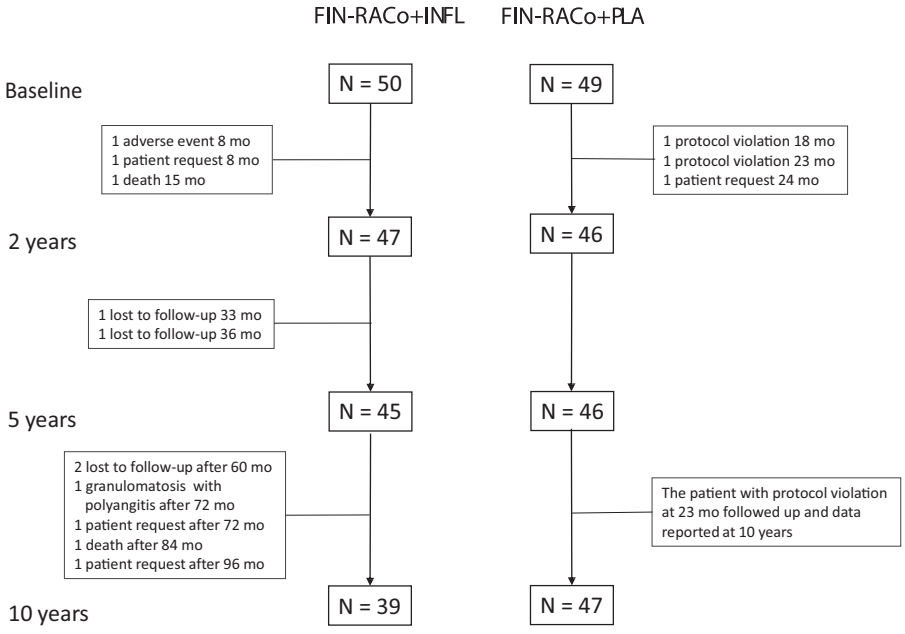

Figure 1. Flow-chart of the patients randomized to receive initial infliximab (the Finnish Rheumatoid Arthritis Combination Therapy Trial [FIN-RACo] + INFL) or placebo (FIN-RACo + PLA) for 6 months in addition to a combination of 3 conventional synthetic disease-modifying antirheumatic drugs and 7.5-mg prednisolone for 2 years and followed up for 10 years. After the 5-year visit, data were available for 43 patients in the FIN-RACo + INFL group, of which 4 patients dropped out by 10 years, and for 47 patients in the FIN-RACo + placebo group, all of which continued throughout the follow-up. mo = months.

addition were double-blind randomized to receive either IFX or placebo infusions at weeks 4, 6, 10, 18, and 26 (8). An active use of intraarticular glucocorticoid injections to all inflamed joints was part of the protocol throughout the follow-up. After the 2-year visit, if the patient was in remission by the strict NEO-RACo criteria (described below), prednisolone was gradually tapered off, followed by gradual reduction of conventional synthetic DMARDs (csDMARDs) as well. If remission was lost, the previous DMARD treatment/dosage was restored (9). If the patient was a nonresponder after dose and drug adjustments (less than a 50% improvement according to ACR criteria for improvement at maximal combination after individual substitutions) at 2 consecutive visits, the evaluation starting after week 26, the patient was regarded as failing treatment, and the therapy was open, including the possibility of using anti–tumor necrosis factor (anti-TNF) blocking agents (9).

After 5 years, study visits took place by protocol once a year, but clinical visits happened as often as needed. At all time points, the treatment was targeted to a strict NEO-RACo definition of remission, characterized as the presence of 5 of the 6 following criteria: morning stiffness <15 minutes, no fatigue, no joint pain, no tender joints (68 joint count), no swelling in joints (66 joint count) or tendons, and erythrocyte sedimentation rate (ESR) <30 mm/hour in women and <20 mm/hour in men. The therapies could be modified according to the judgment of the treating rheumatologist, with the use of all available csDMARDs, biologic DMARDs (bDMARDs), and glucocorticoids, orally as well as intraarticularly.

Outcomes and follow-up. The clinical assessments included evaluation of the number of swollen and tender joints (66 of 68 joints), patient's assessment of pain (10-cm visual analog scale [VAS]), patient's global assessment of disease activity (10-cm VAS), physician's global assessment of disease activity (10-cm VAS), patient's assessment of physical function according to the Health Assessment Questionnaire (HAQ), and acute-phase reactants (C-reactive protein level and ESR). The Disease Activity Score using 28 joints (DAS28) was also calculated. The medications used, the intraarticular glucocorticoid injections given, and the occurrence of adverse effects were carefully elucidated at each visit.

The small joints of the hands and feet were radiographed at 7 and 10 years and scored by an experienced radiologist (LL), who was aware of the chronology of the radiographs, according to the modified Sharp/van der Heijde (SHS) method. The primary outcome measures were the strict NEO-RACo remissions and the radiographic damage in hands and feet at 10 years. The secondary outcome measure was the DAS28 remission. In addition, we report the use of bDMARDs and adverse events.

Statistical analysis. Statistical comparisons between the groups were made using a *t*-test, bootstrap type *t*-test, Mann-Whitney test, chi-square test, or Fisher-Freeman-Halton exact test. The longitudinal remission data were analyzed with generalized estimating equations models with an unstructured correlation structure (binomial distribution with a log link). The bootstrap method (5,000 replications) was used when the theoretical distribution of the test statistics were unknown or in the case of violation of the assumptions (e.g., non-normality). The Kaplan-Meier method was used to estimate the cumulative use of bDMARDs and was compared between groups with the versatile weighted log-rank test. Clinical outcome variables were analyzed by the intent-to-treat principle, with the last observation carried forward. All analyses were performed using Stata software, version 14.1.

RESULTS

The flow chart of the patients is shown in Figure 1. One patient in the original FIN-RACo + placebo group was excluded from the 2- and the 5-year analyses due to a protocol viola-

Table 1. Demographic, clinical, and radiographic findings at baseline in patients randomized to receive initial infliximab (FIN-RACo + IFX) or initial placebo infusions (FIN-RACo + placebo) for 6 months in addition to a combination of 3 csDMARDs and 7.5-mg prednisolone for 2 years*

Finding	FIN-RACo + IFX (n = 43)	FIN-RACo + placebo (n = 47)	*P*
Demographic data at baseline			
Female, no. (%)	30 (70)	29 (62)	0.42
Age, years	48 ± 9	47 ± 11	0.32
Symptom duration, median (IQR) months	4 (2–6)	4 (2–6)	0.99
Rheumatoid factor present, no. (%)	33 (77)	34 (72)	0.63
Measures of disease activity at baseline			
Number of swollen joints (0–66)	15 ± 5	16 ± 8	0.38
Number of tender joints (0–68)	19 ± 10	21 ± 11	0.22
Erythrocyte sedimentation rate, mm/hour	34 ± 22	33 ± 22	0.93
Patient's global assessment (VAS, mm)	51 ± 24	48 ± 27	0.52
Pain (VAS, mm)	55 ± 27	53 ± 27	0.65
Physician's global assessment (VAS, mm)	49 ± 22	55 ± 20	0.17
DAS28	5.54 ± 1.00	5.60 ± 1.39	0.81
Physical function (HAQ)	1.09 ± 0.61	0.91 ± 0.71	0.22
Radiography at baseline			
Erosion score†	2.6 ± 7.2	1.3 ± 2.9	0.30
Narrowing score†	0.5 ± 1.6	0.3 ± 0.6	0.42
Total score†	3.1 ± 8.4	1.6 ± 3.2	0.29
Erosions in hand or foot radiographs, no. (%)	20 (47)	15 (32)	0.16

* Values are the mean ± SD unless indicated otherwise. FIN-RACo = Finnish Rheumatoid Arthritis Combination Therapy Trial; IFX = infliximab; csDMARDs = conventional synthetic disease-modifying antirheumatic drugs; IQR = interquartile range; VAS = visual analog scale; DAS28 = Disease Activity Score using 28 joints; HAQ = Health Assessment Questionnaire.
† Modified Sharp/van der Heijde method.

tion (bDMARD initiation despite an ACR response >50%) and subsequent treatment with a TNF inhibitor, but the patient was included in the 10-year analysis. One patient from the original FIN-RACo + placebo group withdrew consent at the 24-month visit and was included in the 2-year analysis but not after that. A slightly greater number of patients were lost from the original FIN-RACo + IFX group than from the FIN-RACo + placebo group during the 10-year follow-up period, but the baseline data of the dropouts were comparable to data from those patients who continued in the trial (data not shown). The baseline demographics and the measures of disease activity, function, and extent of structural joint damage at baseline are shown in Table 1.

The proportions of patients in NEO-RACo and in DAS28 remissions between 2 and 10 years are shown in Figures 2A and 2B. At 2 years, more patients in the FIN-RACo + IFX group had reached the very strict NEO-RACo remission, but after that, the differences leveled out. In addition, even though at 10 years a slightly higher proportion of patients in the FIN-RACo + IFX group reached the NEO-RACo remission, the difference was not statistically significant. Regarding the DAS28 remission, most of the patients in both groups reached this target throughout the follow-up (Figure 2B). The proportions of patients reaching various HAQ scores at 10 years are shown in Figure 2C. The HAQ score of 0 was reached by 66% of patients in the FIN-RACo + IFX group, and by 61% of patients

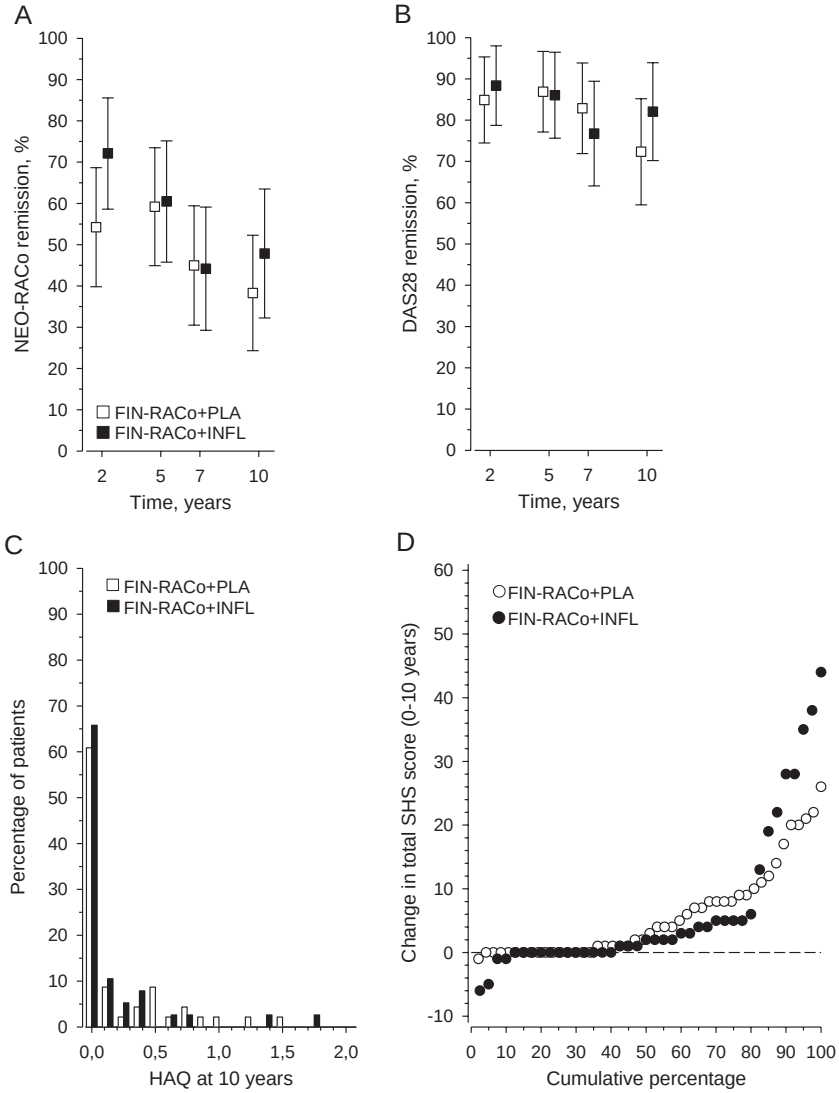

Figure 2. **A**, The proportions of patients in remission according to the TNF-Blocking Therapy in Combination With Disease-Modifying Antirheumatic Drugs in Early Rheumatoid Arthritis study; **B**, The proportions of patients in remission according to the Disease Activity Score using 28 joints between 2–10 years; **C**, The proportions of patients reaching various Health Assessment Questionnaire scores at 10 years; and **D**, Probability plot of radiographic progression from baseline to 10 years in patients randomized to receive initial infliximab (the Finnish Rheumatoid Arthritis Combination Therapy Trial [FIN-RACo] + INFL) or placebo (FIN-RACo + PLA) for 6 months in addition to a combination of 3 conventional synthetic disease-modifying antirheumatic drugs and 7.5-mg prednisolone for 2 years. SHS = Sharp/van der Heijde.

Table 2. csDMARDs, bDMARDs, and prednisolone use at the 5–10-year check-up visits in both original treatment groups participating in the NEO-RACo trial and initially treated with an intensified FIN-RACo regimen for 2 years, then double-blind randomized to receive either infliximab (FIN-RACo + IFX) or placebo (FIN-RACo + placebo) infusions for 6 months*

	5 years		6 years		7 years		8 years		9 years		10 years	
Medications	+ IFX (n = 43)	+ Placebo (n = 47)	+ IFX (n = 43)	+ Placebo (n = 46)	+ IFX (n = 43)	+ Placebo (n = 46)	+ IFX (n = 41)	+ Placebo (n = 47)	+ IFX (n = 39)	+ Placebo (n = 47)	+ IFX (n = 39)	+ Placebo (n = 47)
No DMARD	1 (2.3)	0 (0.0)	2 (4.7)	0 (0.0)	2 (4.7)	2 (4.4)	3 (7.3)	3 (6.4)	1 (2.6)	4 (8.5)	5 (12.8)	4 (8.5)
Single csDMARD	1 (2.3)	1 (2.1)	1 (2.3)	3 (6.5)	3 (7.0)	7 (15.2)	4 (9.8)	7 (14.9)	6 (15.4)	8 (17.0)	6 (15.4)	9 (19.2)
Combination of csDMARDs	24 (55.8)	28 (59.6)	20 (46.5)	24 (52.2)	16 (37.2)	17 (37.0)	13 (31.7)	15 (31.9)	12 (30.8)	9 (19.2)	10 (25.6)	7 (14.9)
Prednisolone alone	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (2.6)	1 (2.1)	1 (2.6)	0 (0.0)
Single or combination csDMARDs + prednisolone	17 (39.5)	13 (27.7)	17 (39.5)	14 (30.4)	17 (39.5)	16 (34.8)	13 (31.7)	15 (31.9)	11 (28.2)	15 (31.9)	10 (25.6)	18 (38.3)
Single or combination csDMARDs + bDMARD	0 (0.0)	1 (2.1)	0 (0.0)	1 (2.2)	1 (2.3)	0 (0.0)	2 (4.9)	0 (0.0)	2 (5.1)	1 (2.1)	2 (5.1)	1 (2.1)
Prednisolone + bDMARD	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (2.2)	0 (0.0)	1 (2.1)	1 (2.6)	1 (2.1)	1 (2.6)	2 (4.3)
Single or combination csDMARDs + prednisolone + bDMARD	0 (0.0)	4 (8.5)	3 (7.0)	4 (8.7)	4 (9.3)	3 (6.5)	6 (14.6)	6 (12.8)	5 (12.8)	8 (17.0)	4 (10.3)	6 (12.8)

* Values are the number (%). After the 2-year visit, if the patient was in strict remission, the medications could be tapered off, but reinstituted if remission was lost. No statistically significant differences between the groups in the frequencies of various treatment strategies at the check-up visits were found (Fisherman-Freeman-Halton test). csDMARDs = conventional synthetic disease-modifying antirheumatic drugs; bDMARDS = biologic DMARDS; NEO-RACo = TNF-Blocking Therapy in Combination With Disease-Modifying Antirheumatic Drugs in Early Rheumatoid Arthritis; FIN-RACo = Finnish Rheumatoid Arthritis Combination Therapy Trial; IFX = infliximab.

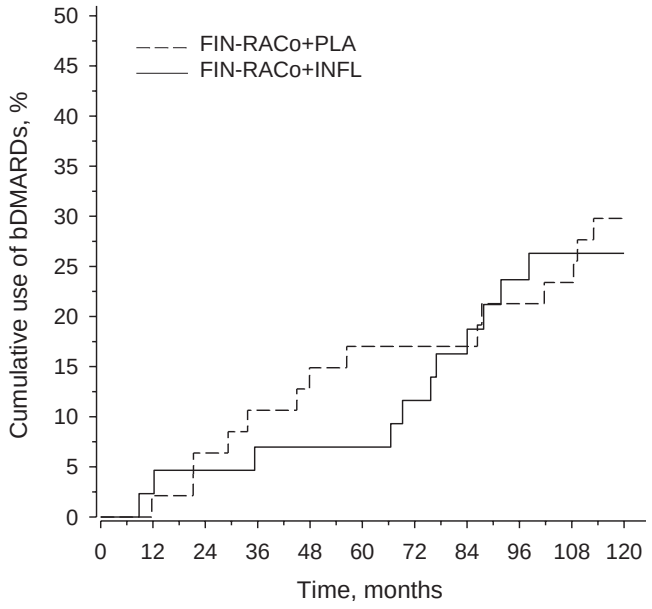

Figure 3. The cumulative use of biologic disease-modifying antirheumatic drugs (bDMARDs) in patients randomized to receive initial infliximab (the Finnish Rheumatoid Arthritis Combination Therapy Trial [FIN-RACo] + INFL) or placebo (FIN-RACo + PLA) for 6 months in addition to a combination of 3 conventional synthetic DMARDs and 7.5-mg prednisolone for 2 years and followed up for 10 years.

in the FIN-RACo + placebo group (P = 0.64). The mean ± SD HAQ score at 10 years was 0.17 ± 0.38 in the FIN-RACo + IFX group and 0.22 ± 0.37 in the FIN-RACo + placebo group (P = 0.59).

The details of radiographic damage scores at baseline are shown in Table 1 and the probability plot of radiographic progression is shown in Figure 2D. The radiographic joint damage progression remained slow in most of the patients up until 10 years, when the mean total SHS score was 9.8 in the FIN-RACo + IFX group and 7.3 in the FIN-RACo + placebo group (P = 0.34). The respective progression rates were 0.65 (95% confidence interval [95% CI] 0.31–1.1) and 0.58 (95% CI 0.39–0.79) units per year. Only 15% of all the patients had a total score higher than 20, and 20% had a total score of 0.

The DMARD and prednisolone treatments used by both patient groups after 5 years are shown in Table 2. There were no statistically significant differences between the groups in the treatment strategies throughout the follow-up. From 5 to 10 years, the use of combinations of csDMARDs was tapered down; the balance was shifted toward the use of single csDMARDs, and at 10 years, as many as 10.5% of the patients were using no DMARD. However, approximately one-third of the patients needed to use various combinations of csDMARDs with prednisolone throughout the follow-up. After 5 years, a total of 55.6% of the patients were at least sporadically using prednisolone. Among those patients using prednisolone for ≥1 period during the study, the mean ± SD daily dose of prednisolone during the study span was 1.8 ± 1.6 mg in the FIN-RACo + IFX group and 1.6 ± 1.4 mg in the FIN-RACo + placebo group (P = 0.65). After the 6-month blinded period, during the follow-up of 10 years, 26.3% (95% CI 15.5–42.5) of patients in the FIN-RACo + IFX group and 29.8% (95% CI 18.8–45.0) of patients in the FIN-RACo + placebo group had at some point been taking bDMARDs (P = 0.74) (Figure 3). After 6 months, the median (interquartile range) time using bDMARDs was 23 (2–63) months for patients in the FIN-RACo + placebo group and 11 (2–28) months for patients in the FIN-RACo + IFX group (P = 0.41). The number of bDMARDs used by the patients ranged between 1 and 3 in both groups; 1 bDMARD was sufficient for 50% of patients in the FIN-RACo + placebo group and for 58% of patients in the FIN-RACo + IFX group. At 10 years, 18.6% of all patients were currently using bDMARDs (Table 2).

Between 5 and 10 years, the occurrence of adverse events is shown in Table 3. There were 5 cases of malignancies (3 breast cancers, 1 metastatic adenomatous cancer, 1 unspecified malignancy) in the FIN-RACo + IFX group and none in the placebo group. Otherwise, the number of any adverse events, serious adverse events, or those adverse

Table 3. Adverse events (AEs) between 5 and 10 years in patients randomized to receive initial infliximab (FIN-RACo + IFX) or initial placebo infusions (FIN-RACo + placebo) for 6 months in addition to a combination of 3 csDMARDs and 7.5-mg prednisolone for 2 years*

AEs	FIN-RACo + IFX (n = 43)	FIN-RACo + placebo (n = 47)	P
Frequency of any AEs, no. (%)	34 (79)	29 (62)	0.073
Number of AEs/ patient	2.3 ± 1.8	2.2 ± 2.6	0.93
Frequency of moderate-serious AEs, no. (%)	28 (65)	26 (55)	0.34
No. of moderate-serious AEs/ patient	1.5 ± 1.6	1.4 ± 1.8	0.91
Malignancies, no. (%)	5 (12)	0 (0)	0.022
AEs leading to change of DMARD, no. (%)	19 (44)	15 (32)	0.23
No. of AEs leading to change of DMARD/patient	0.9 ± 1.2	0.6 ± 1.3	0.45
AEs related to DMARDs, no. (%)	18 (42)	20 (43)	0.95
No. of AEs related to DMARDs/ patient	0.8 ± 1.3	1.1 ± 1.8	0.42

* Values are the mean ± SD unless indicated otherwise. FIN-RACo = Finnish Rheumatoid Arthritis Combination Therapy Trial; IFX = infliximab; csDMARDs = conventional synthetic disease-modifying antirheumatic drugs.

events possibly related to the study medications did not differ between the groups.

DISCUSSION

This study shows that excellent clinical results achieved with early, remission-targeted treatment with a combination of csDMARDs and systemic (supplemented with intraarticular, if needed) glucocorticoid therapy in patients with recent-onset RA were sustained in most patients up until 10 years. At that time, approximately 40% of the patients had no RA symptoms, and 70% fulfilled the DAS28 criteria for remission, and the radiographic joint damage progression remained slow in the majority of patients. Furthermore, most of the patients preserved good functional capacity.

However, only 10% of the patients reached these goals without any DMARD at 10 years, and the majority needed active medications throughout the follow-up, with up to 25–30% of the patients in both groups requiring bDMARD treatment at some point in their disease course. This result is in accordance with real-life data (11) and agrees with the treatment protocol aiming at sustained remission. While the use of csDMARDs can be considered as self-evidently necessary, oral glucocorticoids raise contradictory opinions (12), and the use of bDMARDs is by no means straightforward and often confronts medical, social, and especially economic obstacles (13). Nevertheless, in different reports, approximately 50% of patients with established RA are currently treated with glucocorticoids, and depending on the patient population, 20–40% are treated with bDMARDs (14).

Earlier long-term studies have shown the course of RA with suboptimal treatment (15–17). As expected, the results of the current trial are far superior. To our knowledge, studies with an active, modern treat-to-target strategy and long follow-up times are sparse (5–7). The Dutch Behandel Strategieen (BeSt) trial compared 4 strategies guided by the DAS in 508 patients with early RA (7). In that trial, the mean HAQ score at 10 years was 0.57 and thus higher than in our study. Furthermore, 38% of the patients in the BeSt trial had dropped out from the 10-year follow-up, especially those with a higher baseline HAQ score. The remission rate evaluated by DAS (18) in the BeSt trial at 10 years was 53%, but the different definition of remission makes the comparison to our results difficult. In the BeSt trial, the drug-free remission was a treatment goal, unlike in our trial, and was reached by 14% of the patients participating at 10 years. Comparing the radiographic progression between these 2 trials is somewhat complicated, since more patients in the BeSt trial seem to have had erosive disease at baseline than in our trial. Furthermore, the duration of symptoms of the patients at entry in the BeSt trial was ≤2 years compared with ≤1 year in our study. Nevertheless, the total SHS score at 10 years was somewhat lower in the NEO-RACo patients than in the BeSt patients. Additionally, when comparing the probability plots showing the radiographic progression of each patient in these trials, the scale in the BeSt trial reaches up to 250 instead of 60 in our trial, and the highest outliers appear to have had considerably more progression than in the NEO-RACo trial. Comparison of medications used in these trials is basically impossible due to the heterogeneity of the strategies, and furthermore, approximately 20% of the patients in each group in the BeSt trial at 10 years were using medications outside the protocol. Still, the use of combination csDMARDs and low-dose prednisolone appeared to be more common in the NEO-RACo trial.

When comparing the NEO-RACo results to the long-term outcomes of the original FIN-RACo trial, the NEO-RACo remission rates in the current trial were surprisingly similar to the strict ACR remission rate in the original FIN-RACo combination therapy group at the 11-year visit (45–38% versus 38%, respectively). This result was despite the fact that only 11% of the FIN-RACo patients had been treated with bDMARDs (5). Comparing the radiographic joint damage progression between these 2 trials is complicated due to different methodologies (Larsen versus SHS score). However, evidently the more aggressive continuous treatment with higher doses of MTX and the earlier availability of bDMARDs in the NEO-RACo trial has led to less radiographic progression (mean 7.3–9.8 of a maximum of 448 with the SHS method) than noted in the FIN-RACo trial (mean 17 of a maximum 200 with the Larsen method) (6,19).

When comparing our results to real-life observational data, a Norwegian cross-sectional, observational study on RA patients with the disease duration of approximately 10 years showed that more recent cohorts had lower disease activity and better functional capacity than older ones (14). Nevertheless, compared to our patients, the percentage of patients in remission was lower, implying that the treatment in this real-life setting was not as efficient as in our trial, even though 26.0–34.9% of patients in all Norwegian year-cohorts had been taking bDMARDs.

Evidently, the main limitation of our study is the small size of the study population. The original population was calculated to have the power to demonstrate a 30% difference in the remission rates between the groups at 2 years. Smaller differences, therefore, may not be distinguished, especially at 10 years. Thus, this follow-up study functions best by showing the long-term evolution of this well-defined and actively treated population, regardless of the original randomization group, a strategy used even by larger randomized controlled trials with prolonged follow-ups. Another limitation of our study is that not all patients participated in all follow-up visits. However, the missing data were processed with the last observation carried forward method, and by the end of the trial only 13% of the patients were lost to follow-up, an excellent result considering the long follow-up period.

Even in the current treat-to-target era there appear to be different cultures of treating RA. One is based on the

fear of long-term overtreatment, having drug-free remissions as goals, even if those remissions turn out to be temporary, and then retreating the possible flares. The other strategy, employed also in this trial, tapers down the medications very conservatively and continues the csDMARD treatment even in patients in sustained remission if there are no adverse events. Further, a very strict sustained remission was required before any tapering of the DMARDs was allowed, making the feared overtreatment more likely, which would have made its potential harmful consequences visible in this trial. In spite of this possibility, there were no unexpected safety issues in all patients, and the rate of adverse events, especially serious adverse events, was not striking and was at least comparable to the data published from other long-term studies, mainly carried out on patients receiving biologic treatment (5,20,21). Nevertheless, there was a difference in the cancer incidence after 5 years between the groups. The incidences of lung cancer and lymphoma are known to be increased among RA patients, whereas for breast cancer there appears to be no increase in risk (22). Furthermore, there are several larger studies without signs of elevated risk of malignancies, even after or during long-term IFX treatment (23,24). Therefore, the finding of 5 malignancies in the NEO-RACo IFX group is somewhat unexpected, since the groups had received comparable treatments, including bDMARDs, after the initial double-blind randomized phase of IFX versus placebo infusions. Thus, the malignancies observed in our study population are unlikely to be related to the initial 6-month IFX treatment. Taken together, because the clinical outcomes remained very good, one could conclude that the earliest possible tapering of at least csDMARDs need not be a self-evident goal in treatment of RA.

Ample evidence has thus far shown that RA, as we diagnose it today, is an active and progressive disease requiring continuous and very often lifelong treatment. The current concept of a window of opportunity for early treatment allows us to start the medications before any structural joint damage has appeared. In a real-world setting, the prolonged combination csDMARD therapy has proven to be a cost-effective strategy to maintain remission in many patients (25). Our trial confirms the long-term efficacy of such a strategy in a well-defined follow-up.

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. Rantalaiho 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. Hannonen, Möttönen, Kaipiainen-Seppänen, Kauppi, Laasonen, Kautiaine, Leirisalo-Repo.

Acquisition of data. Rantalaiho, Koski, Hannonen, Möttönen, Kaipiainen-Seppänen, Yli-Kerttula, Kauppi, Uutela, Malmi, Julkunen, Laasonen, Kautiaine, Leirisalo-Repo.

Analysis and interpretation of data. Rantalaiho, Sandström, Laasonen, Kautiaine, Leirisalo-Repo.

REFERENCES

1. Smolen JS, Aletaha D, Bijlsma JW, Breedveld FC, Boumpas D, Burmester G, et al, for the T2T Expert Committee. Treating rheumatoid arthritis to target: recommendations of an international task force. Ann Rheum Dis 2010;69:631–7.
2. Rannio T, Asikainen J, Kokko A, Hannonen P, Sokka T. Early remission is a realistic target in a majority of patients with DMARD-naive rheumatoid arthritis. J Rheumatol 2016;43:699–706.
3. Van Nies JA, Krabben A, Schoones JW, Huizinga TW, Kloppenburg M, van der Helm-van Mil AH. What is the evidence for the presence of a therapeutic window of opportunity in rheumatoid arthritis? A systematic literature review. Ann Rheum Dis 2014;73:861–70.
4. Mäkinen H, Hannonen P, Sokka T. Definitions of remission for rheumatoid arthritis and review of selected clinical cohorts and randomised clinical trials for the rate of remission. Clin Exp Rheumatol 2006;24 Suppl 43:S22–8.
5. Rantalaiho V, Korpela M, Hannonen P, Kautiainen H, Järvenpää S, Leirisalo-Repo M, et al, for the FIN-RACo Trial Group. The good initial response to therapy with a combination of traditional disease-modifying antirheumatic drugs is sustained over time: the eleven-year results of the Finnish Rheumatoid Arthritis Combination Therapy trial. Arthritis Rheum 2009;60:1222–31.
6. Rantalaiho V, Korpela M, Laasonen L, Kautiainen H, Järvenpää S, Hannonen P, et al, and the FIN-RACo Trial Group. Early combination disease-modifying antirheumatic drug therapy and tight disease control improve long-term radiologic outcome in patients with early rheumatoid arthritis: the 11-year results of the Finnish Rheumatoid Arthritis Combination Therapy trial. Arthritis Res Ther 2010;12:R122.
7. Markusse IM, Akdemir G, Dirven L, Goekoop-Ruiterman YP, van Groenendael JH, Han KH, et al. Long-term outcomes of patients with recent-onset rheumatoid arthritis after 10 years of tight controlled treatment: a randomized trial. Ann Intern Med 2016;164:523–31.
8. Leirisalo-Repo M, Kautiainen H, Laasonen L, Korpela M, Kauppi MJ, Kaipiainen-Seppänen O, et al, for the NEO-RACo Study Group. Infliximab for 6 months added on combination therapy in early rheumatoid arthritis: 2-year results from an investigator-initiated, randomised, double-blind, placebo-controlled study (the NEO-RACo Study). Ann Rheum Dis 2013;72:851–7.
9. Rantalaiho V, Kautiainen H, Korpela M, Hannonen P, Kaipiainen-Seppänen O, Möttönen T, et al. Targeted treatment with a combination of traditional DMARDs produces excellent clinical and radiographic long-term outcomes in early rheumatoid arthritis regardless of initial infliximab: the 5-year follow-up results of a randomised clinical trial, the NEO-RACo trial. Ann Rheum Dis 2014;73:1954–61.
10. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315–24.
11. Schmajuk G, Solomon DH, Yazdany J. Patterns of disease-modifying antirheumatic drug use in rheumatoid arthritis patients after 2002: a systematic review. Arthritis Care Res (Hoboken) 2013;65:1927–35.
12. Buttgereit F, Bijlsma JW, Strehl C. Will we ever have better glucocorticoids? Clin Immunol 2018;186:64–6.
13. Yelin E, Tonner C, Kim SC, Katz JN, Ayanian JZ, Brookhart MA, et al. Sociodemographic, disease, health system, and contextual factors affecting the initiation of biologic agents in rheumatoid arthritis: a longitudinal study. Arthritis Care Res (Hoboken) 2014;66:980–9.

14. Haugeberg G, Hansen IJ, Soldal DM, Sokka T. Ten years of change in clinical disease status and treatment in rheumatoid arthritis: results based on standardized monitoring of patients in an ordinary outpatient clinic in southern Norway. Arthritis Res Ther 2015;17:219.
15. Lindqvist E, Jonsson K, Saxne T, Eberhardt K. Course of radiographic damage over 10 years in a cohort with early rheumatoid arthritis. Ann Rheum Dis 2003;62:611–6.
16. Courvoisier N, Dougados M, Cantagrel A, Goupille P, Meyer O, Sibilia J, et al. Prognostic factors of 10-year radiographic outcome in early rheumatoid arthritis: a prospective study. Arthritis Res Ther 2008;10:R106.
17. Svensson B, Andersson ML, Forslind K, Ajeganova S, Hafström I, on behalf of the BARFOT Study Group. Persistently active disease is common in patients with rheumatoid arthritis, particularly in women: a long-term inception cohort study. Scand J Rheumatol 2016;45:448–55.
18. Van der Heijde DM, van 't Hof M, van Riel PL, van de Putte LB. Development of a disease activity score based on judgment in clinical practice by rheumatologists. J Rheumatol 1993;20:579–81.
19. Levitsky A, Wick MC, Möttönen T, Leirisalo-Repo M, Laasonen L, Korpela M, et al. Early treatment intensification induces favourable radiographic outcomes according to predicted versus observed radiographic progression in early rheumatoid arthritis: a subanalysis of the randomised FIN-RACo and NEO-RACo trials. Clin Exp Rheumatol 2016;34:1065–71.
20. Keystone EC, Breedveld FC, van der Heijde D, Landewé R, Florentinus S, Arulmani U, et al. Longterm effect of delaying combination therapy with tumor necrosis factor inhibitor in patients with aggressive early rheumatoid arthritis: 10-year efficacy and safety of adalimumab from the randomized controlled PREMIER trial with open-label extension. J Rheumatol 2014;41:5–14.
21. Furst DE, Kavanaugh A, Florentinus S, Kupper H, Karunaratne M, Birbara CA. Final 10-year effectiveness and safety results from study DE020: adalimumab treatment in patients with rheumatoid arthritis and an inadequate response to standard therapy. Rheumatology (Oxford) 2015;54:2188–97.
22. Simon TA, Thompson A, Gandhi KK, Hochberg MC, Suissa S. Incidence of malignancy in adult patients with rheumatoid arthritis: a meta-analysis. Arthritis Res Ther 2015;17:212.
23. Lopez-Olivo MA, Tayar JH, Martinez-Lopez JA, Pollono EN, Cueto JP, Gonzales-Crespo MR, et al. Risk of malignancies in patients with rheumatoid arthritis treated with biologic therapy: a meta-analysis. JAMA 2012;308:898–908.
24. Wadström H, Frisell T, Askling J, for the Anti-Rheumatic Therapy in Sweden (ARTIS) Study Group. Malignant neoplasms in patients with rheumatoid arthritis treated with tumor necrosis factor inhibitors, tocilizumab, abatacept, or rituximab in clinical practice: a nationwide cohort study from Sweden. JAMA Intern Med 2017;177:1605–12.
25. Sokka T, Haugeberg G, Asikainen J, Widding Hansen IJ, Kokko A, Rannio T, et al. Similar clinical outcomes in rheumatoid arthritis with more versus less expensive treatment strategies: observational data from two rheumatology clinics. Clin Exp Rheumatol 2013;31:409–14.

Arthritis Care & Research
Vol. 71, No. 11, November 2019, pp 1459–1472
DOI 10.1002/acr.23785

American College of Rheumatology
Empowering Rheumatology Professionals

Correlation of the Multi-Biomarker Disease Activity Score With Rheumatoid Arthritis Disease Activity Measures: A Systematic Review and Meta-Analysis

Tate M. Johnson,[1] Kyle A. Register,[2] Cynthia M. Schmidt,[3] James R. O'Dell,[1] Ted R. Mikuls,[1] Kaleb Michaud,[4] and Bryant R. England[1]

Objective. There are conflicting reports on the validity of the multi-biomarker disease activity (MBDA) score for assessing rheumatoid arthritis (RA) disease activity. Our aim was to perform a systematic review of the MBDA and a meta-analysis of the correlation between the MBDA and other RA disease activity measures.

Methods. A systematic review was performed by searching Medline, Embase, Scopus, Google Scholar, and the Cochrane Library from inception to March 7, 2017. Study details, MBDA performance, and study quality were assessed by independent reviewers. Correlations of the MBDA with composite RA disease activity measures were pooled using random-effects meta-analyses.

Results. A total of 22 studies were identified in the systematic review, of which 8 (n = 3,242 assays) reported correlations of the MBDA with RA disease activity measures. Pooling results from these 8 studies in the meta-analysis, the MBDA demonstrated modest correlations with the Disease Activity Score in 28 joints using the C-reactive protein level (DAS28-CRP; r = 0.41, 95% confidence interval [95% CI] 0.36–0.46) and the Disease Activity Score using the erythrocyte sedimentation rate (DAS28-ESR; r = 0.48, 95% CI 0.38–0.58), with weaker correlations observed with the Simplified Disease Activity Index (SDAI; r = 0.35, 95% CI 0.26–0.43), Clinical Disease Activity Index (CDAI; r = 0.26, 95% CI 0.19–0.33), and Routine Assessment of Patient Index Data 3 (RAPID3; r = 0.23, 95% CI 0.19–0.27). Correlations between change in MBDA and change in disease activity measures ranged from r = 0.53 for the DAS28-ESR to r = 0.26 for the CDAI.

Conclusion. The MBDA demonstrates moderate convergent validity with the DAS28-CRP and the DAS28-ESR but weaker correlations with the SDAI, CDAI, and RAPID3. While it appears to complement existing RA disease activity measures, further assessment of the performance characteristics of the MBDA is warranted.

INTRODUCTION

Rheumatoid arthritis (RA) is an autoimmune inflammatory arthritis characterized by synovitis, progressive damage, functional disability, extraarticular manifestations, and premature mortality (1). The 2015 American College of Rheumatology (ACR) and 2016 European Union League Against Rheumatism (EULAR) treatment guidelines recommend early treatment with disease-modifying antirheumatic drug (DMARD) therapy in a treat-to-target strategy to obtain sustained remission or low disease activity (2,3). This approach has been shown to improve clinical outcomes, decrease cost, and limit radiographic progression (RP) (4). By definition, adhering to a treat-to-target strategy for RA management requires regular assessment of RA disease activity.

Numerous RA disease activity measures have been developed that include various patient-reported measures, provider assessments, and laboratory measurements of inflammation (5). Following a critical review of the literature (including psychometric properties,

Dr. Mikuls' work was supported by the NIH (National Institute of General Medical Sciences grant U54-GM-115458, National Institute on Alcohol Abuse and Alcoholism grant R25-AA-020818, and National Institute of Arthritis and Musculoskeletal and Skin Diseases grant 2P50-AR-60772). Dr. Michaud is recipient of a research grant from the Rheumatology Research Foundation. Dr. England's work was supported by the UNMC Physician-Scientist Training Program, the UNMC Mentored Scholars Program, and the UNMC Internal Medicine Scientist Development Award.

[1]Tate M. Johnson, MD, James R. O'Dell, MD, Ted R. Mikuls, MD, MSPH, Bryant R. England, MD: University of Nebraska Medical Center and Veterans Affairs Nebraska-Western Iowa Health Care System, Omaha; [2]Kyle A. Register, MD: University of Utah, Salt Lake City; [3]Cynthia M. Schmidt, MD, MLS: McGoogan Library of Medicine, University of Nebraska Medical Center, Omaha; [4]Kaleb Michaud, PhD: University of Nebraska Medical Center, Omaha, and FORWARD, The National Databank for Rheumatic Diseases, Wichita, Kansas.

Dr. O'Dell has received consulting fees from Medac (less than $10,000). Dr. Mikuls has received consulting fees from Pfizer (less than $10,000) and research support from Bristol-Myers Squibb, Horizon Pharma, and Ironwood Pharmaceuticals. Dr. Michaud has received research support from Pfizer. No other disclosures relevant to this article were reported.

Address correspondence to Bryant R. England, MD, 986270 Nebraska Medical Center, Omaha, NE 68198. E-mail: Bryant.england@unmc.edu.

Submitted for publication June 1, 2018; accepted in revised form October 9, 2018.

SIGNIFICANCE & INNOVATIONS

- Through a systematic review and meta-analysis, the multi-biomarker disease activity (MBDA) score demonstrated moderate convergent validity with the Disease Activity Score in 28 joints using the C-reactive protein level and the Disease Activity Score using the erythrocyte sedimentation rate but weaker correlations with other rheumatoid arthritis disease activity measures.
- Additional performance characteristics of the MBDA score, such as predicting radiographic changes, discriminating disease activity states, and predicting treatment response, are summarized through a systematic review.

feasibility, and cost), survey of practicing rheumatologists, and expert opinion, the ACR has recommended the use of the Disease Activity Score with 28-joint count (DAS28), the Simplified Disease Activity Index (SDAI), the Clinical Disease Activity Index (CDAI), the Routine Assessment of Patient Index Data 3 (RAPID3), the Patient Activity Scale (PAS), or PAS-II (6). All of these measures incorporate subjective assessments of disease activity by the patient or provider, which can be influenced by factors other than RA disease activity, such as noninflammatory pain (7). Laboratory measures such as erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP) level are objective but nonspecific, insensitive, and may only measure 1 domain of disease activity (8). Given the aforementioned limitations in RA disease activity assessment, there remains a need for the development of improved measures of RA disease activity.

The Multi-Biomarker Disease Activity (MBDA) score is a novel, commercially available blood test developed to assess RA disease activity (9). The algorithm used to calculate the MBDA was initially derived to predict simultaneously collected DAS28 using the C-reactive protein level (DAS28-CRP) scores. The algorithm combines 12 individual serum biomarkers involved in the pathogenesis of RA (interleukin-6, tumor necrosis factor receptor type I, vascular cell adhesion molecule 1, epidermal growth factor, vascular endothelial growth factor A, YKL-40, MMP-1 (matrix metalloproteinase 1), MMP-3, CRP, serum amyloid A, leptin, and resistin) to produce a disease activity score with values ranging from 0–100. As an objective measure of disease activity, the MBDA score may be useful in routine clinical practice or complement clinical disease activity assessment in the challenging comorbid patient. However, concerns have been raised regarding the validity of the MBDA in measuring RA disease activity (10,11).

Given the discrepant validity reported for the MBDA, the purpose of our study was to perform a systematic review of the MBDA score in RA and determine the convergent validity, or degree to which 2 measures that are measuring the same construct agree, of the MBDA with ACR-endorsed RA disease activity measures through a meta-analysis.

MATERIALS AND METHODS

We conducted a systematic review and meta-analysis following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (12) and registered the protocol with PROSPERO (ID: CRD42017060181), an international prospective registry of systematic reviews.

Search strategy. A full description of the systematic literature review search strategy is available in Supplementary Appendix 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23785/abstract. Briefly, led by a medical librarian (CMS), we searched Medline, Embase, Scopus, Google Scholar, and the Cochrane library from the inception of each database through March 7, 2017 using combinations of terms for the MBDA score and RA.

Study selection. For this systematic review, we included all published manuscripts in the English language reporting original observations relevant to MBDA performance in RA. To contribute to the meta-analysis, studies were required to report (or have the necessary data to report) a correlation between the MBDA and an ACR-endorsed RA disease activity measure (DAS28, SDAI, CDAI, PAS, PAS-II, or RAPID3). Although included in the initial protocol, study heterogeneity in evaluating RP (study population, cut-off values for RP, duration of study follow-up, modeling of the MBDA score, and statistical analyses utilized) precluded meaningful meta-analysis of RP. Two authors (KAR, BRE) separately reviewed titles, abstracts, and full text to determine eligibility for inclusion. There was perfect agreement between reviewers for study inclusion in both the systematic review and meta-analysis.

Data extraction. Two authors (TMJ, KAR) independently extracted study data in duplicate, including patient characteristics (age, sex, disease activity, and serologic status), study characteristics (study design, country of origin, inclusion/exclusion criteria, sample size, funding source, sample handling, and duration of follow-up), and study outcomes. Items extracted for the meta-analysis included correlation of the MBDA with composite RA disease activity measures and correlation of change in the MBDA with change in RA disease activity measures. Corresponding authors were contacted to provide missing data or to provide overall correlations for cohorts that were reported in multiple studies in order to avoid duplication. In instances where the corresponding authors could not provide the requested data, we requested data directly from study sponsors.

Quality assessment. Study quality was independently assessed by 2 authors (TMJ, KAR) using a 13-item assessment

tool adapted from tumor biomarker reporting guidelines (13,14). There was 96.6% agreement in the assessment of study quality items, with differences settled by third author (BRE) review. Quality scores were reported as a percentage of quality items fulfilled.

Statistical analysis. We calculated pooled correlation coefficients with 95% confidence intervals (95% CI) of the MBDA with RA disease activity measures by performing a random-effects meta-analysis using the DerSimonian-Laird model (15). Studies that assessed MBDA performance in multiple cohorts or time points (e.g., baseline and follow-up) were modeled separately in the meta-analysis. In sensitivity analysis, we transformed correlation coefficients to Fisher's Z scores prior to meta-analysis. Results were consistent with using untransformed correlation coefficients (data not shown). Stratified analyses were completed based on study time-point (baseline versus follow-up), with baseline analyses also restricting each study cohort to 1 contribution in the meta-analysis. Finally, we performed analyses stratified by study design (observational versus randomized controlled trial). Heterogeneity across studies was assessed using I^2. Publication bias was assessed using a funnel plot for correlation with the DAS28-CRP since this was the most frequently reported disease activity measure in our analyses. All analyses were performed using Stata statistical software, version 14.0.

RESULTS

Study selection. Our search strategy identified an initial 718 studies, with 470 remaining after excluding duplicates (Figure 1). Full-text review of 121 studies was completed with exclusion of those reported only in abstract form (n = 98) or without original data (n = 1), resulting in 22 articles included in the systematic review (9–11,16–34). Eight studies showed correlations with RA disease activity measures and were included in the meta-analyses (10,11,16,17,19,22,23,25).

Study characteristics from systematic review. Study and patient characteristics for all cohorts included in the systematic review are detailed in Table 1 and Table 2. Most studies (n = 22) identified were secondary analyses, with original investigations including both observational studies and randomized controlled trials. Patients in the systematic review were predominantly female (60–91%) in their fifth to sixth decade (mean/median age 51/61 years), seropositive (61–97% rheumatoid factor [RF]; 55–98% anti–cyclic citrullinated peptide antibody [anti-CCP]), and had moderate to high disease activity (DAS28 ranged from 3.2–6.0, except for one study of patients in remission). Patients also included in the meta-analysis were similar (67–84% female; age 51–61 years; 62–93% RF positive; 57–88% anti-CCP positive; DAS28 3.5–5.7). Sample sizes from individual studies ranged from 24 to 524 patients, with assessments of 94 to 558 serum samples.

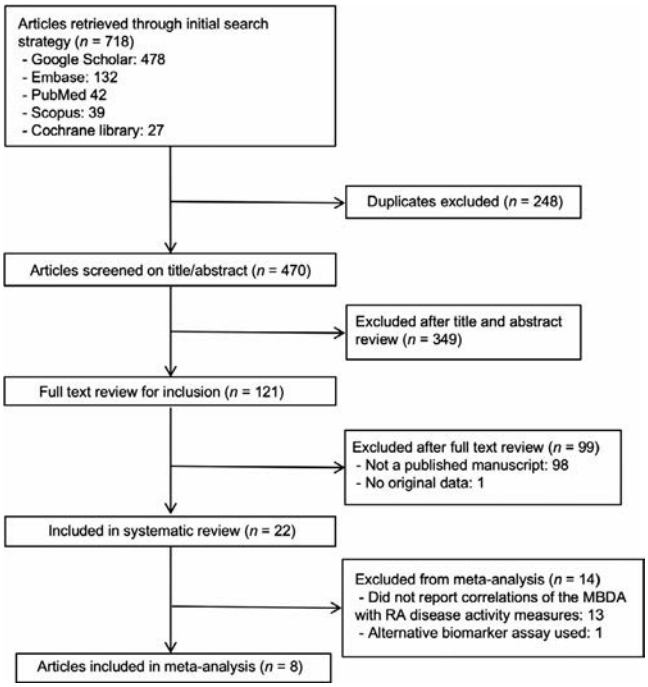

Figure 1. Flow diagram of study selection. Search strategy identified 718 articles with 470 remaining after removing duplicates. After title and abstract review, 121 full-text manuscripts were reviewed with 22 fulfilling criteria for inclusion into the systematic review and 8 further fulfilling criteria for inclusion in meta-analysis. MBDA = Multi-Biomarker Disease Activity; RA = rheumatoid arthritis.

Crescendo Bioscience was the most frequently noted funding source and provided funding or support (ranging from MBDA measurement to full study support) in 18 of 22 studies. Specific contributions of the study sponsors are listed in Table 1.

Quality assessment. After assessment of 13 quality items, all studies included in the systematic review fulfilled >75% of quality items (range 77–100%) (Table 1). Complete details of study quality assessment are provided in Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23785/abstract. The most common quality items missing were descriptions of sample handling (i.e., collection, preservation, and storage; n = 7), descriptions of confounding variables considered (n = 10), and assessments of biomarker performance in both univariate and multivariate analyses (n = 10).

Outcomes of studies identified in systematic review. Of the 22 studies, 8 demonstrated MBDA correlations with a composite RA disease activity measure as a primary or secondary outcome with 6 identifying significant positive correlations (Table 3) (16,17,19,22,23,25). The strength of correlations of MBDA score with RA disease activity measures are detailed in the meta-analysis section. The MBDA discriminated between low and moderate-high disease activity categories

Table 1. Characteristics of studies reporting on the Multi-Biomarker Disease Activity (MBDA) score in rheumatoid arthritis*

Study: author, year	Country	Study design	Sample handling; sample assay	Funding source	Study sponsor contributions	Quality score†
Systematic review/ meta analysis						
Bakker, 2012	Netherlands	POS	Standard separator tubes, frozen after collection, stored at −20° C until analysis; Meso Scale Discovery, ELISA	Crescendo Bioscience	Conception and design, data collection, analysis, manuscript preparation	100
Curtis, 2012	Multiple	POS	Serum separator tubes, processed at study site, shipped overnight using NanoCool shippers (2–8° C); Meso Scale Discovery	Crescendo Bioscience; Biogen Idec; NIH; Agency for Healthcare Research and Quality	Conception and design, data collection, analysis, manuscript preparation	100
Hirata, 2013	Netherlands	RCT	Serum separated, dispensed, and stored at −70° C; Meso Scale Discovery	Crescendo Bioscience	Study support	77
Hambardzumyan, 2015	Sweden	RCT	Not reported; Meso Scale Discovery	Crescendo Bioscience; Swedish Rheumatism Association; Stockholm County (ALF funds); Schering-Plough Sweden	MBDA score analysis at no cost	92
Hirata, 2015	Japan	ROS	Samples stored at −40° C after collection and −70° C after transport to site of analysis; Meso Scale Discovery	Crescendo Bioscience; Ministry of Health, Labor, and Welfare of Japan; Ministry of Education, Culture, Sports, Science, and Technology of Japan; University of Occupational and Environmental Health	Shipment of samples and MBDA analysis at no cost	85
Fleischmann, 2016	Multiple	RCT	Not reported	Bristol-Myers Squibb	Study design, manuscript review	77
Reiss, 2016	Multiple	RCT	Not reported; used reported Vectra DA algorithm	Genentech; F. Hoffmann-La Roche	Study design, data collection and analysis, manuscript preparation	77
Krabbe, 2017	Denmark	POS	Samples stored at −80° C; same reagents and immunoassay instruments as Vectra DA test	Not reported	Not reported	85
Systematic review only						
Eastman, 2012	US, Canada	POS	Aliquoted into single-use vials, stored at −80° C until thawed for assay; Meso Scale Discovery	Crescendo Bioscience	Conducted study	100
Centola, 2013	US, UK	POS, RCT	Serum separator tubes maintained at 2–8° C until frozen at −80° C, BRASS cohort shipped samples at ambient temperature prior to serum separation; Luminex-based assays, Meso Scale Discovery, ELISA	Crescendo Bioscience; Biogen Idec	Conception and design, data collection, analysis, manuscript preparation	100
Li, 2013	US	POS	Not applicable	Crescendo Bioscience	Conducted study	85

(Continued)

Table 1. *(Cont'd)*

Study: author, year	Country	Study design	Sample handling; sample assay	Funding source	Study sponsor contributions	Quality score†
Peabody, 2013	Germany	RCT	Not applicable	Crescendo Bioscience	Study design, manuscript preparation	91
van der Helm-van Mil, 2013	Netherlands	POS	Not reported; same reagents and immunoassay as Vectra DA	Netherlands Organization for Health Research and Development; Dutch Arthritis Association; Crescendo Bioscience	Conception and design, data collection, analysis, manuscript preparation	77
Markusse, 2014	Netherlands	RCT	Samples stored at −70° C; Meso Scale Discovery	Dutch insurance companies; Schering Plough B.V.; Janssen B.V.	Study support	100
Michaud, 2015	US	ROS	Not applicable	Crescendo Bioscience	Study support	100
van Vollenhoven, 2015	Sweden	RCT	De-identified, frozen serum samples; not reported	Crescendo Bioscience; Schering-Plough Sweden	Editorial, graphic, and statistical support; data analysis, manuscript preparation	77
Hambardzumyan, 2016	Sweden	RCT	Not reported; Meso Scale Discovery	Crescendo Bioscience; Swedish Rheumatism Association; Stockholm County (ALF funds); Schering-Plough Sweden	MBDA score analysis at no cost	85
Hirata, 2016	Japan	POS	Samples stored at −40° C until transfer to Crescendo Bioscience, then stored at −70° C; Meso Scale Discovery	Crescendo Bioscience; Research Grant-In Aid for Scientific Research by the Ministry of Health, Labor and Welfare of Japan; Ministry of Education, Culture, Sports, Science and Technology of Japan; University of Occupational and Environmental Health, Japan.	Biomarker analysis and statistical support	92
Lee, 2016	US	POS	De-identified, frozen, serum samples; Meso Scale Discovery	Crescendo Bioscience	Generation of biomarker data, statistical analysis, manuscript formatting	100
Li, 2016	Netherlands	POS	De-identified, frozen, serum samples; Meso Scale Discovery	Crescendo Bioscience	Funded sample handling, generation of biomarker data and statistical analysis	100
Rech, 2016	Germany	RCT	Serum stored at −80° C; Meso Scale Discovery	Multiple‡	MBDA score analysis at no cost	100
Hambardzumyan, 2017	Sweden	ROS	Not reported; Meso Scale Discovery	Crescendo Bioscience; Swedish Rheumatism Association; Stockholm County (ALF funds); Schering-Plough Sweden	MBDA score analysis at no cost	77

* POS = prospective observational study; ELISA = enzyme-linked immunosorbent assay; RCT = randomized controlled trial; ROS = retrospective observational study; DA = disease activity; BRASS = Brigham and Women's Rheumatoid Arthritis Sequential Study.
† Quality score = percentage of 13 quality items reported.
‡ Supported by Crescendo Bioscience, the Deutsche Forschungsgemeinschaft (SPP1468-IMMUNOBONE), the Bundesministerium für Bildung und Forschung (BMBF; project METARTHROS), the Marie Curie project OSTEOIMMUNE, the TEAM and MASTERSWITCH projects of the European Union and the Innovative Medicines Initiative funded project BTCure.

Table 2. Baseline characteristics of patients in studies reporting on the Multi-Biomarker Disease Activity score in rheumatoid arthritis*

Study: author, year	Country	Study design	Patients/samples: baseline (total), no.†	Age, years‡	Female, %	DAS28-CRP‡	RF/anti-CCP+ (%)
Systematic review/ meta-analysis							
Bakker, 2012	Netherlands	POS	72/72 (74/120)	53 ± 15	70	5.6 ± 1.0§	68/–
Curtis, 2012	Multiple	POS					
Seropositive validation			230/–	58 (48–66)	77	4.1 (2.3–5.8)	93/88
Seronegative validation			141/–	57 (46–65)	82	3.7 (2.4–4.9)	0
Seronegative performance			141/–	58 (49–65)	79	3.5 (2.4–4.7)	0
Treatment response			45/45 (45/144)	54 (39–64)	84	5.5 (4.9–6.4)	73/–
Hirata, 2013	Netherlands	RCT	91/91 (125/179)	53 ± 14	74	5.5 ± 0.9	62/57
Hambardzumyan, 2015	Sweden	RCT	235/235	–	72	5.4 ± 1.0	65/57
Hirata, 2015	Japan	ROS	147/147 (147/378)	60 (50–68)	84	5.0 (4.3–5.7)	86/–
Fleischmann, 2016	Multiple	RCT	496/496 (524/–)	Ref.	Ref.	5.5 (–)	Ref.
Reiss, 2016	Multiple	RCT	48/48 (78/107)	51 ± 14	82	5.7 ± 0.9	
Krabbe, 2017	Denmark	POS	50/50 (50/284)	61 (50–70)	67	4.9 (4.2–5.6)§	
Systematic review only							
Eastman, 2012	US, Canada	POS	512/–	60 (20–91)¶	76	3.2 (1.1–8.2)	76/–
Centola, 2013	US, UK	POS, RCT					
Study I			128/128	60 ± 13	82	5.8 (4.7–6.5)	83/63
Study II			320/320	59 ± 14	80	4.0 (2.9–5.3)	83/62
Study III			85/255	59 ± 13	91	3.8 (2.7–5.0)	64/62
Study IV			119/119	60 ± 14	77	5.2 (4.1–6.2)	97/61
Pilot Imaging Study			24/107	56 ± 13	75	3.3 (2.2–4.4)	
Li, 2013	US	POS	101/101	62 ± 13	82	–	63/45
Peabody, 2013	Germany	RCT		–	–	–	
van der Helm-van Mil, 2013	Netherlands	POS	163/271	55 ± 14	67	–	65/66
Markusse, 2014	Netherlands	RCT	91/91 (125/180)	53 ± 14	75	5.8 ± 1.0	62/56
Michaud, 2015	US	ROS	Ref.	Ref.	Ref.	Ref.	Ref.
van Vollenhoven, 2015	Sweden	RCT	347/347 (347/474)	–	–	–	
MTX-naive cohort			220/220 (220/220)	Ref.	Ref.	5.7 (–)§	Ref.
MTX-IR cohort			127/127 (127/254)	Ref.	Ref.	4.9 (–)§	Ref.
Hambardzumyan, 2016	Sweden	RCT	220/220 (220/558)	–	71	5.7 ± 1.0)	65/57
Hirata, 2016	Japan	POS	83/83 (83/249)	59 ± 14	84	5.7 ± 1.2	87/–
Lee, 2016	US	POS	198/198	58 ± 11	85	–	63/62
Li, 2016	Netherlands	POS	163/271	55 ± 14	67	3.3 (2.3–4.3)	66/67
Rech, 2016	Germany	RCT	94/94	55 ± 19	60	1.9 ± 0.8§	61/56
Hambardzumyan, 2017	Sweden	ROS	157/157	–	80	6.0 ± 1.0§	62/55

* If left blank, the study did not report the item. DAS28-CRP = Disease Activity Score using 28 joints with C-reactive protein level; RF = rheumatoid factor; anti-CCP = anti–cyclic citrullinated peptide antibody; POS = prospective observational study; RCT = randomized controlled trial; ROS = retrospective observational study; Ref. = authors listing refers to original study; MTX-IR = methotrexate incomplete response.
† If single set of data is provided, only total patients/samples were available for report.
‡ Mean/median ± SD or (interquartile range).
§ DAS28 without modifier is reported.
¶ Full age range reported.

based on the DAS28-CRP (area under the receiver operating characteristic curve [AUROC] 0.70–0.86) (16,17) and ACR/EULAR remission (AUROC 0.83) (22). There were conflicting results specific to the validity of the MBDA following DMARD initiation. The MBDA did not correlate with the DAS28-CRP, CDAI, SDAI, or RAPID3 over 2 years of follow-up in patients with active RA treated with adalimumab or abatacept in a randomized controlled trial (10). Correlations of the MBDA with the DAS28-CRP decreased over 24 weeks of treatment with tocilizumab (r = 0.50 at baseline, r = 0.19–0.33 between weeks 4–24), as did agreement in disease activity categories (77.1% at baseline, 23.7% at week 24) (11). In contrast, a change in the MBDA correlated with a change in the DAS28-CRP or the DAS28-ESR after initiation of tumor necrosis factor inhibitors in a Japanese cohort (22).

In contrast with other composite ACR-recommended RA disease activity measures, MBDA scores were not influenced by comorbid fibromyalgia in 1 study (26). Additional comorbidities including hypertension, osteoarthritis, degenerative joint disease, osteoporotic bone fractures, diabetes mellitus, and asthma also did not affect MBDA performance (9). Exclusion of CRP level, a common component between the MBDA score

Table 3. Results of published literature on the Multi-Biomarker Disease Activity (MBDA) score in rheumatoid arthritis*

Study: author, year	Primary aim	Secondary aim	Results summary
Systematic review/ meta-analysis			
Bakker, 2012	Correlation with disease activity measures	Contribution of non-CRP biomarkers, MBDA response to treatment, ability to predict RP	MBDA correlated with DAS28-CRP and discriminated between remission/low and mod./high DAS28-CRP disease activity categories (AUROC 0.86); non-CRP biomarkers independently associated with SJC28, TJC28, and VAS-GH; MBDA decreased with 6 months treatment (53[18] to 39[16]), more significantly in intensive treatment arm; MBDA did not predict radiographic progression
Curtis, 2012	Establish criterion and discriminant validity	Contribution of non CRP MBDA biomarkers; characterize performance in seropositive vs. seronegative patients	MBDA correlated with DAS28-CRP, CDAI, SDAI, and RAPID3, and discriminated low vs. mod./high DAS28-CRP (AUROC 0.70–0.77 across cohorts studied); ΔMBDA correlated with ΔDAS28 CRP, ACR response criteria and discriminated clinical response (DAS28-CRP AUROC 0.77; ACR50 AUROC 0.69); MBDA better correlated with disease activity measures in seropositive (vs. seronegative) patients; SDAI (r = 0.55 vs. 0.29), CDAI (r = 0.48 vs. 0.21), RAPID3 (r = 0.47 vs. 0.26); non-CRP MBDA biomarkers predicted DAS28-CRP
Hirata, 2013	Correlation with disease activity measures	Ability to discriminate EULAR disease activity categories	MBDA significantly correlated with DAS28-ESR, SDAI, CDAI, and HAQ DI; MBDA correlated with ΔDAS28-ESR and SDAI (not CDAI) over 1-year follow-up; remission by MBDA associated with ACR/EULAR (AUROC 0.83), DAS28-ESR, CDAI, and SDAI remission criteria
Hambardzumyan, 2015	Ability of baseline MBDA to predict radiographic progression (ΔSHS >5)	–	Baseline MBDA higher (P <0.001) in patients with RP; MBDA independent predictor of RP as continuous (OR 1.05 [95% CI 1.02–1.08]) or categorical variable (OR 3.86 [95% CI 1.04–14.26] for high vs. low/mod. MBDA)
Hirata, 2015	Correlation with change in disease activity measures	Comparison between anti-TNF therapies	ΔMBDA correlated with ΔDAS28-ESR and ΔDAS28-CRP, but not ΔCDAI or ΔSDAI; no difference in correlations between anti-TNF therapies
Fleischmann, 2016	Correlation with disease activity measures	Correlation with radiographic progression	Not associated with DAS28-CRP, CDAI, SDAI, RAPID3, or radiographic progression over 2-year follow-up
Reiss, 2016	Effect of TCZ on correlation of MBDA with disease activity	Effect of TCZ on individual biomarkers in MBDA	Correlation of MBDA with DAS28-CRP decreased (Spearman's P = 0.50 at baseline to P = 0.19–0.33) between weeks 4 and 24, and agreement between low/mod./high MBDA and DAS28-CRP categories decreased (77.1% to 23.7%) with 24 weeks of TCZ treatment; individual analyte changes following TCZ treatment included an increase in IL-6 and a decrease in CRP and serum amyloid A
Krabbe, 2017	Correlation with imaging measures of inflammation	Correlation with DAS28-CRP	MBDA did not correlate with imaging inflammation at baseline or week 52, and in general did not predict change in imaging inflammation; correlated modestly with MRI synovitis (r = 0.43), MRI bone marrow edema (r = 0.36), and US power Doppler score (r = 0.35) at week 26; MRI/US were concordant with MBDA in detecting disease activity for patients in DAS28-CRP remission; MBDA correlated with DAS28-CRP at baseline and week 26; ΔMBDA correlated with ΔDAS28-CRP from baseline to 26 weeks, but not baseline to 52 weeks
Systematic review only		–	
Eastman, 2012	Analytical performance of MBDA multiplex assay	–	MBDA biomarker assays were precise, with minimal interference or cross-reactivity
Centola, 2013	Development of MBDA score	Impact of comorbidities on MBDA	MBDA algorithm developed through biomarker screening, feasibility studies, and assay optimization; comorbidities assessed (hypertension, osteoarthritis, osteoporotic bone fractures, degenerative joint disease, diabetes mellitus, asthma) were not associated with the MBDA

(Continued)

Table 3. *(Cont'd)*

Study: author, year	Primary aim	Secondary aim	Results summary
Li, 2013	Effect on provider treatment choices	Effect on overall drug use, correlation with PrGA	Treatment plans changed in 38% of patients with MBDA; no effect on overall drug use; modest correlation with PrGA (r = 0.35)
Peabody, 2013	Impact on quality scores using clinical performance and value vignettes	Appropriate use of DMARDs, number of labs or imaging tests ordered, use of other resources	Quality scores improved 12% with MBDA; appropriate use of DMARDS improved with comorbid patients; no effect on number of labs or imaging tests ordered, or use of health care resources
van der Helm-van Mil, 2013	Frequency of radiographic progression (ΔSHS >3) in MBDA remission	Detection of subclinical disease activity	Greater rate of nonprogression in MBDA remission vs. nonremission (93% vs. 70%). +LR of nonprogression in MBDA remission 4.73 (95% CI 1.67–15.0); high MBDA score in DAS28-CRP remission increased risk of radiographic progression (RR 2.28 [95% CI 1.13 3.68])
Markusse, 2014	Ability to predict radiographic progression (ΔSHS >5)	–	MBDA at baseline discriminates radiographic progressors vs. nonprogressors better than DAS (AUROC 0.767 [95% CI 0.639 0.896]) and predicts RP based on MBDA at baseline (RR 1.039 [95% CI 1.018–1.059]) and 1 year (RR 1.037 [95% CI 1.009–1.065]) associated with increased RP
Michaud, 2015	Outcomes and cost when used in RA management	–	Decreased HAQ scores (0.09 in 1 year, 0.02 over 10 years), increase quality-adjusted life years 0.08 and decreased overall cost US $457
van Vollenhoven, 2015	Impact on recruitment to clinical trials based on data from SWEFOT trial	–	High MBDA (>44) enhanced recruitment in low CRP (<10) patients; additional 24% MTX-naive patients and 47% MTX-incomplete responders included
Hambardzumyan, 2016	Ability of MBDA at multiple time points to predict radiographic progression (ΔSHS >5)	Ability of MBDA to predict RP in triple therapy (TT) vs. anti-TNF treated patients	Persistently low/mod. MBDA was predictive of less RP; MBDA was numerically (but not statistically) superior to CRP, ESR, and DAS28 for identifying RP; patients with high MBDA scores on TT had increased risk of RP compared to anti-TNF therapy (45% vs. 25% at baseline and 57% vs. 32% at month 3)
Hirata, 2016	Correlation with radiographic progression	–	MBDA correlated with ΔSHS (r = 0.47 at week 24; AUROC 0.44 over 52 weeks). High MBDA increased risk of ΔSHS >3 (RR 14.3 [95% CI 2.5–85.5]) at week 24 compared to low MBDA; in patients with low or mod./high DAS28, MBDA further discriminated risk of radiographic progression
Lee, 2016	Correlation with disease activity measures	Utility in RA patients with fibromyalgia (FM)	MBDA correlated with CRP in RA patients with (r = 0.89) or without (r = 0.73) concomitant FM; composite indices (DAS28-CRP, SDAI, CDAI, RAPID3) all greater in patients with concomitant FM, though no difference in MBDA between these groups
Li, 2016	Correlation with radiographic progression	–	High MBDA increased risk of ΔSHS >3 (RR 3.4 if MBDA 45–51; 4.3 if MBDA 52–59; 5.2 if MBDA ≥60) and ΔSHS >5 (RR 12.4, 12.0, and 17.4) compared with low MBDA; MBDA independent risk factor for radiographic progression after adjustment for SJC28, DAS28-CRP, CRP, and pre-existing joint damage
Rech, 2016	Ability to predict disease relapse in patients tapering DMARDs	–	Baseline MBDA scores significantly (P = 0.0001) higher in patients with subsequent relapse; MBDA and anti-CCP independent predictors of disease relapse; able to predict >80% of relapses using anti-CCP plus MBDA
Hambardzumyan, 2017	Predicting response to triple therapy (TT) vs. anti-TNF	–	More patients with low MBDA responded to TT vs. anti-TNF (88% vs. 18%); more patients with high MBDA responded to anti-TNF (35% vs. 58%)

* CRP = C-reactive protein; RP = radiographic progression; DAS28 = Disease Activity Score using 28 joints; AUROC = area under the receiver operating characteristic curve; SJC28 = swollen 28-joint count; TJC28 = tender 28-joint count; VAS-GH = visual analog scale of patients' assessment of general health; CDAI = Clinical Disease Activity Index; SDAI = Simple Disease Activity Index; RAPID3 = Routine Assessment of Patient Index Data 3; ACR50 = American College of Rheumatology response criteria 50% improvement; EULAR = European League Against Rheumatism; ESR = erythrocyte sedimentation rate; HAQ DI = Health Assessment Questionnaire disability index; SHS = Sharp/van der Heijde score; RP = radiographic progression; OR = odds ratio; 95% CI = 95% confidence interval; anti-TNF = anti–tumor necrosis factor; TCZ = tocilizumab; IL-6 = interleukin-6; MRI = magnetic resonance imaging; US = ultrasound; PrGA = provider global assessment of disease activity; DMARD = disease-modifying antirheumatic drug; +LR = positive likelihood ratio; RR = relative risk; RA = rheumatoid arthritis; SWEFOT = Swedish Pharmacotherapy trial; MTX = methotrexate; anti-CCP = anti–cyclic citrullinated peptide antibody.

and the DAS28-CRP and SDAI, did not attenuate MBDA score performance in 2 independent studies (16,17).

Nine studies investigated the ability of the MBDA to predict RP (Table 3). In secondary analyses of both observational studies and randomized controlled trials, the MBDA predicted RP in 6 studies (19,20,24,28,29,33). Sharp/van der Heijde score (SHS) cut-offs, analytic methods, and resulting effect sizes were highly variable. Relative to patients with low disease activity, patients with high disease activity by MBDA score had a relative risk of RP range of 1.04–14.30 and an odds of RP range of 1.03–3.86. RA patients in MBDA remission were less likely to demonstrate RP, with a positive likelihood ratio for non-progression (change SHS <3) of 4.73 (33). The MBDA more effectively discriminated radiographic progressors from

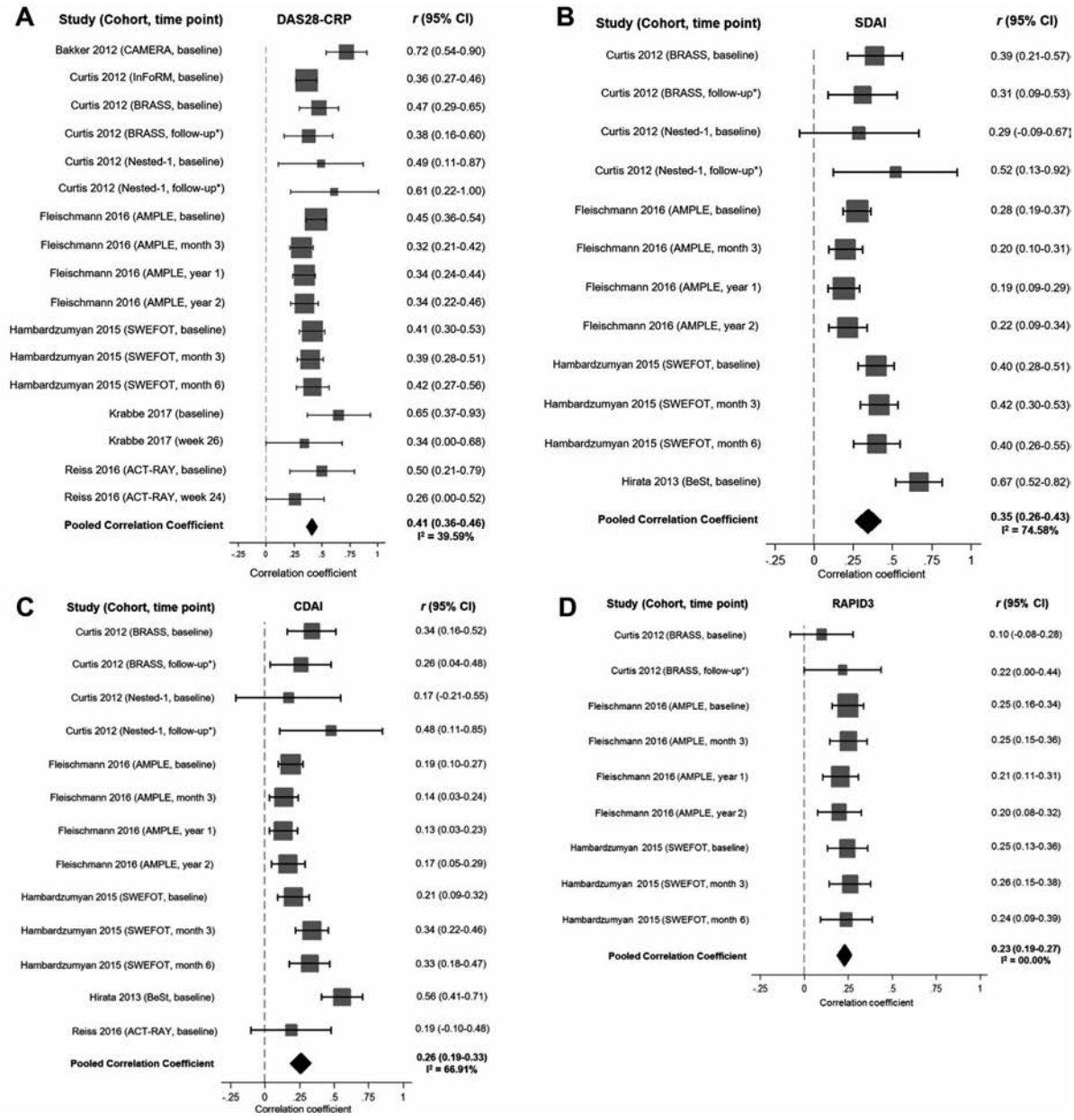

Figure 2. Correlation of the multi-biomarker disease activity score with rheumatoid arthritis disease activity measures. Forest plots demonstrating the correlation of the multi-biomarker disease activity score with RA disease activity measures including **A**, Disease Activity Score using 28 joints and the C-reactive protein level (DAS28-CRP); **B**, Simple Disease Activity Index (SDAI); **C**, Clinical Disease Activity Index (CDAI); and **D**, Routine Assessment of Patient Index Data 3 (RAPID3). Studies assessing correlation at multiple time points are modeled separately for meta-analysis. * Patients with initial follow-up at 6 weeks and again at 12 weeks if adequate treatment response not obtained in the BRASS (Brigham and Women's Rheumatoid Arthritis Sequential Study) and Nested-1 cohorts (17). CAMERA = Computer Assisted Management in Early Rheumatoid Arthritis study; InFoRM = Index for Rheumatoid Arthritis Measurement; AMPLE = Abatacept Versus Adalimumab Comparison in Biologic-Naive RA Subjects With Background Methotrexate; SWEFOT = Swedish Pharmacotherapy trial; ACT-RAY = Study of Tocilizumab and MTX Treatment Strategies in Patients With Active RA With Inadequate Response to Prior MTX Treatment; BeSt = Treatment Strategies for Rheumatoid Arthritis.

non-progressors than the 44-joint DAS (AUROC 0.767 versus 0.521) (29) as well as CRP and ESR (20). Additionally, among those in low or moderate/high disease activity as determined by the DAS28, MBDA scores further discriminated risk of RP (24). In contrast, 2 studies demonstrated less capability of the MBDA score to predict RP (10,16). In an observational study of RA patients treated with methotrexate (MTX), the MBDA was not predictive of RP (odds ratio 1.033, 95% CI 0.995–1.072) (16). Likewise, in a randomized controlled trial comparing abatacept and adalimumab in MTX-inadequate responders, MBDA categories were not associated with radiographic non-progression (10). Finally, the MBDA did not demonstrate consistent correlation with novel imaging measures of inflammation such as magnetic resonance imaging–based synovitis/bone marrow edema or power Doppler ultrasound score (25).

Initial studies of the MBDA on quality of care in RA have yielded favorable results. A decision analysis projecting cost-effectiveness estimated an overall decrease in Health Assessment Questionnaire scores, increase in quality-adjusted life years, and decreased net total cost (savings in labor force participation and work productivity offsetting an increase in direct medical costs) accompanying use of the MBDA score (30). These effects were largely driven by the effect of the MBDA on provider treatment choices, based on data from a previous study that showed that treatment plans changed in 38% of patients when an MBDA score was provided (27). Overall quality scores improved on clinical performance and value vignettes in providers receiving MBDA scores (31).

An emerging area of research poses the question whether the MBDA predicts treatment response. Among RA patients with an inadequate response to MTX in the Swedish Pharmacotherapy trial (SWEFOT), more patients with low MBDA responded to triple therapy (MTX, hydroxychloroquine, and sulfasalazine) than the addition of infliximab (88% versus 18%; $P = 0.006$), defined as achieving a DAS28 <3.2 or a EULAR good response. Patients with high MBDA scores responded more frequently to the addition of infliximab rather than triple therapy (58% versus 35%; $P = 0.040$) (21). In a separate analysis from the SWEFOT trial, patients with high MBDA scores receiving triple therapy were more likely to have RP after 2 years of follow-up than patients receiving MTX and infliximab (20). Additional outcomes reported on the MBDA included its potential to increase recruitment to clinical trials (i.e., identifying patients with low CRP who would otherwise be excluded) (34), as well as its ability to predict disease relapse (32).

Meta-analysis. Correlations of the MBDA with composite RA disease activity measures were available from 8 studies (Figure 2 and Supplementary Table 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23785/abstract). Using a random-effects meta-analysis with 3,242 samples from 6 studies (8 cohorts and 17 time points), the MBDA score was moderately correlated with the DAS28-CRP (r = 0.41, 95% CI 0.36–0.46) with low-to-moderate heterogeneity ($I^2 = 39.59\%$, $P = 0.07$). Performance was stronger at baseline (r = 0.48, 95% CI 0.39–0.56) compared to follow-up time points (r = 0.36, 95% CI 0.31–0.40). Performance was also numerically superior in observational studies (r = 0.49, 95% CI 0.38–0.61) compared to randomized controlled trials (r = 0.38, 95% CI 0.34–0.42), but there was substantially greater heterogeneity detected in observational studies ($I^2 = 54.7\%$, $P = 0.03$) than randomized controlled trials ($I^2 = 0.0\%$, $P = 0.57$). The MBDA score performed similarly in its correlation with the DAS28-ESR (r = 0.48, 95% CI 0.38–0.58), although this was reported in fewer studies (3 studies with 3 cohorts, 5 different time points, n = 1,367) and with greater heterogeneity ($I^2 = 67.81\%$, $P = 0.01$). In 4 studies (5 cohorts, 12 time points, n = 2,664), the MBDA score demonstrated a low-to-moderate correlation with the SDAI (r = 0.35, 95% CI 0.26–0.43) with high heterogeneity ($I^2 = 74.58\%$, $P < 0.001$). Correlation with the CDAI was less robust (r = 0.26, 95% CI 0.19–0.33), with moderate heterogeneity ($I^2 = 66.91\%$, $P < 0.001$) among the 5 contributing studies (6 cohorts, 13 time points, n = 2,719). The MBDA score was weakly correlated with the RAPID3 (r = 0.23, 95% CI 0.19–0.27) in 2 studies (3 cohorts, 9 time points, n = 2,416) with very low study heterogeneity ($I^2 = 0.00\%$, $P = 0.92$).

Correlation of MBDA change with change in RA disease activity was also determined with random-effects meta-analysis (Figure 3). Correlations of MBDA change with change in the DAS28-CRP were reported in 5 studies (6 cohorts, n = 1,857), with a moderate correlation observed (r = 0.42, 95% CI 0.37–0.48). Change in MBDA score also moderately correlated with change in the DAS28-ESR (r = 0.53, 95% CI 0.46–0.60) in 3 studies (3 cohorts and 5 time points, n = 825). Weaker correlations were seen with change in the SDAI (r = 0.33, 95% CI 0.26–0.40) in 4 studies (5 cohorts, 9 time points, n = 1,710), the CDAI (r = 0.26, 95% CI 0.20–0.33) in 4 studies (5 cohorts, 9 time points, n = 1,718), and the RAPID3 (r = 0.31, 95% CI 0.25–0.38) in 3 studies (3 cohorts, 7 time points, n = 1,617). Heterogeneity observed for comparisons of change in MBDA score with change in RA disease activity was low-to-moderate ($I^2 = 0.00$–43.44%, $P = 0.08$–0.77).

Publication bias was assessed by funnel plot analysis with the DAS28-CRP, the most frequently reported RA disease activity measure (see Supplementary Figure 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23785/abstract). Study distribution was largely symmetric, which suggests that there was no substantial publication bias among studies included in the meta-analysis.

DISCUSSION

The MBDA score, a composite score of 12 serum biomarkers that was initially derived to predict the DAS28-CRP, provides an objective measure of RA disease activity by eliminating subjective assessments from the patient or provider, a potentially useful

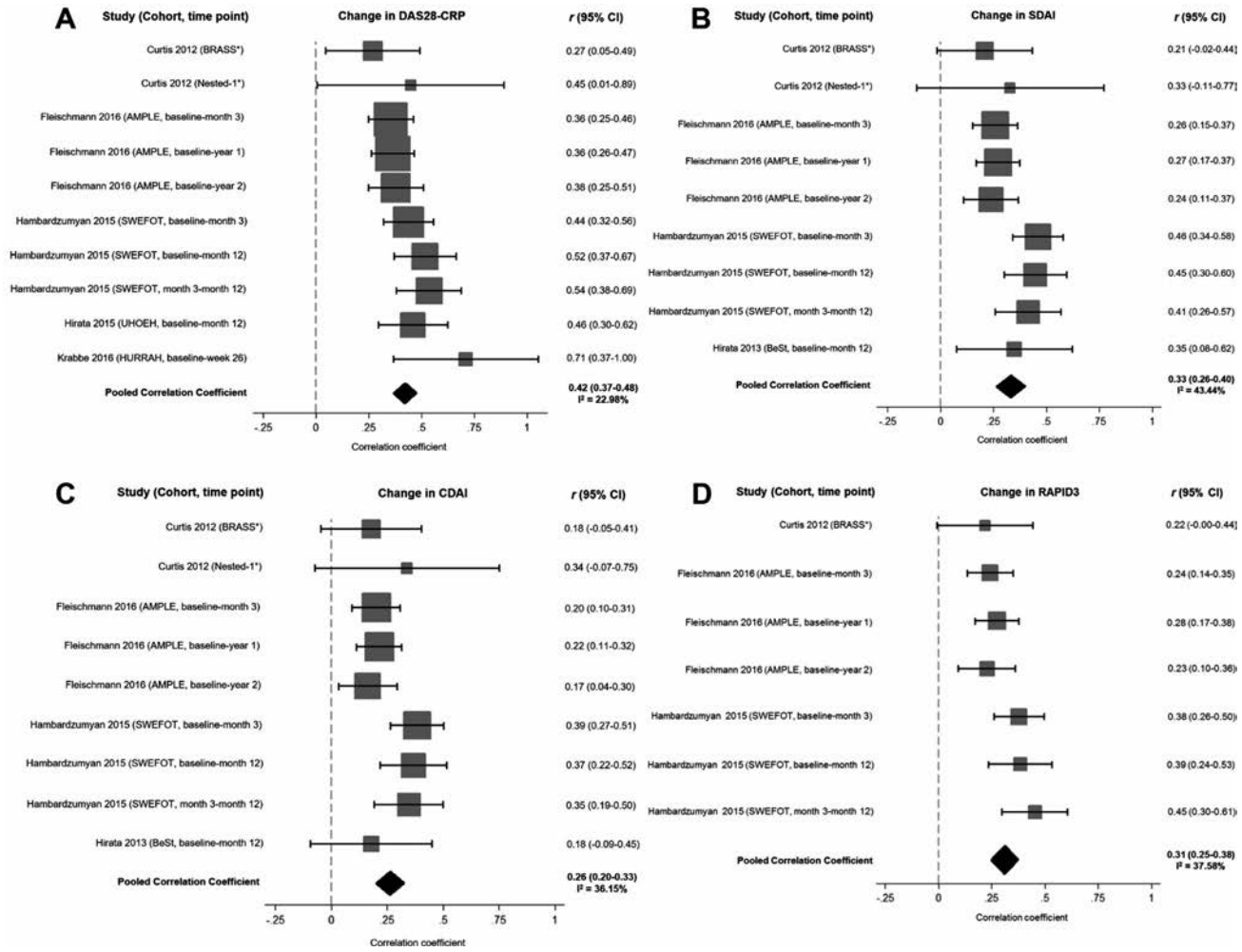

Figure 3. Correlation of the change in multi-biomarker disease activity score with change in rheumatoid arthritis disease activity measures over time. Forest plots demonstrating correlation of the change in multi-biomarker disease activity score with change in RA disease activity measures including **A**, Disease Activity Score using 28 joints and the C-reactive protein level (DAS28-CRP); **B**, Simple Disease Activity Index (SDAI); **C**, Clinical Disease Activity Index (CDAI); **D**, Routine Assessment of Patient Index Data 3 (RAPID3). Studies assessing correlation at multiple follow-up points are modeled separately for meta-analysis. * Patients followed up initially at 6 weeks and again at 12 weeks if adequate treatment response not obtained in the Brigham and Women's Rheumatoid Arthritis Sequential Study and Nested-1 cohorts (17). UHOEH = University Hospital of Occupational and Environmental Health; HURRAH = Humira in Rheumatoid Arthritis. See Figure 2 for other definitions.

tool when caring for RA patients. However, there is conflicting data regarding its convergent validity with other RA disease activity measures. In this study, we report the first systematic review to assess the performance of the MBDA score in RA across multiple outcomes and the first meta-analysis of the convergent validity with ACR-endorsed RA disease activity measures.

The DAS28 using either ESR or CRP level is often considered the gold standard in RA disease activity measure and is frequently used for disease activity measurement in clinical trials. Using a random-effects meta-analysis including 3,242 MBDA measurements, we found moderate correlations between the MBDA score with both the DAS28-CRP (r = 0.41) and DAS28-ESR (r = 0.48). These were the strongest correlations observed between the MBDA and composite RA disease activity measures, perhaps anticipated given that the MBDA score was derived to predict the DAS28-CRP (9). The SDAI, another ACR-endorsed RA disease activity measure that includes patient, physician, and acute-phase reactant components, demonstrated a weaker correlation with the MBDA (r = 0.35) compared to DAS28 measures but a stronger correlation when compared with the CDAI (r = 0.26) and RAPID3 (r = 0.23). Because the CDAI differs from the SDAI only by the exclusion of CRP, this suggests that the common component of acute-phase reactants is unlikely to account entirely for its performance in measuring RA disease activity. Further supporting this are 2 studies identified in our systematic review in which MBDA scores were associated with RA disease activity after exclusion of CRP (16,17).

Variable heterogeneity was observed in the meta-analyses of MBDA scores with RA disease activity measures. There was

moderate-to-high heterogeneity for cross-sectional correlations examining MBDA correlations with the DAS28-ESR, SDAI, and CDAI. The majority of this heterogeneity appeared to be related to exceptionally strong performance in the Treatment Strategies for Rheumatoid Arthritis study (22). Exclusion of this study reduced variability (I^2) by 40% for the DAS28-ESR, 24% for the SDAI, and 34% for the CDAI. The remaining cross-sectional and longitudinal correlations between MBDA scores and RA disease activity measures had low-to-moderate heterogeneity by I^2. Although there was only moderate heterogeneity for correlations between the MBDA scores and the DAS28-CRP, sensitivity analyses found this heterogeneity to be limited to observational studies (I^2 = 54.7% in observational studies) and minimal heterogeneity to exist in studies from randomized controlled trials (I^2 = 0.0%).

Our systematic review identified additional studies characterizing the performance of the MBDA score in assessing RA disease outcomes. The MBDA score was able to discriminate between low versus moderate/high disease activity in 3 studies (16,17,22), an important characteristic with the goal of treating to remission or low disease activity (2). MBDA scores were also predictive of RP in several studies and independent cohorts, although study heterogeneity precluded a formal meta-analysis of this performance characteristic. The ability of this tool to predict a well-established complication of active RA further supports the validity of the MBDA as a measure of RA disease activity. Noninflammatory pain such as that resulting from comorbid fibromyalgia can influence traditional RA disease activity measures, complicating disease activity measurement. However, as an objective serum biomarker measurement, MBDA scores were not influenced by fibromyalgia (26).

As with any test being implemented in clinical use, the cost of testing is an important consideration to patients, providers, and the health care system. Our systematic review identified limited study of cost-effectiveness, with a simulated analysis suggesting cost-effectiveness when balancing improvement in quality of life and increased labor force participation (30). Further analysis of its cost-effectiveness from actual patient data is needed to confirm these findings.

Our study has limitations. Only 22 studies of the MBDA score in RA were published during the search period, and only 8 of these fulfilled eligibility criteria to be included in the meta-analysis. We included only published manuscripts because patient cohorts were often used in multiple studies, and the inclusion of results from gray literature would have increased the probability of duplicate inclusion of subjects/samples. Additionally, the MBDA manufacturer supported 18 of the 22 included studies, although in 5 of these, support was only provided for MBDA measurement in the absence of any further involvement in study design or analysis. Details of how serum samples were collected, processed, and stored prior to analysis were inconsistently reported. Because sample handling may affect biomarker measurement (35), this represents an unknown confounder in this meta-analysis. Finally, a novel scoring algorithm for the MBDA score has been developed recently that accounts for age and body mass index (36). Because of our search dates, studies using this revised score were not included.

There are a number of strengths to our study. We conducted a systematic review of the MBDA score in RA, searching 5 databases for eligible studies. Rigorous methodology was used with duplicate assessment of study eligibility, data abstraction, and quality assessment. Moreover, study quality assessment was completed using an adapted tool from tumor biomarker reporting guidelines (13,14). Finally, corresponding authors and companies were contacted to provide additional data on studies that did not initially report correlations with composite RA disease activity measures and to prevent duplication of subjects/samples in the meta-analysis.

In summary, this is the first systematic review and meta-analysis to examine the performance of the MBDA score in RA. The MBDA score demonstrates moderate convergent validity with the DAS28-CRP and DAS28-ESR, less robust correlation with the SDAI, and weak convergent validity with the CDAI and RAPID3, composite measures lacking acute-phase reactants. It also appears to predict RP and influence provider decision-making, although these findings need further validation in light of high levels of variability and low effect sizes observed across studies. While the MBDA score represents another tool to measure RA disease activity, further assessment of its ability to improve RA management (such as the ability to predict treatment response or comparisons of patient outcomes for individuals treated to target with the MBDA score versus other RA disease activity measures), validation of its performance characteristics, evaluation of a recently proposed scoring modification, as well as appraisal through independently funded efforts are necessary.

ACKNOWLEDGMENTS

Data from Bristol-Myers Squibb was acquired by way of a data sharing agreement through SOAR (Supporting Open Access for Researchers). The authors thank Jeffrey Curtis, Roy Fleischman, Karen Hambardzumyan, Bristol-Myers Squibb, and Crescendo Bioscience for assisting with the acquisition of and/or providing additional data.

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. England 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. Johnson, Register, Schmidt, O'Dell, Mikuls, Michaud, England.

Acquisition of data. Johnson, Register, Schmidt, England.

Analysis and interpretation of data. Johnson, Register, Schmidt, O'Dell, Mikuls, Michaud, England.

REFERENCES

1. Sokka T, Abelson B, Pincus T. Mortality in rheumatoid arthritis: 2008 update. Clin Exp Rheumatol 2008;26:S35–61.
2. Smolen JS, Landewe R, Bijlsma J, Burmester G, Chatzidionysiou K, Dougados M, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. Ann Rheum Dis 2017;76:960–77.
3. Singh JA, Saag KG, Bridges SL Jr, Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. Arthritis Care Res (Hoboken) 2016;68:1–25.
4. Goswami RP, Basu K, Das S, Mondal S, Ghosh P, Ghosh A. Evidence for treating rheumatoid arthritis to target: results of a systematic literature search update. Ann Rheum Dis 2016;75:e35.
5. Anderson JK, Zimmerman L, Caplan L, Michaud K. Measures of rheumatoid arthritis disease activity: patient (PtGA) and provider (PrGA) global assessment of disease activity, disease activity score (DAS) and disease activity score with 28-joint counts (DAS28), simplified disease activity index (SDAI), clinical disease activity index (CDAI), patient activity score (PAS) and patient activity score-II (PASII), routine assessment of patient index data (RAPID), rheumatoid arthritis disease activity index (RADAI) and rheumatoid arthritis disease activity index-5 (RADAI-5), chronic arthritis systemic index (CASI), patient-based disease activity score with ESR (PDAS1) and patient-based disease activity score without ESR (PDAS2), and mean overall index for rheumatoid arthritis (MOI-RA). Arthritis Care Res (Hoboken) 2011;63 Suppl 11:S14–36.
6. Anderson J, Caplan L, Yazdany J, Robbins ML, Neogi T, Michaud K, et al. Rheumatoid arthritis disease activity measures: American college of rheumatology recommendations for use in clinical practice. Arthritis Care Res (Hoboken) 2012;64:640–7.
7. Joharatnam N, McWilliams DF, Wilson D, Wheeler M, Pande I, Walsh DA. A cross-sectional study of pain sensitivity, disease-activity assessment, mental health, and fibromyalgia status in rheumatoid arthritis. Arthritis Res Ther 2015;17:11.
8. Keenan RT, Swearingen CJ, Yazici Y. Erythrocyte sedimentation rate and C-reactive protein levels are poorly correlated with clinical measures of disease activity in rheumatoid arthritis, systemic lupus erythematosus and osteoarthritis patients. Clin Exp Rheumatol 2008;26:814–9.
9. Centola M, Cavet G, Shen Y, Ramanujan S, Knowlton N, Swan KA, et al. Development of a multi-biomarker disease activity test for rheumatoid arthritis. PLoS One 2013;8:e60635.
10. Fleischmann R, Connolly SE, Maldonado MA, Schiff M. Estimating disease activity using multi-biomarker disease activity scores in rheumatoid arthritis patients treated with abatacept or adalimumab. Arthritis Rheumatol 2016;68:2083–9.
11. Reiss WG, Devenport JN, Low JM, Wu G, Sasso EH. Interpreting the multi-biomarker disease activity score in the context of tocilizumab treatment for patients with rheumatoid arthritis. Rheumatol Int 2016;36:295–300.
12. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2016;354:i4086.
13. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM, et al. Reporting recommendations for tumor MARKer prognostic studies (REMARK). Breast Cancer Res Treat 2006;100:229–35.
14. Chen M, Huang J, Zhu Z, Zhang J, Li K. Systematic review and meta-analysis of tumor biomarkers in predicting prognosis in esophageal cancer. BMC Cancer 2013;13:539.
15. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177–88.
16. Bakker MF, Cavet G, Jacobs JW, Bijlsma JW, Haney DJ, Shen Y, et al. Performance of a multi-biomarker score measuring rheumatoid arthritis disease activity in the CAMERA tight control study. Ann Rheum Dis 2012;71:1692–7.
17. Curtis JR, van der Helm-van Mil AH, Knevel R, Huizinga TW, Haney DJ, Shen Y, et al. Validation of a novel multibiomarker test to assess rheumatoid arthritis disease activity. Arthritis Care Res (Hoboken) 2012;64:1794–803.
18. Eastman PS, Manning WC, Qureshi F, Haney D, Cavet G, Alexander C, et al. Characterization of a multiplex, 12-biomarker test for rheumatoid arthritis. J Pharm Biomed Anal 2012;70:415–24.
19. Hambardzumyan K, Bolce R, Saevarsdottir S, Cruickshank SE, Sasso EH, Chernoff D, et al. Pretreatment multi-biomarker disease activity score and radiographic progression in early RA: results from the SWEFOT trial. Ann Rheum Dis 2015;74:1102–9.
20. Hambardzumyan K, Bolce RJ, Saevarsdottir S, Forslind K, Wallman JK, Cruickshank SE, et al. Association of a multibiomarker disease activity score at multiple time-points with radiographic progression in rheumatoid arthritis: results from the SWEFOT trial. RMD Open 2016;2:e000197.
21. Hambardzumyan K, Saevarsdottir S, Forslind K, Petersson IF, Wallman JK, Ernestam S, et al. A multi-biomarker disease activity score and the choice of second-line therapy in early rheumatoid arthritis after methotrexate failure. Arthritis Rheumatol 2017;69:953–63.
22. Hirata S, Dirven L, Shen Y, Centola M, Cavet G, Lems WF, et al. A multi-biomarker score measures rheumatoid arthritis disease activity in the BeSt study. Rheumatology (Oxford) 2013;52:1202–7.
23. Hirata S, Li W, Defranoux N, Cavet G, Bolce R, Yamaoka K, et al. A multi-biomarker disease activity score tracks clinical response consistently in patients with rheumatoid arthritis treated with different anti-tumor necrosis factor therapies: a retrospective observational study. Mod Rheumatol 2015;25:344–9.
24. Hirata S, Li W, Kubo S, Fukuyo S, Mizuno Y, Hanami K, et al. Association of the multi-biomarker disease activity score with joint destruction in patients with rheumatoid arthritis receiving tumor necrosis factor-α inhibitor treatment in clinical practice. Mod Rheumatol 2016;26:850–6.
25. Krabbe S, Bolce R, Brahe CH, Dohn UM, Ejbjerg BJ, Hetland ML, et al. Investigation of a multi-biomarker disease activity score in rheumatoid arthritis by comparison with magnetic resonance imaging, computed tomography, ultrasonography, and radiography parameters of inflammation and damage. Scand J Rheumatol 2017;46:353–8.
26. Lee YC, Hackett J, Frits M, Iannaccone CK, Shadick NA, Weinblatt ME, et al. Multibiomarker disease activity score and C-reactive protein in a cross-sectional observational study of patients with rheumatoid arthritis with and without concomitant fibromyalgia. Rheumatology (Oxford) 2016;55:640–8.
27. Li W, Sasso EH, Emerling D, Cavet G, Ford K. Impact of a multibiomarker disease activity test on rheumatoid arthritis treatment decisions and therapy use. Curr Med Res Opin 2013;29:85–92.
28. Li W, Sasso EH, van der Helm-van Mil AH, Huizinga TW. Relationship of multi-biomarker disease activity score and other risk factors with radiographic progression in an observational study of patients with rheumatoid arthritis. Rheumatology (Oxford) 2016;55:357–66.
29. Markusse IM, Dirven L, van den Broek M, Bijkerk C, Han KH, Ronday HK, et al. A multibiomarker disease activity score for rheumatoid arthritis predicts radiographic joint damage in the BeSt study. J Rheumatol 2014;41:2114–9.
30. Michaud K, Strand V, Shadick NA, Degtiar I, Ford K, Michalopoulos SN, et al. Outcomes and costs of incorporating a multibiomarker disease activity test in the management of patients with rheumatoid arthritis. Rheumatology (Oxford) 2015;54:1640–9.

31. Peabody JW, Strand V, Shimkhada R, Lee R, Chernoff D. Impact of rheumatoid arthritis disease activity test on clinical practice. PLoS One 2013;8:e63215.
32. Rech J, Hueber AJ, Finzel S, Englbrecht M, Haschka J, Manger B, et al. Prediction of disease relapses by multibiomarker disease activity and autoantibody status in patients with rheumatoid arthritis on tapering DMARD treatment. Ann Rheum Dis 2016;75:1637–44.
33. Van der Helm-van Mil AH, Knevel R, Cavet G, Huizinga TW, Haney DJ. An evaluation of molecular and clinical remission in rheumatoid arthritis by assessing radiographic progression. Rheumatology (Oxford) 2013;52:839–46.
34. Van Vollenhoven RF, Bolce R, Hambardzumyan K, Saevarsdottir S, Forslind K, Petersson IF, et al. Enhancement of patient recruitment in rheumatoid arthritis clinical trials using a multi-biomarker disease activity score as an inclusion criterion. Arthritis Rheumatol 2015;67:2855–60.
35. Zhao X, Qureshi F, Eastman PS, Manning WC, Alexander C, Robinson WH, et al. Pre-analytical effects of blood sampling and handling in quantitative immunoassays for rheumatoid arthritis. J Immunol Methods 2012;378:72–80.
36. Curtis JR, Greenberg JD, Harrold LR, Kremer JM, Palmer JL. Influence of obesity, age, and comorbidities on the multi-biomarker disease activity test in rheumatoid arthritis. Semin Arthritis Rheum 2018;47:472–7.

Arthritis Care & Research
Vol. 71, No. 11, November 2019, pp 1473–1481
DOI 10.1002/acr.23784

American College of Rheumatology
Empowering Rheumatology Professionals

Factors Associated With Hand and Upper Arm Functional Disability in People With Rheumatoid Arthritis: A Systematic Review

Hisham Arab Alkabeya,[1] Ann-Marie Hughes,[2] and Jo Adams[2]

Objective. This original systematic review aimed to summarize evidence within observational studies on the factors associated with hand functional disability in adults with rheumatoid arthritis (RA).

Methods. A rigorous extensive systematic literature search was conducted in 6 medical databases for peer-reviewed English language observational studies that explore the factors associated with hand function for people with RA. Factors were critically classified under the domains of the International Classification of Functioning, Disability and Health (ICF) framework and health-related factors. The methodologic quality was determined using the appraisal tool for cross-sectional studies. Factors related to hand function that were investigated in ≥2 studies were explored using a best-evidence synthesis.

Results. Twenty articles from 1,271 citations met the inclusion criteria. All presented cross-sectional data (5 high-quality and 15 low-quality articles), resulting in limited evidence in the best-evidence synthesis. For the factors classified under the ICF domains, the best-evidence synthesis indicated that a diverse range of positive and negative factors were associated with hand function. However, key factors were hand strength, disease activity, and pain intensity. It is evident that few sociodemographic factors have been explored for the association with hand function.

Conclusion. Although the level of evidence was limited, modifiable factors such as grip strength, disease activity, and pain were identified as the most influential factors on hand function in people with RA. The findings of the present review indicate that important sociodemographic factors that impact hand function in individuals with RA have not yet been considered or reported in clinical research.

INTRODUCTION

Rheumatoid arthritis (RA) is an inflammatory, systemic autoimmune and chronic disease that affects approximately 1% of individuals worldwide. The disease pathogenesis remains unknown (1). Hand involvement is typically present in 80–90% of the people with RA (2) and results in stiffness, swelling, pain, range of motion (ROM) limitation, deformity, and muscle weakness (3). These impairments have a formidable impact on hand function and daily life activities (4), causing hand functional disability for a substantial percentage (81%) of people with RA (5).

Current management of RA focuses on early diagnosis and early intensive intervention with disease-modifying antirheumatic drugs (DMARDs) together with biologic medication. These new generation drugs have delivered substantial improvements in decreasing disease activity and minimizing disability (6). However, with recent analysis of cohorts of people with RA who are receiving DMARDs and biologic treatments, it is evident that hand impairments and activity limitations remain as significant problems (7). Moreover, hand problems exacerbate progressively even in patients in remission or with low disease activity (8), and hand function was reported to be substantially worse when compared to referents, despite low disease activity (9). Despite new drug advances and targeted medical treatment, hand function problems for people with RA still persist. Hand function is an important component of disability in people with RA (10). Fortuitously, hand function assessments can be sensitive tools for assessing change in hand functional status (11).

Presented by Mr. Arab Alkabeya in partial fulfillment of the requirements for a PhD degree, Southampton University, Southampton, UK.
[1]Hisham Arab Alkabeya, MSc, BSc: University of Southampton, Southampton, UK, and Arab American University, Jenin, Palestine; [2]Ann-Marie Hughes, PhD, MSc, PGDip, BSc, Jo Adams, PhD, MSc, DipCOT: University of Southampton, Southampton, UK.

No potential conflicts of interest relevant to this article were reported.
Address correspondence to Hisham Arab Alkabeya, MSc, BSc, School of Health Sciences, Building 67, University of Southampton, Highfield Campus, Southampton SO17 1BJ, UK. E-mail: H.Arab-Alkabeya@soton.ac.uk.
Submitted for publication April 24, 2018; accepted in revised form October 9, 2018.

SIGNIFICANCE & INNOVATIONS

- Observational studies have focused predominantly on body structure and function factors, which highlights a lack of consideration and investigation into personal and environmental factors when considering the impact of rheumatoid arthritis (RA) on hand function.
- Modifiable factors such as grip or pinch strength, disease activity, and pain are the most influential factors on hand function in people with RA.
- Well-designed longitudinal, preferably cohort, studies are now needed to better understand the influence of sociodemographic factors on hand functional disability in people with RA.

Since the focus of rehabilitation interventions is to maintain and improve hand function abilities for people with RA (12), it is important to identify the factors that influence the impact of RA on hand function in daily living activities. Consequently, more knowledge about the factors influencing hand functional outcome in people with RA is needed. Hand function interventions for people with RA can be improved by understanding and considering these factors upon planning and delivering treatment intervention. No review has yet reported an overview of the factors associated with hand functional disability for people with RA. Therefore, this study aimed to provide a comprehensive synthesis of the evidence reported within observational studies for the factors associated with hand functional disability in patients with RA in a real-world setting rather than in randomized controlled trials.

MATERIALS AND METHODS

Protocol registration and eligibility criteria. The protocol for this systematic review was registered with the International Prospective Register of Systematic Reviews in May 2017 (protocol reference: CRD42017065856). Studies were included if they fulfilled all of the following criteria: 1) full-length, peer-reviewed studies published in English, 2) observational studies that explored and reported factors associated with hand functional disability, 3) studies that involved participants with the diagnosis of RA, either according to the American College of Rheumatology (ACR) criteria (13) or the 2010 ACR/European League Against Rheumatism criteria for RA (14), or 4) studies that have used hand functional disability outcome measures (either self-reported or objective measures) commonly used with persons with rheumatic diseases, have psychometric support, and evaluate hand-related activity limitations and/or participation restrictions.

Articles that have only used self-reported hand function subscales from generic disability measures or hand functional disability outcome measures of impairment were excluded. This is because generic disability measures are not designed to provide detailed feedback on hand function and include insufficient coverage on hand use. Studies including participants diagnosed with seropositive criteria were excluded, because seropositive criteria are mainly based on the rheumatoid factor (RF) which can occur in other autoimmune conditions and chronic infection.

Information sources and search strategy. A computerized literature search was performed in MEDLINE, EMBASE, CINAHL, PsycINFO, AMED, and Web of Sciences databases. Medical subject headings (MeSH) and free text search keywords were utilized to develop the search for this review. The search strategy was formulated in MEDLINE (Table 1) and adapted for use in other databases after consultation with an experienced medical librarian. Published filters were used to identify studies published in English and from January 1990 to March 2017. Reference lists of all studies meeting the inclusion criteria have been checked. Using Google Scholar, forward citation searching was performed in the current review. Key studies that have been identified by the database searches and selected as meeting the inclusion criteria have been used to carry out citation searching. All citations were imported into EndNote (version X7) library for data management.

Study selection and data collection. Following removal of duplicates (using EndNote software), the study selection

Table 1. Search strategy in MEDLINE through EBSCOhost

#	Search terms
S1	((I hand or TI hands) N3 (TI activit* OR TI abilit* or TI function* OR TI perform* OR TI skill* OR TI impair* OR TI disabilit*)) OR ((AB hand OR AB hands) N3 (AB activit* OR AB abilit* OR AB function* OR AB perform* OR AB skill* or AB impair* OR AB disabilit*))
S2	((MH "Hand+") OR (MH "Hand Deformities") OR (MH "Hand Strength")) AND ((TI ADL OR TI "daily activit*" OR TI "activity limitation*" OR TI "activities of daily living") OR (AB ADL OR AB "daily activit" OR AB "activity limitation*" OR AB "activities of daily living"))
S3	S1 OR S2
S4	(MH "Arthritis, Rheumatoid") OR (TI RA) or (AB RA) OR (TI "Rheumatoid Arthritis") OR (AB "Rheumatoid Arthritis")
S5	(MH "Arthritis, Juvenile") OR (TI "Juvenile Arthritis") OR (AB "Juvenile Arthritis")
S6	S4 NOT S5
S7	S3 AND S6
S8	limit S7 (English language, yr = "1990-Current")

process was completed in 2 stages. The first stage included examining only the titles and abstracts of the search results to eliminate all clearly ineligible publications. Secondly, a full-text review of articles that appeared to meet the inclusion criteria or in cases when a decision could not be made based on the title and abstract alone was conducted. The selection process was completed entirely by the first author. The research team was consulted where any ambiguity arose.

Pertinent data were extracted and documented by the first author, and cross-checked by the research team for completion and accuracy. A predesigned data extraction form was used to extract general information (author and year of publication), characteristics of participants (sample size, disease duration, age, and sex), study characteristics, and hand-function outcome measures, factors, and the association between factors and outcome. Factors tested for association with hand function have been categorized under the domains of the International Classification of Functioning, Disability and Health framework and health-related factors.

Assessment of methodologic quality. Three reviewers independently assessed the methodologic quality of the included articles. The first author (HA) assessed all studies included in the review, and each one of the other reviewers (JA and AMH) assessed half each of the included articles. The quality and risk of bias of the included studies were assessed using a critical appraisal checklist, to assess the quality of cross-sectional studies (the appraisal tool for cross-sectional studies [AXIS]) (15). The AXIS comprises 20 items that focus mainly on the presented methods and results. Seven questions of the AXIS are related to the quality of reporting, 7 questions are related to the study design quality, and 6 questions are related to the risk of biases. Each item was scored by the mean using the following scoring system: "yes" (Y) = 1; and "no" (N) or "don't know" (DK) = 0. The overall score is a percentage score of all 20 items. Studies with an overall score of ≥60% were rated as high quality (16). Disagreements regarding quality assessments between the reviewers were resolved by discussion.

Best-evidence synthesis. Included studies exhibited marked heterogeneity in terms of patient characteristics, outcome measures, statistical analysis, and reporting of results. Consequently, meta-analysis was not possible and the best-evidence synthesis approach was used instead, as recommended by Slavin (17). Only factors tested for association with hand functional disability, which have been measured and reported in the same manner and investigated in ≥2 studies, were included in the best-evidence synthesis. For studies that used 2 tools to evaluate hand function and reported an association between a factor and 1 tool but no association with another tool, the following conditions were applied: If the study used a generic hand function tool and a hand specific tool, then only the results of the latter were considered. If the study used 2 specific hand function tools, then the results of the tool, which has been used more frequently in the included studies, were considered. The Van Tulder ranking system for the level of evidence (18) was used as this is widely used and contemporaneous (16) (Table 2). Initially, the studies were categorized according to the type of study design. The favored design was cohort study followed by case-control design and, at last, cross-sectional design. After that, the studies were ranked according to their methodologic quality overall score. A result was consistent if the factor was significantly associated to hand function with the same direction of the association.

Table 2. Best-evidence synthesis

Level of evidence	Description
Strong	Generally consistent findings were presented in multiple high-quality cohort studies
Moderate	One high-quality cohort study and at least 2 high-quality case–control studies, or when at least 3 high-quality case–control studies show generally consistent findings
Limited	Generally consistent findings were found in a single cohort study, or in maximum 2 case–control studies, or in multiple cross-sectional studies
Conflicting	Less than 75% of the studies reported consistent findings
No evidence	No study could be found

RESULTS

Study selection. The search of the selected databases resulted in the retrieval of 1,254 citations (MEDLINE 395; EMBASE 566; CINAL 122; AMED 54; PsychINFO 18; Web of Sciences 99), and another source search yielded 17 citations. After the removal of duplicate citations, 764 articles remained. Screening of citation titles and abstracts excluded 703 citations from the review. Out of the remaining 61 citations, 41 were excluded with reasons as presented in Figure 1. Finally, 20 articles met all inclusion criteria and were included in the present review. Hand searching for these articles resulted in the retrieval of 1 additional article, which was published in the Turkish language; thus, it was excluded. Forward citation tracking did not yield any further articles for inclusion in the review.

Study characteristics. The articles in the review were based on 19 independent studies of people with RA. Fifteen of the 20 articles (75%) were cross-sectional (2,3,19–31), 2 were case-control (32,33), and 3 were cohort studies (11,34,35). Case-control and cohort studies included in this review presented cross-sectional data on the association between factors and hand function, therefore all studies were considered to be cross-sectional, resulting in limited evidence in the best-evidence synthesis. A full overview of study characteristics of the included studies is presented in Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr23784/abstract.

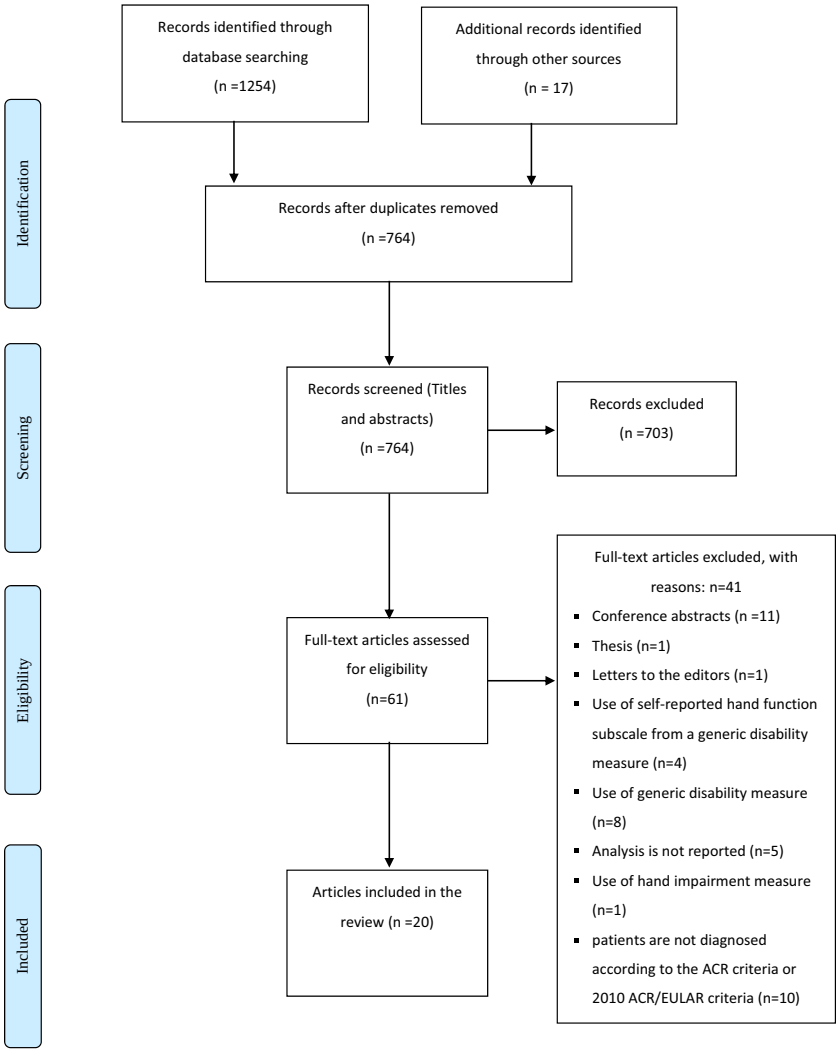

Figure 1. PRISMA flow chart of search results. ACR = American College of Rheumatology; EULAR = European League Against Rheumatism.

Methodologic quality. There was initial disagreement between the leading author (HA) and the second author (AMH) on 19% of the methodologic quality items scored, and between HA and the third author (JA) on 27% of the items scored. Almost all disagreements were due to reading errors or a difference in interpretation of the methodologic quality criteria. After 4 consensus meetings, no disagreement persisted, and a third reviewer was not required to achieve consensus. Overall quality, assessed by the reviewers as the total percentage of quality appraisal items endorsed for each study, was high (≥60%) for 5 studies (25%) (2,3,21,22,29). The mean quality score for the 20 included articles was 49.5% (range 25–75%). The risk of bias items (6,7,13,14) were inadequately met by the included studies, even for the studies with high overall quality scores. Unlike risk of bias, 80% (n = 16) of the included studies have a high score for reporting quality. The mean quality score for reporting was 75% (range 29–100%). The overall scores of methodologic quality, and quality scores for reporting, design and risk of bias domains of the included studies are presented in Supplementary Table 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr23784/abstract.

Factors related to hand function. A summary of all factors considered for best-evidence analysis is presented in Tables 3 and 4. Regarding body structure and function factors, limited hand function was found to be associated with weak hand strength measures (power, lateral pinch, tip pinch and tripod pinch strength), increase of dominant hand fingers flexion deficit, high disease activity (composite measure, tender joints count, C-reactive protein and patient global assessment of disease activity), presence of deformities in dominant hand, increase of ulnar deviation angle of dominant and nondominant hand, low mental health status, high pain intensity (bodily pain and pain during activity), and more hand structural damage. Also, limited evidence was found to support nondominant hand fingers flexion deficit is not associated with hand function. Conflicting evidence was found for the association between the factors swollen joint count, erythrocyte

Table 3. Overview of findings regarding associations of body structure and function factors with hand function*

Body structure and function factors	Association found (references)	No association found	Level of evidence
Strength			
Power grip (mean value of both hands)	1 HQ (22) and 3 LQ (23,33,35)		Limited
Power grip (dominant hand)	2 HQ (21,29) and 1 LQ (27)		Limited
Power grip (nondominant hand)	2 HQ (21,29)		Limited
Lateral pinch (mean value of both hands)	1 HQ (22) and 1 LQ (23)		Limited
Tip pinch (mean value of both hands)	1 HQ (22) and 1 LQ (23)		Limited
Tripod pinch (mean value of both hands)	1 HQ (22) and 1 LQ (23)		Limited
Range of motion			
Dominant hand fingers flexion deficit	1 HQ (21) and 1 LQ (30)		Limited
Nondominant hand fingers flexion deficit		1 HQ (21) and 1 LQ (30)	Limited
Disease activity			
Composite measure	2 HQ (2,21) and 5 LQ (19,23,24,26,34)	1 LQ (11)	Limited
Swollen joint count	1 HQ (2) and 1 LQ (24)	1 LQ (31)	Conflicting
Tender joint count	1 HQ (2) and 2 LQ (24,31)		Limited
ESR	2 LQ (19,24)	1 HQ (2) and 2 LQ (25,35)	Conflicting
CRP	1 HQ (2) and 2 LQ (24,25)		Limited
PGA	1 HQ (2) and 1 LQ (24)		Limited
Deformity			
Presence of deformities in dominant hand	2 LQ (20,32)		Limited
Presence of deformities in both hands	1 LQ (19)	1 HQ (21)	Conflicting
Ulnar deviation of dominant hand	1 HQ (29) and 2 LQ (27,34)		Limited
Ulnar deviation of nondominant hand	1 HQ (29) and 1 LQ (34)		Limited
Mental health	1 HQ (2) and 2 LQ (20,24)		Limited
Vitality	1 HQ (2) and 1 LQ (24)	One LQ (20)	Conflicting
Structural damage (radiographic)	1 HQ (22) and 3 LQ (23,25,27)		Limited
Pain			
Bodily pain (VAS)	2 LQ (23,35)		Limited
Bodily pain (SF-36)	1 HQ (2) and 1 LQ (24)		Limited
Hand pain during activity (SODA tasks)	2 LQ (30,35)	One LQ (20)	Conflicting
Hand pain during activity (VAS)	2 HQ (2,21)		Limited
Hand pain at rest (VAS)	1 HQ (2) and 1 LQ (25)	One HQ (21) and 1 LQ (30)	Conflicting
Stiffness			
Duration	1 HQ (21)	One LQ (25)	Conflicting
Intensity	1 HQ (21)	One LQ (20)	Conflicting

* HQ = high quality; LQ = low quality; ESR = erythrocyte sedimentation rate; CRP = C-reactive protein; PGA = patient global assessment; VAS = visual analog scale; SF-36 = Short-Form 36 Health Survey; SODA = Sequential Occupational Dexterity Assessment.

sedimentation rate (ESR), presence of deformities in both hands, vitality, sum of painful Sequential Occupational Dexterity Assessment tasks, hand pain intensity at rest, stiffness (intensity and duration), and hand function.

For functional status factors, limited evidence was found that would suggest that reduced hand function is associated with low functional status level (physical, social, and emotional function). In considering personal factors, there was conflicting evidence for the association between age and hand function. Seven studies reported that the difference between men and women with regard to hand function was not statistically significant; consequently, limited evidence is documented in the best-evidence synthesis. Regarding environmental factors, limited evidence was found that work activity is not associated with hand function. Finally, for health-related factors there is conflicting evidence for the association between the factors of disease duration, general health status, and hand function. Also, limited evidence was found that the level of RF is not associated with hand function.

DISCUSSION

This is the first systematic review that provides an overview of factors associated with hand functional disability in people with RA. From reviewing the literature, there is a lack of consistency with the variation in measures used in reporting hand impairments, leading to a limited ability to make comparison between studies. For instance, measuring and reporting ROM was inconsistent between the included studies, and a majority of the studies did not provide a clear description of what is being measured (i.e., active or passive ROM). In addition, there were deficiencies associated with hand impairment outcome measurements, such as subjectively reporting hand

Table 4. Overview of findings regarding associations of functional status, personal, environmental, and health-related factors with hand function*

Factors	Association found	No association found	Level of evidence
Functional status			
Physical functioning (HAQ)	2 HQ (2,21) and 5 LQ (11,23,24,27,30)	1 LQ (19)	Limited
Physical functioning (SF-36)	1 HQ (2) and 1 LQ (24)		Limited
Social functioning	1 HQ (2) and 1 LQ (24)		Limited
Emotional role	1 HQ (2) and 1 LQ (24)		Limited
Personal factors			
Age	3 LQ (19,28,31)	4 LQ (20,26,30,34,35)†	Conflicting
Sex		1 HQ (21) and 6 LQ (19,20,26,30,31,34,35)†	Limited
Environmental factors			
Work activity		3 LQ (19,20,32)	Limited
Health related factors			
Health condition			
Disease duration	7 LQ (19,23,25,28,30,31,34,35)†	2 HQ (3,21) and 2 LQ (20,26)	Conflicting
Rheumatoid factor		3 LQ (19,23,32)	Limited
General state of health	1 HQ (2) and 1 LQ (24)	1 LQ (20)	Conflicting

* HAQ = Health Assessment Questionnaire; HQ = high quality; LQ = low quality; SF-36 = Short Form 36 Health Survey.
† Studies 34 and 35 and were considered as 1 body of evidence because both studies reported the findings from the same sample of RA patients with regard to the association between hand function and disease duration, age, and sex.

deformities with lack of detail about assessment or grading methods. Based on these observations, it is evident that protocols for assessments of hand impairments in the RA population need to be agreed and implemented. Consistency in reporting hand function is also now required.

Although quality of reporting was satisfactory for the majority of the studies identified in this review, almost all studies failed to account for and minimize systematic errors. Therefore, conclusions from this review could be at risk of bias due to weaknesses in those studies included. Improving selection and reporting of study participants, especially response rates and information about nonrespondents would address these biases and should be incorporated into future research.

This review showed that studies that consider hand disability in people with RA reported predominantly on body structure and function factors. There was a lack of consideration of, and investigation into, personal and environmental factors when considering the impact of RA on hand function. Many factors of body structure and function were significantly associated with hand functional disability. Importantly, grip strength is routinely recorded in rheumatology clinical trials. This is appropriate and relevant because grip strength is a valid indicator of disability (36), has been shown to predict later hand function (37), and contributes to hand function improvements (38) in people with RA. In this review, more than half of the included studies assessed the association between power grip strength and hand function, and all reported statistically significant relationships, regardless of the measurement or reporting method. This confirms what has been suggested by the findings of the present review—that power grip strength is a valid and reliable indicator of hand function in RA population and clinicians can have confidence in this finding for using it in clinical practice.

Disease activity variables have been found to be associated with hand function, except for ESR and swollen joint count, for which there was conflicting evidence. An explanation of this observation may be due to the fact that different hand function assessment tools cover different spectrum of functioning (39), and people with RA show unique and different clinical presentations; thus, no single disease activity variable can accurately detect every patient's disease activity at any given point in time (40). The results of the current review suggest that disease activity is a modifiable parameter that significantly contributes to hand function.

Pain in RA is the main treatment target for patients and clinicians (41). Results from studies of RA cohorts that were conducted during an era when biologic treatments were available demonstrated that pain still remains a problem and influences the performance of valued life activities (7,42). In the present review, limited evidence was found that higher intensity of bodily pain and hand pain during activity were associated with an increase of hand functional disability, and conflicting evidence was found for the association with hand pain at rest. This indicates that hand pain during activity may substantially contribute to hand functional disability. In line with these results, a recent longitudinal report on a Swedish RA cohort indicated that general pain was higher than hand pain during activity, which in turn was higher than hand pain at rest (7).

The studies included in the present review indicate an association between structural damage and hand function, and that an increase of radiographic joint damage is correlated with an increase of hand functional limitations. However, in agreement with recent evidence that radiographic joint damage is less influential in the context of modern treatment (43), the relative importance of structural damage may be of less importance in future research.

The fact that conflicting evidence was found for the association between hand function and hand stiffness duration and intensity is remarkable, since stiffness is a symptom widely experienced by patients with RA. Besides methodologic explanations (i.e., only cross-sectional studies with relatively small sample size), evidence from a systematic review of stiffness measures demonstrated that there is limited evidence to support the validity of the currently available stiffness measures (44). Furthermore, qualitative evidence suggested that patients with RA experience stiffness differently and reported stiffness in terms of impact rather than by duration or severity (45).

In examining functional status, there was limited evidence stating that reduced hand function is associated with poor physical function, social function, and emotional role. The association found between hand function and functional status measures may indicate that hand disability influences both the activity and participation level of functioning. This is because physical function measures such as Health Assessment Questionnaire (HAQ), are measuring activity limitations, whereas social functioning and emotional role scales measure participation restrictions (46).

Few personal factors have been explored for the association with hand function. Out of 12 personal factors identified as meaningful for general functioning in RA (47), only coping could be categorized as a personal factor, and was included in 1 study as a factor for hand function (34); thus, it was not included in the best-evidence synthesis. Important personal factors in relation to specific hand functional outcome are not identified. Identifying the role of these factors as determinants and modifiers of hand function can facilitate the process when evaluating and planning interventions for people with RA.

The findings of a qualitative study recruiting RA patients with hand deformity proposed that environmental factors play a significant role in hand-related activity limitations and participation restriction (48). However, in this review the impact of only a few environmental factors have been explored in relation to hand function. One factor, namely work activity, was included in the best-evidence synthesis, since it was assessed in 3 independent studies (19,20,32). However, the relative importance and influence of environmental factors might vary according to the settings and culture. For instance, low-income countries tend to have limited or fewer resources in terms of health care system, compared with high-income countries. Moreover, social support and beliefs about health disability may differ across countries. Considering these issues, important environmental factors in relation to hand functional outcomes in specific cultures and settings should be identified.

For health-related factors, conflicting evidence was found regarding the relationship between hand function and disease duration and general health status. Long disease duration was expected to be significantly associated with poor hand function, since hand impairments are prevalent and deteriorate over time in patients with long disease duration (3,8). Alongside, the limitations mentioned earlier concerning the methodologies of the included studies, a possible explanation for this finding might be related to the fact that patients with long disease duration may have adapted to their situation and they do not expect any effective treatment to be available (3). Cross-sectional studies have concluded that disease activity is the major explanatory factor for activity limitations in RA patients, with disease duration less than 10 years (36). Accordingly, disease duration may be an irrelevant factor to consider when evaluating hand function, particularly with disease duration <10 years.

This review is not without limitations. Only the first author screened the titles and abstracts. However, citations were only considered irrelevant if the title or abstract did not include any information on hand function outcomes. Moreover, the review team were consulted where any ambiguity arose during the selection process. Therefore, the possibility of removing relevant studies was low. This review was limited by the wide variation in the included studies' sample sizes (range 25–200). Accordingly, sample size may affect the results of associations reported in the included studies; small associations are significant in studies with a large sample size and not in studies with a small sample size. The studies included in this review did not all present the size of the association within their statistical analysis and reporting, so it is difficult to preclude that the results are biased by this. The studies included have used self-reported and performance-based measures of hand function or both. This probably influenced the results of this review, since, performance-based measures cover a narrow spectrum of hand functioning (39), and may not accurately reflect hand abilities (49). Furthermore, performance based and self-reported measures of hand function are not strongly associated (50). Future research may benefit from stratifying outcomes rather than combining them. The quality assessment tool (AXIS) used in this review was developed based on literature and methodologic standards; however, further studies are required to explore its test–retest reliability. Attention should be given to the disagreement (27% and 19%) between the reviewers on the methodologic quality of the study. Reducing the scoring options into "yes," "no," instead of including "don't know" may increase the simplicity of use of the AXIS and may minimize the disagreements between reviewers. The grey literature or unpublished studies were not searched as there are few studies focusing on hand function in RA patients (19); therefore, the number of extra studies identified by grey literature would also be small. Studies written in English were selected and included in the review. The percentage of all articles written in other languages was small (8%); consequently, it is unlikely that this percentage would introduce language bias into the review. Finally, of the 20 articles included in the review, 1 author (JA) involved in the present review authored 2 articles. However, to ensure that the quality assessment process was unbiased, quality assessment of these 2 articles were completed by the first and second author.

This systematic review has summarized current evidence for the factors associated with hand function in RA patients. It has also underlined areas where methodology is lacking and potential directions for future research. There are numerous factors where current evidence is limited or conflicting. These factors can be classified as modifiable (e.g., disease activity, hand strength, psychosocial factors) and nonmodifiable factors (e.g., age, sex, structural damage). Focusing on nonmodifiable factors offers little added value to improve hand function in people with RA. Therefore, modifiable factors should be of key concern as some of these factors can be modified with specific strategies and interventions. Before new strategies and interventions are established to improve hand function in people with RA, well-designed longitudinal studies need to be performed to get more understanding in the causality between factors and hand function. Important sociodemographic factors in relation to hand function in patients with RA need more considerations by future research.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Mr. Arab Alkabeya 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. Arab Alkabeya, Hughes, Adams.
Acquisition of data. Arab Alkabeya, Hughes, Adams.
Analysis and interpretation of data. Arab Alkabeya, Hughes, Adams.

REFERENCES

1. Gibofsky A. Overview of epidemiology, pathophysiology, and diagnosis of rheumatoid arthritis. Am J Manag Care 2012;18:S295–302.
2. Durmus D, Uzuner B, Durmaz Y, Bilgici A, Kuru O. Michigan Hand Outcomes Questionnaire in rheumatoid arthritis patients: relationship with disease activity, quality of life, and handgrip strength. J Back Musculoskelet Rehabil 2013;26:467–73.
3. Horsten NC, Ursum J, Roorda LD, van Schaardenburg D, Dekker J, Hoeksma AF. Prevalence of hand symptoms, impairments and activity limitations in rheumatoid arthritis in relation to disease duration. J Rehabil Med 2010;42:916–21.
4. Vliet Vlieland TP, van der Wijk TP, Jolie IM, Zwinderman AH, Hazes JM. Determinants of hand function in patients with rheumatoid arthritis. J Rheumatol 1996;23:835–40.
5. Bodur H, Yilmaz Ö, Keskin D. Hand disability and related variables in patients with rheumatoid arthritis. Rheumatol Int 2006;26:541–4.
6. Yamanaka H, Inoue E, Singh G, Tanaka E, Nakajima A, Taniguchi A, et al. Improvement of disease activity of rheumatoid arthritis patients from 2000 to 2006 in a large observational cohort study IORRA in Japan. Mod Rheumatol 2007;17:283–9.
7. Thyberg I, Dahlstrom O, Bjork M, Stenstrom B, Adams J. Hand pains in women and men in early rheumatoid arthritis, a one year follow-up after diagnosis. The Swedish TIRA project. Disabil Rehabil 2016;39:291–300.
8. Toyama S, Tokunaga D, Fujiwara H, Oda R, Kobashi H, Okumura H, et al. Rheumatoid arthritis of the hand: a five-year longitudinal analysis of clinical and radiographic findings. Mod Rheumatol 2014;24:69–77.
9. Romero-Guzman AK, Menchaca-Tapia VM, Contreras-Yanez I, Pascual-Ramos V. Patient and physician perspectives of hand function in a cohort of rheumatoid arthritis patients: the impact of disease activity. BMC Musculoskelet Disord 2016;17:392.
10. Bjork MA, Thyberg IS, Skogh T, Gerdle BU. Hand function and activity limitation according to health assessment questionnaire in patients with rheumatoid arthritis and healthy referents: 5-year followup of predictors of activity limitation (The Swedish TIRA Project). J Rheumatol 2007;34:296–302.
11. Eberhardt K, Sandqvist G, Geborek P. Hand function tests are important and sensitive tools for assessment of treatment response in patients with rheumatoid arthritis. Scand J Rheumatol 2008;37:109–12.
12. Hammond A. What is the role of the occupational therapist? Best Pract Res Clin Rheumatol 2004;18:491–505.
13. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315–24.
14. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO III, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2010;62:2569–81.
15. Downes MJ, Brennan ML, Williams HC, Dean RS. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). BMJ Open 2016;6:e011458.
16. Veenhof C, Huisman PA, Barten JA, Takken T, Pisters MF. Factors associated with physical activity in patients with osteoarthritis of the hip or knee: a systematic review. Osteoarthritis Cartilage 2012;20:6–12.
17. Slavin RE. Best evidence synthesis: an intelligent alternative to meta-analysis. J Clin Epidemiol 1995;48:9–18.
18. Van Tulder M, Furlan A, Bombardier C, Bouter L, Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. Spine 2003;28:1290–9.
19. Belghali S, Ben Abderrahim K, Mahmoud I, Baccouche K, El Amri N, Zeglaoui H, et al. Brief Michigan Hand Outcomes Questionnaire in rheumatoid arthritis: a cross-sectional study of 100 patients. Hand Surg Rehabil 2017;36:24–9.
20. Andrade JA, Brandão MB, Pinto MR, Lanna CC. Factors associated with activity limitations in people with rheumatoid arthritis. Am J Occup Ther 2016;70:1–7.
21. Bircan Ç, Erdinç Gündüz N, Tekgül A, Çetin P, Önen F, Kizil R, et al. Grip ability test in rheumatoid arthritis patients: relationship with disease activity and hand-specific self-report questionnaires. Arch Rheumatol 2014;29:160–6.
22. Dogu B, Kuran B, Yilmaz F, Usen A, Sirzai H. Is hand bone mineral density a marker for hand function in patients with established rheumatoid arthritis? The correlation among bone mineral density of the hand, radiological findings and hand function. Clin Rheumatol 2013;32:1177–83.
23. Dedeoğlu M, Gafuroğlu Ü, Yilmaz Ö, Bodur H. The relationship between hand grip and pinch strengths and disease activity, articular damage, pain, and disability in patients with rheumatoid arthritis. Turk J Rheumatol 2013;28:69–77.
24. Aktekin LA, Eser F, Başkan BM, Sivas F, Malhan S, Öksüz E, et al. Disability of Arm Shoulder and Hand Questionnaire in rheumatoid arthritis patients: relationship with disease activity, HAQ, SF-36. Rheumatol Int 2011;31:823–6.
25. Özeri Z, Duyur Çakýt B, Taþkýn S, Genç H, Saraçoðlu M, Rana Erdem H. The relationships among functional impairment, disability and articular damage in rheumatoid hand. J PMR Sci 2008;2:53–8.

26. Birtane M, Kabayel DD, Uzunca K, Unlu E, Tastekin N. The relation of hand functions with radiological damage and disease activity in rheumatoid arthritis. Rheumatol Int 2008;28:407–12.
27. Sahin F, Kotevoglu N, Taspinar S, Yilmaz F, Kuran B. Comparison of functional disability scales and their relevance to radiological progression in patients with rheumatoid arthritis in remission. Clin Exp Rheumatol 2006;24:540–5.
28. Adams J, Burridge J, Mullee M, Hammond A, Cooper C. Self-reported hand functional ability measured by the DASH in individuals with early rheumatoid arthritis. Hand Ther 2005;10:21–4.
29. Adams J, Burridge J, Mullee M, Hammond A, Cooper C. Correlation between upper limb functional ability and structural hand impairment in an early rheumatoid population. Clin Rehabil 2004;18:405–13.
30. O'Connor D, Kortman B, Smith A, Ahern M, Smith M, Krishnan J. Correlation between objective and subjective measures of hand function in patients with rheumatoid arthritis. J Hand Ther 1999;12:323–9.
31. Jonsson B, Larsson SE. Hand function and total locomotion status in rheumatoid arthritis: an epidemiologic study. Acta Orthop Scand 1990;61:339–43.
32. Erol AM, Ceceli E, Ramadan SU, Borman P. Effect of rheumatoid arthritis on strength, dexterity, coordination and functional status of the hand: relationship with magnetic resonance imaging findings. Acta Reumatologica Portuguesa 2016;41:328–37.
33. Kinikli Gİ, Şahin A, Güney H, Yüksel İ, Kinikli G. Investigation of grip strength and upper extremity functional disability in patients with rheumatoid arthritis. J Exerc Ther Rehabil 2016;3:60–5.
34. Van Lankveld W, Näring G, van 't Pad Bosch P, van de Putte L. Behavioral coping and physical functioning: the effect of adjusting the level of activity on observed dexterity. J Rheumatol 1999;26:1058–64.
35. Van Lankveld WG, van 't Pad Bosch P, van de Putte L. Predictors of changes in observed dexterity during one year in patients with rheumatoid arthritis. Br J Rheumatol 1998;37:733–9.
36. Toussirot E. Predictive factors for disability as evaluated by the health assessment questionnaire in rheumatoid arthritis: a literature review. Inflamm Allergy Drug Targets 2010;9:51–9.
37. Björk MA, Thyberg IS, Skogh T, Gerdle BU. Hand function and activity limitation according to health assessment questionnaire in patients with rheumatoid arthritis and healthy referents: 5-year followup of predictors of activity limitation (the Swedish TIRA project). J Rheumatol 2007;34:296–302.
38. Hall AM, Copsey B, Williams M, Srikesavan C, Lamb SE, on behalf of the Sarah Trial Team. Mediating effect of changes in hand impairments on hand function in patients with rheumatoid arthritis: exploring the mechanisms of an effective exercise program. Arthritis Care Res (Hoboken) 2017;69:982–8.
39. Stamm TA, Cieza A, Machold KP, Smolen JS, Stucki G. Content comparison of occupation-based instruments in adult rheumatology and musculoskeletal rehabilitation based on the International Classification of Functioning. Disability and Health. Arthritis Rheum 2004;51:917–24.
40. Salomon-Escoto KI, Gravallese EM, Kay J. Assessment of control of rheumatoid arthritis disease activity. Best Pract Res Clin Rheumatol 2011;25:497–507.
41. Carr A, Hewlett S, Hughes R, Mitchell H, Ryan S, Carr M, et al. Rheumatology outcomes: the patient's perspective. J Rheumatol 2003;30:880–3.
42. Ahlstrand I, Bjork M, Thyberg I, Falkmer T. Pain and difficulties performing valued life activities in women and men with rheumatoid arthritis. Clin Rheumatol 2015;34:1353–62.
43. Carpenter L, Norton S, Nikiphorou E, Jayakumar K, McWilliams DF, Rennie KL, et al. Reductions in radiographic progression in early rheumatoid arthritis over twenty-five years: changing contribution from rheumatoid factor in two multicenter UK inception cohorts. Arthritis Care Res (Hoboken) 2017;69:1809–17.
44. Van Tuyl LH, Lems WF, Boers M. Measurement of stiffness in patients with rheumatoid arthritis in low disease activity or remission: a systematic review. BMC Musculoskelet Disord 2014;15:28.
45. Halls S, Dures E, Kirwan J, Pollock J, Baker G, Edmunds A, et al. Stiffness is more than just duration and severity: a qualitative exploration in people with rheumatoid arthritis. Rheumatology (Oxford) 2015;54:615–22.
46. Stucki G, Cieza A. The International Classification of Functioning, Disability and Health (ICF) core sets for rheumatoid arthritis: a way to specify functioning. Ann Rheum Dis 2004;63:40–5.
47. Dur M, Coenen M, Stoffer MA, Fialka-Moser V, Kautzky-Willer A, Kjeken I, et al. Do patient-reported outcome measures cover personal factors important to people with rheumatoid arthritis? A mixed methods design using the International Classification of Functioning, Disability and Health as frame of reference. Health Qual Life Outcomes 2015;13:27.
48. Nicklasson M, Jonsson H. Experience of participation as described by people with hand deformity caused by rheumatic disease. Br J Occup Ther 2012;75:29.
49. Fowler NK, Nicol AC. Functional and biomechanical assessment of the normal and rheumatoid hand. Clin Biomech (Bristol, Avon) 2001;16:660–6.
50. Rallon CR, Chen CC. Relationship between performance-based and self-reported assessment of hand function. Am J Occup Ther 2008;62:574–9.

Arthritis Care & Research
Vol. 71, No. 11, November 2019, pp 1482–1487
DOI 10.1002/acr.23825

American College of Rheumatology
Empowering Rheumatology Professionals

BRIEF REPORT

Translating Treatment Effects Between Rheumatoid Arthritis Activity Measures and American College of Rheumatology Responses in Direct Comparison Trials

Abhijit Dasgupta and Michael M. Ward

Objective. Direct comparison trials in rheumatoid arthritis (RA) increasingly use changes in continuous disease activity measures as endpoints. However, the between-arm differences in these scores that are clinically meaningful are uncertain. To aid interpretation of clinical trials that use the Disease Activity Score in 28 joints (DAS28) or Simplified Disease Activity Index (SDAI) as endpoints, we developed statistical equivalences between changes in these measures and American College of Rheumatology (ACR) responses.

Methods. For superiority trials, we computed the minimal detectable difference in DAS28 changes and SDAI changes that correspond to the ACR criteria for 20% improvement (ACR20) and 50% improvement (ACR50) responses at the same type I and type II errors and same sample size. For noninferiority trials, we computed noninferiority margins that were statistically equivalent across measures. Standard deviations of the changes in the DAS28 and SDAI from a recent observational study were used as the basis of calculations in our examples.

Results. In the base scenario with type 1 error 0.05 and power 0.80, a trial with 300 subjects per arm would detect a 0.31-point difference in mean DAS28 change scores and 3.71-point difference in mean SDAI change scores as statistically equivalent to an absolute difference of 11% in ACR20 between treatment arms. We developed a web-based utility that provides equivalent differences among these measures for customized sample sizes, error rates, and standard deviations of the DAS28 and SDAI between-arm differences.

Conclusion. The DAS28 and SDAI responses can be related to statistically equivalent changes in ACR responses, which can aid the interpretation of trials that use these measures.

INTRODUCTION

With the proliferation of new treatments for rheumatoid arthritis (RA), comparative trials that directly assess the relative efficacy of 2 or more active drugs or treatment strategies have assumed great importance in informing treatment decisions (1). Many such trials have used the American College of Rheumatology (ACR) response criteria as the primary endpoint, even though these criteria were developed to distinguish active treatments from placebo (2). Increasingly, direct comparison trials have used changes in an RA activity measure as the primary endpoint, based in part on the perception that continuous measures may afford greater statistical power (3,4). Continuous measures also allow distinction of responses among patients who have the same level of ACR response.

Comparative trials often test the difference in changes between study arms over the duration of the trial. This comparison entails 2 mental steps: appreciation of the change over time within study arms, and then comparison of these changes between arms. The degree of difference in these change scores that is clinically meaningful has not been established, which complicates the interpretation of these trials. This comparison is even more challenging in direct comparison trials than in placebo-controlled trials because each treatment is likely efficacious to some degree. In contrast, threshold outcomes such as the ACR criteria for 20% improvement (ACR20) directly provide a measure of within-arm improvement, and the only task is to compare the proportion of responders between study arms. Years of use have provided clinicians with intuition on the clinical meaning of ACR responses.

Supported by the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases Intramural Research Program grant ZIA-AR-041153).

Abhijit Dasgupta, PhD, Michael M. Ward, MD, MPH: National Institute of Arthritis and Musculoskeletal and Skin Diseases, Bethesda, Maryland.

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

Address correspondence to Michael M. Ward, MD, MPH, NIAMS/NIH, Building 10 CRC, Room 4-1339, 10 Center Drive, Bethesda, MD 20892. E-mail: wardm1@mail.nih.gov.

Submitted for publication May 11, 2018; accepted in revised form December 18, 2018.

SIGNIFICANCE & INNOVATIONS

- To our knowledge, this is the first study to report how changes in the Disease Activity Score in 28 joints (DAS28) and the Simplified Disease Activity Index (SDAI) translate to equivalent American College of Rheumatology (ACR) responses in rheumatoid arthritis.
- We developed a web-based utility that allows users to derive statistical equivalences between DAS28 changes, SDAI changes, and ACR response criteria for both superiority trials and noninferiority trials.

Translating the treatment effects associated with continuous RA activity measures to corresponding ACR responses could aid the interpretation of trials that use an RA activity measure as the endpoint. Here, we used statistical equivalences to develop comparable detectable treatment effects between changes in the Disease Activity Score in 28 joints (DAS28) and Simplified Disease Activity Index (SDAI) and the ACR response criteria to provide context for changes in the continuous measures. When reading trials that compare treatments using the DAS28 or SDAI, rheumatologists can use these tools to produce ACR responses of the same effect.

MATERIALS AND METHODS

The statistical design of a typical 2-arm trial involves 4 quantities: the number of subjects (N) per arm; the type I error (α); the type II error (β); and the true difference (Δ) in outcome between treatment arms to be detected (5). Here, Δ denotes either the true difference in average DAS28 or SDAI changes or the true difference in ACR responses between trial arms. In the base analysis, we set α at 0.05 and β at either 0.2 or 0.1. For a range of sample sizes (N), we determined the minimum detectable differences in the ACR20, the ACR criteria for 50% improvement (ACR50), the DAS28, and the SDAI between the 2 arms that would give the stated error rates. This correspondence provides a data-driven way to understand equivalent magnitudes of improvement among RA measures.

Clinical trials may test either superiority or noninferiority (i.e., drug B is no worse than drug A within some margin). The noninferiority test is defined by H_0: $\Delta \leq \delta$ versus H_1: $\Delta > \delta$ for some preassigned noninferiority margin $\delta < 0$. The noninferiority margin is the maximal difference between treatments that one would still accept as indicating that one treatment (typically a new drug) was not inferior to the comparison treatment (typically the standard drug). Treatment differences more extreme than this margin would imply that the new treatment is not noninferior to the standard treatment (see Supplementary Figure 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23825/abstract).

We computed equivalences among RA outcome measures for both superiority and noninferiority designs. For simplicity, we considered 2-arm randomized trials with equal-sized arms. We examined outcomes at 1 timepoint, comparing between-arm differences in the mean change in DAS28 or SDAI from the start to end of the trial and the proportion of subjects meeting ACR20 and ACR50 criteria at the end of the trial. We used the observed

Table 1. Equivalent differences in Disease Activity Score in 28 joints (DAS28) change, Simplified Disease Activity Index (SDAI) change, and American College of Rheumatology criteria for 20% improvement (ACR20) responses between arms in a superiority trial for different sample sizes at statistical power levels of 80% and 90%*

				Arm 1 ACR20			Arm 2 ACR20		
Power	N	ΔDAS28	ΔSDAI	p_0	p_1	Δ	p_0	p_1	Δ
0.80	20	1.22	13.93	50	90	40	70	100	30
	30	0.99	11.27	50	83	33	70	96	26
	40	0.85	9.72	50	79	29	70	94	24
	50	0.76	8.67	50	77	27	70	92	22
	100	0.53	6.10	50	69	19	70	86	16
	200	0.38	4.30	50	64	14	70	82	12
	300†	0.31†	3.51†	50†	61†	11†	70†	80†	10†
	400	0.27	3.04	50	60	10	70	79	9
	500	0.24	2.72	50	59	9	70	78	8
0.90	20	1.41	16.12	50	94	44	70	100	30
	30	1.14	13.04	50	87	37	70	99	29
	40	0.98	11.25	50	83	33	70	96	26
	50	0.88	10.03	50	80	30	70	94	24
	100	0.62	7.06	50	72	22	70	88	18
	200	0.44	4.98	50	66	16	70	84	18
	300†	0.36†	4.06†	50†	63†	13†	70†	81†	11†
	400	0.31	3.52	50	61	11	70	80	10
	500	0.27	3.14	50	60	10	70	79	9

* N denotes sample size per arm, and p_0 and p_1 denote the percentages for the reference drug and the comparator drug, respectively. Calculations were based on standard deviations of 1.34 and 15.32 for change in DAS28 and SDAI, respectively. Associations will differ with the use of other values for standard deviation.
† Example from the text.

standard deviations of the changes in DAS28 and SDAI from a prospective longitudinal study of treatment responses in patients with active RA for the power calculations in our examples (6). Our study examined longitudinal changes in RA activity in 250 patients with active RA before and after escalation of antirheumatic medications. The standard deviations of the DAS28 change and SDAI change were the only data from this study needed for the present analysis.

We initially set the ACR20 response of the reference drug (p_0) at 50% and repeated the analysis at p_0 of 40%, 60%, and 70% to determine the effect of different levels of response on the power relationships, given that responses in this range are commonly observed in trials. Similarly, we set the initial ACR50 response of the reference drug at 20% and subsequently varied this to 30%, 40%, and 50%. We used the 2-sample *t*-test for continuous outcomes and the 2-sample test of proportions for the ACR20 and ACR50 (5). All computations were based on 2-sided tests.

For noninferiority trials, we computed the noninferiority margin δ for each outcome to establish noninferiority at the specified error rates, assuming there truly was no difference between the 2 drugs (i.e., $\Delta_{ACR20}=\Delta_{ACR50}=\Delta_{DAS28}=\Delta_{SDAI}=0$). This established correspondences of the noninferiority margins for the different measures assuming no true difference. We used R packages pwr and TrialSize for computations (7–9).

RESULTS

Superiority trials. The mean ± SD DAS28 change with treatment was 1.31 ± 1.34, and mean SDAI change was 14.91 ± 15.32. These standard deviations were used in the power calculations for the base analysis, the results of which are presented in the tables and figures.

Table 1 shows the minimum detectable differences for each outcome for different sample sizes for a null hypothesis of no differences between treatments. For example, based on the

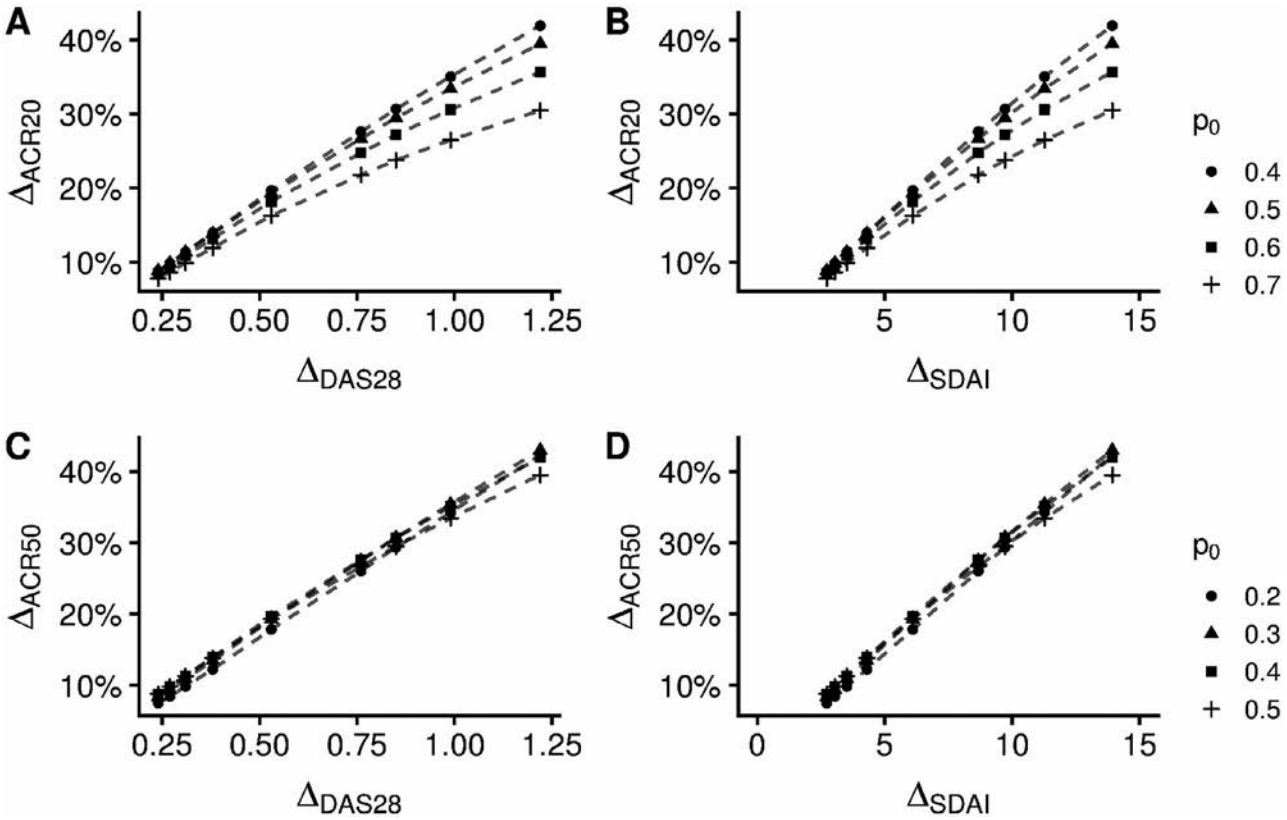

Figure 1. Equivalent detectable differences at 80% statistical power between the American College of Rheumatology criteria for 20% improvement (ACR20) responses and between-arm differences in the Disease Activity Score in 28 joints (DAS28) (**A**) and between-arm differences in the Simplified Disease Activity Index (SDAI) at varying levels of ACR20 response in the reference treatment arm (p_0) (**B**). The association between change in mean DAS28 difference and change in mean SDAI difference and corresponding changes in ACR20 when ACR20 response was either 50% (▲) or 70% (+) correspond to data from Table 1. For example, a between-group difference in DAS28 change of 0.53 would correspond to an absolute difference of 19% in ACR20 responses when the ACR20 response in the reference group was 50%, and it would correspond to an absolute difference of 16% in ACR20 responses when the ACR20 response in the reference group was 70%. Equivalent detectable differences at 80% statistical power between the ACR criteria for 50% improvement (ACR50) responses and between-arm differences in the DAS28 (**C**) and between-arm differences in the SDAI at varying levels of ACR50 response in the reference treatment arm (p_0) (**D**).

observed variability of the outcomes, a trial with 300 subjects per arm would be able to detect a 0.31-point difference in mean DAS28 change scores (i.e., mean DAS28 change in arm 1 minus mean DAS28 change in arm 2 = 0.31) as significantly different with a type I error of 5% and a power of 80%. A trial of the same size with an ACR20 response of 50% in the reference drug arm (i.e., p_0 = 50%) would detect a difference of no less than 11% in ACR20 between the 2 arms as statistically significant (i.e., ACR20 responses of 50% versus 61%). In other words, a detectable difference in mean DAS28 change of 0.31 would be statistically equivalent to an absolute difference in ACR20 responder proportions of 11%. Similarly, a between-arm mean difference in SDAI of 3.51 points was equivalent to an absolute difference of 11% in the proportion of ACR20 responders between groups. Increasing power from 80% to 90% had only a small effect on equivalences among measures (Table 1).

If the ACR20 response for the reference drug were 70% instead of 50%, a trial of 300 subjects per arm would be able to detect a 0.31-point difference in mean DAS28 change scores as significantly different, with a type 1 error of 5% and a power of 80%, and equivalently a difference in ACR20 responses of 10% as statistically significant (Table 1). The proportion of ACR20 responders in the reference arm had little effect on the equivalences at smaller DAS28 or SDAI differences, but it had modest effects at larger detectable differences (Figure 1). At large detectable differences in mean DAS28 changes (or mean SDAI changes), the ACR response associated with a given DAS28 (or SDAI) change would be smaller in studies where the reference drug had larger ACR responses.

Equivalent detectable differences were similar between ACR20 and ACR50 responses in absolute terms, although ACR50 differences were larger in relative terms (Table 2). For example, a detectable difference in DAS28 of 0.30 between trial arms was equivalent to an absolute difference in ACR50 responses of 10%, assuming an ACR50 response of 20% in the reference arm. The proportion of ACR50 responders in the reference arm had little effect on the equivalences regardless of the DAS28 or SDAI difference (Figure 1).

The equivalent differences of DAS28, SDAI, and ACR responses also depend on the standard deviations of the changes in DAS28 and SDAI. With larger standard deviations, the change in DAS28 or SDAI that was statistically equivalent to a given ACR response was larger (see Supplementary Figure 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23825/abstract).

Noninferiority trials. Equivalent noninferiority margins for ACR20 and ACR50 responses and changes in the DAS28 and SDAI are shown in Table 3, based on the assumption of no difference between treatments. With 50 subjects per arm, to show noninferiority between drugs with 80% power ACR20 would require a noninferiority margin of –25%, while the DAS28 would need a margin of –0.67 and the SDAI a margin of –7.62. This means that with 50 subjects per arm, a difference between responses as large as –7.62 in SDAI would not be sufficient to declare the 2 arms not

Table 2. Equivalent differences in Disease Activity Score in 28 joints (DAS28) change, Simplified Disease Activity Index (SDAI) change, and American College of Rheumatology criteria for 50% improvement (ACR50) responses between arms in a superiority trial for different sample sizes at statistical power levels of 80% and 90%*

				Arm 1 ACR50			Arm 2 ACR50		
Power	N	ΔDAS28	ΔSDAI	p_0	p_1	Δ	p_0	p_1	Δ
0.80	20	1.22	13.93	20	62	42	40	82	42
	30	0.99	11.27	20	54	34	40	75	35
	40	0.85	9.72	20	49	29	40	71	31
	50	0.76	8.67	20	46	26	40	68	28
	100	0.53	6.10	20	38	18	40	60	20
	200	0.38	4.30	20	32	12	40	54	14
	300†	0.31†	3.51†	20†	30†	10†	40†	51†	11†
	400	0.27	3.04	20	28	8	40	50	10
	500	0.24	2.72	20	28	8	40	49	9
0.90	20	1.41	16.12	20	68	48	40	87	47
	30	1.14	13.04	20	60	40	40	80	40
	40	0.98	11.25	20	54	34	40	75	35
	50	0.88	10.03	20	50	30	40	72	32
	100	0.62	7.06	20	41	21	40	63	23
	200	0.44	4.98	20	34	14	40	56	16
	300†	0.36†	4.06†	20†	32†	12†	40†	53†	13†
	400	0.31	3.52	20	30	10	40	51	11
	500	0.27	3.14	20	29	9	40	50	10

* N denotes sample size per arm, and p_0 and p_1 denote the percentages for the reference drug and the comparator drug, respectively. Calculations were based on standard deviations of 1.34 and 15.32 for change in DAS28 and SDAI, respectively. Associations will differ with the use of other values for standard deviation.
† Example from the text.

Table 3. Noninferiority margins required for American College of Rheumatology criteria for 20% improvement (ACR20) and ACR criteria for 50% improvement (ACR50) responses, change in Disease Activity Score in 28 joints (DAS28), and change in Simplified Disease Activity Index (SDAI) to detect noninferiority if there is truly no difference between the study arms, assuming a type I error of 5% and power of either 80% or 90%*

Power	N	δACR20	δACR50	δDAS28	δSDAI
0.80	50	−25	−20	−0.67	−7.62
	100	−18	−14	−0.47	−5.39
	200	−12	−10	−0.33	−3.81
	300	−10	−8	−0.27	−3.11
	400	−9	−7	−0.24	−2.69
	500	−8	−6	−0.21	−2.41
	600	−7	−6	−0.19	−2.20
	700	−7	−5	−0.18	−2.04
	800	−6	−5	−0.17	−1.91
	900	−6	−5	−0.16	−1.80
	1,000	−6	−4	−0.15	−1.70
0.90	50	−29	−23	−0.78	−8.97
	100	−21	−17	−0.55	−6.34
	200	−15	−12	−0.39	−4.48
	300	−12	−10	−0.32	−3.66
	400	−10	−8	−0.28	−3.17
	500	−9	−7	−0.25	−2.84
	600	−8	−7	−0.23	−2.59
	700	−8	−6	−0.21	−2.40
	800	−7	−6	−0.20	−2.24
	900	−7	−6	−0.18	−2.11
	1,000	−7	−5	−0.18	−2.01

* N = sample size per arm.

noninferior to each other. With 500 subjects per arm, these margins reduce to –8% for the ACR20, –0.21 for the DAS28, and –2.41 for the SDAI. Varying the ACR20 response in the reference arm from 40% to 70% had little effect (see Supplementary Figure 3, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23825/abstract). Power equivalences between ACR responses and DAS28 and SDAI changes for noninferiority studies with different margins are presented in Supplementary Appendix 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23825/abstract.

Customizable associations. We have developed a web-based utility that allows users to generate equivalences among these measures for a range of sample sizes, changes in RA activity measures, and reference group ACR responses (URL: https://webbedfeet.shinyapps.io/PowerApp/). In addition, the standard deviations for the change in DAS28 and SDAI can be modified based on the user's data. This utility can be used in planning direct comparison trials. Statistical equivalences with the ACR criteria for 70% improvement are also provided in the utility.

DISCUSSION

Investigators are increasingly using RA activity measures as the primary endpoint in direct comparison trials despite uncertainty about the most appropriate target treatment effect or noninferiority margin (1). Here we provided estimates of between-arm differences in DAS28 and SDAI change scores that were statistically equivalent to ACR responses, which are more familiar to rheumatologists. These crosswalks can be used to improve the interpretability of trials that examine changes in the DAS28 or SDAI as the outcome.

In a superiority trial, a 19% absolute difference in ACR20 responses, arguably an important treatment effect, was equivalent to a difference in DAS28 change scores of 0.53 and a difference in SDAI change scores of 6.1 and would require 100 patients per arm. However, treatments in a direct comparison trial might be expected to have ACR20 differences of only ~10%. Differences of this degree would correspond to DAS28 change score differences of 0.3 and SDAI change score differences of 3.0 to 3.5 and would require 300 to 400 subjects per arm.

An additional consideration in noninferiority trials is setting the noninferiority margin. The margin is largely based on clinical judgment, although the treatment effects in placebo-controlled studies can provide some guidance (10). Due to greater familiarity, it may be easier to judge noninferiority margins for an ACR20 response than for DAS28 or SDAI responses. Our crosswalks provide equivalences in these margins for the DAS28 and SDAI.

We provided equivalences at selected sample sizes and type I and type II error rates, but investigators may be interested in modifying the sample size or in powering a trial to detect a specific difference in DAS28 or SDAI changes. Therefore, we developed a utility that allows users to vary the sample size, error rates, treatment effects, change score standard deviations, and ACR responses in the reference group to provide estimates specific to their needs.

Our study has some limitations. We examined changes in RA activity measures as the outcome because these are analogous to ACR responses. Rather than examining changes from baseline, investigators may choose to compare end-of-trial DAS28 or SDAI scores between arms. We did not include other RA activity measures, such as the Clinical Disease Activity Index, because these have not often been used as primary trial endpoints. We based our calculations on the standard deviations of changes in the DAS28 and SDAI observed in a large prospective study of patients with active RA. These changes were very similar to those reported in recent clinical trials (11–13). For example, the standard deviations of the DAS28 change in the 2 active-treatment arms in the RA-BEAM trial were 1.26 and 1.39 (compared to 1.34 in our study), while the standard deviations for the SDAI change were 14.2 and 15.1 (compared to 15.32 in our study). Should different standard deviations be of interest, investigators can input these in the web-based utility. Using a new perspective of statistical information, this study provides calibration across RA response measures that can aid the planning and interpretation of trials.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Hazlewood GS, Barnabe C, Tomlinson G, Marshall D, Devoe D, Bombardier C. Methotrexate monotherapy and methotrexate combination therapy with traditional and biologic disease modifying antirheumatic drugs for rheumatoid arthritis: abridged Cochrane systematic review and network meta-analysis. BMJ 2016;i1777.
2. Felson D, Anderson J, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727–35.
3. Moreland LW, O'Dell JR, Paulus HE, Curtis JR, Bathon JM, St. Clair EW, et al. A randomized comparative effectiveness study of oral triple therapy versus etanercept plus methotrexate in early aggressive rheumatoid arthritis: the Treatment of Early Aggressive Rheumatoid Arthritis trial. Arthritis Rheum 2012;64:2824–35.
4. O'Dell JR, Mikuls TR, Taylor TH, Ahluwalia V, Brophy M, Warren SR, et al. Therapies for active rheumatoid arthritis after methotrexate failure. N Engl J Med 2013;369:307–18.
5. Chow SC, Shao J, Wang H. Sample size calculations in clinical research. 2nd ed. Boca Raton: Taylor & Francis; 2007.
6. Ward MM, Guthrie LC, Alba MI. Clinically important changes in individual and composite measures of rheumatoid arthritis activity: thresholds applicable in clinical trials. Ann Rheum Dis 2015;74:1691–6.
7. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2016. URL: https://www.R-project.org/.
8. Champely S, Ekstrom C, Dalgaard P, Gill J, Weibelzahl S, Anandkumar A, et al. pwr: basic functions for power analysis. 2017. URL: https://CRAN.R-project.org/package=pwr.
9. Zhang E, Wu VQ, Chow SC, Zhang HG. TrialSize: functions and examples in sample size calculation in clinical research. 2013. URL: https://CRAN.R-project.org/package=TrialSize.
10. US Food and Drug Administration. Non-inferiority clinical trials to establish effectiveness: guidance for industry. November 2016. URL: https://www.fda.gov/downloads/Drugs/Guidances/UCM202140.pdf.
11. Taylor PC, Keystone EC, van der Heijde D, Weinblatt ME, Del Carmen Morales L, Reyes Gonzaga J, et al. Baricitinib versus placebo or adalimumab in rheumatoid arthritis. N Engl J Med 2017;376:652–62.
12. Porter D, van Melckebeke J, Dale J, Messow CM, McConnachie A, Walker A, et al. Tumour necrosis factor inhibition versus rituximab for patients with rheumatoid arthritis who require biological treatment (ORBIT): an open-label, randomised controlled, non-inferiority, trial. Lancet 2016;388:239–47.
13. Fleischmann RM, Damjanov NS, Kivitz AJ, Legedza A, Hoock T, Kinnman N. A randomized, double-blind, placebo-controlled, twelve-week, dose-ranging study of decernotinib, an oral selective JAK-3 inhibitor, as monotherapy in patients with active rheumatoid arthritis. Arthritis Rheumatol 2015;67:334–43.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 71, No. 11, November 2019, pp 1488–1494
DOI 10.1002/acr.24028

Racial Disparities in Total Knee Replacement Failure As Related to Poverty

Anne R. Bass,[1] Bella Mehta,[1] Jackie Szymonifka,[2] Jackie Finik,[1] Stephen Lyman,[1] Emily Ying Lai,[1] Michael Parks,[1] Mark Figgie,[1] Lisa A. Mandl,[1] and Susan M. Goodman[1]

Objective. To determine whether racial disparities in total knee replacement (TKR) failure are explained by poverty.

Methods. Black and white New York state residents, enrolled in a prospective single-institution TKR registry January 1, 2008 to February 6, 2012, who underwent primary unilateral TKR (n = 4,062) were linked to the New York Statewide Planning and Research Cooperative System database (January 1, 2008 to December 31, 2014) to capture revisions performed at outside institutions. Patients were linked by geocoded addresses to residential census tracts. Multivariable Cox regression was used to assess predictors of TKR revision. Multivariable logistic regression was used to analyze predictors of TKR failure, defined as TKR revision in New York state ≤2 years after surgery, or as Hospital for Special Surgery (HSS) TKR quality of life score "not improved" or "worsened" 2 years after surgery.

Results. The mean ± SD age was 68.4 ± 10 years, 64% of patients were female, 8% lived in census tracts with >20% of the population under the poverty line, and 9% were black. Median follow-up time was 5.3 years. A total of 3% of patients (122 of 4,062) required revision a median 454 days (interquartile range 215–829) after surgery. TKR revision risk was higher in blacks than whites, with a hazard ratio of 1.69 (95% confidence interval 1.01–2.81), but in multivariable analysis, only younger age, male sex, and constrained prosthesis were predictors of TKR revision. TKR failure occurred in 200 of 2,832 cases (7%) with 2-year surveys. Risk factors for TKR failure were non-osteoarthritis TKR indication, low surgeon volume, and low HSS Expectations Survey score, but not black race. Community poverty was not associated with TKR revision or failure.

Conclusion. There was a trend toward higher TKR revision risk in blacks, but poverty did not modify the relationship between race and TKR revision or failure.

INTRODUCTION

Blacks are at increased risk of total knee replacement (TKR) revision compared to whites (1–6), and blacks report significantly more pain and worse function 2 years after TKR (7). We previously showed that racial disparities in patient-reported outcomes are strongly influenced by community poverty and education (7,8). Whether disparities in TKR revision risk are related to poverty is unknown. In this study, we leveraged a large single-institution TKR registry, linked to a statewide discharge database, to identify TKR revisions that occurred at other hospitals. TKR cases were also linked to residential census tracts to identify those patients living in impoverished neighborhoods. We introduce the concept of TKR failure, which encompasses both TKR revision and failure to improve knee-related quality of life (QoL) 2 years after surgery. The goal of the study was to determine whether racial disparities in TKR failure are explained by poverty.

PATIENTS AND METHODS

Patients. All black and white residents of New York state who were enrolled in the Hospital for Special Surgery (HSS) Knee Replacement Registry and who underwent a primary unilateral TKR during the enrollment period were included in the study. The registry was established by HSS in cooperation with Weill Cornell Medical College through a grant from the Agency

Supported by the Hospital for Special Surgery Complex Joint Reconstruction Center. The Hospital for Special Surgery Total Knee Replacement Registry was established through a grant from the Agency for Healthcare Research and Quality (grant AHRQ U18 HS16075-01).

[1]Anne R. Bass, MD, Bella Mehta, MBBS, Jackie Finik, MPH, Stephen Lyman, PhD, Emily Ying Lai, MS, Michael Parks, MD, Mark Figgie, MD, Lisa A. Mandl, MD, MPH, Susan M. Goodman, MD: Hospital for Special Surgery, New York, New York; [2]Jackie Szymonifka, MS: New York University, New York.

Dr. Parks has received consulting fees from Zimmer Biomet (more than $10,000). Dr. Goodman has received consulting fees from Celgene and UCB (less than $10,000 each) and Novartis (more than $10,000), research support from Horizon and Novartis, and owns stock or stock options in Reginosine. No other disclosures relevant to this article were reported.

Address correspondence to Anne R. Bass, MD, Hospital for Special Surgery, 535 East 70th Street, New York, NY 10021. E-mail: bassa@hss.edu.

Submitted for publication December 23, 2018; accepted in revised form July 9, 2019.

SIGNIFICANCE & INNOVATIONS

- In patients undergoing surgery at a high-volume hospital, community poverty does not increase the risk of total knee replacement (TKR) revision or failure in blacks or whites.
- Young age, male sex, and a constrained prosthesis increase the risk of TKR revision.
- Low surgeon volume and low preoperative expectations for improvement after surgery are associated with a higher risk of TKR failure.

for Healthcare Research and Quality. Baseline data collected on all patients included age, sex, body mass index, race, ethnicity, insurance status, and education. Patient address, race (if not self-identified by the patient), indication for surgery, Charlson-Deyo comorbidities score, implant type, and operative time were obtained from the medical record. Implant types were defined as constrained (constrained condylar, hinge), or nonconstrained (cruciate retaining, cruciate sparing, posterior cruciate substituting). Surgeon volume was measured at the attending level, including cases performed with trainees. Patient-reported measures collected included the preoperative Knee Injury and Osteoarthritis Outcome Score (KOOS) (9). Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) scores were derived from the KOOS (WOMAC scale 0–100, where a higher score indicates better status) (9–11). Patients also completed the HSS Expectations Survey, a well-validated survey with 17 questions graded on a 5-point Likert scale in the domains of pain relief, functional improvement, and psychosocial well-being. The score is reported as a number from 0 to 100, with 100 being the highest expectations for improvement, and a difference of 7 is considered clinically significant (12). Registry patients completed follow-up surveys via mail or an emailed web-enabled link at 2 and 5 years post-TKR. Follow-up rates were 80% for 2-year surveys and 70% for 5-year surveys.

Statewide Planning and Research Cooperative System (SPARCS). SPARCS collects information on all discharges from nonfederal acute-care hospitals in the state. We used this system's data from January 1, 2008 through December 31, 2014 to identify HSS registry patients revised elsewhere. To minimize the potential effect of missing out-of-state follow-ups on our estimates of revision after TKR, we limited our analysis to New York state residents. We previously verified that SPARCS captured 99% of all TKRs performed at our institution in 2009 (3). Procedure codes used to identify revision TKR cases in SPARCS are listed in Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24028/abstract. The reason for revision was determined using International Classification of Diseases, Ninth Revision diagnostic codes and categorized as septic or aseptic (mechanical, fracture, or other) (see Supplementary Table 2, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24028/abstract). A requirement of using SPARCS-identified data is that no cells containing <11 individuals may be reported in order to maintain patient confidentiality. Cells with <11 individuals have been censored in tables using SPARCS data.

Census tracts. We used geocoding of individual case addresses to link registry patients to specific US census tracts and census blocks. Census tracts are small geographic areas containing approximately 4,000 individuals, while census blocks are census-tract subdivisions containing approximately 1,000 individuals. These geographic units are designed to be homogeneous with respect to population characteristics, economic status, and living conditions (13,14). We obtained census-tract– and block-level socioeconomic variables from the American Community Survey/US Census using the Geographic Information Systems. The percentage of individuals below the poverty line at the census-tract level is highly sensitive to gradients in health (15).

Primary and secondary outcomes. The primary outcome was TKR failure, defined as 1 or both of the following: revision occurring ≤2 years from the initial surgery, or a knee-related QoL rating of 5 or 6. The scale is based on the answer to the following question: "How much did your knee surgery improve the quality of your life? 1) More improvement than I ever dreamed possible, 2) Great improvement, 3) Moderate improvement, 4) A little improvement, 5) No improvement, or 6) The quality of my life is worse." Knee-related QoL is a global measure of the patient's assessment of the success of surgery not specifically related to pain, function, stiffness, or alignment, which can vary based on presurgical status (16). More than 90% of TKR patients identify knee-related QoL as very or extremely important as an outcome measure (9). Patients who did not have a 2-year QoL survey AND who did not have a revision prior to 2 years were excluded from the primary analysis. Our secondary outcome was time to TKR revision, counted as the number of days between the date of initial surgery to date of revision. For patients who did not have a revision prior to December 31, 2014, time to TKR revision was calculated as the number of days between the date of initial surgery to December 31, 2014. Finally, we also analyzed factors associated with septic versus aseptic revision TKR.

Statistical analysis. Baseline patient characteristics were summarized using descriptive statistics. Continuous variables were summarized as mean ± SD and compared using *t*-tests. Categorical variables were summarized as frequency and compared using chi-square tests. We measured the time to TKR revision using Kaplan-Meier curves and reported log-rank tests. We used Cox regression to assess predictors of TKR revision and logistic regression to analyze predictors of TKR failure. Stepwise selection was used, requiring 0.10 significance for model entry

Table 1. Patient characteristics by race*

Variable	Overall (n = 4,062)	Black (n = 358)	White (n = 3,704)	*P*
Patient demographics				
Age at surgery, mean ± SD years	68.4 ± 9.9	65.5 ± 10.2	68.7 ± 9.8	<0.001
Female	2,610 (64.3)	291 (81.3)	2,319 (62.6)	<0.001
Body mass index, mean ± SD kg/m^2	30.8 ± 6.4	33.8 ± 6.9	30.5 ± 6.3	<0.001
Comorbidities				
Sleep apnea	387 (9.5)	36 (10.1)	351 (9.5)	0.723
Coronary artery disease	566 (13.9)	33 (9.2)	533 (14.4)	0.007
Chronic obstructive pulmonary disease	476 (11.7)	53 (14.8)	423 (11.4)	0.058
Diabetes mellitus	536 (13.2)	97 (27.1)	439 (11.9)	<0.001
Hypertension	2,292 (56.4)	256 (71.5)	2,036 (55.0)	<0.001
Obesity	750 (18.5)	103 (28.8)	647 (17.5)	<0.001
Renal disease	139 (3.4)	20 (5.6)	119 (3.2)	0.018
Arthritis				
Rheumatoid arthritis	127 (3.1)	22 (6.1)	105 (2.8)	<0.001
Osteoarthritis	3,944 (97.1)	348 (97.2)	3,596 (97.1)	0.917
Any infection index surgery	120 (3.0)	13 (3.6)	107 (2.9)	0.429
Education level				<0.001
Some high school	128 (3.3)	37 (11.1)	91 (2.6)	–
High school graduate	670 (17.4)	62 (18.7)	608 (17.3)	–
Some college	784 (20.3)	98 (29.5)	686 (19.5)	–
College graduate	947 (24.6)	73 (22.0)	874 (24.8)	–
Masters, professional or doctorate	1,326 (34.4)	62 (18.7)	1,264 (35.9)	–
Insurance payer†				
Medicaid	116 (2.9)	72 (20.1)	44 (1.2)	<0.001
Medicare	2,717 (66.9)	214 (59.8)	2,503 (67.6)	0.003
Commercial	3,905 (96.2)	280 (78.2)	3,625 (97.9)	<0.001
Self-pay	1,406 (34.6)	135 (37.7)	1,271 (34.3)	0.198
Prosthesis type				0.987
Unconstrained	3,594 (90.2)	315 (90.3)	3,279 (90.2)	–
Constrained	389 (9.8)	34 (9.7)	355 (9.8)	–
Surgeon volume, cases per year				<0.001
<50	360 (8.9)	28 (7.8)	332 (9.0)	–
50–99	756 (18.6)	82 (22.9)	674 (18.2)	–
100–199	1,111 (27.4)	181 (50.6)	930 (25.1)	–
≥200	1,834 (45.2)	67 (18.7)	1,767 (47.7)	–
Community census characteristics				
Census-tract poverty, %				<0.001
<10	3,071 (75.7)	120 (33.5)	2,951 (79.8)	–
10 to <20	659 (16.)	87 (24.3)	572 (15.5)	–
20 to <30	200 (4.9)	83 (23.2)	117 (3.2)	–
30 to <40	87 (2.1)	39 (10.9)	48 (1.3)	–
≥40	38 (0.9)	29 (8.1)	–	–
Census-tract poverty ≥20%	325 (8.0)	151 (42.2)	174 (4.7)	<0.001
Baseline questionnaires				
HSS expectations score, mean ± SD	77.3 ± 18.2	75.2 ± 17.8	77.4 ± 18.2	0.21
WOMAC pain, mean ± SD	54.7 ± 17.9	47.5 ± 19.7	55.2 ± 17.6	<0.001
WOMAC function, mean ± SD	53.7 ± 17.6	46.2 ± 18.9	54.2 ± 17.4	<0.001

* Values are the number (%) unless indicated otherwise. We do not report values for cells containing <11 individuals.
HSS = Hospital for Special Surgery; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index.
† More than one insurance may be selected.

and 0.10 significance to remain in the final model, with the exception of race, age, and prosthesis types, which were forced into corresponding multivariable models as noted in the results. The one exception was "disruption of operation wound" during index admission, which was excluded despite its statistical significance, because it occurred in only 1 patient who underwent aseptic revision, making it difficult to interpret. Interaction between percentage of the census tract under the poverty line and race was examined in both models (TKR failure and TKR revision). For our power calculation, we used multiple logistic regression, assuming an 8.5% failure rate and a 7% black population. A sample size of 2,832 achieved 80% power at a 0.05 significance level to detect an 80% increase in risk of TKR failure (odds ratio [OR] 1.8). To gage potential multicollinearity between race and census-tract poverty, the variance inflation factor (VIF) was used to assess model distortion. The VIF did not exceed 1.2, indicating minimal model distortion due to multicollinearity. This study was approved by our local institutional review board.

RESULTS

There were 7,237 cases who consented to be in the registry and provided a baseline questionnaire between January 1, 2008 and February 6, 2012. Of that number, 7,207 had geocodable addresses, 4,926 were New York state residents, and 4,062 had a unilateral TKR and were either black or white (see Supplementary Figure 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24028/abstract). The 4,062 included cases, when compared to those who were excluded (n = 3,145), were older, more likely to be female, less likely to be black, and more likely to have comorbidities (see Supplementary Table 3, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24028/abstract). Median follow-up was 1,919 days (5.3 years). Mean ± SD age was 68 ± 10 years, 64% were female, 8% resided in census tracts with >20% of the population under the poverty line (census-tract poverty), and 9% were black (Table 1). Compared to whites, blacks were younger, more likely to be female, to have comorbidities, and to have a non-osteoarthritis indication for surgery ($P < 0.001$ for all). Blacks also had less education and were more likely to have Medicaid insurance (and less likely to have Medicare) and to live in communities with high census-tract poverty ($P < 0.001$ for all) (Table 1). A similar percentage of blacks and whites required constrained prostheses and had self-pay insurance.

Revision TKR. Three percent of cases (122 of 4,062) required TKR revision a median of 454 days (interquartile range [IQR] 215, 829) after the index surgery. In univariate analysis, the risk of revision was higher in blacks than in whites, with a hazard ratio (HR) of 1.69 (95% confidence interval [95% CI] 1.01–2.81) (Figure 1 and Supplementary Tables 4 and 5, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24028/abstract).

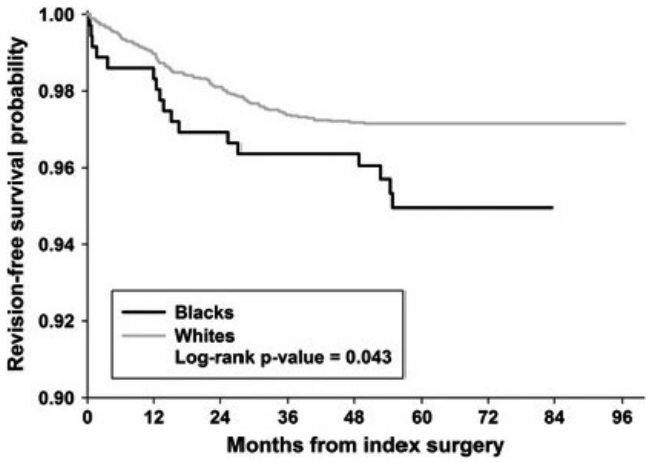

Figure 1. Time to revision of total knee replacement in blacks versus whites.

Table 2. Multivariable model for time to total knee replacement revision*

Multivariable model	HR (95% CI)	*P*
Age at surgery, per 5 years	0.80 (0.74–0.88)	<0.001
Female	0.62 (0.43–0.89)	0.009
Race		
White	1.00 (Ref.)	–
Black	1.56 (0.91–2.67)	0.105
Prosthesis type		
Unconstrained	1.00 (Ref.)	–
Constrained	2.31 (1.42–3.76)	<0.001

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

Other factors associated with an increased risk of TKR revisions were constrained prosthesis type and self-pay insurance, while older age, female sex, and Medicare insurance were associated with a lower risk of TKR revision.

TKR revision, race, and poverty. The following variables were retained following stepwise selection: age, sex, race, insurance, and prosthesis type. Race was forced into the final model. As shown in Table 2, race was not associated with time to TKR revision. There was no association between census-tract poverty and time to TKR revision, and there was no interaction effect of race and census-tract poverty on time to TKR revision (see Supplementary Table 6, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24028/abstract). Risk factors for TKR revision were older age, female sex, and constrained prosthesis type.

TKR failure. Of 2,832 patients who returned 2-year knee-related QoL surveys or underwent TKR revision within 2 years, 7% (200 of 2,832) experienced TKR failure (93 revised and 107 nonrevised). Patients included in the TKR failure cohort were less likely to be black (7% versus 12%; $P < 0.001$) or live in a census tract with >20% of the population under the poverty level (7% versus 11%; $P < 0.001$) than those excluded from this analysis. In univariate analysis, factors associated with the risk of TKR failure were age, non-osteoarthritis indication for surgery, insurance, surgeon volume, prosthesis type, census-tract poverty, and HSS expectations score (see Supplementary Tables 7 and 8, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24028/abstract). There was a trend toward an increased risk of TKR failure in blacks, with OR 1.53 (95% CI 0.95–2.45).

TKR failure, race, and poverty. The following variables were retained following stepwise selection: sex, osteoarthritis diagnosis, insurance, census-tract poverty, surgeon volume, and HSS expectations score. Age, race, and prosthesis type were forced into the final model. As shown in Table 3, race was not associated with the risk of TKR failure. No interaction effect between race and census-tract poverty on odds of TKR failure was observed (see

Table 3. Multivariable model, odds of total knee replacement failure*

Multivariable model	OR (95% CI)	*P*
Age	0.96 (0.86–1.07)	0.454
Race		
Black	1.03 (0.43–2.47)	0.942
White	1.00 (Ref.)	–
Preop knee index surgery for osteoarthritis	0.41 (0.18–0.93)	0.032
Surgeon volume, cases per year		
<50	3.00 (1.56–5.76)	0.001
50–99	1.25 (0.64–2.42)	0.515
100–199	2.12 (1.23–3.65)	0.007
≥200	1.00 (Ref.)	–
Census-tract poverty, %		
<10	0.32 (0.067–1.62)	0.169
10 to <20	0.66 (0.13–3.39)	0.618
20 to <30	0.17 (0.02–1.45)	0.105
30 to <40	0.23 (0.02–2.95)	0.259
≥40	1.00 (Ref.)	–
HSS expectations, per 10 units	0.84 (0.75–0.94)	0.002
Prosthesis type		
Unconstrained	1.00 (Ref.)	–
Constrained	1.14 (0.55–2.34)	0.72

* OR = odds ratio; 95% CI = 95% confidence interval; Ref. = reference; HSS = Hospital for Special Surgery.

Supplementary Table 9, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24028/abstract). Risk factors for TKR failure included non-osteoarthritis diagnosis, low surgeon volume, and low HSS expectations score.

Septic versus aseptic revision TKR. Twenty-three of the 122 TKR revisions (19%) in this study were septic, and 99 of 122 (81%) were aseptic (see Supplementary Table 10, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24028/abstract). Median time to septic versus aseptic revision was 186 days (IQR 50, 586) versus 472 days (IQR 323, 847) ($P = 0.01$). Factors associated with septic (versus aseptic) revision included older age ($P = 0.02$) and having a surgeon with a low annual volume of surgery ($P = 0.02$).

DISCUSSION

In this study, we analyzed more than 4,000 TKR cases followed for a median of 5.3 years, linked them to a discharge database to capture revisions performed anywhere in the state, and also linked them to US residential census tracts to analyze the relationship between race, community poverty, and 2-year TKR failure (TKR revision, or no improvement/worsening of knee-related QoL). In our study, 3% of cases required TKR revision and 7% experienced TKR failure. Although black race was associated with time to TKR revision in univariate analysis (and 4.7% of blacks versus 2.8% of whites underwent TKR revision), race did not reach statistical significance in the final model (HR 1.56 [95% CI 0.91–2.67]). In our analysis of TKR failure, there was a trend toward a higher risk of revision in blacks in univariate analysis (OR 1.53 [95% CI 0.95–2.45]), but this trend was not observed in the final adjusted model. There was no interaction between race and poverty for either TKR revision or TKR failure.

These results differ significantly from our prior study, in which we analyzed predictors of poor WOMAC pain and function scores 2 years after TKR and found a strong association between both race and census-tract poverty and WOMAC pain and function (7). In that study, we also found a strong interaction between race and census-tract poverty whereby blacks living in wealthy neighborhoods fared as well as whites, whereas in disadvantaged communities, there were large racial disparities. In the current study, poverty was not associated with TKR revision/failure risk, and race was only a significant risk factor in univariate analysis. This finding suggests that TKR revision/failure is an outcome distinct from pain and function after TKR, and that TKR revision/failure risk may be more related to features of the original surgery (e.g., prosthesis type, surgeon volume, hospital volume) than to characteristics of the patient. The fact that all patients in the current study had their surgery at a single high-volume orthopedic hospital may have mitigated racial disparities in our models since hospital characteristics likely contribute significantly to racial disparities in TKR revision risk (3,17–21).

Factors associated with TKR revision in our study were younger age, male sex, and constrained prosthesis type. Factors associated with the TKR failure were low surgeon volume and low baseline patient expectations for the outcome of surgery. Neither obesity nor the presence of comorbidities, both of which have been linked to black race (22) and to TKR revision (23) were significant risk factors for TKR revision or failure. The link between a constrained prosthesis type and TKR revision has been demonstrated in other studies (6,24,25), but, of note, blacks and whites in our study had similar utilization of constrained prostheses. Having a surgeon with a low annual volume of surgery (<50 cases/year) was associated with triple the risk of TKR failure in our study, but blacks were no more likely to have a surgeon with a low annual volume of surgery than whites. We also found that TKR revisions that occurred in patients of surgeons with a low annual volume of surgery were disproportionately septic revisions. Other studies have demonstrated a stepwise increase in TKR complication rates, including revisions, as surgeon volume declines (18,26). Complication rates may also be higher when surgery is performed by trainees. For example, in 1 study of unicompartmental knee replacements, there were more reoperations for bearing dislocation in the trainee group, although there were similar TKR survival rates for attending and trainee surgeons (27).

Blacks in our study were more likely to be female than were whites (81% versus 63%), which is protective, but were also younger (age 66 versus 69 years), a revision risk factor. Our finding that female sex and older age were protective against TKR revision is consistent with other studies (3,28–32). Specifically, women in our study were 36% less likely to undergo TKR revision, and for

every 5 years of increased age, the risk of TKR revision declined 20%. Sex differences in TKR outcomes have been ascribed to higher rates of wear and osteolysis in men, greater physical activity in men, or differences in knee biomechanics (32). The elevated risk in TKR revision in younger individuals has been attributed in part to higher rates of posttraumatic secondary osteoarthritis in this cohort (29).

High baseline patient expectations for the outcome of surgery reduced the likelihood of TKR failure in our study, but there were no significant differences in baseline expectations between blacks and whites. Specifically, for every 10-unit increase in the HSS expectations score, the risk of TKR failure declined 16%. A recent study showed an association between HSS expectations scores and patient-reported outcome measures and activity levels in the year after surgery (33), and prior studies have suggested that blacks have lower preoperative expectations (34). Patient expectations have been shown to correlate with health literacy (35), but the relationship between health literacy and health disparities has not yet been well established (36).

Despite the large size of our TKR registry, the relatively smaller number of patients who resided in New York and returned 2-year knee-related QoL surveys may have limited our power to analyze the impact of race on TKR failure. However, this weakness is balanced by many strengths. To our knowledge, this is the first study to link an institutional registry to a statewide discharge database. This powerful approach enabled ascertainment of TKR revisions that occurred at other institutions and avoided the loss to follow-up seen in other TKR registries. In addition, by linking registry patients to their residential census tracts, we were able to analyze socioeconomic factors such as community poverty as they impact TKR revision and TKR failure risk. The large size of the registry, the availability of detailed information about patients' baseline pain, function, and expectations for surgery, and knowledge of prosthesis types are other study strengths. Analyzing patients from a single institution allowed us to address the impact of patient and community factors on TKR failure risk, independent of hospital volume and quality. This methodology is an important feature of this study because blacks are generally more likely to choose low-volume orthopedic hospitals (21), and having surgery at a low-volume hospital increases the risk of TKR revision (3,19). Finally, 8.8% of the patients in our study were black, a high percentage given that blacks account for only 5.4% of TKRs performed in the US (37).

In conclusion, our study did not show an association between TKR revision or TKR failure and community poverty in either blacks or whites. Factors associated with TKR revision and/or TKR failure were age, sex, prosthesis type, surgeon volume, and patients' preoperative expectations for the outcome of surgery. Directing black patients to high-volume orthopedic hospitals may help minimize disparities in TKR survival.

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. Bass 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. Bass, Mehta, Lyman, Parks, Goodman.
Acquisition of data. Bass, Mehta, Ying Lai, Goodman.
Analysis and interpretation of data. Bass, Mehta, Szymonifka, Finik, Figgie, Mandl, Goodman.

REFERENCES

1. Bass AR, McHugh K, Fields K, Goto R, Parks ML, Goodman SM. Higher total knee arthroplasty revision rates among United States blacks than whites: a systematic literature review and meta-analysis. J Bone Joint Surg Am 2016;98:2103–8.
2. Roche M, Law TY, Sultan AA, Umpierrez E, Khlopas A, Rosas S, et al. Racial disparities in revision total knee arthroplasty: analysis of 125,901 patients in national US private payer database. J Racial Ethn Health Disparities 2019;6:101–9.
3. Dy CJ, Marx RG, Bozic KJ, Pan TJ, Padgett DE, Lyman S. Risk factors for revision within 10 years of total knee arthroplasty. Clin Orthop Relat Res 2014;472:1198–207.
4. Blum MA, Singh JA, Lee GC, Richardson D, Chen W, Ibrahim SA. Patient race and surgical outcomes after total knee arthroplasty: an analysis of a large regional database. Arthritis Care Res (Hoboken) 2013;65:414–20.
5. Bolognesi MP, Greiner MA, Attarian DE, Watters TS, Wellman SS, Curtis LH, et al. Unicompartmental knee arthroplasty and total knee arthroplasty among Medicare beneficiaries, 2000 to 2009. J Bone Joint Surg Am 2013;95:e174.
6. Namba RS, Cafri G, Khatod M, Inacio MC, Brox TW, Paxton EW. Risk factors for total knee arthroplasty aseptic revision. J Arthroplasty 2013;28:122–7.
7. Goodman SM, Mandl LA, Parks ML, Zhang M, McHugh KR, Lee Y, et al. Disparities in TKA outcomes: census tract data show interactions between race and poverty. Clin Orthop Relat Res 2016;474:1986–95.
8. Goodman SM, Mandl LA, Mehta B, Navarro-Millan I, Russell LA, Parks ML, et al. Does education level mitigate the effect of poverty on total knee arthroplasty outcomes? Arthritis Care Res (Hoboken) 2018;70:884–91.
9. Roos EM, Toksvig-Larsen S. Knee injury and osteoarthritis outcome score (KOOS): validation and comparison to the WOMAC in total knee replacement. Health Qual Life Outcomes 2003;1:17.
10. Bellamy N. Outcome measurement in osteoarthritis clinical trials. J Rheumatol Suppl 1995;43:49–51.
11. Nilsdotter AK, Toksvig-Larsen S, Roos EM. A 5 year prospective study of patient-relevant outcomes after total knee replacement. Osteoarthritis Cartilage 2009;17:601–6.
12. Mancuso CA, Sculco TP, Wickiewicz TL, Jones EC, Robbins L, Warren RF, et al. Patients' expectations of knee surgery. J Bone Joint Surg Am 2001;83:1005–12.
13. Subramanian SV, Chen JT, Rehkopf DH, Waterman PD, Krieger N. Comparing individual- and area-based socioeconomic measures for the surveillance of health disparities: a multilevel analysis of Massachusetts births, 1989–1991. Am J Epidemiol 2006;164:823–34.
14. Subramanian SV, Chen JT, Rehkopf DH, Waterman PD, Krieger N. Racial disparities in context: a multilevel analysis of neighborhood variations in poverty and excess mortality among black populations in Massachusetts. Am J Public Health 2005;95:260–5.

15. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures: the public health disparities geocoding project. Am J Public Health 2003;93:1655–71.
16. Lyman S, Lee Y, Franklin PD, Li W, Cross MB, Padgett DE. Validation of the KOOS, JR: a short-form knee arthroplasty outcomes survey. Clin Orthop Relat Res 2016;474:1461–71.
17. Jeschke E, Citak M, Günster C, Matthias Halder A, Heller K, Malzahn J, et al. Are TKAs performed in high-volume hospitals less likely to undergo revision than TKAs performed in low-volume hospitals? Clin Orthop Relat Res 2017;475:2669–74.
18. Wilson S, Marx RG, Pan T, Lyman S. Meaningful thresholds for the volume-outcome relationship in total knee arthroplasty. J Bone Joint Surg Am 2016;98:1683–90.
19. Badawy M, Espehaug B, Indrekvam K, Engesæter LB, Havelin LI, Furnes O. Influence of hospital volume on revision rate after total knee arthroplasty with cement. J Bone Joint Surg Am 2013;95:e131.
20. Cai X, Cram P, Vaughan-Sarrazin M. Are African American patients more likely to receive a total knee arthroplasty in a low-quality hospital? Clin Orthop Relat Res 2012;470:1185–93.
21. FitzGerald JD, Soohoo NF, Losina E, Katz JN. Potential impact on patient residence to hospital travel distance and access to care under a policy of preferential referral to high-volume knee replacement hospitals. Arthritis Care Res (Hoboken) 2012;64:890–7.
22. Piccolo RS, Duncan DT, Pearce N, McKinlay JB. The role of neighborhood characteristics in racial/ethnic disparities in type 2 diabetes: results from the Boston area community health (BACH) survey. Soc Sci Med 2015;130:79–90.
23. Bozic KJ, Lau E, Ong K, Chan V, Kurtz S, Vail TP, et al. Risk factors for early revision after primary TKA in Medicare patients. Clin Orthop Relat Res 2014;472:232–7.
24. Pitta M, Esposito CI, Li Z, Lee Y, Wright TM, Padgett DE. Failure after modern total knee arthroplasty: a prospective study of 18,065 knees. J Arthroplasty 2018;33:407–14.
25. Vertullo CJ, Lewis PL, Lorimer M, Graves SE. The effect on long-term survivorship of surgeon preference for posterior-stabilized or minimally stabilized total knee replacement: an analysis of 63,416 prostheses from the Australian Orthopaedic Association National Joint Replacement registry. J Bone Joint Surg Am 2017;99:1129–39.
26. Liddle AD, Pandit H, Judge A, Murray DW. Effect of surgical caseload on revision rate following total and unicompartmental knee replacement. J Bone Joint Surg Am 2016;98:1–8.
27. Bottomley N, Jones LD, Rout R, Alvand A, Rombach I, Evans T, et al. A survival analysis of 1084 knees of the Oxford unicompartmental knee arthroplasty: a comparison between consultant and trainee surgeons. Bone Joint J 2016;98-B:22–7.
28. Bayliss LE, Culliford D, Monk AP, Glyn-Jones S, Prieto-Alhambra D, Judge A, et al. The effect of patient age at intervention on risk of implant revision after total replacement of the hip or knee: a population-based cohort study. Lancet 2017;389:1424–30.
29. Julin J, Jämsen E, Puolakka T, Konttinen YT, Moilanen T. Younger age increases the risk of early prosthesis failure following primary total knee replacement for osteoarthritis: a follow-up study of 32,019 total knee replacements in the Finnish Arthroplasty register. Acta Orthop 2010;81:413–9.
30. Meehan JP, Danielsen B, Kim SH, Jamali AA, White RH. Younger age is associated with a higher risk of early periprosthetic joint infection and aseptic mechanical failure after total knee arthroplasty. J Bone Joint Surg Am 2014;96:529–35.
31. Stambough JB, Clohisy JC, Barrack RL, Nunley RM, Keeney JA. Increased risk of failure following revision total knee replacement in patients aged 55 years and younger. Bone Joint J 2014;96-B:1657–62.
32. Singh JA, Kwoh CK, Richardson D, Chen W, Ibrahim SA. Sex and surgical outcomes and mortality after primary total knee arthroplasty: a risk-adjusted analysis. Arthritis Care Res (Hoboken) 2013;65:1095–102.
33. Jain D, Nguyen LL, Bendich I, Nguyen LL, Lewis CG, Huddleston JI, et al. Higher patient expectations predict higher patient-reported outcomes, but not satisfaction, in total knee arthroplasty patients: a prospective multicenter study. J Arthroplasty 2017;32:S166–70.
34. Ibrahim SA, Siminoff LA, Burant CJ, Kwoh CK. Differences in expectations of outcome mediate African American/white patient differences in "willingness" to consider joint replacement. Arthritis Rheum 2002;46:2429–35.
35. Hadden KB, Prince LY, Bushmiaer MK, Watson JC, Barnes CL. Health literacy and surgery expectations in total hip and knee arthroplasty patients. Patient Educ Couns 2018;101:1823–7.
36. Mantwill S, Monestel-Umaña S, Schulz PJ. The relationship between health literacy and health disparities: a systematic review. PLoS One 2015;10:e0145455.
37. Singh JA, Lu X, Rosenthal GE, Ibrahim S, Cram P. Racial disparities in knee and hip total joint arthroplasty: an 18-year analysis of national Medicare data. Ann Rheum Dis 2014;73:2107–15.

Arthritis Care & Research
Vol. 71, No. 11, November 2019, pp 1495–1503
DOI 10.1002/acr.23852

American College of Rheumatology
Empowering Rheumatology Professionals

Modifiable Determinants of Exercise Use in a Diverse Ethnic Population With Osteoarthritis

Ernest R. Vina,[1] Michael J. Hannon,[2] Leslie R. M. Hausmann,[3] Said A. Ibrahim,[4] Jazmin Dagnino,[1] Andrea Arellano,[1] and C. Kent Kwoh[1]

Objective. To determine the extent of ethnic differences in the use of exercise for therapy and identify relevant modifiable determinants of exercise use among patients with knee/hip osteoarthritis (OA).

Methods. Knee/hip OA study participants were identified. Surveys were administered to collect patient sociodemographic and clinical information, and beliefs and attitudes about providers and treatments. Final multivariable logistic regression models were created using a fully conditional method.

Results. Hispanic participants (n = 130), compared to non-Hispanic participants (n = 232), were less likely to have private medical insurance (9.2% versus 31.0%) or to report having excellent/very good health (40.7% versus 52.6%). They were also less likely to report using exercise for OA treatment in the last 6 months (56% versus 73%; $P = 0.003$). When adjusted for age and disease severity, the difference in exercise use among ethnicities remained significant (odds ratio [OR] 0.59 [95% confidence interval (95% CI) 0.36–0.99]). In a multivariable logistic regression model designed to determine the most important determinants of exercise use for OA treatment, in the last 6 months the following were all associated with exercise use: having knee instead of hip OA (OR 2.83 [95% CI 1.51–5.29]), having family/friends who exercise (OR 3.20 [95% CI 1.76–5.84]), having a good understanding of what happens after exercise (OR 2.19 [95 CI 1.15–4.19]), and higher perceived benefit of exercise (OR 2.24 [95% CI 1.64–3.04]).

Conclusion. Among patients with knee/hip OA, Hispanics were less likely to exercise for OA treatment. Increased knowledge about the benefits of exercise for treatment and improved familiarity with exercise as treatment for OA may increase exercise use.

INTRODUCTION

Osteoarthritis (OA) is the most common rheumatic disease in the US and worldwide (1). Approximately 14 million people in the US, including >3 million racial/ethnic minorities, have symptomatic knee OA (1). The exact prevalence of OA in Hispanics is unknown, but research suggests that 5–22% have arthritis, of which OA is the most common form (2,3). According to national surveys, the prevalence of arthritis-attributable activity limitation, arthritis-attributable work limitation, and severe joint pain are also significantly higher among Hispanics than among non-Hispanic whites (4,5).

While there is no known cure for OA, impaired muscle function and reduced fitness are potentially amenable to exercise, also known as leisure time physical activity (PA) (6,7). Evidence suggests that exercise programs may result in immediate and short-term benefits, including reduced knee pain, improved physical function, and better quality of life among people with knee and hip OA (8,9). The benefit of exercise for pain is, in fact, comparable with reported estimates of benefit from simple analgesics taken for knee pain (8). The American College of Rheumatology (ACR) and other professional organizations also recommend that all patients with knee and hip OA participate in aerobic and strengthening exercises to help treat the disease (10). Unfortunately, among the

Dr. Vina's work was supported in part by the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases grant K23-AR-067226). Dr. Ibrahim's work was supported in part by a K24 Mid-Career Development Award and by the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases grant K24-AR-055259). Dr. Kwoh's work was supported by the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases grant R01-AR-066601).

[1]Ernest R. Vina, MD, MS, Jazmin Dagnino, BS, Andrea Arellano, BS, C. Kent Kwoh, MD: University of Arizona, College of Medicine and/or Arthritis Center, Tucson; [2]Michael J. Hannon, MA: University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; [3]Leslie R. M. Hausmann, PhD: University of Pittsburgh School of Medicine and Veterans Affairs Pittsburgh Healthcare System, Pittsburgh, Pennsylvania; [4]Said A. Ibrahim, MD, MPH, MBA: Weill Cornell Medicine, New York, New York.

Dr. Kwoh has received honoraria from AbbVie (less than $10,000) and EMD Serono (more than $10,000) and has provided consulting services for Astellas, Thusane, and Novartis (less than $10,000 each), and EMD Serono and Express Scripts (more than $10,000 each). No other disclosures relevant to this article were reported.

Address correspondence to Ernest R. Vina, MD, MS, University of Arizona Arthritis Center, 1501 N. Campbell Ave., PO Box 245093, Tucson, AZ 85724-5093. E-mail: evina@email.arizona.edu.

Submitted for publication October 30, 2018; accepted in revised form February 12, 2019.

SIGNIFICANCE & INNOVATIONS

- This study found that Hispanics were less likely than non-Hispanics to utilize exercise as treatment for osteoarthritis (OA).
- After adjustment for age and OA disease severity, ethnic differences in exercise use in the last 6 months remained significant.
- The most important predictors of exercise use for OA included having family/friends who exercise, having a good understanding of what exercise entails, and higher perceived benefit of the therapy.

various ethnic/racial groups in the US, Hispanics have the highest rates of inactivity and the lowest rates of meeting guidelines of sufficient PA (11,12).

Although some evidence suggests that social determinants of health contribute to ethnic disparities in treatment utilization (13), socioeconomic status is unlikely to fully explain the higher prevalence of physical inactivity during leisure time among Hispanics, regardless of whether they exercise to treat OA or not. A study by Crespo et al (14) demonstrated that Mexican American adults reported less leisure time PA than non-Hispanic whites, irrespective of income, education, occupation, and marital status. There are other determinants of treatment utilization, including patient clinical characteristics (e.g., disease severity), attitudes toward providers, and knowledge of and attitudes toward treatments (15). Moreover, a previous study showed that ethnic/racial minority status is associated with greater experience of OA-related symptoms (16). A study by Armstrong et al (17) additionally demonstrated that Hispanics are more likely than non-Hispanic whites to report higher levels of distrust of physicians. We also previously showed ethnic differences in familiarity with and perceptions of exercise for OA treatment (18).

Whether ethnic differences in exercise use for OA treatment persist despite adjustment for these clinical factors and patient beliefs and values is unknown. It is also unclear as to whether patient beliefs and attitudes toward providers and exercise are independently associated with eventual use of exercise for therapy in OA. This is particularly relevant, as patient beliefs and attitudes toward specific treatment approaches are potentially modifiable at the point of clinical care (19). Therefore, the primary objective of this study was to determine the extent of ethnic differences in the current and past use of exercise as therapy for patients with knee or hip OA. The secondary objective was to identify the most relevant modifiable variables that are significantly associated with exercise use among those with OA.

PATIENTS AND METHODS

Study setting and sample. The study protocol was approved by the Institutional Review Board of the University of Arizona. Patients were recruited from the Banner University medical center rheumatology, internal medicine, and sports medicine clinics. Medical record reviews were conducted to identify potential subjects. Our target sample included patients scheduled for a clinic visit within the next 45 days with the following characteristics: ≥50 years of age; self-identified as Hispanic, African American, or white; had a diagnosis of knee or hip OA; and did not have a cognitive dysfunction diagnosis. Patients with knee and hip OA were identified, and those with cognitive dysfunction were excluded using the International Classification of Diseases, Ninth Revision, Clinical Modification codes (see Supplementary Figure 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23852/abstract).

Screening and recruitment. Patients identified through medical record reviews and those included in a research registry with OA were screened by telephone for eligibility. In order to assess the presence of chronic, frequent pain due to knee or hip OA, questions from the Arthritis Supplement of the National Health and Nutrition Examination Survey (NHANES) I (20) and NHANES III (21), respectively, were used. Patients with inflammatory arthritis or those who had both knee and hip joint replacement surgery were excluded. The presence or absence of knee or hip OA was determined using the ACR classification criteria (22,23). Knee OA criteria were based on the presence of chronic frequent knee pain in patients who were ages ≥50 years and radiographic evidence of OA (22). Hip OA criteria were based on presence of hip pain and having femoral and/or acetabular osteophytes on radiograph (23). Those who met the appropriate criteria for inclusion gave appropriate consent and were given a survey to complete at home and mail back using a stamped envelope. English and Spanish surveys were available.

Primary outcome. The primary outcome variable was current use of exercise as a form of treatment for OA. This was determined by the question, "Are you currently using or participating in exercise for joint pain or arthritis?" (yes/no). The secondary outcome variable was utilization of exercise in the last 6 months ("Have you used or participated in exercise for joint pain or arthritis in the last 6 months?" [yes/no]).

Study variables. *Sociodemographic*. Information that was collected based on self-report included ethnicity, age, sex, race, educational attainment, employment, marital status, annual household income, and medical insurance status. Structural social support was obtained by determining the number of close friends/relatives. Functional social support was assessed using a 5-item Medical Outcomes Study social support scale (range 0–100) (24). Level of acculturation was measured among Hispanics using the 5-item Short Acculturation Scale for Hispanics (SASH) (25). Hispanic research participants were dichotomized between those

Table 1. Sociodemographic and clinical characteristics by ethnicity*

	Hispanic (n = 130)	Non-Hispanic (n = 232)	*P*†
Age, mean ± SD years	61.8 ± 8.4	65.4 ± 8.1	<0.001
Women	87 (67.4)	171 (73.7)	0.225
Race			<0.001
White	40 (30.8)	202 (87.1)	
African American	1 (0.8)	25 (10.8)	
American Indian or Alaskan Native	10 (7.7)	0 (0.0)	
Other	51 (39.2)	4 (1.7)	
Missing/refuse to answer	28 (21.54)	1 (0.4)	
Education			<0.001
<High school diploma	27 (20.8)	7 (3.0)	
High school/GED	63 (48.5)	78 (33.6)	
≥Associate's degree	35 (26.9)	144 (62.1)	
Employment			<0.001
Full-time	22 (16.9)	43 (18.5)	
Part-time	7 (5.4)	21 (9.1)	
Unemployed	17 (13.1)	8 (3.4)	
Disabled	40 (30.8)	37 (16.0)	
Retired	35 (26.9)	120 (51.7)	
Missing/refuse to answer	9 (6.9)	3 (1.29)	
Marital status, married	57 (43.8)	102 (44.0)	1.000
Annual household income			<0.001
<$20,000	71 (54.6)	60 (25.9)	
$20,000–$39,999	16 (12.3)	38 (16.4)	
≥$40,000	24 (18.5)	115 (49.6)	
Missing/refuse to answer/don't know	19 (14.6)	19 (8.2)	
Insurance			
Medicaid	35 (26.9)	40 (17.2)	0.031
Medicare	58 (44.6)	132 (56.9)	0.028
Medigap	12 (9.2)	58 (25.0)	<0.001
Private	12 (9.2)	72 (31.0)	<0.001
HMO	14 (10.8)	21 (9.1)	0.584
Other	34 (26.2)	26 (11.2)	<0.001
Social support, mean ± SD	72.3 ± 27.7	71.2 ± 26.1	0.730
No. of close friends/relatives, mean ± SD	30.0 ± 37.8	14.1 ± 23.6	<0.001
PHQ-8, mean ± SD	6.3 ± 5.7	5.2 ± 5.3	0.089
Health literacy‡			<0.001
Adequate	89 (69.1)	219 (94.8)	
Inadequate	39 (30.5)	12 (5.2)	
Overall quality of life			0.009
Excellent	13 (10.2)	30 (12.9)	
Very Good	39 (30.5)	92 (39.7)	
Good	32 (25.0)	62 (26.7)	
Fair	32 (25.0)	36 (15.5)	
Poor	12 (9.4)	12 (5.2)	
Comorbidity score, mean ± SD	2.8 ± 2.3	2.6 ± 2.2	0.606
Arthritis self-efficacy, mean ± SD	5.1 ± 2.2	5.6 ± 2.1	0.038
WOMAC-pain, mean ± SD	59.2 ± 20.3	44.0 ± 19.8	<0.001
WOMAC-stiffness, mean ± SD	61.1 ± 21.6	53.1 ± 22.5	0.001
WOMAC-disability, mean ± SD	58.6 ± 21.4	42.9 ± 19.8	<0.001
Kellgren/Lawrence grade			
1	25 (20.2)	57 (25.3)	0.196
2	43 (34.7)	80 (35.6)	
3	37 (29.8)	60 (26.8)	
4	19 (15.3)	28 (12.4)	
Knee (versus hip) participant	104 (80.0)	166 (71.2)	0.080

* Values are the number (%) unless indicated otherwise. GED = general equivalency diploma; HMO = health maintenance organization; PHQ-8 = Patient Health Questionnaire-8; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index.
† Fisher's exact test for categorical variables, Wilcoxon-Mann-Whitney for ordinal variables, and *t*-test for continuous variables.
‡ Based on the question: "How confident are you filling out medical forms by yourself?" (27).

who were less acculturated (average SASH score <3.0) versus more acculturated (average SASH score ≥3.0).

Clinical. Quality of life was assessed by asking, "How would you rate your overall quality of life?" The question is scored on a 5-point ordinal scale ranging from poor to excellent (26). Health literacy was determined by asking, "How confident are you filling out medical forms by yourself?" (27). Participants were dichotomized between those with adequate (extremely, quite a bit) versus inadequate (somewhat, a little bit, or not at all) health literacy. Depression was assessed using the Patient Health Questionnaire-8 (PHQ-8; range 0–24) (28). Medical comorbidity was assessed by self-report using a modified Charlson Comorbidity Index (29).

OA-related disease severity was measured using the 24-item Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). This measure consists of 3 subscales of OA-related disease severity with regard to pain, stiffness, and disability (30). Subscale scores were rescaled to 0–100. Arthritis self-efficacy, or patients' confidence to manage arthritis-related pain, was also determined (range 1–10) (31). Radiographic knee or hip OA disease severity was determined using the Kellgren/Lawrence grading system (32).

Beliefs and attitudes about providers and treatments. Physician participatory decision-making style was assessed by having patients rate their physician's propensity to involve them in treatment decisions (range 0–100) (33). Hall's Trust in Physicians scale was administered (range 11–55) (34). Familiarity with exercise as a treatment for OA was assessed by determining whether patients were aware of the treatment as an option for arthritis, aware of friends or family who exercise, and had an adequate understanding of what happens to someone when he/she exercises (yes or no for each item) (35). Perceived benefit and risk of exercise were assessed using measures of benefit (4 item) and risk (3 item) of total joint replacement surgery, adapted for exercise (36). Responses were averaged to obtain a scale of 1 to 5; higher values indicate increased perception of benefit or risk, respectively.

Statistical analysis. Parametric and nonparametric tests were used to compare patient sociodemographic information, clinical characteristics, and beliefs and attitudes about providers and treatments by ethnicity. Categorical variables were compared using Fisher's exact test. Ordinal variables were compared using the Wilcoxon-Mann-Whitney test. Continuous variables were compared using a 2-sample *t*-test. In the same manner, these characteristics were contrasted by exercise use (current and last 6 months).

A series of logistic regression analyses were conducted to examine the relationship between ethnicity and exercise use (current). The initial model included ethnicity as the only independent variable. The second set tested for the effect of ethnicity after further adjustment for age and OA disease severity, using the WOMAC total score (categorized into tertiles). A final multivariable model was then constructed. All study covariates/mediators that were significantly associated ($P < 0.10$) with the outcome (current exercise use) based on bivariate analyses were considered for inclusion in the model. A fully conditional method was used to create an imputed full model (37). Only variables that were significantly associated ($P < 0.15$) with the outcome in the full model were included in the final model. The entire process was repeated using exercise in the last 6 months as the outcome variable.

RESULTS

Compared to non-Hispanic study participants with OA (n = 232), Hispanic study participants with OA (n = 130) were more likely to be unemployed/disabled and less likely to have private medical insurance (Table 1). Hispanic participants were also more likely to have fair/poor health and had worse WOMAC scores. The majority of Hispanic study participants filled out an English (74.6%) rather than a Spanish (25.4%) version of the survey. Among Hispanics, 62.5% were found to be more acculturated and the rest were found to be less so. Hispanics, in comparison to non-Hispanics, were less likely to

Table 2. Health beliefs about exercise and providers by ethnicity*

	Hispanic (n = 130)	Non-Hispanic (n = 232)	*P*†
Beliefs about exercise			
Familiarity with exercise, no. (%)			
Heard of use of it to treat OA	90 (70.9)	207 (90.8)	<0.001
Have family/friends that received it for OA treatment	41 (37.6)	117 (60.3)	<0.001
Have a good understanding of what happens after treatment	81 (74.3)	181 (83.8)	0.053
Perception of benefit (1, lowest benefit; 5, highest benefit)	3.0 ± 1.0	3.3 ± 1.0	<0.001
Perception of risk (1, lowest risk; 5, highest risk)	2.0 ± 0.8	1.8 ± 0.8	0.083
Beliefs about providers			
Physician participatory decision-making style	53.5 ± 30.2	67.9 ± 28.0	<0.001
Trust in physicians	37.6 ± 8.8	37.24 ± 8.1	0.499

* Values are the mean ± SD unless indicated otherwise. OA = osteoarthritis.
† Fisher's exact test for familiarity with exercise questions; Wilcoxon-Mann-Whitney for others

Table 3. Unadjusted and adjusted odds of currently exercising for OA treatment*

Variable	Unadjusted model 1 (n = 349)	Adjusted model 2 (n = 326)	Final adjusted model 3 (n = 349)
Hispanic	0.52 (0.33–0.81)†	0.64 (0.39–1.06)	0.86 (0.49–1.50)
Age		1.01 (0.98–1.04)	0.99 (0.96–1.02)
OA disease severity (WOMAC total)‡			
Minimal–mild		Ref.	Ref.
Moderate		0.50 (0.28–0.89)†	0.60 (0.31–1.19)
Severe		0.35 (0.19–0.63)†	0.62 (0.30–1.27)
Married			1.64 (0.97–2.77)
PHQ-8 score			1.05 (0.99–1.12)
Quality of life, poor or fair (vs. excellent, very good, or good)			0.37 (0.18–0.77)†
Have family/friends that exercise for OA treatment			2.22 (1.27–3.89)†
Have good understanding of what happens after exercise			1.76 (0.89–3.48)
Perceived benefit of exercise score			1.87 (1.40–2.50)†
Perceived risk of exercise score			0.75 (0.53–1.05)

* Values are the odds ratio (95% confidence interval). OA = osteoarthritis; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index; Ref. = reference; PHQ-8 = Patient Health Questionnaire-8.
† $P < 0.05$.
‡ Categories created based on WOMAC total tertile scores with a higher score indicating higher OA disease severity.

have heard about exercise as treatment for OA and to have family/friends who received exercise therapy for treatment (Table 2). Among Hispanics, the mean perceived benefit of exercise score was also lower while the mean perceived risk of exercise score was marginally higher.

Exercise use for OA by ethnicity. Hispanics, compared to non-Hispanics, were less likely to report current exercise use (50% versus 65%; OR 0.52 [95% CI 0.33–0.81]) and exercise use in the past 6 months (56% versus 73%; OR 0.49 [95% CI 0.31–0.77]) for OA treatment (Tables 3 and 4). When adjusted for age and OA disease severity, the ethnic differences in current exercise use was attenuated and was no longer statistically significant (OR 0.64 [95% CI 0.39–1.06]) but the ethnic differences in exercise use in the last 6 months remained statistically significant (OR 0.59 [95% CI 0.36–0.99). When further adjusted for specific lower extremity joint involvement, reported knowledge of having family or friends who exercise for OA, reported understanding of what happens after exercise, and perceived benefit score, the association between ethnicity and exercise use in the last 6 months was further attenuated and was no longer statistically significant (OR 0.74 [95% CI 0.41–1.33]).

Determinants of exercise use: current and last 6 months. Regardless of ethnicity, patients with OA who were currently exercising had higher annual income and were more likely to have private medical insurance compared with OA patients not currently exercising for OA treatment (Table 5). Current exercise use was associated with lower PHQ-8 score, adequate health literacy, excellent/very good overall quality of life, lower comorbidity score, higher arthritis self-efficacy score,

Table 4. Unadjusted and adjusted odds of exercising in the last 6 months for OA treatment*

Variable	Unadjusted model 1 (n = 354)	Adjusted model 2 (n = 330)	Final adjusted model 3 (n = 354)
Hispanic	0.49 (0.31–0.77)†	0.59 (0.36–0.99)†	0.74 (0.41–1.33)
Age		1.00 (0.97–1.03)	0.98 (0.95–1.02)
OA disease severity (WOMAC total)‡			
Minimal-mild		Ref.	Ref.
Moderate		0.59 (0.32–1.08)	0.71 (0.34–1.46)
Severe		0.40 (0.21–0.74)†	0.60 (0.28–1.29)
Knee (vs. hip) OA			2.83 (1.51–5.29)†
Have family/friends that exercise for OA treatment			3.20 (1.76–5.84)†
Have good understanding of what happens after exercise			2.19 (1.15–4.19)†
Perceived benefit of exercise score			2.24 (1.64–3.04)†

* Values are the odds ratio (95% confidence interval). OA = osteoarthritis; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index; Ref. = reference.
† $P < 0.05$.
‡ Categories created based on WOMAC total tertile scores, with a higher score indicating higher OA disease severity.

Table 5. OA patient characteristics and beliefs associated with current exercise use*

	Currently exercising (n = 209)	Not currently exercising (n = 140)	*P*†
Race			0.019
White	156 (74.6)	82 (58.6)	
Black or African American	9 (4.3)	13 (9.3)	
American Indian or Alaskan Native	5 (2.4)	5 (3.6)	
Other	23 (11.0)	28 (20.0)	
Missing/refuse to answer	16 (7.7)	12 (8.6)	
Education			0.053
<High school diploma	15 (7.2)	15 (10.7)	
High school/GED	73 (34.9)	62 (44.3)	
≥Associate's degree	118 (56.5)	59 (42.1)	
Other	3 (1.4)	4 (2.9)	
Marital status, married	103 (49.3)	54 (38.6)	0.062
Annual household income			<0.001
<$20,000	61 (29.2)	61 (43.6)	
$20,000–39,999	31 (14.8)	22 (15.7)	
≥$40,000	102 (48.8)	37 (26.4)	
Missing/refuse to answer/don't know	15 (7.2)	20 (14.3)	
Insurance			
Medicaid	34 (16.3)	38 (27.1)	0.015
Private	60 (28.7)	23 (16.4)	0.010
Social support, mean ± SD	74.4 ± 23.8	68.7 ± 30.1	0.070
PHQ-8, mean ± SD	4.9 ± 5.3	6.4 ± 5.6	0.022
Health literacy‡			0.007
Adequate	188 (90.4)	110 (79.7)	
Inadequate	20 (9.6)	28 (20.3)	
Overall quality of life			<0.001
Excellent	29 (13.9)	13 (9.4)	
Very good	92 (44.0)	35 (25.4)	
Good	55 (26.3)	37 (26.8)	
Fair	25 (12.0)	39 (28.3)	
Poor	8 (3.8)	14 (10.1)	
Comorbidity score, mean ± SD	2.4 ± 1.9	3.0 ± 2.4	0.006
Arthritis self-efficacy, mean ± SD	5.7 ± 2.1	5.0 ± 2.1	0.002
WOMAC-pain, mean ± SD	45.3 ± 21.6	54.3 ± 19.6	<0.001
WOMAC-stiffness, mean ± SD	53.7 ± 22.0	58.9 ± 23.0	0.037
WOMAC-disability, mean ± SD	44.3 ± 21.7	54.0 ± 20.9	<0.001
Familiarity with exercise			
Heard of use of it to treat OA	200 (96.2)	92 (67.6)	<0.001
Have family/friends that received it for OA treatment	115 (64.6)	40 (33.9)	<0.001
Have a good understanding of what happens after treatment	179 (87.8)	79 (69.9)	<0.001
Perception of benefit, mean ± SD	3.5 ± 0.9	2.7 ± 1.0	<0.001
Perception of risk, mean ± SD	1.7 ± 0.7	2.1 ± 0.9	<0.001
Physician participatory decision-making style, mean ± SD	67.9 ± 27.0	55.5 ± 31.6	<0.001

* Values are the number (%) unless indicated otherwise ($P < 0.10$). OA = osteoarthritis; GED = general equivalency diploma; PHQ-8 = Patient Health Questionnaire-8; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index.
† Fisher's exact test for categorical variables, Wilcoxon-Mann-Whitney for ordinal variables, and *t*-test for continuous variables.
‡ Based on the question: "How confident are you filling out medical forms by yourself?" (27).

and better WOMAC pain and disability scores. Reporting "yes" to the measure items regarding familiarity with exercise as treatment for OA, and higher perceived benefit and lower perceived risk of exercise use were all significantly associated with current exercise use for OA treatment. In the final multivariable logistic regression model with current exercise use as the outcome variable, having family or friends who exercise for OA treatment (OR 2.22 [95% CI 1.27–3.89]), higher perceived benefit of exercise (OR 1.87 [95% CI 1.40–2.50]), and overall quality of life (OR 0.37 [95% CI 0.18–0.77]) were the only variables that were statistically significantly associated with exercise use for OA (Table 3).

The sociodemographic and clinical information, and beliefs about treatments and providers associated with exercise use for OA in the last 6 months were largely the same as that of current exercise use for OA (see Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.

com/doi/10.1002/acr.23852/abstract). In the final multivariable logistic regression model with exercise use in the last 6 months as the outcome variable, having family or friends who exercise for OA treatment (OR 3.20 [95% CI 1.76–5.84]), having a good understanding of what happens after exercise (OR 2.19 [95 CI 1.15–4.19]), higher perceived benefit of exercise score (OR 2.24 [95% CI 1.64–3.04]), and having knee instead of hip OA (OR 2.83 [95% CI 1.51–5.29]) were all statistically significantly associated with exercise use for OA in the last 6 months (Table 4). Multivariable logistic regression models with current exercise or exercise in the last 6 months as the outcome variable, but without imputation, were created with similar results (data not shown).

DISCUSSION

Our findings on ethnic differences in exercise use for OA are consistent with what other studies on ethnic differences in PA have found (11,12). According to the 2010 National Health Interview Survey, approximately 42% of Hispanics versus 29% of non-Hispanic whites were classified as being "inactive" based on the 2008 PA Guidelines for Americans (11). Also, only 38% of Hispanics, compared to 49% of non-Hispanic whites, were considered "sufficiently active" in this national survey. Our study uniquely contributes to the literature by demonstrating the extent of ethnic disparities in exercise use for OA treatment, after controlling for relevant clinical information and patient attitudes and beliefs. The results of our study suggest that, after adjustment for age, disease severity, specific lower extremity joint involvement, familiarity with exercise as treatment for OA, and perceived benefits of exercise, the ethnic differences in exercise use in the last 6 months was no longer statistically significant. Our findings suggest that these factors may be most relevant in explaining the ethnic differences in exercise use among those with OA.

Demonstrating that the belief in the benefit of exercise and a good understanding of the effects of exercise were both associated with increased exercise use in OA, regardless of ethnicity, has important implications. While certain factors that determine exercise use may be difficult to modify, patient knowledge and attitudes toward exercise may be modified at the point of care. Becoming educated about the benefits of exercise may influence views about exercise. In a study by Bopp et al (38), Hispanic participants who received culturally and spiritually relevant education materials that promoted the health benefits of PA were more likely to identify reasons for exercise and to accurately describe PA recommendations. In a qualitative study of Mexican Americans (39), sense of well-being that was derived from PA was identified as a primary motivator to increase PA.

With greater knowledge about exercise, patients with OA may have greater confidence in their ability to exercise. Self-efficacy, the confidence in one's own ability to successfully carry out courses of action, has been found to be a strong predictor of PA among Hispanics (40–43). A review by Marquez et al (40) identified self-efficacy as the most commonly reported correlate of PA among Hispanics. This was found with self-efficacy measures specific to exercise (41,42) and measures of general efficacy (43). In a study of older Mexican American women (42), for example, exercise self-efficacy was related to leisure/sport activities and daily and habitual activities. We also found an association between arthritis self-efficacy and current exercise use in our cohort.

Similarly, the finding in the present study that having family/friends who exercise for OA was also associated with increased use of exercise suggests that exposing patients with OA to social networks that regularly exercise may motivate them to follow suit. Other studies have demonstrated that knowing people who exercise or seeing people who exercise in the neighborhood has been associated with increased PA (44,45). Family or friends who exercise may provide exercise-related social support (e.g., having someone with whom patients with OA could exercise). Such type of social support is a common correlate of exercise among Hispanics (41,46). Hovell et al (46) reported that social support from friends specifically for exercise was strongly correlated with minutes of walking, walking for exercise, and vigorous activity among Hispanics. Many argue that social support is particularly important for Hispanics, given the emphasis on family and interpersonal relationships within the Hispanic culture (47). The measure specifically regarding social support in our study was not specific to leisurely activities, though, and we did not find a significant association between general social support and exercise use.

Previous studies have shown that exercise use was less common among Hispanics with poorer quality of life. Generalized fatigue/tiredness has been previously linked to being less physically active among Hispanics (48) and identified as a barrier to starting an exercise regimen (49). Latinas from North Carolina who reported excellent or very good health were also more likely to meet PA guideline recommendations than others (44). Our study adds to the literature by demonstrating that greater OA disease severity and poorer quality of life were associated with decreased exercise use among patients with OA. It is unclear as to why exercise use in the last 6 months was more common among those with knee OA than among those with hip OA. In our study cohort, hip OA was only assessed among those who had knee replacement surgery or among those who did not have knee OA symptoms.

There are important limitations to consider in interpreting our findings. First, while Hispanics were found to be less physically active in our study and other studies (11,12), significant variability in PA exists among different Hispanic subgroups. According to a national survey (12), Cubans and Dominicans were found to be the least active among different Hispanic subgroups in the US. We primarily recruited from Arizona, where Hispanics are mostly of Mexican descent (50). Second, our study and other published studies (40) only measured leisure time PA and neglected adequate examination of occupational or domestic activity. Yet, for some Hispanics, these nonleisure time activities may be the domi-

nant forms of PA in their daily lives (40). Third, other cultural beliefs and attitudes may also determine exercise and were unmeasured in this study. Cultural attitudes toward weight and body shape may act as barriers to exercise. In Hispanic culture, for instance, being overweight symbolizes wealth or good health, and some Hispanic women may not prioritize exercising due to reduced concerns about being overweight (51). Finally, the extent and quality of exercise were not specifically measured.

Our study showed differences in the use of exercise for the treatment of lower extremity OA between Hispanics and non-Hispanics. However, the ethnic differences in exercise use were no longer significant after adjustment for OA-specific clinical characteristics and patient familiarity with and perceived benefits of exercise for OA treatment. We also demonstrated that potentially modifiable factors, including patients' familiarity with exercise as therapy for OA and their perceived benefits of exercise, were major determinants of exercise use. These findings may have important implications in terms of modifying risk factors for physical inactivity. Behavioral programs designed to increase exercise may need to be tailored to different ethnic groups. Training programs directed toward providers may need to include culturally appropriate strategies to motivate patients to exercise. Additional research should be conducted to determine whether the implementation of such programs could decrease the gap in the utilization of exercise between Hispanics and non-Hispanics with OA.

ACKNOWLEDGMENTS

The authors thank all study staff and participants.

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. Vina 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. Vina, Hausmann, Ibrahim, Kwoh.
Acquisition of data. Dagnino, Arellano.
Analysis and interpretation of data. Vina, Hannon, Hausmann, Ibrahim, Kwoh.

REFERENCES

1. Deshpande BR, Katz JN, Solomon DH, Yelin EH, Hunter DJ, Messier SP, et al. Number of persons with symptomatic knee osteoarthritis in the US: impact of race and ethnicity, age, sex, and obesity. Arthritis Care Res (Hoboken) 2016;68:1743–50.
2. Centers for Disease Control and Prevention (CDC). Prevalence of doctor-diagnosed arthritis and arthritis-attributable effects among Hispanic adults, by Hispanic subgroup–United States, 2002, 2003, 2006, and 2009. MMWR Morb Mortal Wkly Rep 2011;60:167–71.
3. Wright NC, Riggs GK, Lisse JR, Chen Z. Women's Health Initiative. Self-reported osteoarthritis, ethnicity, body mass index, and other associated risk factors in postmenopausal women-results from the Women's Health Initiative. J Am Geriatr Soc 2008;56:1736–43.
4. Bolen J, Schieb L, Hootman JM, Helmick CG, Theis K, Murphy LB, et al. Differences in the prevalence and severity of arthritis among racial/ethnic groups in the United States, National Health Interview Survey, 2002, 2003, and 2006. Prev Chronic Dis 2010;7:A64. E-pub 2010 Apr 15.
5. Barbour KE, Helmick CG, Boring M, Brady TJ. 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.
6. Buchner DM, Beresford SA, Larson EB, LaCroix AZ, Wagner EH. Effects of physical activity on health status in older adults. II. Intervention studies. Annu Rev Public Health 1992;13:469–88.
7. Fiatarone MA, Evans WJ. The etiology and reversibility of muscle dysfunction in the aged. J Gerontol 1993;48:77–83.
8. Fransen M, McConnell S, Harmer AR, Van der Esch M, Simic M, Bennell KL. Exercise for osteoarthritis of the knee. Cochrane Database Syst Rev 2015;1:CD004376.
9. Fransen M, McConnell S, Hernandez-Molina G, Reichenbach S. Exercise for osteoarthritis of the hip. Cochrane Database Syst Rev 2014;CD007912.
10. Hochberg MC, Altman RD, April KT, Benkhalti M, Guyatt G, McGowan J, et al. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. Arthritis Care Res (Hoboken) 2012;64:465–74.
11. Schiller JS, Lucas JW, Ward BW, Peregoy JA. Summary health statistics for U.S. adults: National Health Interview Survey, 2010. Vital Health Stat 10 2012:1–207.
12. Neighbors CJ, Marquez DX, Marcus BH. Leisure-time physical activity disparities among Hispanic subgroups in the United States. Am J Public Health 2008;98:1460–4.
13. Office of Disease Prevention and Health Promotion. Healthy People 2020: Social determinants of health–2015. URL: https://www.healthypeople.gov/2020/topics-objectives/topic/social-determinants-of-health.
14. Crespo CJ, Smit E, Andersen RE, Carter-Pokras O, Ainsworth BE. Race/ethnicity, social class and their relation to physical inactivity during leisure time: results from the Third National Health and Nutrition Examination Survey, 1988-1994. Am J Prev Med 2000;18:46–53.
15. Andersen R, Newman JF. Societal and individual determinants of medical care utilization in the United States. Milbank Mem Fund Q Health Soc 1973;51:95–124.
16. Allen KD, Helmick CG, Schwartz TA, DeVellis RF, Renner JB, Jordan JM. Racial differences in self-reported pain and function among individuals with radiographic hip and knee osteoarthritis: the Johnston County Osteoarthritis Project. Osteoarthritis Cartilage 2009;17:1132–6.
17. Armstrong K, Ravenell KL, McMurphy S, Putt M. Racial/ethnic differences in physician distrust in the United States. Am J Public Health 2007;97:1283–9.
18. Vina ER, Ran D, Hannon MJ, Kwoh CK. Understanding ethnic differences in the utilization of exercise for osteoarthritis. Ann Rheum Dis 2018;77 Suppl 2:A519.
19. Canino G, Koinis-Mitchell D, Ortega AN, McQuaid EL, Fritz GK, Alegria M. Asthma disparities in the prevalence, morbidity, and treatment of Latino children. Soc Sci Med 2006;63:2926–37.
20. Davis MA, Ettinger WH, Neuhaus JM. Obesity and osteoarthritis of the knee: evidence from the National Health and Nutrition Examination Survey (NHANES I). Semin Arthritis Rheum 1990;20 Suppl 1:34–41.
21. Christmas C, Crespo CJ, Franckowiak SC, Bathon JM, Bartlett SJ, Andersen RE. How common is hip pain among older adults? Results from the Third National Health and Nutrition Examination Survey. J Fam Pract 2002;51:345–8.

22. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. Arthritis Rheum 1986;29:1039–49.
23. Altman R, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum 1991;34:505–14.
24. Sherbourne CD, Stewart AL. The MOS social support survey. Soc Sci Med 1991;32:705–14.
25. Marin G, Sabogal F, Marin BV, Otero-Sabogal R, Perez-Stable EJ. Development of a short acculturation scale for Hispanics. Hisp J Behav Sci 1987;9:183–205.
26. Covinsky KE, Wu AW, Landefeld CS, Connors AF Jr, Phillips RS, Tsevat J, et al. Health status versus quality of life in older patients: does the distinction matter? Am J Med 1999;106:435–40.
27. Chew LD, Griffin JM, Partin MR, Noorbaloochi S, Grill JP, Snyder A, et al. Validation of screening questions for limited health literacy in a large VA outpatient population. J Gen Intern Med 2008;23:561–6.
28. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatr Ann 2002;32:1–7.
29. Chaudhry S, Jin L, Meltzer D. Use of a self-report-generated Charlson Comorbidity Index for predicting mortality. Med Care 2005;43:607–15.
30. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833–40.
31. Lorig K, Chastain RL, Ung E, Shoor S, Holman HR. Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis. Arthritis Rheum 1989;32:37–44.
32. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957;16:494–502.
33. Kaplan SH, Gandek B, Greenfield S, Rogers W, Ware JE. Patient and visit characteristics related to physicians' participatory decision-making style. Results from the Medical Outcomes Study. Med Care 1995;33:1176–87.
34. Anderson LA, Dedrick RF. Development of the Trust in Physician scale: a measure to assess interpersonal trust in patient-physician relationships. Psychol Rep 1990;67:1091–100.
35. Ibrahim SA, Siminoff LA, Burant CJ, Kwoh CK. Understanding ethnic differences in the utilization of joint replacement for osteoarthritis: the role of patient-level factors. Med Care 2002;40 Suppl 1:I44–51.
36. Suarez-Almazor ME, Souchek J, Kelly PA, O'Malley K, Byrne M, Richardson M, et al. Ethnic variation in knee replacement: patient preferences or uninformed disparity? Arch Intern Med 2005;165:1117–24.
37. Liu Y, De A. multiple imputation by fully conditional specification for dealing with missing data in a large epidemiologic study. Int J Stat Med Res 2015;4:287–95.
38. Bopp M, Fallon EA, Marquez DX. A faith-based physical activity intervention for Latinos: outcomes and lessons. Am J Health Promot 2011;25:168–71.
39. Mier N, Medina AA, Ory MG. Mexican Americans with type 2 diabetes: perspectives on definitions, motivators, and programs of physical activity. Prev Chronic Dis 2007;4:A24.
40. Marquez DX, McAuley E, Overman N. Psychosocial correlates and outcomes of physical activity among Latinos: a review. Hisp J Behav Sci 2004;26:195–229.
41. Marquez DX, McAuley E. Social cognitive correlates of leisure time physical activity among Latinos. J Behav Med 2006;29:281–9.
42. Laffrey SC. Physical activity among older Mexican American women. Res Nurs Health 2000;23:383–92.
43. Duffy ME, Rossow R, Herandez M. Correlates of health-promotion activities in employed Mexican American women. Nurs Res 1996;45:18–24.
44. Evenson KR, Sarmiento OL, Tawney KW, Macon ML, Ammerman AS. Personal, social, and environmental correlates of physical activity in North Carolina Latina immigrants. Am J Prev Med 2003;25 Suppl 1:77–85.
45. Gallant MP. The influence of social support on chronic illness self-management: a review and directions for research. Health Educ Behav 2003;30:170–95.
46. Hovell M, Sallis J, Hofstetter R, Barrington E, Hackley M, Elder J, et al. Identification of correlates of physical activity among Latino adults. J Community Health 1991;16:23–36.
47. Sabogal F, Marín G, Otero-Sabogal R, Marín BV, Perez-Stable EJ. Hispanic familism and acculturation: what changes and what doesn't? Hisp J Behav Sci 1987;9:397–412.
48. King AC, Castro C, Wilcox S, Eyler AA, Sallis JF, Brownson RC. Personal and environmental factors associated with physical inactivity among different racial-ethnic groups of U.S. middle-aged and older-aged women. Health Psychol 2000;19:354–64.
49. Heesch KC, Brown DR, Blanton CJ. Perceived barriers to exercise and stage of exercise adoption in older women of different racial/ethnic groups. Women Health 2000;30:61–76.
50. Ennis SR, Rios-Vargas M, Albert NG. The Hispanic population: 2010. 2010 Census Briefs. United States Census Bureau; 2011. URL: https://www.census.gov/prod/cen2010/briefs/c2010br-04.pdf.
51. Juarbe TC. Cardiovascular disease-related diet and exercise experiences of immigrant Mexican women. West J Nurs Res 1998;20:765–82.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 71, No. 11, November 2019, pp 1504–1515
DOI 10.1002/acr.23786

Impact of Exercise Therapy on Molecular Biomarkers Related to Cartilage and Inflammation in Individuals at Risk of, or With Established, Knee Osteoarthritis: A Systematic Review and Meta-Analysis of Randomized Controlled Trials

Alessio Bricca,[1] André Struglics,[2] Staffan Larsson,[2] Martijn Steultjens,[3] Carsten B. Juhl,[4] and Ewa M. Roos[1]

Objective. To investigate the impact of exercise therapy on molecular biomarkers related to cartilage and inflammation in individuals at risk of, or with established, knee osteoarthritis by conducting a systematic review of randomized controlled trials (RCTs).

Methods. We conducted a literature search up to September 2017 in 5 major databases with no restriction on publication year or language. Data were extracted from the first available follow-up time point, and we performed a narrative synthesis for the effect of exercise therapy on molecular biomarkers related to cartilage and inflammation. A subset of studies reporting sufficient data was combined in a meta-analysis, using an adjusted random-effects model.

Results. Twelve RCTs involving 57 study comparisons at 4 to 24 weeks following an exercise-therapy intervention were included. Exercise therapy decreased molecular biomarkers in 17 study comparisons (30%), had no effect in 36 (63%), and increased molecular biomarkers in 4 study comparisons (7%). Meta-analyses of 9 biomarkers showed that exercise therapy was associated with nonsignificant reductions of the C-reactive protein level, C-terminal crosslinking telopeptide of type II collagen, tumor necrosis factor (TNF), soluble TNF receptors 1 and 2, C2C neoepitope of type II collagen, and cartilage oligomeric matrix protein, compared to nonexercising control groups, and exercise therapy had no effect on interleukin-6 and soluble interleukin-6 receptor.

Conclusion. Exercise therapy is not harmful, because it does not increase the concentration of molecular biomarkers related to cartilage turnover and inflammation, implicated in osteoarthritis progression. The overall quality of evidence was downgraded to low because of the limited number of RCTs available.

INTRODUCTION

Osteoarthritis (OA) is the most common joint disease, and its prevalence in the western world has doubled since the mid-twentieth century (1). OA represents one of the main reasons for disability, where the knee accounts for >80% of the disease burden (2). It is broadly agreed that OA is driven by a combination of biomechanic and proinflammatory factors, ultimately leading to osteochondral changes, with cartilage breakdown being one of the hallmarks of OA (3). Exercise is essential for the health of the knee joint, with cartilage being able to adapt its structure, composition, and metabolism to a wide range of activities (4–6). However, very high doses of exercise such as playing sports at an elite level (4,7) as well as the absence of exercise, in the forms of sedentary behavior (8) or immobilization (9), are associated with cartilage loss and OA development. Therapeutic exercise is a cornerstone in the management of OA, (10,11). When prescribed for specific therapeutic goals, exercise has been shown to be clinically safe (12–14) and effective in reducing pain and improving function (15,16). Nevertheless, some patients with knee OA still believe that therapeutic exercise may be detrimental to their knee joints (17), constituting a barrier to exercise.

Supported by the European Union Seventh Framework Programme (project KNEEMO, FP7-PEOPLE-2013-ITN, grant 607510).

[1]Alessio Bricca, PhD, Ewa M. Roos, PhD, PT: University of Southern Denmark, Odense, Denmark; [2]André Struglics, PhD, Staffan Larsson, PhD: Lund University, Lund, Sweden; [3]Martijn Steultjens, PhD: Glasgow Caledonian University, Glasgow, UK; [4]Carsten B. Juhl, PhD, PT: University of Southern Denmark, Odense, and Herlev and Gentofte Hospitals, University of Copenhagen, Copenhagen, Denmark.

Dr. Roos is developer of the Knee Injury and Osteoarthritis Outcome Score (KOOS) and several other freely available patient-reported outcome measures. No other disclosures relevant to this article were reported.

Address correspondence to Alessio Bricca, PhD, Research Unit for Musculoskeletal Function and Physiotherapy, Department of Sports Science and Clinical Biomechanics, Faculty of Health Sciences, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark. E-mail: abricca@health.sdu.dk.

Submitted for publication May 11, 2018; accepted in revised form October 9, 2018.

SIGNIFICANCE & INNOVATIONS

- To the best of our knowledge, this is the first study summarizing the effect of exercise therapy on molecular biomarkers related to cartilage turnover and inflammation in individuals at risk of, or with established, knee osteoarthritis.
- Based on the available evidence, individuals at risk of, or with established, knee osteoarthritis can be told that exercise therapy is not harmful to their knee joints.
- Future studies should preferably obtain synovial fluid from individuals at risk of, or at early stages of, osteoarthritis and focus on a set of biomarkers, rather than on single biomarkers.

A moderate mechanical loading of the knee joint from exercise therapy is thought to slow down cartilage breakdown by balancing anabolic and catabolic reactions in the extracellular matrix (18). Molecular biomarkers in blood, urine, and joint fluids are promising disease markers in predicting structural OA progression and in assessing therapeutic response related to cartilage and inflammation (19,20). Individuals with knee OA have higher levels of circulating cartilage-derived biomarkers compared to healthy controls (21,22). Systematic reviews that included overweight and normal-weight youth (23) and adult (24) participants, with or without cardiovascular diseases (25), have shown a beneficial effect of exercise on reducing the C-reactive protein (CRP) level, a molecular biomarker related to systemic inflammation also involved in OA progression. In the OA population, individual studies indicate that single bouts of exercise therapy promote immediate changes to molecular biomarkers related to cartilage extracellular matrix turnover (e.g., cartilage oligomeric matrix protein [COMP]) and inflammation (e.g., interleukin-10 [IL-10]) that, in general, return to baseline levels after a short period of rest (6,26–28). However, whether therapeutic exercise interventions have an impact on the molecular biomarker concentration has previously only been investigated in individual studies, and the effect has not been summarized in a systematic review and meta-analysis. We aimed to investigate the impact of exercise therapy interventions on molecular biomarkers related to articular cartilage and inflammation, by systematically reviewing published randomized controlled trials in individuals at risk of, or with established, knee OA.

MATERIALS AND METHODS

Protocol. Study selection, eligibility criteria, data extraction, and statistical analysis were performed according to the Cochrane Collaboration guidelines (29). The study has been reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines and the study protocol was registered at PROSPERO (CRD42017055850).

Eligibility criteria. We included randomized controlled trials that investigated the impact of exercise therapy on molecular biomarkers related to cartilage and inflammation in individuals at risk of, or with established, knee OA. Studies were excluded when no full text was available, or when the active treatment arm involved other joint-loading interventions besides exercise therapy.

Literature search. A systematic literature search was performed with no restriction on publication year and language in Medline via PubMed, Embase via Ovid, CINAHL (including pre-CINAHL) via EBSCO, the Cochrane Central Register of Controlled Trials, and Web of Science up to January 2017. The search was re-run in these databases up to September 2017.

Search methods and study selection. The studies were identified by performing a customized search strategy (see Supplementary Appendix 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23786/abstract). All terms were searched, if possible, both as keywords (Medical Subject Headings) and as text words in titles and abstracts. In MEDLINE and EMBASE, animal studies were identified and removed before screening all the studies, using a validated animal filter (30,31). At first, 2 of the authors (AB and CBJ) independently screened titles and abstracts, and all studies deemed eligible by at least 1 of the authors were checked independently in full text by the same reviewers. Disagreements regarding inclusion were discussed between the 2 reviewers until consensus was reached.

Data collection. Data were extracted by 2 of the authors (AB and CBJ) from the studies, including tables and graphs of published articles. A customized data extraction form was developed for each of the molecular biomarkers. The molecular biomarkers were grouped by bio-fluid source into synovial fluid, serum, plasma, and urine. The following data extraction was mandatory: authors of the study, year of publication, design of the trial, intervention characteristics, location of the trial (in the case of multicenter studies, the primary investigator's affiliation applied), the number of patients allocated (to the exercise and control groups), the number of patients in the intent-to-treat (ITT) population (in the intervention and control groups), the average patient age, average body mass index (BMI), the number of females within the ITT population, duration of the study, intervention characteristics, site of collection of bio-fluid, and analysis method of molecular biomarkers.

Molecular biomarkers classification. We classified molecular biomarkers based on their main function and grouped them into biomarkers of either inflammation or cartilage extracellular matrix turnover. Molecular biomarkers related to inflammation were subgrouped into markers of inflammation (CRP, CRP

degradation, IL-6, tumor necrosis factor [TNF], and transforming growth factor β1) and cytokine receptors (soluble IL-6 receptor [sIL-6r] and soluble TNF receptors 1 and 2 [TNFR1 and TNFR2]). Molecular biomarkers related to cartilage extracellular matrix turnover were subgrouped into: 1) proteases (matrix metalloprotease 3); 2) turnover of collagens (type II collagen synthesis [type II collagen carboxy propeptide (CPII)], type II collagen degradation [C2C neoepitope, C2M neoepitope, and C-terminal crosslinking telopeptide (CTX-II)], and total collagen [hydroxyproline (HP)]); 3) glycoproteins (cartilage oligomeric matrix protein [COMP]); and 4) glycosaminoglycans (total glycosaminoglycans using the dimethylmethylene blue assay, chondroitin sulfate [epitopes 3B3 and 7D4], keratan sulfate [epitope 5D4], and hyaluronic acid [HA]).

Meta-analysis of a subset of molecular biomarkers. We performed meta-analyses when at least 2 study comparisons were available for an outcome of interest. We estimated the standardized mean difference (SMD) as the difference between mean change values (or post-intervention values when only post-intervention data were available) in the intervention and control groups, divided by the pooled SD, using a random-effects model and Hedges' correction. The SD was extracted or estimated from the SEM, the 95% confidence interval (95% CI), *P* value, or other methods recommended by the Cochrane Collaboration (29). Between-study heterogeneity was calculated using the I^2 statistic (32), measuring the proportion of variation (i.e., inconsistency) in the combined estimates due to between-study variance (33). An I^2 value of 0% indicates no inconsistency among the results of individual trials, while an I^2 value of 100% indicates maximum inconsistency. When several intervention groups were compared to 1 control group, the number of participants in the control group was divided by the number of intervention groups and each was analyzed as a separate study comparison.

Narrative synthesis of results. For the effect of exercise therapy on molecular biomarkers, we reported a statistically significant ($P < 0.05$) decrease or increase in molecular biomarker concentrations for the exercise therapy group compared to the control group or no difference (no change) in biomarker concentrations between the 2 groups. The effect estimates derived from meta-analyses were included in the overall narrative synthesis of the results and reported as a decrease if the SMD was less than –0.2; no difference if the SMD was between –0.2 and 0.2; and as an increase if the SMD was higher than 0.2 (29). We performed subgroup analyses on molecular biomarker localization (i.e., synovial fluid, blood, and urine), reporting the number of studies that

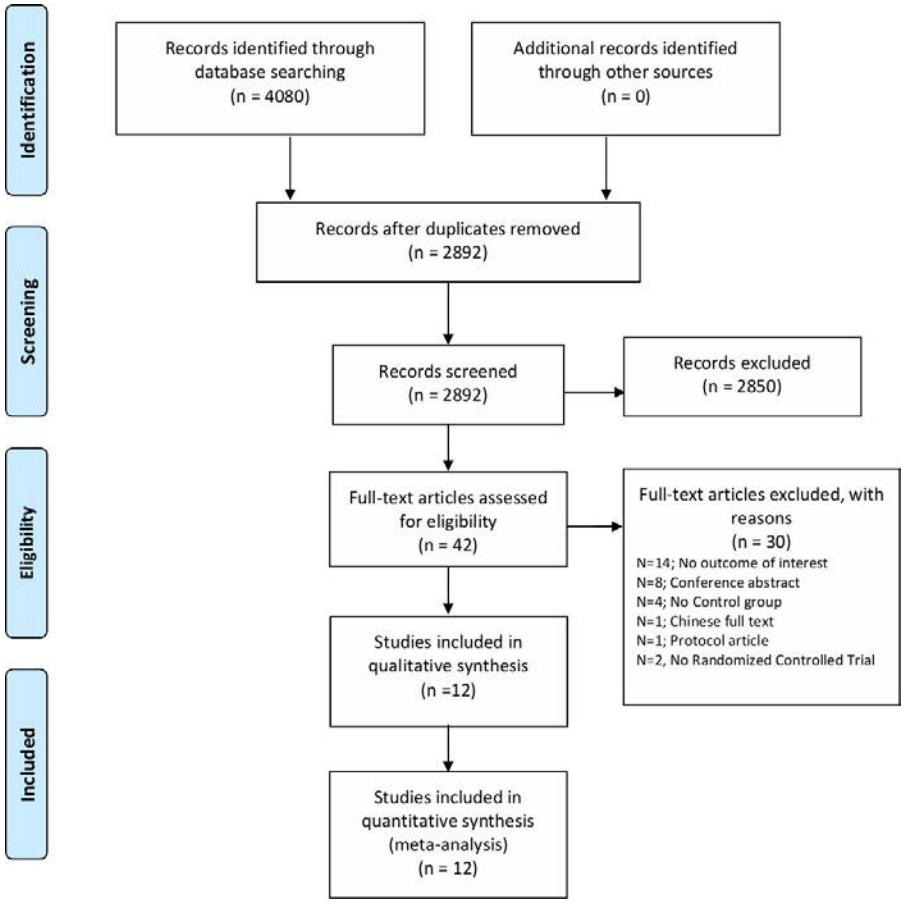

Figure 1. Flow chart of the included studies.

showed a change in concentration of the molecular biomarkers of interest.

When several intervention groups were included in a study, the between-group difference was reported for each possible comparison. For example, when a study had 2 intervention groups (A and B) and 1 control group (C), we compared A versus C and B versus C and reported the results as 2 separate study comparisons. Although including multiple comparisons from the same study does not completely rule out dependence between estimates of effect in meta-analysis, this procedure is in accordance with the Cochrane handbook (29).

Sensitivity analysis and quality of evidence. In addition, we performed a sensitivity analysis on the studies not included in the meta-analysis by calculating their effect size when sufficient data were available. The Grading of Recommendations, Assessment, Development and Evaluations (GRADE) technique is a systematic approach to rate the overall quality of evidence, from high to very low. The presence of high-quality evidence indicates that "future research is very unlikely to change the estimates of effect" while very low-quality evidence indicates that "any estimate of effect is very uncertain." The GRADE assessment involves the following domains: risk of bias (i.e., the methodologic flaws of the studies); inconsistency (i.e., the heterogeneity of results across studies), indirectness (i.e., the generalizability of the findings to the target population), the precision of the estimates, and the risk of publication bias (34).

Risk of bias and the overall quality of evidence was independently assessed by 2 authors (AB and CBJ) using the GRADE approach (34). Disagreements in initial ratings of methodologic quality assessment were discussed between 2 of the authors (AB and CBJ) until consensus was reached. The risk of bias was assessed with regard to the risk of selection bias, performance bias, detection bias, attrition bias, reporting bias, and other sources of bias. Each of the following listed domains was assessed as adequate, unclear, or inadequate: sequence generation, allocation concealment, blinding, incomplete outcome data addressed, selective outcome reporting, or other bias (i.e., funding) (32) (Figure 1).

RESULTS

Study selection and characteristics. The literature search resulted in 4,080 publications, of which 42 individual studies were identified as potentially eligible and checked in full text. Ultimately, we included 12 articles involving 57 study comparisons (Figure 1). One study was reported in 2 different articles (35,36). We included both articles and counted them as 1 study with 2 study comparisons, as suggested in the Cochrane guidelines. A subset of 31 study comparisons involving the molecular biomarkers of inflammation (CRP, IL-6, and TNF), cytokine receptors (sIL-6r, TNFR1, and TNFR2), type II collagen degradation (C2C and CTX-II), and glycoproteins (COMP) were included in the meta-analyses (Figure 2).

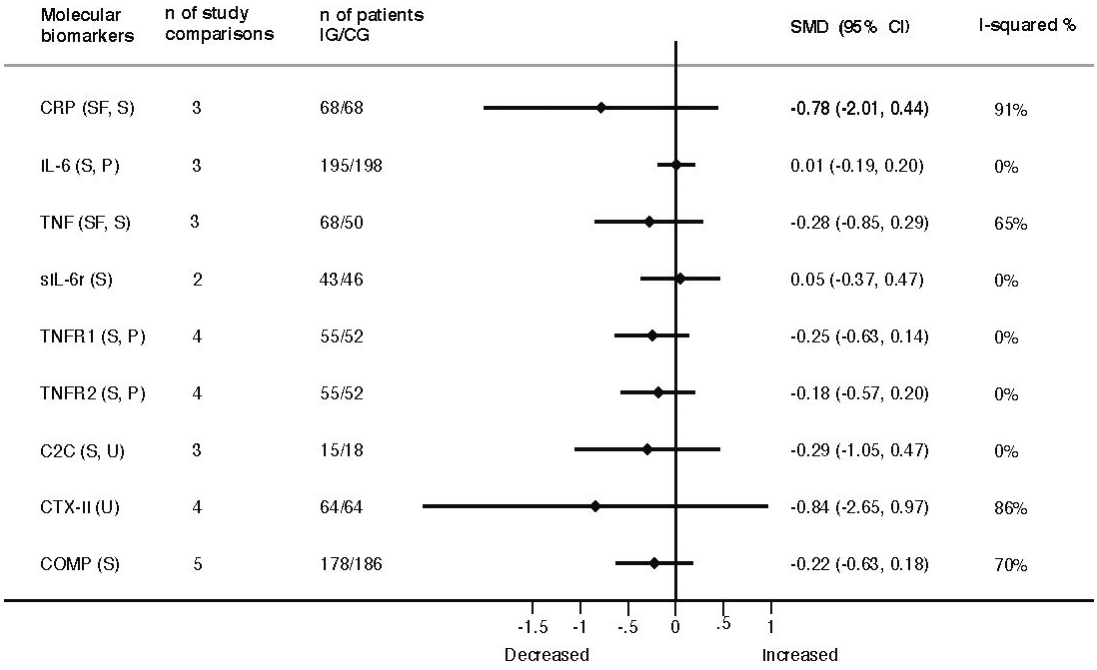

Figure 2. Cumulative forest plot for the effect of exercise therapy on molecular biomarkers. IG = intervention group; CG = control group; SMD = standardized mean difference; 95% CI = 95% confidence interval; I-squared = statistical heterogeneity; CRP = C-reactive protein; SF = synovial fluid; S = serum; IL-6 = interleukin-6; P = plasma; TNF = tumor necrosis factor; sIL-6r = soluble interleukin-6 receptor; TNFR1 = soluble TNF receptor 1; TNFR2 = soluble TNF receptor 2; C2C = C2C neoepitope of type II collagen; U = urine; CTX-II = C-terminal crosslinking of type II collagen; COMP = cartilage oligomeric matrix protein.

Participants. In the 12 articles, a total of 1,114 participants were included, of which 70% were women. Participant mean ± SD age was 65 ± 5.7 years, with a mean ± SD BMI of 29.7 ± 3.2 kg/m^2. One study showed only the age range, which was from 41 to 63 years (37) and 1 study included only those with a BMI <35 (38). One study included participants at risk of OA (i.e., no radiographic signs of OA, sedentary behavior, age >60 years, and a BMI >27) (39), and the remaining 11 studies included participants with or without pain but with radiographic knee OA, ranging from Kellgren/Lawrence (K/L) grades 1 to 4 (36–38,40–47) (Table 1).

Types of exercise therapy interventions and molecular biomarker outcomes. The types of exercise therapy interventions used were strengthening exercise in 5 studies (37,38,42–44), aerobic exercise in 3 studies (42,46,47), and a combination of strengthening and aerobic exercise in 5 studies (36,39–41,45) (Table 2). Biomarker samples were obtained at 4 to 24 weeks following the exercise therapy intervention in all the studies. Additionally, 3 studies reported in 4 articles included 1 additional follow-up assessment at 18 months (35,36,39,45); see Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23786/abstract for a narrative synthesis of these results. However, we included the first available follow-up time point in our analyses to allow for a more homogeneous time to follow-up, ranging from 1 to 6 months across the included studies.

Of the 12 biomarker studies, 5 investigated markers of inflammation (36,37,39,42,45), 2 investigated cytokine receptors (39,44), 1 investigated proteases (37), 3 investigated turnover of collagens (38,41,46), 4 investigated glycoproteins (38,40,43,45), and 5 investigated glycosaminoglycans (38,41,45–47) (Table 3). Some studies investigated >1 molecular biomarker. Detailed characteristics of molecular biomarkers are reported in Supplementary Tables 2 and 3, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23786/abstract.

Overall narrative synthesis of results. Twelve studies included 57 study comparisons, of which 63% (36 study comparisons) did not differ in molecular biomarker concentrations between the intervention and control groups. In total, 30% (17 study comparisons) showed a decrease and 7% (4 study comparisons) showed an increase in molecular biomarker concentration, all in favor of the exercise therapy intervention group. Results from individual studies are shown in Table 3.

Meta-analyses of a subset of molecular biomarkers. Meta-analyses showed statistically nonsignificant reductions of the molecular biomarkers CRP (SMD –0.78 [95% CI –2.01, 0.44]), CTX-II (SMD –0.84 [95% CI –2.65, 0.97]), TNF (SMD –0.28 [95% CI –0.85, 0.29]), TNFR1 (SMD –0.25 [95% CI –0.63, 0.14]), TNFR2 (SMD –0.18 [95% CI –0.57, 0.20]), C2C (SMD –0.29 [95% CI –1.05, 0.47]), and COMP (SMD –0.22 [95% CI –0.63, 0.18]), all in favor of exercise therapy. Analyses showed no effect for IL-6 (SMD 0.01 [95% CI –0.19, 0.20]) and sIL-6r (SMD 0.05 [95% CI –0.37, 0.47]) (Figure 2).

Subgroup analysis on molecular biomarker collection site. For molecular biomarkers in synovial fluid, blood (serum and plasma), and urine, exercise therapy was associated with a change in biomarker concentrations in 50% (n = 4), 36% (n = 16), and 20% (n = 1) of the study comparisons, respectively (calculated from Table 3). Further, 97% (n = 28) of the studies on molecular biomarkers related to inflammation, and 89% (n = 25) of the studies on molecular biomarkers of cartilage extracellular matrix turnover, were either unchanged or decreased after exercise (calculated from Table 3).

Table 1. Study participant characteristics*

Author, year (ref.)	Location	Inclusion criteria†	K/L grade	Women, %	Age, years	BMI kg/m^2
Andersson et al, 2006 (40)	Sweden	Pain and radiographic OA	3/4	51	56 ± 6	29.5 ± 4.8
Bautch et al, 1997 (46)	US	ACR criteria for OA	2/3/4	72	69 ± 2	28.8 ± 2.2
Bautch et al, 2000 (47)	US	ACR criteria for OA	2/3/4	67	69.7 ± 1.9	28.6 ± 1
Chua et al, 2008 (45)	US	Pain, radiographic OA, and BMI >30	2/3	66	68.7 ± 0.8	33.5 ± 0.7
Hunt et al, 2013 (38)	Canada	ACR criteria for OA	NA	52	66.1 ± 11.3	<35
Messier et al, 2013 (36); Loeser et al, 2017 (35)‡	US	Pain, radiographic OA, BMI from 27 to 41, and sedentary	2/3	72	66 ± 6	33.6 ± 3.7
Nagaoka et al, 2010 (41)	Japan	Radiographic OA	1/2/3/4	81	62.8 ± 10.8	23.3 ± 3.2
Nicklas et al, 2004 (39)	US	At risk of OA: BMI >27 and sedentary	NA	71.8	68.5 ± 5	34.3 ± 5.3
Samut et al, 2015 (42)	Turkey	Radiographic OA and sedentary	1/2/3	90	60.3 ± 6	31.6 ± 5.4
Simao et al, 2012 (44)	Brazil	ACR criteria for OA	2/3/4	87.5	71.6 ± 4.5	27.9 ± 49
Wang et al, 2016 (43)	China	ACR criteria and BMI <30	2/3	71.8	61.3 ± 9.3	26.4 ± 1.3
Zhang et al, 2013 (37)	China	ACR criteria for OA	1/2/3	62.0	Range 41–63	NA

* Values are the mean ± SD unless indicated otherwise. K/L= Kellgren/Lawrence; BMI = body mass index; OA = osteoarthritis; NA = not applicable or not assessed.
† When American College of Rheumatology (ACR) criteria for OA are used, they are from 1986 (51).
‡ Both articles reported on the same study.

Table 2. Exercise therapy intervention characteristics in the different studies*

Author, year (ref.)	Study comparison	IG/CG, no.	Exercise type	Time/week, minutes, total weeks	Intensity	Exercise sessions attended/ total, no.	IG dropout/ IG total, no.	IG adverse events
Andersson et al, 2006 (40)	Supervised and home exercise vs. nonexercising control group	29/29	Weight-bearing exercises (strengthening and neuromuscular)	2, 60, 6	60% maximum heart rate	11/12	3/29	n = 1 increased knee symptom
Bautch et al, 1997 (46)	Supervised exercise vs. nonexercising group (healthy advice)	15/15	Aerobic and flexibility	3, 60, 12	50% Vo_2 max	NA	9/15	NA
Bautch al, 2000 (47)	Supervised exercise vs. nonexercising group (healthy advice)	11/10	Aerobic and flexibility	3, 60, 12	50% Vo_2 max	NA	NA	NA
Chua et al, 2008 (45): study 1	Supervised exercise vs. nonexercising group (healthy advice)	45/52	Aerobic and strengthening exercise	3, 60, 24	50–75 HRR	NA	NA	NA
Chua et al, 2008 (45): study 2	Supervised exercise and diet (weight loss) vs. nonexercising control group (diet weight loss)	45/52	Aerobic and strengthening exercise	3, 60, 24	50–75 HRR	NA	NA	NA
Hunt et al, 2013 (38)	Supervised and home exercise vs. nonexercising control group	9/8	Strengthening exercise of lower extremity	4, NA, 10	Additional resistance with ankle cuff weights	36/40	0/9	None
Messier et al, 2013 (36); Loeser et al, 2017 (35)†	Supervised exercise and diet (weight loss) vs. nonexercising control group (diet weight loss)	152/152	Aerobic and strengthening exercise	3, 60, 24	NA	50/72	14/152	n = 3; 1 muscle strain and 2 trips/falls
Nagaoka et al, 2010 (41): study 1	Exercise vs. nonexercising control group	11/11	Aerobic, strengthening, and pool exercise	NA, NA, NA	NA	NA	NA	NA
Nagaoka et al, 2010 (41): study 2	Exercise and chicken comb extract vs. nonexercising control group and chicken comb extract	11/10	Aerobic, strengthening, and pool exercise	NA, NA, NA	NA	NA	NA	NA
Nicklas et al, 2004 (39): study 1	Exercise vs. nonexercising control group	67/70	Aerobic, strengthening, and pool exercise	3, 60, 16	Aerobic: 50–75% HRR	29/48 (exercise)	9/67	NA
Nicklas et al, 2004 (39): study 2	Exercise and diet (weight loss) vs. nonexercising control group (diet weight loss)	64/71	Aerobic, strengthening, and pool exercise	3, 60, 16	Strength: additional resistance with cuff weights and weighted vests	31/48 (exercise + diet)	6/64	NA
Samut et al, 2015 (42): study 1	Exercise vs. nonexercising control group	15/12	Aerobic or strengthening exercise	3, NA, 6	Aerobic: 70–75% HRR	NA	NA	NA
Samut et al, 2015 (42): study 2	Exercise vs. nonexercising control group	14/12	Aerobic or strengthening exercise	3, NA, 6	Strength: 5 concentric flexion and extension at angular velocities of 60, 90, 120, and 180 degrees/ second	NA	NA	NA

(Continued)

Table 2. *(Cont'd)*

Author, year (ref.)	Study comparison	IG/CG, no.	Exercise type	Time/week, minutes, total weeks	Intensity	Exercise sessions attended/ total, no.	IG dropout/ IG total, no.	IG adverse events
Simao et al, 2012 (44): study 1	Exercise vs. nonexercising control group	10/12	Squat training exercise	3, NA, 12	From 10° to 60° of knee flexion	35/36	1/10	NA
Simao et al, 2012 (44): study 2	Exercise vs. nonexercising control group	10/12	Vibratory platform training exercise	3, NA, 12	35–40 Hz, amplitude 4 mm, acceleration range 2.78–3.26 g-force	35/36	1/10	NA
Wang et al, 2016 (43)	Supervised exercise and vibratory platform vs. exercise	49/50	Quadriceps exercise and vibratory platform vs. quadriceps exercise	5, 30, 24	Platform (35 Hz; amplitude of 4-mm to 6-mm displacement)	NA	0/49	n = 1 increased knee pain
Zhang et al, 2013 (37)	Exercise and diclofenac sodium vs. nonexercising control group and diclofenac sodium	50/50	Flexibility and isometric quadriceps exercise; diclofenac (75 mg twice daily)	4, NA, 4	10-second isometric contraction of the quadriceps at 0° and 90° knee joint angle and 10-second rest; repeated 10 times for 1 exercise circle	NA	0/50	NA

* IG = intervention group; CG = control group; Vo_2 max = maximal oxygen uptake; NA = not assessed; HRR = heart rate reserve.
† Both articles reported on the same study.

Table 3. Impact of exercise therapy on molecular biomarkers related to inflammation and cartilage extracellular matrix turnover*

	Related to inflammation								Related to cartilage extracellular matrix turnover											
Author, year (ref.)	CRPM	CRP	IL-6	TNF	TGFβ1	sIL-6r	TNFR1	TNFR2	Protease: MMP-3	Type II synthesis: CPII†	Type II degradation: C2C†	Type II degradation: C2M†	Type II degradation: CTX-II†	Total collagen: HP†	COMP‡	Total GAG: DMMB§	CS: 3B3§	CS: 7D4§	KS: 5D4§	HA§
Andersson et al, 2006 (40)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	=S	–	–	–	–	–
Bautch et al, 2000 (47)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	↓SF	= SF	= SF	–	–
Bautch et al, 1997 (46)	–	–	–	–	–	–	–	–	–	–	–	–	–	= SF	–	–	–	–	= SF	
Chua et al, 2008 (45): study 1	–	–	–	–	↓S	–	–	–	–	–	–	–	–	–	↑S	–	–	–	↑S	=S
Chua et al, 2008 (45): study 2	–	–	–	–	↑S	–	–	–	–	–	–	–	–	–	↓S	–	–	–	=S	=S
Hunt et al, 2013 (38)	–	–	–	–	–	–	–	–	–	↓S	=U	–	↓U	–	↓S	–	–	–	–	↓S
Messier et al, 2013 (36); Loeser et al, 2017 (35): study 1¶	–	–	=S	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Messier et al, 2013 (36); Loeser et al, 2017 (35): study 2¶	=S	–	–	–	–	–	–	–	–	–	–	=S	–	–	–	–	–	–	–	–
Nagaoka et al, 2010 (41): study 1	–	–	–	–	–	–	–	–	–	↓S	↑S	–	=U	–	–	–	–	–	–	=S
Nagaoka et al, 2010 (41): study 2	–	–	–	–	–	–	–	–	–	=S	↓S	–	=U	–	–	–	–	–	–	=S
Nicklas et al, 2004 (39): study 1	–	↓S	=S	=S	–	=S	↓S	=S	–	–	–	–	–	–	–	–	–	–	–	–
Nicklas et al, 2004 (39): study 2	–	=S	=S	=S	–	=S	=S	=S	–	–	–	–	–	–	–	–	–	–	–	–

(Continued)

Table 3. *(Cont'd)*

	Related to inflammation								Related to cartilage extracellular matrix turnover											
Author, year (ref.)	CRPM	CRP	IL-6	TNF	TGFβ1	sIL-6r	TNFR1	TNFR2	Protease: MMP-3	Type II synthesis: CPII†	Type II degradation: C2C†	Type II degradation: C2M†	Type II degradation: CTX-II†	Total collagen: HP†	COMP‡	Total GAG: DMMB§	CS: 3B3§	CS: 7D4§	KS: 5D4§	HA§
Samut et al, 2015 (42): study 1	–	=S	=S	=S	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Samut et al, 2015 (42): study 2	–	=S	=S	=S	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Simao et al, 2012 (44): study 1	–	–	–	–	–	–	↓P	↓P	–	–	–	–	–	–	–	–	–	–	–	–
Simao et al, 2012 (44): study 2	–	–	–	–	–	–	↓P	↓P	–	–	–	–	–	–	–	–	–	–	–	–
Zhang et al, 2013 (37)	–	↓SF	–	↓SF	–	–	–	–	↓SF	–	–	–	–	–	–	–	–	–	–	–
Wang et al, 2016 (43)	–	–	–	–	–	–	–	–	–	–	–	–	↓U	–	↓S	–	–	–	–	–

* Table shows a statistically significant ($P < 0.05$) decrease (↓), increase (↑), or no difference (=) in the concentration of molecular biomarkers in the intervention compared to the control group. From meta-analyses and individual studies the results are reported as a decrease if the standardized mean difference (SMD) is less than –0.2 r, as no difference if the SMD is –0.2 to 0.2, and as an increase if the SMD is >0.2. Biomarkers in the different articles were analyzed in synovial fluid (SF), serum (S), plasma (P), and urine (U). CRPM = C-reactive protein degradation; CRP = C-reactive protein; IL-6 = interleukin-6; TNF = tumor necrosis factor; TGFβ1 = transforming growth factor β1; CR = cytokine receptor; sIL-6r = soluble interleukin-6 receptor; TNFR1 = soluble TNF receptor 1; TNFR 2 = soluble TNF receptor 2; MMP-3 = matrix metalloprotease 3; CPII = type II collagen carboxy propeptide; C2C = C2C neoepitope of type II collagen; C2M = neoepitope of type II collagen; CTX-II = C-terminal crosslinking of type II collagen; HP = hydroxyproline; COMP = cartilage oligomeric matrix protein; GAG = glycosaminoglycan; DMMB = 1,9-dimethyl-methylene blue assay; CS = chondroitin sulfate; 3B3 = CS epitope; 7D4 = CS epitope; KS = keratan sulfate; 5D4 = KS epitope; HA = hyaluronic acid.
† Turnover of collagens.
‡ Glycoprotein.
§ Glycosaminoglycans.
¶ Both articles reported on the same study.

Sensitivity analysis for the effect of exercise on molecular biomarkers. Data were available to calculate an effect size for 51 of 57 individual studies. Overall, the effect sizes for 44 of 51 study comparisons supported our main analysis. The remaining 7 study comparisons from 3 studies (38,46,47) changed from being classified as no effect to being decreased for serum CPII, urine CTX-II, serum HP, synovial fluid 3B3, synovial fluid 7D4, synovial fluid 5D4, and serum HA (see Supplementary Table 4, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23786/abstract).

Quality of evidence. The majority of the studies applied proper randomization and allocation, although some studies failed to clearly describe or adequately address dropouts of participants in the analyses (attrition bias) and failed to describe whether outcome assessors (the individuals responsible for analyzing the samples) were blinded to the outcomes of interest (detection bias) (see Supplementary Figure 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23786/abstract). However, the high heterogeneity reported in some of the meta-analyses and the too few studies investigating the same outcomes made us downgrade the quality of evidence for inconsistency (substantial heterogeneity) and imprecision (large 95% CIs of the estimates).

DISCUSSION

We summarized the impact of exercise therapy on cartilage biomarkers in individuals at risk of, or with established, knee OA participating in randomized controlled trials. Our results suggest that exercise therapy is not harmful because it does not increase the concentration of molecular biomarkers related to inflammation and cartilage turnover, associated with cartilage breakdown. All in all, this finding was consistent in both the main and sensitivity analyses, because the majority of studies point to either a decrease or an unchanged level. However, due to substantial heterogeneity and large CIs in the meta-analysis estimates, the overall quality of evidence was downgraded to low.

Systematic reviews and meta-analyses have shown that exercise is a safe treatment associated with few and only minor adverse events, such as temporary flares in pain, in individuals with knee OA (12–16). Additionally, in a previous systematic review of exercise trials, we have shown that exercise is not harmful to cartilage when evaluated by magnetic resonance imaging (48). The findings of the current study, evaluating the effect from exercise therapy on molecular biomarkers related to inflammation and cartilage extracellular matrix turnover, are in line with these previous findings and support exercise therapy as a safe treatment for knee joint cartilage in individuals at risk of, or with established, knee OA.

Molecular biomarkers obtained from the synovial fluid may be more sensitive for detecting changes from exercise therapy, due to the proximity of the synovial fluid to the joint tissues (49). In agreement, we found higher rates of biomarker concentration change in synovial fluid (50%) compared to that of blood (36%) or urine (20%). Therefore, the origin of the fluid is important information when interpreting results from therapeutic studies.

As our meta-analyses indicate, a reduction across the molecular biomarkers associated with inflammation (i.e., CRP and TNF) and cartilage breakdown or turnover (i.e., C2C, CTX-II, and COMP) favors exercise therapy, and we can say that exercise therapy is, if anything, beneficial for cartilage assessed via molecular biomarkers; however, this hypothesis needs further investigation.

This study has limitations. We could not perform meta-analyses of all the molecular biomarkers investigated due to the low number of studies reporting the same markers. Neither could we perform additional analysis, which is considered an important step in exploring relationships in evidence synthesis. Also, due to large heterogeneity and large CIs, the overall quality of evidence was downgraded to low. To properly interpret this evaluation, one should note that the included studies followed the available guidelines in conducting and reporting the studies, and therefore, the low quality of evidence, rather than being related to methodologic flaws of the studies, was caused by the limited number of randomized controlled trials in the literature.

These results highlight the need for more high-quality randomized controlled trials to further investigate the impact of knee-joint loading exercise on cartilage and inflammation related to molecular biomarkers. Such studies should preferably include individuals at risk of, or at early stages of, OA, when the anabolic and catabolic reactions in the cartilage extracellular matrix are better balanced and a therapeutic exercise intervention theoretically may have the ability to prevent or slow down the catabolic activities driving OA progression.

As no single biomarker has been shown to explain OA development and progression, defined by osteophyte formation or joint space narrowing, we recommend that future studies focus on a set of biomarkers, rather than single biomarkers, using established commercial biomarker assays such as those used in the OA Initiative (50). The clinical implication of our findings is that individuals at risk of, or with established, knee OA can be told that exercise therapy is not harmful, and if anything, is positive for the turnover of articular cartilage and inflammation.

The therapeutic exercise commonly prescribed to prevent and treat symptomatic knee OA appears safe for knee-joint cartilage, because it does not increase the molecular biomarkers related to inflammation and cartilage turnover associated with OA. However, due to the limited number of randomized studies, the overall quality of the evidence supporting this conclusion was downgraded to low.

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. Bricca 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. Bricca, Struglics, Larsson, Steultjens, Juhl, Roos.

Acquisition of data. Bricca, Struglics.

Analysis and interpretation of data. Bricca, Struglics, Larsson, Juhl, Roos.

REFERENCES

1. Wallace IJ, Worthington S, Felson DT, Jurmain RD, Wren KT, Maijanen H, et al. Knee osteoarthritis has doubled in prevalence since the mid-20th century. Proc Natl Acad Sci U S A 2017;114:9332–6.
2. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380:2163–96.
3. Griffin TM, Guilak F. The role of mechanical loading in the onset and progression of osteoarthritis. Exerc Sport Sci Rev 2005;33:195–200.
4. Nia HT, Bozchalooi IS, Li Y, Han L, Hung HH, Frank E, et al. High-bandwidth AFM-based rheology reveals that cartilage is most sensitive to high loading rates at early stages of impairment. Biophys J 2013;104:1529–37.
5. Mündermann A, Klenk C, Billich C, Nuesch C, Pagenstert G, Schmidt-Trucksass A, et al. Changes in cartilage biomarker levels during a transcontinental multistage footrace over 4486 km. Am J Sports Med 2017;45:2630–6.
6. Cattano NM, Driban JB, Cameron KL, Sitler MR. Impact of physical activity and mechanical loading on biomarkers typically used in osteoarthritis assessment: current concepts and knowledge gaps. Ther Adv Musculoskelet Dis 2017;9:11–21.
7. Fernandes GS, Parekh SM, Moses J, Fuller C, Scammell B, Batt ME, et al. Prevalence of knee pain, radiographic osteoarthritis and arthroplasty in retired professional footballers compared with men in the general population: a cross-sectional study. Br J Sports Med 2018;52:678–83.
8. Munukka M, Waller B, Rantalainen T, Hakkinen A, Nieminen MT, Lammentausta E, et al. Association between leisure time physical activity level and articular cartilage in postmenopausal women with mild knee osteoarthritis: a 12-month follow-up study after 4-month intervention. Osteoarthritis Cartilage 2016;24:1708–17.
9. Leong DJ, Gu XI, Li Y, Lee JY, Laudier DM, Majeska RJ, et al. Matrix metalloproteinase-3 in articular cartilage is upregulated by joint immobilization and suppressed by passive joint motion. Matrix Biol 2010;29:420–6.
10. McAlindon TE, Bannuru RR, Sullivan MC, Arden NK, Berenbaum F, Bierma-Zeinstra SM, et al. OARSI guidelines for the non-surgical management of knee osteoarthritis. Osteoarthritis Cartilage 2014;22:363–88.
11. Fernandes L, Hagen KB, Bijlsma JW, Andreassen O, Christensen P, Conaghan PG, et al. EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis. Ann Rheum Dis 2013;72:1125–35.
12. Quicke JG, Foster NE, Thomas MJ, Holden MA. Is long-term physical activity safe for older adults with knee pain? A systematic review. Osteoarthritis Cartilage 2015;23:1445–56.
13. Sandal LF, Roos EM, Bogesvang SJ, Thorlund JB. Pain trajectory and exercise-induced pain flares during 8 weeks of neuromuscular exercise in individuals with knee and hip pain. Osteoarthritis Cartilage 2016;24:589–92.
14. Ageberg E, Link A, Roos EM. Feasibility of neuromuscular training in patients with severe hip or knee OA: the individualized goal-based NEMEX-TJR training program. BMC Musculoskelet Disord 2010;11:126.
15. Fransen M, McConnell S, Harmer AR, Van der Esch M, Simic M, Bennell KL. Exercise for osteoarthritis of the knee. Cochrane Database Syst Rev 2015;1:CD004376.
16. Juhl C, Christensen R, Roos EM, Zhang W, Lund H. Impact of exercise type and dose on pain and disability in knee osteoarthritis: a systematic review and meta-regression analysis of randomized controlled trials. Arthritis Rheumatol 2014;66:622–36.
17. Hendry M, Williams NH, Markland D, Wilkinson C, Maddison P. Why should we exercise when our knees hurt? A qualitative study of primary care patients with osteoarthritis of the knee. Fam Pract 2006;23:558–67.
18. Bricca A, Juhl CB, Grodzinsky AJ, Roos EM. Impact of a daily exercise dose on knee joint cartilage: a systematic review and meta-analysis of randomized controlled trials in healthy animals. Osteoarthritis Cartilage 2017;25:1223–37.
19. Lotz M, Martel-Pelletier J, Christiansen C, Brandi ML, Bruyere O, Chapurlat R, et al. Republished: value of biomarkers in osteoarthritis. Current status and perspectives. Postgrad Med J 2014;90:171–8.
20. Van Spil WE, DeGroot J, Lems WF, Oostveen JC, Lafeber FP. Serum and urinary biochemical markers for knee and hip-osteoarthritis: a systematic review applying the consensus BIPED criteria. Osteoarthritis Cartilage 2010;18:605–12.
21. Valdes AM, Meulenbelt I, Chassaing E, Arden NK, Bierma-Zeinstra S, Harat D, et al. Large scale meta-analysis of urinary C-terminal telopeptide, serum cartilage oligomeric protein and matrix metalloprotease degraded type II collagen and their role in prevalence, incidence and progression of osteoarthritis. Osteoarthritis Cartilage 2014;22:683–9.
22. Hoch JM, Mattacola CG, McKeon JM, Howard JS, Lattermann C. Serum cartilage oligomeric matrix protein (sCOMP) is elevated in patients with knee osteoarthritis: a systematic review and meta-analysis. Osteoarthritis Cartilage 2011;19:1396–404.
23. Garcia-Hermoso A, Sanchez-Lopez M, Escalante Y, Saavedra JM, Martinez-Vizcaino V. Exercise-based interventions and C-reactive protein in overweight and obese youths: a meta-analysis of randomized controlled trials. Pediatr Res 2016;79:522–7.
24. Swardfager W, Herrmann N, Cornish S, Mazereeuw G, Marzolini S, Sham L, et al. Exercise intervention and inflammatory markers in coronary artery disease: a meta-analysis. Am Heart J 2012;163:666–76.
25. Fedewa MV, Hathaway ED, Ward-Ritacco CL. Effect of exercise training on C reactive protein: a systematic review and meta-analysis of randomised and non-randomised controlled trials. Br J Sports Med 2017;51:670–6.
26. Helmark IC, Mikkelsen UR, Borglum J, Rothe A, Petersen MC, Andersen O, et al. Exercise increases interleukin-10 levels both intraarticularly and peri-synovially in patients with knee osteoarthritis: a randomized controlled trial. Arthritis Res Ther 2010;12:R126.
27. Neidhart M, Muller-Ladner U, Frey W, Bosserhoff AK, Colombani PC, Frey-Rindova P, et al. Increased serum levels of non-collagenous matrix proteins (cartilage oligomeric matrix protein and melanoma inhibitory activity) in marathon runners. Osteoarthritis Cartilage 2000;8:222–9.
28. Erhart-Hledik JC, Favre J, Asay JL, Smith RL, Giori NJ, Mundermann A, et al. A relationship between mechanically-induced changes in serum cartilage oligomeric matrix protein (COMP) and changes in cartilage thickness after 5 years. Osteoarthritis Cartilage 2012;20:1309–15.
29. Higgins JP, Green S, editors. Cochrane handbook for systematic reviews of interventions, version 5.1.0. The Cochrane Collaboration; 2011. URL: www.handbook.cochrane.org.
30. De Vries RB, Hooijmans CR, Tillema A, Leenaars M, Ritskes-Hoitinga M. Updated version of the Embase search filter for animal studies. Lab Anim 2014;48:88.
31. Hooijmans CR, Tillema A, Leenaars M, Ritskes-Hoitinga M. Enhancing search efficiency by means of a search filter for find-

ing all studies on animal experimentation in PubMed. Lab Anim 2010;44:170–5.

32. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21:1539–58.
33. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557–60.
34. Schünemann H, Brożek J, Guyatt G, Oxman A, editors. GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013. The GRADE Working Group; 2013. URL: guidelinedevelopment.org/handbook.
35. Loeser RF, Beavers DP, Bay-Jensen AC, Karsdal MA, Nicklas BJ, Guermazi A, et al. Effects of dietary weight loss with and without exercise on interstitial matrix turnover and tissue inflammation biomarkers in adults with knee osteoarthritis: the Intensive Diet and Exercise for Arthritis trial (IDEA). Osteoarthritis Cartilage 2017;25:1822–8.
36. Messier SP, Mihalko SL, Legault C, Miller GD, Nicklas BJ, DeVita P, et al. Effects of intensive diet and exercise on knee joint loads, inflammation, and clinical outcomes among overweight and obese adults with knee osteoarthritis: the IDEA randomized clinical trial. JAMA 2013;310:1263–73.
37. Zhang SL, Lu HQ, Xu XZ, Zhi J, Geng JJ, Chen J. Effects of exercise therapy on knee joint function and synovial fluid cytokine levels in patients with knee osteoarthritis. Mol Med Rep 2013;7:183–6.
38. Hunt MA, Pollock CL, Kraus VB, Saxne T, Peters S, Huebner JL, et al. Relationships amongst osteoarthritis biomarkers, dynamic knee joint load, and exercise: results from a randomized controlled pilot study. BMC Musculoskelet Disord 2013;14:115.
39. Nicklas BJ, Ambrosius W, Messier SP, Miller GD, Penninx BW, Loeser RF, et al. Diet-induced weight loss, exercise, and chronic inflammation in older, obese adults: a randomized controlled clinical trial. Am J Clin Nutr 2004;79:544–51.
40. Andersson ML, Thorstensson CA, Roos EM, Petersson IF, Heinegard D, Saxne T. Serum levels of cartilage oligomeric matrix protein (COMP) increase temporarily after physical exercise in patients with knee osteoarthritis. BMC Musculoskelet Disord 2006;7:98.
41. Nagaoka I, Nabeshima K, Murakami S, Yamamoto T, Watanabe K, Tomonaga A, et al. Evaluation of the effects of a supplementary diet containing chicken comb extract on symptoms and cartilage metabolism in patients with knee osteoarthritis. Exp Ther Med 2010;1:817–27.
42. Samut G, Dincer F, Ozdemir O. The effect of isokinetic and aerobic exercises on serum interleukin-6 and tumor necrosis factor alpha levels, pain, and functional activity in patients with knee osteoarthritis. Mod Rheumatol 2015;25:919–24.
43. Wang P, Yang L, Liu C, Wei X, Yang X, Zhou Y, et al. Effects of whole body vibration exercise associated with quadriceps resistance exercise on functioning and quality of life in patients with knee osteoarthritis: a randomized controlled trial. Clin Rehabil 2016;30:1074–87.
44. Simao AP, Avelar NC, Tossige-Gomes R, Neves CD, Mendonca VA, Miranda AS, et al. Functional performance and inflammatory cytokines after squat exercises and whole-body vibration in elderly individuals with knee osteoarthritis. Arch Phys Med Rehabil 2012;93:1692–700.
45. Chua SD, Messier SP, Legault C, Lenz ME, Thonar EJ, Loeser RF. Effect of an exercise and dietary intervention on serum biomarkers in overweight and obese adults with osteoarthritis of the knee. Osteoarthritis Cartilage 2008;16:1047–53.
46. Bautch JC, Malone DG, Vailas AC. Effects of exercise on knee joints with osteoarthritis: a pilot study of biologic markers. Arthritis Care Res (Hoboken) 1997;10:48–55.
47. Bautch JC, Clayton MK, Chu QL, Johnson KA. Synovial fluid chondroitin sulphate epitopes 3B3 and 7D4, and glycosaminoglycan in human knee osteoarthritis after exercise. Ann Rheum Dis 2000;59:887–91.
48. Bricca A, Juhl CB, Steultjens M, Wirth W, Roos EM. Impact of exercise on articular cartilage in people at risk of, or with established, knee osteoarthritis: a systematic review of randomized controlled trials. Br J Sports Med 2019;53:940–7.
49. Larsson S, Struglics A, Lohmander LS, Frobell R. Surgical reconstruction of ruptured anterior cruciate ligament prolongs trauma-induced increase of inflammatory cytokines in synovial fluid: an exploratory analysis in the KANON trial. Osteoarthritis Cartilage 2017;25:1443–51.
50. Hunter DJ, Nevitt M, Losina E, Kraus V. Biomarkers for osteoarthritis: current position and steps towards further validation. Best Pract Res Clin Rheumatol 2014;28:61–71.
51. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. Arthritis Rheum 1986;29:1039–49.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 71, No. 11, November 2019, pp 1516–1524
DOI 10.1002/acr.23798

Short- or Long-Term Treatment of Spinal Disability in Older Adults With Manipulation and Exercise

Michele Maiers,[1] Jan Hartvigsen,[2] Roni Evans,[3] Kristine Westrom,[1] Qi Wang,[3] Craig Schulz,[3] Brent Leininger,[3] and Gert Bronfort[3]

Objective. Back and neck pain are associated with disability and loss of independence in older adults. Whether long-term management using commonly recommended treatments is superior to shorter-term treatment is unknown. This randomized clinical trial compared short-term treatment (12 weeks) versus long-term management (36 weeks) of back- and neck-related disability in older adults using spinal manipulative therapy (SMT) combined with supervised rehabilitative exercises (SRE).

Methods. Eligible participants were ages ≥65 years with back and neck disability for ≥12 weeks. Coprimary outcomes were changes in Oswestry Disability Index (ODI) and Neck Disability Index (NDI) scores after 36 weeks. An intent-to-treat approach used linear mixed-model analysis to detect between-group differences. Secondary analyses included other self-reported outcomes, adverse events, and objective functional measures.

Results. A total of 182 participants were randomized. The short-term and long-term groups demonstrated significant improvements in back disability (ODI score –3.9 [95% confidence interval (95% CI) –5.8, –2.0] versus ODI score –6.3 [95% CI –8.2, –4.4]) and neck disability (NDI score –7.3 [95% CI –9.1, –5.5] versus NDI score –9.0 [95% CI –10.8, –7.2]) after 36 weeks, with no difference between groups (back ODI score 2.4 [95% CI –0.3, 5.1]; neck NDI score 1.7 [95% CI 0.8, 4.2]). The long-term management group experienced greater improvement in neck pain at week 36, in self-efficacy at weeks 36 and 52, and in functional ability, and balance.

Conclusion. For older adults with chronic back and neck disability, extending management with SMT and SRE from 12 to 36 weeks did not result in any additional important reduction in disability.

INTRODUCTION

Back and neck pain are common symptoms in the elderly and are associated with significant disability that can negatively impact general health, functional independence, and quality of life (1–5). Back and neck pain often occur together and are the leading causes of years lived with disability globally (6,7). Spine pain ranks as the second most expensive chronic noncancer pain among Medicare recipients in the US (8). Because the population of adults age >65 years is predicted to double in size by 2050, mitigating the effects of back- and neck-related disability is an important public health priority (9).

More than half of patients presenting for back or neck pain care report continued or recurrent pain after 1 year (10,11). This fact raises the question of whether longer-term management strategies may be effective in sustaining improvement over time (12). While this question has been explored in small studies of adults with low-back pain, the evidence remains inconclusive, particularly in relation to the elderly (13,14). Spinal manipulative therapy (SMT) and supervised rehabilitative exercise (SRE) are both recommended first-line, nonpharmacologic treatments for back and neck pain in the general population (15–18). Combining both approaches may result in superior pain relief and function (19). In a previous study by our team, a combination of SMT and

ClinicalTrials.gov identifier: NCT01057706.

The content and conclusions herein are those of the authors and should not be construed as the official position or policy of, nor should any endorsements be inferred by, the US government, the Department of Health and Human Services, the Health Resources and Services Administration, the Bureau of Health Professions, or the Division of Medicine and Dentistry.

Supported by the Department of Health and Human Services (Health Resources and Services Administration grant R18HP15127), the Bureau of Health Professions, and the Division of Medicine and Dentistry. Dr. Leininger's work was supported by the NIH (National Center for Complementary and Integrative Health grant K01-AT-00896).

[1]Michele Maiers, DC, MPH, PhD, Kristine Westrom, MD, MS: Northwestern Health Sciences University, Bloomington, Minnesota; [2]Jan Hartvigsen, DC, PhD: University of Southern Denmark and Nordic Institute of Chiropractic and Clinical Biomechanics, Odense, Denmark; [3]Roni Evans, DC, PhD, Qi Wang, MS, Craig Schulz, DC, MS, Brent Leininger, DC, MS, Gert Bronfort, DC, PhD: University of Minnesota, Minneapolis.

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

Address correspondence to Michele Maiers, DC, MPH, PhD, 2501 West 84th Street, Bloomington, MN 55431. E-mail: mmaiers@nwhealth.edu.

Submitted for publication April 30, 2018; accepted in revised form October 16, 2018.

SIGNIFICANCE & INNOVATIONS

- Back and neck pain are not only common among older adults but result in significant disability and loss of independence.
- Both short-term treatment and long-term management with a combination of spinal manipulative therapy and exercise resulted in similar improvement in disability, with the greatest improvement achieved at the time of treatment completion.
- Long-term management resulted in greater improvement in neck pain and self-efficacy, as well as balance and physical performance.
- While mild transient side effects were common, no serious adverse events were reported by participants.

SRE resulted in greater improvements among older adults experiencing neck pain and disability than exercise alone (20); however, robust research in this area is still lacking, and little attention has been paid to the effectiveness of longer-term management.

The primary aim of this randomized, observer-blinded, comparative effectiveness trial was to compare the effectiveness of SMT plus SRE, delivered via either a short-term treatment (12 weeks) or long-term management (36 weeks) strategy, on older adults with spine-related disability. Neck and back disability were coprimary outcomes. Secondary outcomes included neck and back pain, general health, improvement, self-efficacy, kinesiophobia, satisfaction, falls, medication use, and biomechanic functional measures.

PATIENTS AND METHODS

Study design. This randomized clinical trial was approved by the Institutional Review Board at Northwestern Health Sciences University. An in-depth description of the study design, treatment protocols, and outcome measures has been previously published (21).

Participants were recruited from a metropolitan area in the upper midwest US. They were screened for general inclusion criteria by certified study personnel using a computer-guided questionnaire. Those who qualified after initial screening attended a series of 2 baseline evaluations comprised of informed consent, patient self-reported questionnaires, health history, physical evaluation, cervical and lumbar radiographs, and a functional assessment.

Participants were ages ≥65 years, English speaking, and community dwelling, with self-reported back and neck disability ≥12 weeks in duration. Disability was defined as scoring 10% or higher on both the Neck Disability Index (NDI) and Oswestry Disability Index (ODI), with the addition of both scores totaling ≥25. Participants also had the ability to ambulate without the aid of a wheelchair or motorized scooter and had stable prescription pain medication use in the 4 weeks prior to enrollment. Exclusion criteria included significant comorbid conditions and frank contraindications to either SMT or SRE (21).

Randomization. Eligible participants were randomly assigned to treatment, using a 1:1 computer-generated random block permutation allocation scheme under the direction of an independent statistician. The randomization scheme was concealed from study staff, who opened sequentially numbered, opaque, sealed envelopes containing treatment assignments in the presence of participants as they became eligible.

Originally a 3-arm study, the third comparison group (SRE alone for 36 weeks) was discontinued after enrolling 18 randomized participants due to slower than projected enrollment and award reductions from the funding agency. The modified, 2-arm design was approved by the steering committee, funding agency, institutional review board, and data and safety monitoring board.

Interventions. Participants received either 12 or 36 weeks of SMT plus SRE. The same treatments were delivered in both study groups, with the only exception being length of care. SMT was delivered by chiropractors with at least 5 years of experience. SRE was delivered by exercise therapists with ≥4 years of experience instructing pain patients in therapeutic exercise. The exercise therapists were trained to follow a standardized program and delivered care under the supervision of study chiropractors.

SMT focused on high-velocity, low-amplitude manipulation (22). Low-amplitude mobilization, manual distraction, gentle soft-tissue massage, heat or cold therapy, and active or passive muscle stretching were permitted to facilitate SMT. The frequency of care, spinal regions treated, and type of therapy used were left to the discretion of the treating chiropractor, based on clinical experience, patient preferences, and response to care. Visits were approximately 20–30 minutes in length. Each participant received SMT a minimum of once per month, with treatments not exceeding twice per week.

SRE consisted of an aerobic warm up, followed by a standardized program of stretching, strengthening, and balance exercises used in previous research studies and described at length elsewhere (21). The selection of exercises, progressions, and number of repetitions was individualized to accommodate participants' abilities and tolerance. All participants received standardized advice to stay active as well as self-care tips for pain management. Participants received 1-hour instructional sessions twice in the first month, then once per month through the duration of the randomly assigned intervention phase, with daily home exercises encouraged between sessions. Standardized forms documented treatment

visits, including examination findings, treatment used, adverse events, and compliance.

Data collection. Sociodemographic data were collected at baseline. Patient self-reported outcomes were collected via questionnaires at baseline and weeks 4, 12, 34, 36, 52, and 78. Functional outcome measures were recorded at baseline and week 37. Qualitative interviews were collected post-intervention (at week 12 or 36, depending on the group). Physical activity assessed with accelerometry, and participants' perspectives of treatment as ascertained during qualitative interviews, will be reported in subsequent publications.

Coprimary outcomes. Back and neck disability were measured using the ODI, version 2.0, and the NDI (23–25). The NDI was derived from the ODI, therefore having similar measurement properties and aiding in the comparison of results. Each instrument asks participants to rate 10 functional activities on a scale of 0–5 (where 0 = no disability with activity and 5 = maximal disability), with high scores indicating increasing disability.

Secondary outcomes. Neck and low-back pain in the past week, days of prescription or over-the-counter medication, improvement, satisfaction, kinesiophobia, self-efficacy, quality of life, and expectations for improvement were secondary outcomes. The incidence of falls was collected by asking participants if they have fallen and landed on the floor or ground or have fallen and hit an object like a table or chair during the previous 4 weeks (26). Functional ability was assessed by hand grip strength and the Short Physical Performance Battery (SPPB) score (27,28). A more detailed description of data collection has been given in a previous publication (21).

Adverse events. Active and passive surveillance methods were used to collect information on adverse events. Patient self-reported questionnaires asked, "Since you started treatment in the study have you experienced any of the following?" followed by a list of side effects known to be associated with SMT and exercise. Participants were asked to rate yes responses on a 0–10 bothersomeness scale, where 0 = not at all bothersome and 10 = extremely bothersome (29). Chiropractors and therapists queried patients about side effects since the last visit and documented responses on standardized treatment forms. Adverse events were categorized by investigators according to standards defined by the US Department of Health and Human Services.

Sample size. For a 2-arm design, 85 participants were needed to detect a minimally important between-group difference of 10% in the coprimary outcome ODI at week 36, with a power of 0.90 and alpha level of 0.025. Assuming a 15% loss to follow-up rate, 100 participants were sought for each treatment group.

Statistical analysis. Analyses followed a prespecified plan and were performed using SAS software, version 9.3. An intent-to-treat approach included all participants in their originally assigned group. Baseline demographic and clinical characteristics of groups were compared using 2-sample *t*-tests, Wilcoxon's rank sum test, and chi-square and Fisher's exact tests where appropriate. Baseline demographic or clinical variables considered relevant by the investigators based on the literature, or those correlated at 0.5 or greater with the primary outcomes, were included as covariates in all primary and secondary analyses. Sex and expectations have been shown to be predictive of persistent musculoskeletal-related disability in the elderly (30); these variables were subsequently included as covariates.

Primary analysis used linear mixed models (PROC MIXED in SAS) to compare between-group differences in the coprimary outcomes of ODI and NDI between baseline and week 36. All participants scored both, regardless of where their primary symptom was, and each was analyzed separately. The model adjusted for the fixed effects of treatment group, time, and treatment × time interaction and included a random intercept to account for within-patient correlations. The alpha level was reduced to 0.025 to account for testing 2 independent primary outcomes.

The proportion of participants in each group demonstrating ≥15%, ≥30%, and ≥50% improvement in ODI and NDI from baseline were compared using logistic regression (31,32). Finally, area under the curve (AUC) analyses of neck and back disability were conducted using linear regression models. Secondary outcome measures were similarly analyzed using linear mixed models. Analyses included between-group differences at all time points, as well as within-group change from baseline at each time point. Baseline outcome measures were included as covariates when available.

Descriptive statistics of baseline characteristics and outcome measures in the 36-week exercise only arm of the study (n = 18) were not part of the primary analyses and are reported separately in Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23798/abstract.

RESULTS

Of 612 patients screened, 182 were randomized (91 to each group). An additional 18 participants were randomized to a discontinued third intervention group of 36 weeks of SRE (see Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23798/abstract). Figure 1 shows participant flow through the study. Follow-up rates were high, with an overall 97% collection rate of the coprimary outcomes at week 36. Treatment groups were similar at baseline, with the exception of lower expectations for improvement and a greater proportion of women in the short-term treat-

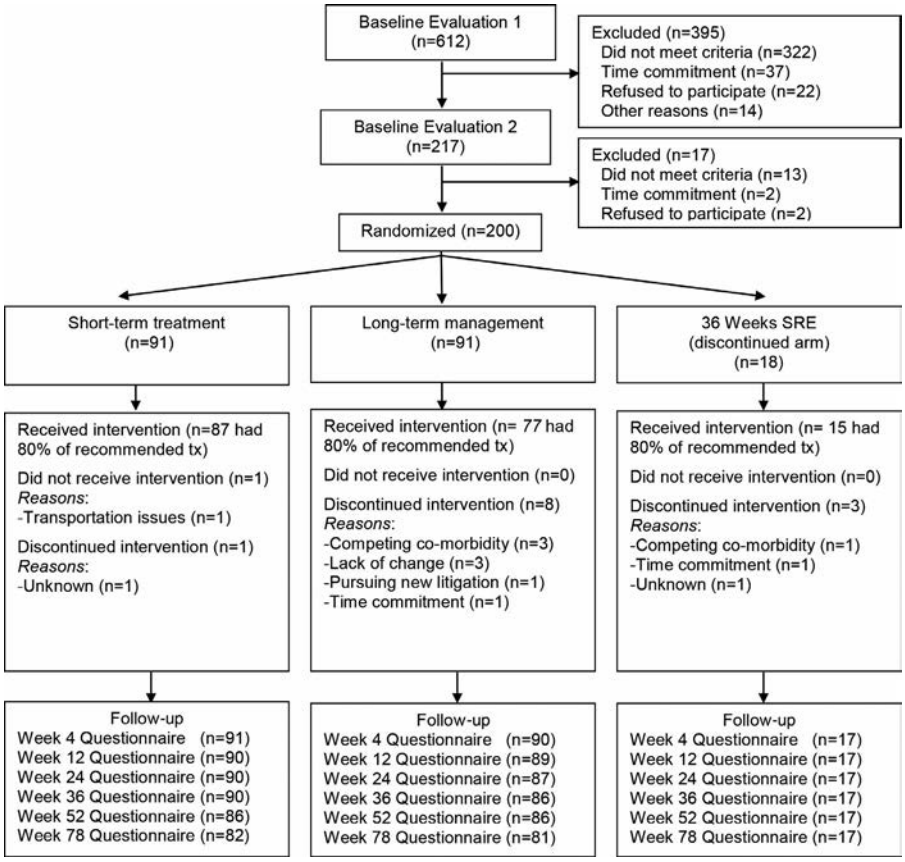

Figure 1. Study participant flow chart. tx = treatment; SRE = supervised rehabilitative exercise.

ment group (Table 1). On average, study participants reported moderate neck-related (NDI score 25.9) and back-related (ODI score 26.2) disability, and moderate neck (4.6 on a scale of 0–10) and back (5.0 on a scale of 0–10) pain.

Participants in the short-term treatment group attended an average of 10 SMT and 4 exercise instruction visits. Those in the long-term management group attended an average of 19 SMT and 9 exercise instruction visits, with an average of 11 SMT and 4 exercise visits occurring in the first 12 weeks. Participants in both groups reported performing the exercises at home an average of 4 times per week at the week 12 visit. That number decreased in both groups to an average of 3 times per week at the week 36 visit. While both groups demonstrated improvement, there was no statistically significant difference between groups in the coprimary outcomes of neck and back disability between baseline and week 36 (Table 2). This result was confirmed by the AUC analysis (ODI score 38.5 [95% CI –141.2, 218.2], $P = 0.67$; NDI score –10.0 [95% CI –180.8, 160.7], $P = 0.91$). The responder analysis demonstrated no statistically significant differences in the proportions who reached 15%, 30%, or 50% improvement in disability at either the 12- or 36-week time point (see Supplementary Figures 1 and 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23798/abstract). The only exception was the proportion reaching 50% improvement in neck disability, which favored the short-term treatment group at week 12 ($P = 0.05$). Overall, both groups actualized the greatest proportion of responders for each threshold at the end of their respective treatment periods.

With few exceptions, there were no statistically significant differences between groups in secondary patient self-reported outcomes (Table 3). Those in the long-term management group did experience greater improvement in neck pain at week 36, and greater gains in self-efficacy at weeks 36 and 52. Expectations for improvement decreased in both groups over time, although more significantly so in the short-term treatment group. There were statistically significant between-group differences in the SPPB overall score and in the SPPB balance test subscore in favor of the long-term management group (see Supplementary Table 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23798/abstract). No between-group differences were observed for the rate of self-reported falls during the follow-up period (see Supplementary Table 3, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.23798/abstract).

Nonstudy treatment. At 12 weeks, 11 participants (6 in the short-term treatment group, 5 in the long-term management group) reported visits to nonstudy health care providers for their back or neck problem in the past month. At

Table 1. SMT plus SRE participant demographics and baseline clinical characteristics*

Characteristic	Overall (n = 182)	Short-term treatment (n = 91)	Long-term management (n = 91)	*P*
Age, mean ± SD, median (range) years	71.1 ± 5.3, 69 (65–87)	71.5 ± 5.4, 70 (65–86)	70.7 ± 5.2, 69 (65–87)	0.35
Female	108 (59)	61 (67)	47 (52)	0.035†
Duration of neck pain, mean ± SD, median (range) years	12.7 ± 13.1, 10 (0.2–58)	13.6 ± 12.6, 10 (0.2–50)	11.7 ± 13.6, 7.5 (0.2–58)	0.10
Duration of back pain, mean ± SD, median (range) weeks	17.6 ± 15.8, 14.5 (0.2–60)	18.0 ± 15.0, 5 (0.3–60)	17.2 ± 16.7, 13 (0.2–60)	0.31
Ethnicity				1
Hispanic or Latino	1 (1)	0 (0)	1 (1)	–
Not Hispanic or Latino	177 (99)	88 (100)	89 (99)	–
Race				0.12
White	175 (97)	90 (99)	85 (94)	–
All others‡	6 (3)	1 (1)	5 (6)	–
Physical activity level§	3.5 (0.8)	3.5 (0.7)	3.5 (0.8)	0.85
Neck Disability Index (range 0–100)	25.9 (8.5)	25.2 (7.5)	26.6 (9.3)	0.26
Neck pain (range 0–10)	4.6 (1.7)	4.5 (1.6)	4.7 (1.8)	0.42
Back Disability Index (range 0–100)	26.2 (9.2)	25.6 (8.1)	26.7 (10.2)	0.39
Back pain (range 0–10)	5.0 (1.9)	5.1 (1.8)	5.0 (2.0)	0.70
Tobacco use (yes)	14 (8)	6 (7)	8 (9)	0.58
Alcohol use (yes)	102 (56)	47 (52)	55 (60)	0.23
Body mass index	28.8 (5.8)	28.7 (5.8)	28.9 (5.8)	0.79
Quebec Task Force classification (neck)¶				0.56
1	41 (23)	20 (22)	21 (23)	–
2	101 (55)	47 (52)	54 (59)	–
3a	17 (9)	9 (10)	8 (9)	–
3b	16 (9)	11 (12)	5 (5)	–
3c	7 (4)	4 (4)	3 (3)	–
Quebec Task Force classification (back)#				0.69
1	118 (65)	61 (67)	57 (63)	–
2	34 (19)	16 (18)	18 (20)	–
3	22 (12)	9 (10)	13 (14)	–
4	8 (4)	5 (5)	3 (3)	–
Expectations for improvement**	2.5 (0.8)	2.4 (0.7)	2.7 (0.8)	0.028†

* Values are the number (%) unless indicated otherwise. SMT = spinal manipulative therapy; SRE = supervised rehabilitative exercises.
† Statistically significant pairwise comparison at ≤0.05.
‡ Includes American Indian or Alaska Native (n = 1), black or African American (n = 4), and other (n = 1).
§ 1 = no physical activity, 6 = very heavy physical activity.
¶ 1 =symptoms of pain, stiffness, or tenderness only; 2 = symptoms and musculoskeletal signs without radiation; 3 = symptoms and a) pain + radiation to extremity, proximally, b) pain + radiation to extremity, distally, c) pain + radiation to upper extremity with neurologic signs.
# 1 = pain without radiation; 2 = pain + radiation to proximal extremity; 3 = pain + radiation to distal extremity; 4 = pain + radiation to lower extremity with neurologic signs.
** Measured by asking: "Three months from now, how do you expect your back and neck problem to be?" (1 = no symptoms/100% improvement, 9 = as bad as it could be/100% worse).

week 36, that number grew to 21 visits reported among those whose study treatment had ended at 12 weeks, and 17 visits for those just finishing their 36 weeks of treatment in the long-term management group. At the week 78 long-term follow-up, similar numbers in each group (25 in the short-term group, 28 in the long-term treatment group) reported nonstudy health care visits.

Adverse events. No serious adverse events were reported. Six participants reported mild-to-moderate adverse events (3 in each group) and included 1 or a combination of increase in neck pain (2 subjects), back pain (1 subject), numbness in the hands (2 subjects) or feet (2 subjects), headache (1 subject), and dizziness with exercise (1 subject). Approximately half of participants reported in questionnaires that they experienced ≥1 side effect over the course of the study (51% of the total sample at 12 weeks; 58% of those in the short-term treatment group, 47% of the long-term management group at 36 weeks). No significant difference between groups was observed in frequency at either time point. An increase or change in neck or back pain was most common.

DISCUSSION

This study is one of the first to examine the relative effectiveness of short-term treatment (12 weeks) versus long-term management (36 weeks) of commonly recommended nonpharmacologic treatments, SMT with SRE, for older spine-pain patients. The study is also novel in that it addresses low-back– and neck pain–related

Table 2. Coprimary outcomes, changes from baseline, and between-group differences in neck and back disability*

Coprimary outcomes	Short-term treatment (n = 91)	Long-term management (n = 91)	Mean difference	*P*
Neck disability (NDI)				
Baseline	25.2 (23.7, 26.8)	26.6 (24.7, 28.6)	–1.4 (–3.9, 1.1)	0.26
Change from baseline				
Week 4	–4.5 (–6.2, –2.7)	–2.9 (–4.7, –1.2)	–1.5 (–4.0, 0.9)	0.23
Week 12	–7.8 (–9.6, –6.1)	–5.5 (–7.3, –3.8)	–2.3 (–4.8, 0.2)	0.07
Week 24	–7.6 (–9.3, –5.8)	–5.7 (–7.4, –3.9)	–1.9 (–4.4, 0.6)	0.14
Week 36	–7.3 (–9.1, –5.5)	–9.0 (–10.8, –7.2)	1.7 (–0.8, 4.2)	0.18
Week 52	–7.2 (–9.0, –5.4)	–7.1 (–8.9, –5.3)	–0.1 (–2.6, 2.4)	0.95
Week 78	–6.8 (–8.6, –5.0)	–8.6 (–10.4, –6.8)	1.8 (–0.8, 4.4)	0.17
Back disability (ODI)				
Baseline	25.6 (23.9, 27.3)	26.7 (24.6, 28.9)	–1.2 (–3.9, 1.5)	0.39
Change from baseline				
Week 4	–1.9 (–3.8, 0)	–2.0 (–3.9, –0.2)	0.1 (–2.5, 2.8)	0.92
Week 12	–6.1 (–8.0, –4.2)	–4.6 (–6.5, –2.7)	–1.5 (–4.2, 1.1)	0.26
Week 24	–5.5 (–7.4, –3.6)	–4.1 (–6.0, –2.3)	–1.4 (–4.0, 1.3)	0.32
Week 36	–3.9 (–5.8, –2.0)	–6.3 (–8.2, –4.4)	2.4 (–0.3, 5.1)	0.08
Week 52	–3.6 (–5.6, –1.7)	–4.4 (–6.3, –2.5)	0.8 (–1.9, 3.5)	0.56
Week 78	–3.5 (–5.5, –1.5)	–5.8 (–7.7, –3.8)	2.3 (–0.5, 5.0)	0.11

* Values are the mean (95% confidence interval) unless indicated otherwise. Higher scores indicate greater disability. NDI = Neck Disability Index; ODI = Oswestry Disability Index.

disability simultaneously, better reflecting patients' real-world pain experiences, in which back and neck pain co-occur (6,33,34). While both groups experienced improvements in disability from baseline to week 36, there were no statistically significant differences and small effect size between groups (Cohen's effect size was 0.22 for NDI and 0.25 for ODI). Both groups achieved the greatest average improvement at the time of treatment completion and generally sustained improvement through the long term. Similar results were observed for the secondary outcomes, with the exception of the long-term management group self-reporting greater improvement in neck pain and self-efficacy and exhibiting increased gains in some objective functional measures. Future-effectiveness and cost-effectiveness studies with long-term follow-up are needed to determine whether these improvements in secondary outcomes are worth the extra time and cost associated with long-term management. Importantly, no serious adverse events were reported. Cumulatively, these findings suggest that SMT and SRE are safe for elderly patients experiencing low-back– and neck pain–related disability and that longer-term management may empower them and result in important functional benefits (e.g., balance, physical performance).

The lack of group differences in primary and most secondary outcomes, and the absence of a no-treatment control, make it difficult to discern the extent to which natural history and regression to the mean impacted the results. However, given the older age, chronicity, and disability of the study sample, and the persistence of observed improvements through the 78-week follow-up, the SMT and SRE probably conferred specific benefits.

Expectations have been shown to be predictive of persistent disability among older adults receiving treatment for musculoskeletal conditions (30). Initially, expectations for improvement were greater among participants randomized to the long-term management group, compared to those receiving a shorter course of care. As a result, this baseline difference was included as a covariate in the analysis. Over the course of the trial, both groups lowered their expectations for improvement, more so in the short-term treatment group. The reasons for this change in expectations, and their impact on improvement, warrants further exploration.

More than one-fourth of the study sample achieved >50% improvement in neck and back disability at the primary time point; 33–55% reported obtaining 30% improvement. Nearly three-fourths of participants reported at least 15% improvement in neck disability, and more than half reported the same magnitude of improvement in back disability. We report a range of thresholds in the responder analysis of this study, because there is some disagreement in the literature as to what constitutes clinically important difference in the NDI (35,36) and ODI (32,37,38). A change of 15% may represent a clinically meaningful improvement among older adults experiencing chronic spine-related disability, especially given the relatively low cost and low risk of SMT and SRE as interventions (31,39,40).

Maintaining good physical function is crucial for the elderly to remain independent. At 36 weeks, those in the long-term management group had significantly increased their score on the SPPB compared to the short-term treatment group, by a degree that could be considered medium in terms of meaningful change (0.76 versus 0.14 points) (41). The Balance Test subscore decreased slightly in the short-term group, whereas it increased significantly in the long-term group (–0.03 versus 0.33).

Table 3. Patient self-reported outcomes at baseline and change at weeks 12, 36, 52, and 78*

Outcome measure and change from baseline	Short-term treatment (n = 91)	Long-term management (n = 91)	Mean difference	*P*
Neck pain†				
Baseline	4.5 (4.2, 4.8)	4.7 (4.3, 5.1)	−0.2 (−0.7, 0.3)	0.42
Week 12	−1.7 (−2.1, −1.3)	−1.3 (−1.7, −0.9)	−0.4 (−0.9, 0.2)	0.17
Week 36	−1.4 (−1.8, −1.1)	−2.1 (−2.5, −1.7)	0.7 (0.1, 1.2)	0.02‡
Week 78	−1.6 (−2.0, −1.2)	−1.9 (−2.3, −1.5)	0.2 (−0.3, 0.8)	0.41
Back pain†				
Baseline	5.1 (4.7, 5.4)	5.0 (4.5, 5.4)	0.1 (−0.5, 0.7)	0.70
Week 12	−1.9 (−2.3, −1.5)	−1.4 (−1.8, −1.0)	−0.5 (−1.1, 0.0)	0.07
Week 36	−1.6 (−2.0, −1.2)	−2.0 (−2.4, −1.6)	0.4 (−0.2, 1.0)	0.19
Week 78	−1.4 (−1.9, −1.0)	−1.7 (−2.2, −1.3)	0.3 (−0.3, 0.9)	0.33
Medication use§				
Baseline	3.0 (2.5, 3.6)	2.7 (2.2, 3.3)	0.3 (−0.5, 1.0)	0.50
Week 12	−0.6 (−1.0, −0.1)	−0.8 (−1.2, −0.3)	0.2 (−0.5, 0.9)	0.55
Week 36	−0.5 (−0.9, 0.0)	−0.9 (−1.4, −0.5)	0.5 (−0.2, 1.1)	0.16
Week 78	−0.8 (−1.3, −0.3)	−0.3 (−0.8, 0.2)	−0.5 (−1.2, 0.2)	0.12
Improvement‖				
12 weeks	3.0 (2.7, 3.3)	3.4 (3.1, 3.7)	−0.4 (−0.8, 0.0)	0.07
36 weeks	3.3 (3.0, 3.6)	3.1 (2.8, 3.4)	0.2 (−0.2, 0.6)	0.31
78 weeks	3.3 (3.0, 3.6)	3.3 (3.0, 3.5)	0.0 (−0.4, 0.4)	0.89
Satisfaction#				
12 weeks	1.8 (1.6, 2.0)	1.8 (1.6, 2.0)	−0.1 (−0.3, 0.2)	0.73
36 weeks	2.0 (1.8, 2.2)	1.8 (1.6, 2.0)	0.2 (−0.1, 0.5)	0.13
78 weeks	2.2 (2.0, 2.3)	2.0 (1.8, 2.2)	0.1 (−0.2, 0.4)	0.35
Kinesiophobia**				
Baseline	34.5 (33.4, 35.6)	34.2 (33.1, 35.4)	0.2 (−1.3, 1.8)	0.77
Week 12	−3.1 (−4.1, −2.2)	−2.1 (−3.1, −1.1)	−1.0 (−2.4, 0.4)	0.14
Week 36	−2.5 (−3.5, −1.5)	−1.9 (−2.9, −0.9)	−0.6 (−2.0, 0.8)	0.39
Week 78	−2.2 (−3.3, −1.2)	−1.7 (−2.7, −0.7)	−0.6 (−2.0, 0.9)	0.43
Self efficacy††				
Baseline	49.0 (47.4, 50.7)	46.3 (44.2, 48.3)	2.8 (0.1, 5.4)	0.04‡
Week 12	3.1 (1.6, 4.6)	1.3 (−0.2, 2.7)	1.8 (−0.3, 3.9)	0.08
Week 36	1.4 (−0.1, 2.9)	3.6 (2.1, 5.1)	−2.2 (−4.3, −0.1)	0.04‡
Week 78	0.9 (−0.7, 2.4)	2.6 (1.1, 5.0)	−1.7 (−3.9, 0.5)	0.12
Quality of life (EQ-5D)‡‡				
Baseline	0.8 (0.8, 0.8)	0.8 (0.8, 0.8)	0.0 (−0.0, 0.0)	0.94
Week 12	0.0 (0.0, 0.1)	0.0 (0.0, 0.0)	0.0 (−0.0, 0.0)	0.15
Week 36	0.0 (0.0, 0.1)	0.0 (0.0, 0.0)	0.0 (−0.0, 0.0)	0.72
Week 78	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.0 (−0.0, 0.0)	0.92
Expectations for improvement§§				
Baseline	2.4 (2.3, 2.6)	2.7 (2.5, 2.8)	−0.3 (−0.5, 0.0)	0.03‡
Week 12	0.6 (0.3, 0.8)	0.1 (−0.1, 0.3)	0.5 (0.1, 0.8)	0.01‡
Week 36	0.9 (0.7, 1.2)	0.6 (0.3, 0.8)	0.4 (0.0, 0.7)	0.05‡

* Values are the mean (95% confidence interval) unless indicated otherwise. EQ-5D = EuroQol 5-domain instrument.
† Neck and low-back pain in the past week (0–10). Lower scores indicate less pain.
‡ Statistically significant pairwise comparison at ≤0.05.
§ Days of prescription or over-the-counter medication use during the past week (0 = have not taken any, 7 = every day).
‖ 1 = no symptoms/100% improvement, 9 = as bad as it could be/100% worse.
1 = completely satisfied/couldn't be better, 7 = completely dissatisfied/couldn't be worse.
** Tampa Scale of Kinesiophobia (17-item instrument). Lower scores indicate lower levels of kinesiophobia.
†† Pain Self-Efficacy Questionnaire (10-item instrument). Higher scores indicate stronger self-efficacy beliefs.
‡‡ EQ-5D was used to determine participants' general health. US population-based preferences were used to calculate an index score. Higher scores indicate greater health status.
§§ Measured by asking: "Three months from now, how do you expect your back and neck problem to be?" (1 = no symptoms/100% improvement, 9 = as bad as it could be/100% worse.)

Using the Oswestry and NDI instruments to measure disability, we observed that improvements were generally larger for neck pain than for low-back pain (–8.28 versus –5.13 for the whole sample), and this finding was true for both men and women (data not shown). The change in neck disability is similar to the 8.4-point improvement observed in a previous trial investigating the same interventions in an elderly population during a 12-week period. Differences in pain, however, were less than the 3-point change

observed in that trial (20). Whether SRE and SMT combined are genuinely more effective for chronic neck pain than for chronic low-back pain in the elderly, or whether this finding is explained by differences in responsiveness between the 2 disability instruments cannot be elucidated from our data.

Originally this trial was designed to include a third arm, which would have allowed us to assess the impact of adding SMT to SRE. Redesign due to poor enrollment and funding did not likely influence the results as reported here. Strengths include a rigorous study design, including self-reported and objective measures; standardized, yet pragmatic interventions; and high follow-up rates. Limitations include the inability to blind clinicians and participants due to the nature of the interventions. While combining both neck- and back-related disability into a single study population may create a more heterogenous sample than either alone, doing so reflects the phenomenon of spine-related disability for many patients and how patients present for care. This trial did not differentiate between specific effects of the intervention and contextual effects, which may play a large role when treating patients with chronic pain. However, the treatments as studied generally reflect how spinal manipulation and rehabilitative exercise are delivered in clinical practice. Finally, while visits to nonstudy health care providers was similar between groups, the impact on the study results is unknown. Our study adds additional support to evidence-based guidelines, which recommend that manual treatment, along with general and specific exercises, should be considered as first-line treatments for patients with back and neck pain (15–18).

For adults ages ≥65 years with chronic back and neck disability, extending management with SMT and SRE from 12 to 36 weeks did not result in any additional important reduction in disability. Statistically significant differences in favor of long-term management were found for the secondary outcomes of self-reported improvement in neck pain and self-efficacy, as well as functional measures of balance and physical performance. These findings may be important for healthy aging and spine care in the elderly and warrant further investigation.

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. Maiers 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. Maiers, Hartvigsen, Westrom, Bronfort.
Acquisition of data. Maiers, Evans, Schulz.
Analysis and interpretation of data. Maiers, Hartvigsen, Wang, Leininger.

REFERENCES

1. Thomas E, Peat G, Harris L, Wilkie R, Croft PR. The prevalence of pain and pain interference in a general population of older adults: cross-sectional findings from the North Staffordshire Osteoarthritis Project (NorStOP). Pain 2004;110:361–8.
2. Hartvigsen J, Christensen K, Frederiksen H. Back pain remains a common symptom in old age: a population-based study of 4486 Danish twins aged 70–102. Eur Spine J 2003;12:528–34.
3. Gill TM, Desai MM, Gahbauer EA, Holford TR, Williams CS. Restricted activity among community-living older persons: incidence, precipitants, and health care utilization. Ann Intern Med 2001;135:313–21.
4. Makris UE, Fraenkel L, Han L, Leo-Summers L, Gill TM. Restricting back pain and subsequent mobility disability in community-living older persons. J Am Geriatr Soc 2014;62:2142–7.
5. Laslett LL, Quinn SJ, Winzenberg TM, Sanderson K, Cicuttini F, Jones G. A prospective study of the impact of musculoskeletal pain and radiographic osteoarthritis on health related quality of life in community dwelling older people. BMC Musculoskelet Disord 2012;13:168.
6. Hartvigsen J, Christensen K, Frederiksen H. Back and neck pain exhibit many common features in old age: a population-based study of 4,486 Danish twins 70–102 years of age. Spine (Phila Pa 1976) 2004;29:576–80.
7. GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016;388:1545–602.
8. Pasquale MK, Dufour R, Schaaf D, Reiners AT, Mardekian J, Joshi AV, et al. Pain conditions ranked by healthcare costs for members of a national health plan. Pain Pract 2014;14:117–31.
9. Ortman JM, Velkoff VA, Hogan H. An aging nation: the older population in the United States. Population estimates and projections. U.S. Department of Commerce Economics and Statistics Administration report P25–1140. 2014. URL: https://www.census.gov/prod/2014pubs/p25-1140.pdf.
10. Hill J, Lewis M, Papageorgiou AC, Dziedzic K, Croft P. Predicting persistent neck pain: a 1-year follow-up of a population cohort. Spine (Phila Pa 1976) 2004;29:1648–54.
11. Itz CJ, Geurts JW, van Kleef M, Nelemans P. Clinical course of nonspecific low back pain: a systematic review of prospective cohort studies set in primary care. Eur J Pain 2013;17:5–15.
12. Carey TS, Garrett JM, Jackman AM. Beyond the good prognosis: examination of an inception cohort of patients with chronic low back pain. Spine (Phila Pa 1976) 2000;25:115–20.
13. Descarreaux M, Blouin JS, Drolet M, Papadimitriou S, Teasdale N. Efficacy of preventive spinal manipulation for chronic low-back pain and related disabilities: a preliminary study. J Manipulative Physiol Ther 2004;27:509–14.
14. Senna MK, Machaly SA. Does maintained spinal manipulation therapy for chronic nonspecific low back pain result in better long-term outcome? Spine (Phila Pa 1976) 2011;36:1427–37.
15. Wong JJ, Cote P, Sutton DA, Randhawa K, Yu H, Varatharajan S, et al. Clinical practice guidelines for the noninvasive management of low back pain: a systematic review by the Ontario Protocol for Traffic Injury Management (OPTIMa) Collaboration. Eur J Pain 2017;21:201–16.
16. Cote P, Wong JJ, Sutton D, Shearer HM, Mior S, Randhawa K, et al. Management of neck pain and associated disorders: a clinical practice guideline from the Ontario Protocol for Traffic Injury Management (OPTIMa) Collaboration. Eur Spine J 2016;25:2000–22.
17. Chou R, Deyo R, Friedly J, Skelly A, Hashimoto R, Weimer M, et al. Nonpharmacologic therapies for low back pain: a systematic review for an American College of Physicians clinical practice guideline. Ann Intern Med 2017;166:493–505.
18. National Institute for Health and Care Excellence. Low back pain and sciatica in over 16s: assessment and management. NICE guideline [NG59]. 2016. URL: https://www.nice.org.uk/guidance/ng59.

19. Balthazard P, de Goumoens P, Rivier G, Demeulenaere P, Ballabeni P, Deriaz O. Manual therapy followed by specific active exercises versus a placebo followed by specific active exercises on the improvement of functional disability in patients with chronic non specific low back pain: a randomized controlled trial. BMC Musculoskelet Disord 2012;13:162.
20. Maiers M, Bronfort G, Evans R, Hartvigsen J, Svendsen K, Bracha Y, et al. Spinal manipulative therapy and exercise for seniors with chronic neck pain. Spine J 2014;14:1879–89.
21. Vihstadt C, Maiers M, Westrom K, Bronfort G, Evans R, Hartvigsen J, et al. Short term treatment versus long term management of neck and back disability in older adults utilizing spinal manipulative therapy and supervised exercise: a parallel-group randomized clinical trial evaluating relative effectiveness and harms. Chiropr Man Therap 2014;22:26.
22. Bergman TF, Peterson DH, Lawrence DJ. Chiropractic technique. New York: Churchill Livingstone; 1993.
23. Fairbank JC, Pynsent PB. The Oswestry Disability Index. Spine (Phila Pa 1976) 2000;25:2940–52.
24. Roland M, Fairbank J. The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. Spine (Phila Pa 1976) 2000;25:3115–24.
25. Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. J Manipulative Physiol Ther 1991;14:409–15.
26. Knudtson MD, Klein BE, Klein R. Biomarkers of aging and falling: the Beaver Dam eye study. Arch Gerontol Geriatr 2009;49:22–6.
27. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. J Gerontol 1994;49:M85–94.
28. Guralnik JM, Ferrucci L, Simonsick EM, Salive ME, Wallace RB. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. N Engl J Med 1995;332:556–61.
29. Maiers MJ, Hartvigsen J, Schulz C, Schulz K, Evans RL, Bronfort G. Chiropractic and exercise for seniors with low back pain or neck pain: the design of two randomized clinical trials. BMC Musculoskelet Disord 2007;8:94.
30. Rundell SD, Sherman KJ, Heagerty PJ, Mock CN, Jarvik JG. The clinical course of pain and function in older adults with a new primary care visit for back pain. J Am Geriatr Soc 2015;63:524–30.
31. Skou ST, Roos EM, Laursen MB. A randomized, controlled trial of total knee replacement. N Engl J Med 2016;374:692.
32. Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine (Phila Pa 1976) 2008;33:90–4.
33. Manchikanti L, Singh V, Falco FJ, Benyamin RM, Hirsch JA. Epidemiology of low back pain in adults. Neuromodulation 2014;17 Suppl:3–10.
34. Hartvigsen J, Natvig B, Ferreira M. Is it all about a pain in the back? Best Pract Res Clin Rheumatol 2013;27:613–23.
35. MacDermid JC, Walton DM, Avery S, Blanchard A, Etruw E, McAlpine C, et al. Measurement properties of the neck disability index: a systematic review. J Orthop Sports Phys Ther 2009;39:400–17.
36. Pool JJ, Ostelo RW, Hoving JL, Bouter LM, de Vet HC. Minimal clinically important change of the Neck Disability Index and the Numerical Rating Scale for patients with neck pain. Spine (Phila Pa 1976) 2007;32:3047–51.
37. Fritz JM, Hebert J, Koppenhaver S, Parent E. Beyond minimally important change: defining a successful outcome of physical therapy for patients with low back pain. Spine (Phila Pa 1976) 2009;34:2803–9.
38. Gatchel RJ, Mayer TG. Testing minimal clinically important difference: consensus or conundrum? Spine J 2010;10:321–7.
39. Hurley MV, Walsh NE, Mitchell H, Nicholas J, Patel A. Long-term outcomes and costs of an integrated rehabilitation program for chronic knee pain: a pragmatic, cluster randomized, controlled trial. Arthritis Care Res (Hoboken) 2012;64:238–47.
40. Villadsen A, Overgaard S, Holsgaard-Larsen A, Christensen R, Roos EM. Immediate efficacy of neuromuscular exercise in patients with severe osteoarthritis of the hip or knee: a secondary analysis from a randomized controlled trial. J Rheumatol 2014;41:1385–94.
41. Perera S, Mody SH, Woodman RC, Studenski SA. Meaningful change and responsiveness in common physical performance measures in older adults. J Am Geriatr Soc 2006;54:743–9.

Arthritis Care & Research
Vol. 71, No. 11, November 2019, pp 1525–1529
DOI 10.1002/acr.23793

American College of Rheumatology
Empowering Rheumatology Professionals

BRIEF REPORT

Increasing Rates of Arthroplasty for Psoriatic Arthritis in the United Kingdom Between 1995 and 2010

Ryan T. Lewinson, Isabelle A. Vallerand, Jeremy M. LaMothe, Laurie M. Parsons, Alexandra D. Frolkis, Mark W. Lowerison, Scott B. Patten, and Cheryl Barnabe

Objective. Arthroplasty requirements among patients with psoriatic arthritis (PsA) are not well known. This information is important to clinical and policy stakeholders for health-system planning and may serve as a surrogate for estimation of the efficacy of disease-modifying therapy.

Methods. We utilized The Health Improvement Network (THIN), a large general practice medical records database in the UK, to assess rates of primary total arthroplasty among patients with PsA and the general population between the years 1995 and 2010. Linear regression was used to estimate arthroplasty rates for the 2 cohorts during the study period, and Poisson regression was used to determine age- and sex-adjusted incidence rate ratios (IRRs) between the PsA and general population cohorts.

Results. We identified 5,619 patients with incident PsA and 5,090,814 eligible patients from the general population between 1995 and 2010. In total, 187 primary total arthroplasties were documented in patients with PsA, and 80,163 primary total arthroplasties were documented in the general population. A trend of increasing arthroplasty rates was observed for both the PsA ($R^2 = 0.809$; $P < 0.0001$) and general population ($R^2 = 0.890$; $P < 0.0001$) cohorts during the study period. After adjustment for age and sex, patients with PsA had a first arthroplasty incidence rate that was twice that of the general population (IRR 2.01 [95% confidence interval 1.73–2.34]; $P < 0.0001$), notably beyond the year 2003 when biologic therapies were introduced.

Conclusion. Both the general population and patients with PsA have experienced increasing rates of first arthroplasty from 1995 to 2010, although the overall incidence rate was significantly higher for those with PsA.

INTRODUCTION

Psoriatic arthritis (PsA), a systemic autoimmune disease with a prevalence estimated at 0.1–0.2%, is characterized by progressive inflammatory arthritis (peripheral joints, axial disease, or both), cutaneous psoriasis, and frequently with tendinopathy (1). Inflammation results in degradation of joint structure, which leads to long-term disability and mobility challenges. For many years, the first-line therapy for patients with PsA were disease-modifying antirheumatic drugs (DMARDs). Although DMARDs have proven efficacy, in a subset of patients they have an insufficient effect or are not tolerated. The resulting uncontrolled inflammation can contribute to severe joint destruction, with patients occasionally having to undergo arthroplasty for symptomatic and functional reasons. Clinical outcomes for patients with severe PsA have markedly improved and continue to do so with the advent of new targeted biologic treatments and intensive treatment strategies (2). However, arthroplasty may still be required in advanced disease (3).

In contrast to rheumatoid arthritis (RA), another systemic autoimmune disease of peripheral joints, rates of arthroplasty among patients with PsA are not well reported or studied. This is surprising, given that PsA tends to affect larger weight-bearing joints compared to RA (1), and because PsA is associated with obesity and other metabolic comorbidities, which are risk factors for osteoarthritis (OA) (4,5). For RA, arthroplasty rates have been used as a proxy for whether advanced medical therapy has contributed to the prevention of advanced-stage disease (6,7), but a

Dr. Lewinson's work was supported by the Canadian Association of Psoriasis Patients and the Canadian Institutes of Health Research (Studentship Award from the Institute of Musculoskeletal Health and Arthritis).

Ryan T. Lewinson, MD, PhD, Isabelle A. Vallerand, MD, PhD, Jeremy M. LaMothe, MD, PhD, Laurie M. Parsons, MD, Alexandra D. Frolkis, MD, PhD, Mark W. Lowerison, MSc, Scott B. Patten, MD, PhD, Cheryl Barnabe, MD, MSc: Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada.

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

Address correspondence to Cheryl Barnabe, MD, MSc, 3330 Hospital Drive NW, Calgary, Alberta T2N 4N1, Canada. E-mail: ccbarnab@ucalgary.ca.

Submitted for publication June 13, 2018; accepted in revised form October 16, 2018.

SIGNIFICANCE & INNOVATIONS

- From 1995 to 2010, the rate of first arthroplasty increased for patients with psoriatic arthritis (PsA) and in the general population.
- The incidence rate of first arthroplasty was twice as high in patients with PsA as compared to the general population.
- The higher incidence rates among PsA patients included the years 2003–2010, when biologic therapies were available.

similar study has not been done for PsA. The types of data that are important for numerous stakeholders include data on patients with PsA who have a vested interest in the musculoskeletal prognosis of their condition, rheumatologists and dermatologists who routinely manage patients with psoriasis and PsA, orthopedic surgeons (because arthroplasty among patients with PsA carries additional risks, given that they may be immunosuppressed in the peri-surgical period and have a broken skin barrier) (3), and health care systems that must consider medical and surgical treatment costs.

Given the rarity of PsA and arthroplasty literature, we sought to utilize a large nationally representative health records database to investigate the rate of first arthroplasty among patients with PsA between 1995 and 2010. We tested the null hypothesis that rates of first arthroplasty would not differ between patients with PsA and in the general population.

PATIENTS AND METHODS

Data source. This study used The Health Improvement Network (THIN) as a data source. THIN is an electronic general practice medical records database that contains long-term health data from nearly 12 million individuals in the UK. The patients registered in THIN have similar age and sex distributions to the general UK population (8,9), and thus estimates derived from THIN data are likely to be representative of the population. While THIN is a general practice database, data from specialists and hospitals are often included in THIN, allowing for detailed long-term follow-up in many patients. THIN has been used previously for the study of PsA (10). Given the large number of patients in THIN with long duration of follow-up, THIN is one of the only suitable data sources to assess the risk of arthroplasty in PsA (a rare outcome in a relatively rare disease).

Study population. THIN was used to identify individuals from the years 1995 to 2010 who were between the ages of 20–90 years and diagnosed with incident PsA, using a one-year washout period and Read codes that have previously been validated for PsA (10). The date on which the first Read code for PsA appears for each patient was taken as their start date in the study. A general population cohort without PsA between the ages of 20–90 years at their start date in THIN was also identified between 1995 and 2010. These data were utilized to estimate the rates of arthroplasty between these cohorts over a 15-year period.

Outcomes. The primary outcome in this study was first instance of primary total arthroplasty; however, patients were followed until the first of the following events: 1) first instance of primary total arthroplasty, 2) transfer out of practice, 3) death, or 4) end of data collection period. Arthroplasty was defined based on the presence of eligible arthroplasty Read codes, which were selected by an orthopedic surgeon (JML). Codes for revision or conversion arthroplasty (i.e., conversion from hemiarthroplasty to total arthroplasty), hip hemiarthroplasty (commonly for fractures), Colonna arthroplasty (for developmental hip dysplasia) and excisional arthroplasty were excluded to ensure focus was on primary arthroplasty (because these other procedures have primary indications other than arthritis). Additionally, in order to ensure that only incident codes for arthroplasty were considered, any patients with codes that met the criteria for arthroplasty within 1 year of the start date in THIN or with arthroplasty codes prior to the code for PsA were excluded from consideration in the study. This established temporality such that arthroplasty was only considered where it occurred after a diagnosis of PsA had been made. We only considered first instance of primary arthroplasty, since this presumably represents a failure to achieve sustained disease control. We elected to not remove any individuals with an existing Read code for OA in either cohort, on the basis that existing literature suggests that the prevalence of OA seems similar between those with PsA and the general population (10,11). Consequently, differences in arthroplasty rates between the groups could be more likely attributed to PsA rather than differing OA rates.

Analysis. All analyses were performed using Stata, version 13.1 with an alpha level of 0.05. First, incidence rates were determined for arthroplasty on an annual basis within the PsA and general population cohorts independently. Linear regressions were conducted for the PsA and general population cohorts to assess the relationship between the incidence rates of arthroplasty and time. In order to compare incidence rates between the PsA and general population cohorts, Poisson regression was used to determine age (dichotomized based on ≥40 years or <40 years at baseline) and sex-adjusted incidence rate ratios (IRRs) and corresponding 95% confidence intervals (95% CIs) for arthroplasty over time. Chi-square tests were used to assess differences in arthroplasty rates between patients with PsA and the general population within each year. A sensitivity analysis was performed whereby Vision Date in THIN (i.e., the date on which Vision software was implemented in each medical practice for electronic data recording) was utilized to mark patient start points in the study, rather than the patient's start date (i.e. date in which they joined a THIN-affiliated practice) to ensure data entry was standardized. This

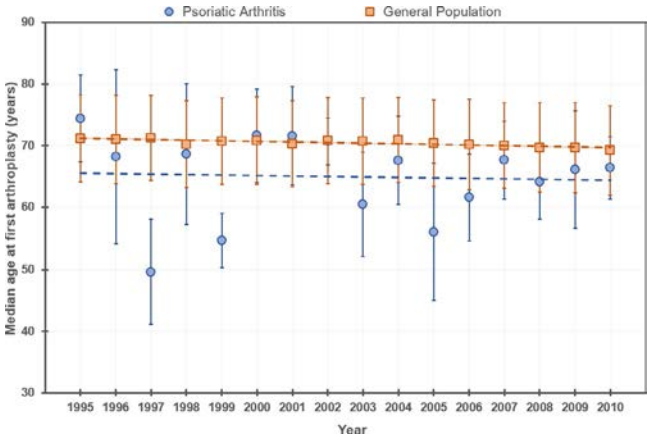

Figure 1. Median age at first arthroplasty across study years for both the psoriatic arthritis and general population cohorts. Error bars show the median and interquartile range. The dotted lines represent the linear regression line of best fit.

manuscript was prepared in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology statement (12), with the Reporting of Studies Conducted using Observational Routinely-Collected Health Data extension (13). Data are available with permission from IMS Health, UK and institutional research board review.

RESULTS

We identified 5,619 patients with incident PsA (34,960 person-years follow-up) and 5,090,814 eligible patients from the general population (44,126,860 person-years follow-up) between 1995 and 2010 for analysis. The median age at PsA diagnosis was 49.7 years (interquartile range [IQR] 18.8). In total, 187 patients with PsA had first instance primary arthroplasties (77 cases for hip, 99 cases for knee, 11 cases for other joints), and 80,163 first instance arthroplasties were documented in the general population (40,759 cases for hip, 34,410 cases for knee, 4,904 cases for other joints, 90 cases where the patient had arthroplasty on more than 1 joint). Patients with PsA were more likely ($P < 0.0001$) to undergo first arthroplasty at a younger age (median 64.7 years [IQR 16.7]) compared to the general population (median 70.2 years [IQR 14.3]). Over the course of the study, there did not appear to be any trends in age at first arthroplasty (Figure 1).

Overall, trends of increasing first arthroplasty rates over time were found for both the general population ($R^2 = 0.890$, $P < 0.0001$) and the PsA cohort ($R^2 = 0.809$, $P < 0.0001$) during the study period. From 1995 to 2010, the rate of first arthroplasty increased from 127/100,000 person-years to 438/100,000 person-years in the PsA group, and from 136/100,000 person-years to 254/100,000 person-years in the general population. The overall age- and sex-adjusted IRR for first arthroplasty in patients with PsA was 2.01 (95% CI 1.73–2.34, $P < 0.0001$) compared to the general population (Figure 2).

Our sensitivity analysis that utilized patient Vision Dates as their start point in the study found largely similar but slightly less conservative results, where first arthroplasty rates increased over time for both the general population ($R^2 = 0.987$; $P < 0.0001$) and the PsA cohort ($R^2 = 0.891$; $P < 0.0001$). Similarly, patients with PsA were found to have a higher rate of first arthroplasty, as determined by age- and sex-adjusted IRR for first arthroplasty (IRR 2.44 [95% CI 2.10–2.84], $P < 0.0001$).

DISCUSSION

Overall, we found an increasing trend in rates of first arthroplasty among patients with PsA and in the general population; however, incidence rates for patients with PsA appear to have increased to a greater degree compared to the general population, mostly due to greater incidence rates from 2003–2010. There are many possibilities to explain this finding, which are outlined herein. The increased rate of arthroplasty in the general population is in agreement with other research (14), often attributed to an aging population, and a similar mechanism could explain a portion of the rise in incidence rates for the PsA population; patients may be living longer with their disease and thus more patients eventually go on to surgery.

One potential explanation for higher rates of first arthroplasty among the PsA group could be increased health care utilization. In particular, patients with PsA may have a greater need to see a physician regularly regarding their drug therapy and disease activity. Thus, through heightened surveillance,

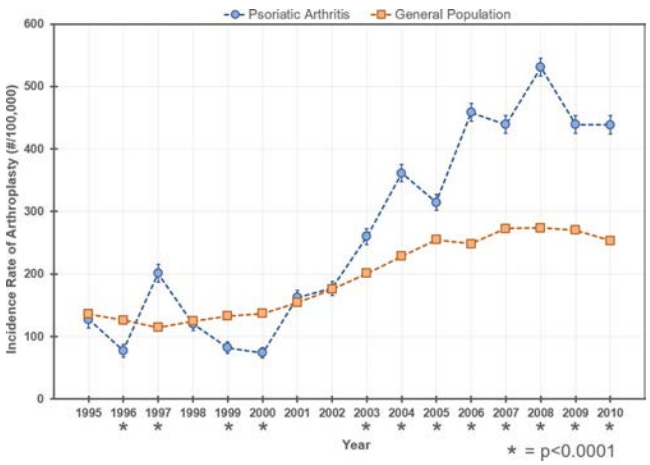

Figure 2. The incidence rate (IRR; 95% confidence intervals [95% CI]) of arthroplasty increased for both the psoriatic arthritis (PsA) cohort and in the general population cohort, from 1995 to 2010. Due to the very large amount of person-time observed in each year, confidence intervals appear very small for the general population cohort. Evaluation of the overall age- and sex-adjusted IRR revealed that the incidence rate of first arthroplasty among patients with PsA was twice that of the general population (IRR 2.01 [95% CI 1.73–2.34]; $P < 0.0001$). *Significant difference in incidence rates by chi-square test.

patients with PsA could also be considered for surgery earlier, potentially explaining the increase in arthroplasty rates observed beyond 2003, when biologics were introduced. We did not match our cohorts by health care utilization as we endeavored to identify differences in the rate of first arthroplasty between PsA patients and the general population on a population level, without restricting to only the sickest people within the general population (i.e., those with the highest health care utilization).

A paradoxical increase in surgeries may also have occurred in the patients with PsA. For instance, etanercept and other biologic therapies began receiving approval for PsA around 2003, and it is possible that these modern treatments effectively modulated disease symptoms. This could have resulted in the heightened ability to observe joints not responding to therapy, and thus the need for arthroplasty. Further, with better disease control and inhibition of joint damage progression, PsA patients with advanced disease may have been better candidates for arthroplasty (i.e., better disease control yields less risk of additional future surgeries or lower surgical risk). In line with this, achieving control of PsA could unmask underlying concurrent OA, making these patients surgical candidates for OA, although this would not likely occur at a greater rate in comparison to the general population (10,11). It is also possible that following the introduction of biologic therapy with aggressive treatment of PsA, perhaps surgical interventions also came to be utilized more aggressively to achieve rapid disease control and symptom management.

Given that the time between the median age at PsA diagnosis and the median age at the time of surgery for PsA was approximately 15 years, it is possible that the study period was too short to detect a discernable reduction in arthroplasty rates in those with PsA following the introduction of biologic therapy and tight disease control–treatment paradigms in 2003. Additionally, while it is possible that the recording of surgical data was not adequately captured in the earlier years of THIN (just as overall data availability and sample size is lower in earlier years of THIN), this would be expected to occur nondifferentially and therefore would not likely account for the relative rise in arthroplasty rates in patients with PsA. Lastly, given that PsA is relatively rare and often difficult to diagnose (3), it is possible that some cases of PsA were misclassified, although again this would also be expected to occur nondifferentially.

As noted from the above discussion, there are many possibilities to explain why patients with PsA might be experiencing greater increases in rates of first arthroplasty in comparison to the general population. Certainly, focused assessments on each of these topics will be needed over the next 10–15 years as more data becomes available, to further elucidate the reasons why arthroplasty rates for patients with PsA increased relative to the general population. In particular, it will be important to gain further understanding on rheumatologist and orthopedic surgeon clinical decision-making patterns with PsA patients, and what aspects might trigger referral and undertaking of arthroplasty. Moreover, with longer-term data collection, it might be revealed whether arthroplasty rates may soon decline for patients with PsA, stemming from the continued use of biologics. Unfortunately, addressing these aspects are not currently possible using the THIN database.

While we identified a large sample size from THIN, including an appropriate number of PsA patients relative to the general population, the total number of arthroplasties was lower than previously documented (14). In the UK, it is reported that nearly 150,000 arthroplasties are performed annually (15); however, this statistic considers all surgeries including primary total arthroplasty, revision arthroplasty, and primary arthroplasty on a contralateral limb in a patient who may have had a previous surgery. In our study, we only considered the first instance of primary arthroplasty, or "first surgery," which accounts for our lower observed total arthroplasty rate. Furthermore, the finding that arthroplasty rates have increased over time in the general population is consistent with other research (14), adding further validity to our data.

This study is limited based on a relatively low number of arthroplasty cases in patients with PsA, affecting the precision of our estimates (as suggested by the slightly wider confidence intervals in this study compared to other THIN studies), yet this study includes one of the largest PsA cohorts considered in an orthopedic surgery context. Based on low sample size numbers for arthroplasty in the PsA cohort, stratification by potentially important covariates such as DMARD/biologic use, obesity, and smoking were not possible. While the data source (THIN) is UK data, and health care practices may differ slightly in other countries, we believe that the results are important to consider in a broader context, given the magnitude of the database needed to address the question of arthroplasty in PsA (a rare outcome in a relatively rare disease). Indeed, few other data sources are available that could address this question. Further work will be needed to understand the generalizability of our results to other countries.

In conclusion, we identified that the incidence rate of first arthroplasty in patients with PsA seems to have increased between 1995 and 2010; however, the reasons for this increase are not entirely clear. Further research and follow-up will be necessary over the next 10–15 years for continued assessment of the role of biologic treatments in prevention of the need for arthroplasty.

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. Lewinson 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. Lewinson, Vallerand, Barnabe.

Acquisition of data. Lewinson, Vallerand, LaMothe, Parsons, Frolkis, Lowerison, Patten, Barnabe.

Analysis and interpretation of data. Lewinson, Vallerand, LaMothe, Parsons, Frolkis, Lowerison, Patten, Barnabe.

REFERENCES

1. Ritchlin CT, Colbert RA, Gladman DD. Psoriatic arthritis. N Engl J Med 2017;376:957–70.
2. McInnes IB, Kavanaugh A, Gottlieb AB, Puig L, Rahman P, Ritchlin C, et al. Efficacy and safety of ustekinumab in patients with active psoriatic arthritis:1 year results of the phase 3, multicentre, double-blind, placebo-controlled PSUMMIT 1 trial. Lancet 2013;382:780–9.
3. Iofin I, Levine B, Badlani N, Klein GR, Jaffe WL. Psoriatic arthritis and arthroplasty: a review of the literature. Bull NYU Hosp Jt Dis 2008;66:41–8.
4. Dubreuil M, Rho YH, Man A, Zhu Y, Zhang Y, Love TJ, et al. Diabetes incidence in psoriatic arthritis, psoriasis and rheumatoid arthritis: a UK population-based cohort study. Rheumatology (Oxford) 2014;53:346–52.
5. Love TJ, Zhu Y, Zhang Y, Wall-Burns L, Ogdie A, Gelfand JM, et al. Obesity and the risk of psoriatic arthritis: a population-based study. Ann Rheum Dis 2012;71:1273–7.
6. Jamsen E, Virta LJ, Hakala M, Kauppi MJ, Malmivaara A, Lehto MU. The decline in joint replacement surgery in rheumatoid arthritis is associated with a concomitant increase in the intensity of anti-rheumatic therapy: a nationwide register-based study from 1995 through 2010. Acta Orthop 2013;84:331–7.
7. Aaltonen KJ, Virkki LM, Jämsen E, Sokka T, Konttinen YT, Peltomaa R, et al. Do biologic drugs affect the need for and outcome of joint replacements in patients with rheumatoid arthritis? A register-based study. Semin Arthritis Rheum 2013;43:52–62.
8. Wallace H, Shorvon S, Tallis R. Age-specific incidence and prevalence rates of treated epilepsy in an unselected population of 2,052,922 and age-specific fertility rates of women with epilepsy. Lancet 1998;352:1970–3.
9. Bhayat F, Das-Gupta E, Smith C, McKeever T, Hubbard R. The incidence of and mortality from leukaemias in the UK: a general population-based study. BMC Cancer 2009;9:252.
10. Ogdie A, Alehashemi S, Love TJ, Jiang Y, Haynes K, Hennessy S, et al. Validity of psoriatic arthritis and capture of disease modifying antirheumatic drugs in the health improvement network. Pharmacoepidemiol Drug Saf 2014;23:918–22.
11. Neogi T, Zhang Y. Epidemiology of osteoarthritis. Rheum Dis Clin North Am 2013;39:1–19.
12. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007;370:1453–7.
13. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med 2015;12:e1001885.
14. Dixon T, Shaw M, Ebrahim S, Dieppe P. Trends in hip and knee joint replacement: socioeconomic inequalities and projections of need. Ann Rheum Dis 2004;63:825–30.
15. National Joint Registry UK. Joint replacement statistics. URL: http://www.njrcentre.org.uk/njrcentre/Patients/Jointreplacementstatistics/tabid/99/Default.aspx.

ARP Announcements

Association of Rheumatology Professionals
2200 Lake Boulevard NE, Atlanta, Georgia 30319
www.rheumatology.org

ACR/ARP Annual Meeting

November 8–13, 2019, Atlanta

Download the New ACR Publications Mobile App

The brand-new ACR Publications app can be downloaded for free from the Apple store or Google Play. ACR members can log in for full-text access to all articles in A*rthritis Care & Research* and *Arthritis & Rheumatology*. Nonmembers can access abstracts of all *AC&R* and *A&R* articles, the full text of articles published more than one year ago, and select open-access articles published recently, as well as the full text of all articles from *ACR Open Rheumatology* and *The Rheumatologist*.

New Division Name

Rheumatology is truly a people specialty; We often develop lifelong relationships with our patients as well as our colleagues. We increasingly recognize that providing the best rheumatologic care requires a team effort. The collegial nature of our specialty is reflected in the ACR's mission statement: To empower rheumatology professionals to excel in their specialty.

In keeping with this mission, we are pleased to announce that our health professionals' membership division is changing its name to Association of Rheumatology Professionals (ARP). This name change highlights the dedication of the ACR to serve the entire rheumatology community. It also reflects our broadened base of interprofessional members (administrators, advanced practice nurses, health educators, nurses, occupational therapists, pharmacists, physical therapists, physician assistants, research teams, and more).

The name is new, but our commitment and promise remain the same: We are here for you, so you can be there for your patients.

ARP Membership

The Association of Rheumatology Professionals (ARP), a division of the American College of Rheumatology, appreciates your continued membership and looks forward to serving you another year. Membership costs range from $30 to $140. ARP welcomes nurse practitioners, nurses, physician assistants, office staff, researchers, physical therapists, occupational therapists, assistants, and students. Student membership is complimentary; the Annual Meeting registration fee is waived for students who submit the required student verification letter. For information, go to www.rheumatology.org and select "Membership" or call 404-633-3777 and ask for an ARP staff member.

New ACR Journal Twitter Account (@ACR_Journals) and Social Media Editor

The ACR journals are heightening our focus on social media, to benefit authors and readers. Among our first activities is the introduction of an official ACR Journals Twitter account: @ACR_Journals. Followers will enjoy special features and the opportunity to engage with authors and other fellow professionals about studies published in *Arthritis Care & Research, Arthritis & Rheumatology,* and *ACR Open Rheumatology*. Authors of published articles will have the opportunity to use @ACR_Journals to share their work and engage in dialogue with others interested in the research. The journals welcome Dr. Paul Sufka of Minneapolis as our first Social Media Editor.