ARP Announcements

Association of Rheumatology Professionals
2200 Lake Boulevard NE, Atlanta, Georgia 30319
www.rheumatology.org

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 *Arthritis 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*.

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.

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 73, No. 10, October 2021

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 73 • October 2021 • NO. 10

Cover image: The image on the cover (from Shultz et al, page 1379) is a computed tomography angiogram of the patient's chest in the axial plane showing evidence of multiple small pulmonary arteriovenous malformations.

Arthritis Care & Research
Vol. 73, No. 10, October 2021, pp 1379–1386
DOI 10.1002/acr.24428

American College of Rheumatology
Empowering Rheumatology Professionals

CLINICOPATHOLOGIC CONFERENCE

A Four-Year-Old Child With Digital Clubbing

Katherine Schultz, Allison Divanovic, Christopher Towe, Alexander Miethke, Katie Wusik, Adrienne Hammill, and Hermine Brunner

CASE PRESENTATION

Chief symptoms

A four-year-old female presented to our pediatric rheumatology clinic with the chief symptom of periodic fevers and abdominal pain.

History of the present illness

The patient's fevers began 4 months prior to presentation following a documented streptococcal pharyngeal infection, with a maximum temperature of 104°F. They occurred every 1–2 weeks and would then last a duration of 4–6 hours. Fevers were followed by severe abdominal pain that prevented eating. She did not have associated emesis, diarrhea, or constipation.

Medical, family, and social history

The past medical history of the patient included allergic rhinitis, characterized by rhinorrhea and congestion, which was treated with nasal spray. Otherwise, she was reported to have normal growth and development. There was no known family history of autoimmunity, including inflammatory bowel disease, systemic lupus erythematosus, vasculitis, psoriasis, or Raynaud's syndrome. Social history was noncontributory. There were no known relevant environmental exposures.

Review of systems

Review of systems was notable for lack of joint pain or swelling, digital clubbing, myalgia, limp, rash, fatigue, or headaches. The patient's parents denied that she had experienced respiratory symptoms, including shortness of breath, dyspnea on exertion, cough, or snoring. There was no history of recurrent respiratory infections.

Physical examination

The girl was a well-appearing child without dysmorphic features. Vital signs were normal, though pulse oximetry was not obtained initially. There were no oral lesions, the oropharynx was clear, and tonsils were without exudate. Findings of ophthalmic examination were normal. Cardiac examination did not reveal murmur, rub, gallop, extra heart sounds, or arrhythmia. Lung examination demonstrated normal and symmetric air entry bilaterally. Her abdomen was soft, nontender to palpation, and without hepatosplenomegaly. Musculoskeletal examination was notable for redness overlying the distal interphalangeal joints and bilateral finger and toe clubbing (Figure 1). There was no evidence of synovitis or muscle tenderness or weakness. The patient had a normal gait without evidence of limp.

Laboratory evaluation

Results of initial laboratory tests were notable for an elevated level of alkaline phosphatase at 1,059 units/liter, but otherwise normal for transaminases, complete metabolic panel (albumin level of 6.5 gm/dl and bilirubin level 0.6 mg/dl), complete blood cell count, sedimentation rate, S100A8/S100A9, S100A12, complement levels, immunoglobulin levels, and C-reactive protein level (obtained between fever episodes). Antinuclear antibodies were negative.

Radiologic evaluation

A chest radiograph demonstrated nonspecific peribronchial thickening.

Katherine Schultz, MD, Allison Divanovic, MD, Christopher Towe, MD, Alexander Miethke, MD, Katie Wusik, LGC, Adrienne Hammill, MD, Hermine Brunner, MD, MSc, MBA: Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio.

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

Address correspondence to Katherine Schultz, MD, Division of Rheumatology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229. Email: kjschult08@gmail.com.

Submitted for publication October 22, 2019; accepted in revised form August 13, 2020.

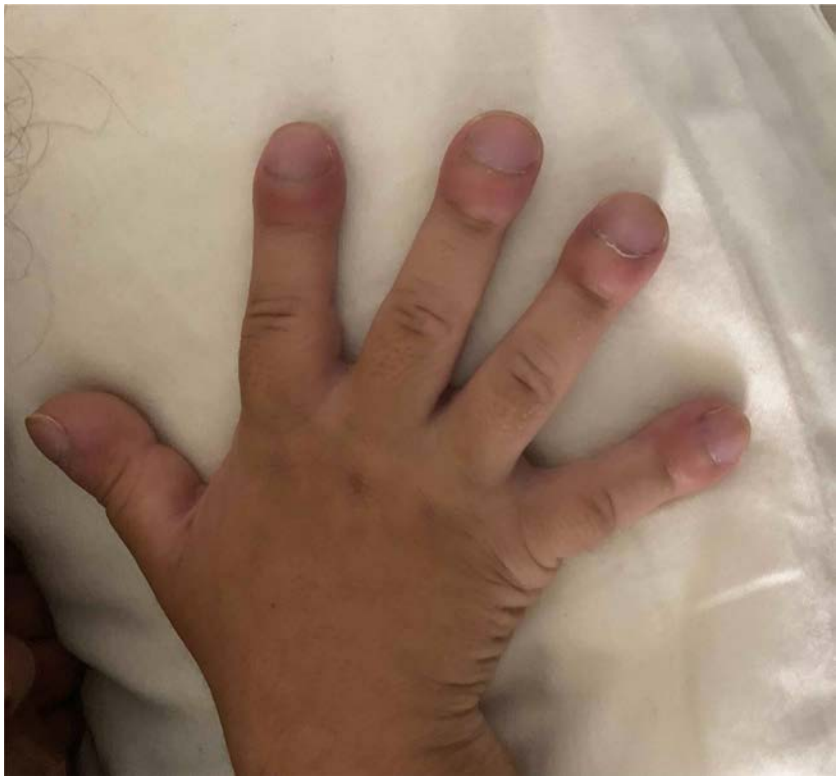


Figure 1. An example of clubbing on a child's hand, with notable redness over the distal phalanges. Photo courtesy of Dr. Grant Schulert (Cincinnati Children's Hospital Medical Center).

CASE SUMMARY

The patient is a 4-year-old female child with a family-reported history of recurrent fevers and associated abdominal pain who was found to have digital clubbing on physical examination.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of clubbing and of recurrent fevers are broad. This review will cover the differential of clubbing, particularly in the setting of recurrent fever.

The exact mechanism behind clubbing remains controversial, though several theories have been proposed (1). One such theory suggests that increased serum concentrations of growth hormone are associated with clubbing, although that has not been demonstrated in all cases of patients with clubbing. It has also been proposed that local vascular changes lead to arteriovenous anastomoses that result in changes of digital microcirculation, though firm evidence in support of this theory is lacking. Another theory proposes that clubbing is a function of platelet production, wherein the normal process of megakaryocyte breakdown into platelets inside the lungs is disrupted, allowing whole megakaryocytes to circulate. The megakaryocytes then become trapped in the finger or toe tip circulation. There, the megakaryocytes release platelet-derived growth factor, ultimately leading to the growth of vascular smooth muscle cells and fibroblasts. However, this theory does not explain the existence of clubbing in all disease entities (1). Thus, the exact pathophysiologic process of clubbing remains unknown.

Despite the pathophysiology being uncertain, clubbing is a definite sign of pathologic changes, and the differential diagnosis includes primary and secondary causes. Primary clubbing can be seen in primary hypertrophic osteoarthropathy (PHO) (2). PHO is also known as pachydermoperiostosis, a rare genetic syndrome affecting the skin and bones. It is characterized by digital clubbing, periosteal new bone formation, coarse facial features, and skin thickening. It typically arises either during the first year of life or during puberty (3). Causes of secondary clubbing are more diverse, as detailed below (Table 1). Causes of secondary clubbing with associated fever are also delineated.

Pulmonary involvement. The presence of clubbing raises concern for pulmonary involvement. This includes underlying pulmonary disease such as cystic fibrosis, pulmonary arteriovenous malformations (AVMs [1]), or causes of diffuse lung disease. Diffuse lung disease is further characterized as being either intrinsic lung disease (e.g., idiopathic interstitial pneumonia, aspiration syndromes), associated with systemic disease (e.g., connective tissue diseases, metabolic storage diseases), or found in infancy (e.g., due to inborn errors of surfactant metabolism) (4).

Our patient was Caucasian, which could raise suspicion for cystic fibrosis, a condition more common in this racial group.

Table 1. Differential diagnoses for digital clubbing*

Pulmonary
Cystic fibrosis†
Diffuse lung disease
Hypersensitivity pneumonitis†
Pulmonary arteriovascular malformation
Hepatopulmonary syndrome
Cardiac
Cyanotic heart disease
Multisystem
COPA syndrome
Sarcoidosis†
CINCA syndrome†
POEMS syndrome†
Gastrointestinal
Inflammatory bowel disease†
Liver disease
Celiac sprue
Juvenile polyposis coli
Neoplastic
Bronchogenic carcinoma†
Pleural tumors†
Lymphoma†
Nasopharyngeal carcinoma†
Mesothelioma†
Infectious disease
Tuberculosis†
Infective endocarditis†
Chronic parasite infection†
HIV†
Vascular
Venous stasis
Psychiatric
Laxative abuse

* Adapted from Spicknall et al (1). CINCA = chronic infantile neurologic cutaneous articular syndrome; POEMS = polyneuropathy, organomegaly, endocrinopathy, monoclonal gammopathy, and skin abnormalities.
† Diagnoses that can be associated with fever.

However, she lacked other signs to suggest cystic fibrosis, such as thick or sticky stools or poor growth. Results on a sweat chloride test were indeterminate, but CFTR gene sequencing did not support the presence of a pathologic mutation.

Pulmonary AVMs can result in digital clubbing (1). Such AVMs can be the result of direct communication between the pulmonary arteries and pulmonary veins. This is most often due to congenital malformations (5). Sometimes, pulmonary AVMs can be acquired, such as in hereditary hemorrhagic telangiectasia (HHT, also known as Rendu-Osler-Weber syndrome) (5,6). They may also be acquired in trauma, metastatic carcinoma, or infections (5).

Lastly, diffuse lung disease was excluded as a computed tomography (CT) scan of the chest did not show parenchymal lung disease.

Cardiac involvement. Cyanotic heart lesions have been associated with clubbing (1,7,8). For example, there is a case report of a child with unilateral finger clubbing. This child was ultimately found to have an absent aortic arch and was relying on a patent ductus arteriosus for blood supply (7). There is also evidence in the literature of patent ductus arteriosus being associated with PHO (8). While PHO is considered a primary cause of clubbing, it is worth mentioning here that clubbing as the finding of a patent ductus arteriosus should not reassure the clinician that another entity, such as PHO, is not involved. These conditions are not associated with fever, however, lowering our initial suspicion for a primary cardiac cause in this patient. Additionally, her chest radiographic was not concerning for cardiomegaly.

Rheumatic diseases. With the presentation of fevers and noted clubbing, concern arose for the possibility of systemic juvenile idiopathic arthritis (JIA) complicated by lung disease, also known as diffuse parenchymal lung disease (9,10). There is increasing awareness of lung disease in systemic JIA. Though some have suggested the involvement of uncontrolled systemic JIA activity or medication exposure as the cause of the increasing rates of systemic JIA–associated lung disease, the exact immunopathologic cause remains unknown. It was recently reported that up to 78% of patients with systemic JIA and lung disease have clubbing (10). Again, these patients are also notably symptomatic with shortness of breath, cough, and/or chest pain (10,11). Our patient did not exhibit pulmonary symptoms, but systemic JIA with interstitial lung disease (ILD) remained on our differential.

When considering periodic fevers, another rheumatic condition to consider is chronic infantile neurologic, cutaneous, articular (CINCA) syndrome, also known as neonatal-onset multisystem inflammatory disease (NOMID) (12). This is a rare autoinflammatory condition characterized by mutation in the *FCAS1* gene and is within the spectrum of other periodic fever syndromes. Historically, CINCA/NOMID has been characterized by neonatal onset (or young infancy onset) of persistent rash, optic disc changes, and progressive neurologic involvement. There have been cases of older children who have clubbing associated with this condition (12). Our patient lacked a rash, and eye or neurologic involvement, to suggest CINCA/NOMID.

Another disease with increasing recognition is COPA syndrome, an autosomal dominant autoinflammatory condition named for mutations in the *COP*α gene (13). COPA syndrome is characterized by restrictive lung disease, arthritis (14), and renal disease (13). Patients with COPA syndrome frequently have clubbing due to their underlying restrictive lung disease (14,15). COPA syndrome is a relatively new and rare disorder, and the incidence and prevalence of this condition are still uncertain. Recurrent or periodic fever does not appear to be a defining feature of COPA syndrome (14,15). This lowered our suspicion that COPA syndrome was the cause of our patient's presentation. She also did not have arthritis or renal involvement to support this diagnosis.

Interestingly, there is a case report of a child with recurrent fevers and hepatosplenomegaly who ultimately developed

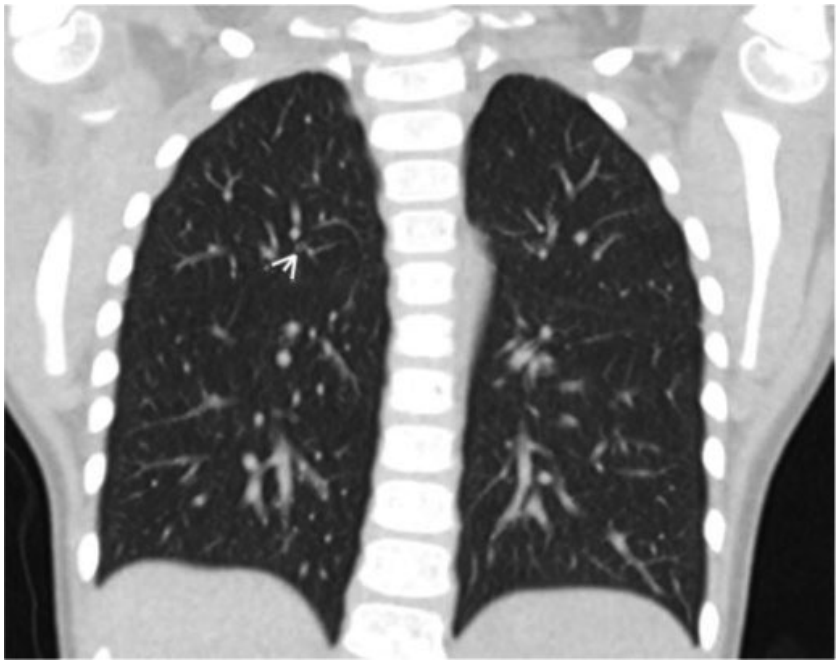


Figure 2. Computed tomography scan of the patient's chest without contrast. There was diffuse but mild nonspecific peribronchial thickening, with one such area of peribronchial thickening highlighted (**arrow**). There was no evidence of interstitial lung disease.

clubbing secondary to portal hypertension due to Takayasu arteritis (16). Given the report of fevers and the initially elevated level of alkaline phosphatase, systemic sarcoidosis was considered. Sarcoidosis is a condition characterized by necrotizing granuloma formation in multiple tissues with manifestations dependent on the organ (17). Our patient did not have lymphadenopathy, parenchymal lung disease, organomegaly, or hypercalcemia that would suggest this diagnosis, and the elevated alkaline phosphatase level normalized spontaneously, further lowering suspicion for this disease.

Gastrointestinal (GI) conditions. GI-related causes of clubbing include inflammatory bowel disease, juvenile polyposis, hepatic disorders (1,18), and celiac disease (1). Though our patient had abdominal pain following fevers, she did not have diarrhea, constipation, mucosal sores, or other findings to suggest primary GI disease. Her initially elevated alkaline phosphatase finding normalized without intervention, also lowering our suspicion for primary GI disease.

Neoplastic disease. Clubbing has been associated with neoplasms for decades, particularly with pulmonary neoplasms (19). However, for our patient, a primary pulmonary neoplasm was felt to be unlikely given her age and lack of risk factors, such as smoking or exposure to smoke. She also lacked systemic symptoms, such as night sweats, coughing, or weight loss, to suggest neoplasm. Her chest radiograph did not demonstrate a mass that would be consistent with primary pulmonary neoplasm.

Infectious disease. Infectious causes of clubbing include more indolent infections, such as tuberculosis, infective endocarditis, HIV, or chronic parasitic infection (1). She lacked risk factors for these infections, lowering suspicion that they were the cause of her symptoms.

CLINICAL COURSE

The patient's fevers resolved without intervention within months of her initial consult visit with rheumatology. It was later discovered she had had dental procedures completed prior to her episodes. It was thought that perhaps some bacterial seeding had contributed to her fevers.

The patient was referred to pulmonology for further evaluation of clubbing noted in the setting of peribronchial thickening seen on chest radiograph. Pulmonology evaluation revealed that pulse oximetry was 97% on room air. She had a normal 6-minute walk test. Chest CT also demonstrated peribronchial thickening but no other abnormality (Figure 2). As a result, a contrast echocardiogram was performed. The echocardiogram showed normal cardiac anatomy and function. It also demonstrated 20–30 bubbles in the left atrium within 4 beats following injection of bubbles into the right heart. This was felt to be concerning for pulmonary arteriovenous malformation (AVM) (Figure 3). She was then referred to cardiology and gastroenterology and evaluated by our GI team due to an elevated level of alkaline phosphatase, though this finding normalized without intervention, and the cause of the original elevation was not determined. Results from a Doppler ultrasound of the patient's liver were normal.

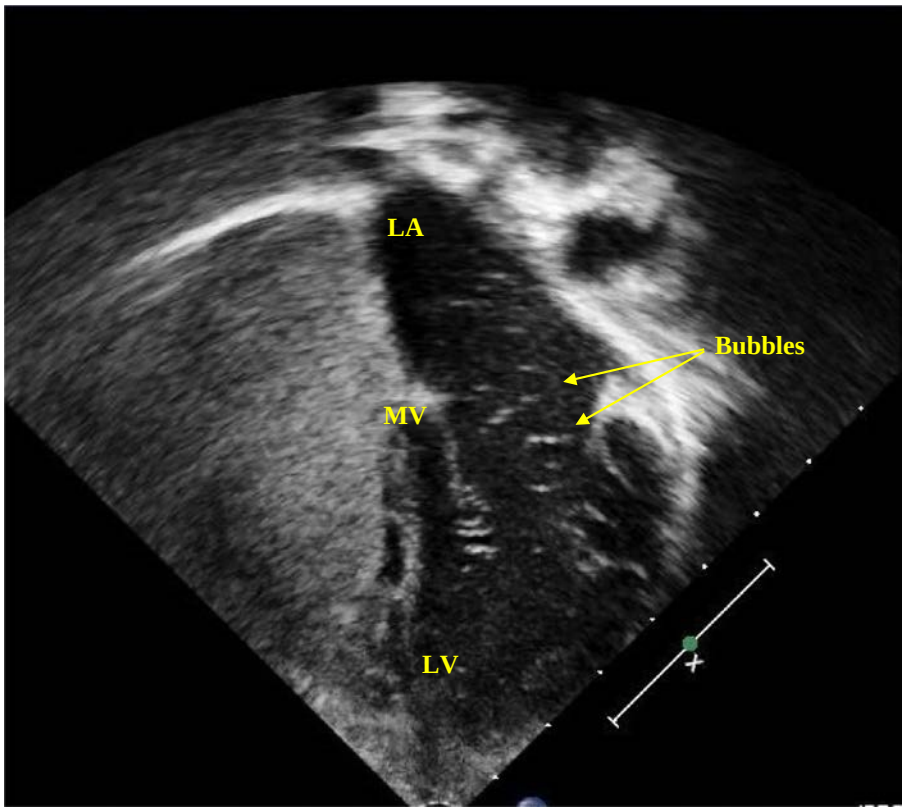


Figure 3. Echocardiogram with contrast demonstrating 20–30 bubbles in the left atrium (**arrows**) within 4 heart beats following injection of bubbles into the right side of the heart. Cardiac anatomy and function were normal. LA = left atrium; LV = left ventricle; MV = mitral valve. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24428/abstract.

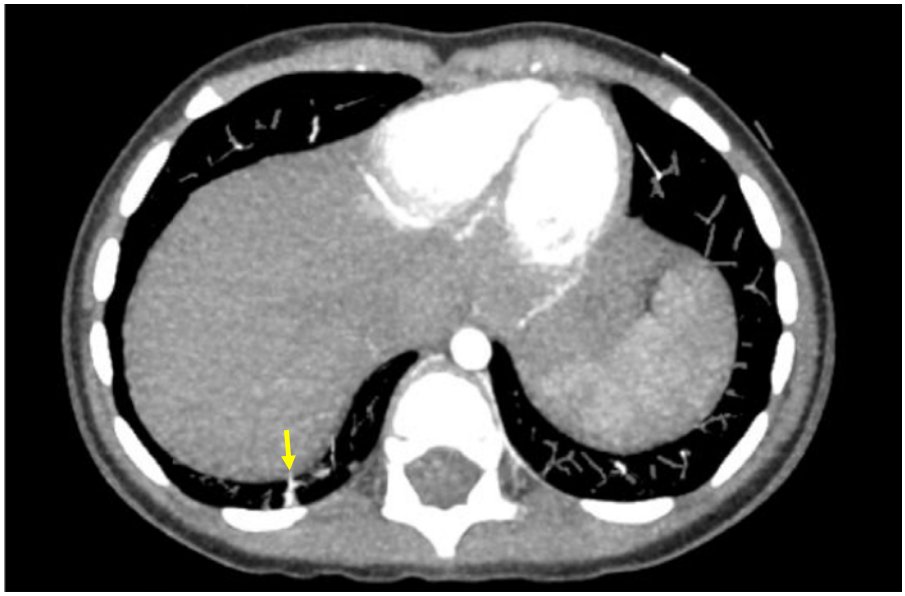


Figure 4. Computed tomography angiogram of the patient's chest in the axial plane. There was evidence of multiple small pulmonary arteriovenous malformations. These are found at the posterior aspects of the bilateral lower lung lobes. They are characterized by enhancement, representing feeding arteries and draining veins. The largest pulmonary arteriovenous malformation is highlighted (**arrow**). Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24428/abstract.

Table 2. Curaçao Criteria for diagnosing hemorrhagic hereditary telangiectasia*

Epistaxis
Spontaneous, recurrent nose bleeds
Telangiectasias
Multiple
At characteristic sites: lips, oral cavity, fingers, and nose
Visceral lesions
Gastrointestinal telangiectasia, pulmonary AVM, hepatic AVM, cerebral AVM, and spinal AVM
Family history
First-degree relative with HHT

* Diagnosis is definite if 3 criteria are present, possible or suspected if 2 criteria are present, or unlikely if fewer than 2 criteria are present. Adapted from Shovlin et al (6). AVM = arteriovenous malformation; HHT = hereditary hemorrhagic telangiectasia.

Eventually she was referred to the Cincinnati HHT Center of Excellence for evaluation and management of HHT. It was then noted at the Center that she had 3 skin telangiectasias, found on the right wrist, right cheek, and left wrist. A subsequent CT angiogram of the chest demonstrated multiple small pulmonary AVMs at the lung bases (Figure 4). A magnetic resonance imaging (MRI) screen of the brain was normal. Genetic testing for HHT—endoglin gene (*ENG*), activin A receptor type II–like 1 gene (*ACVRL1*), mothers against DPP homolog 4 (*SMAD4*), growth differentiation factor 2 (*GDF2*) (20), ephrin type-B receptor 4 (*EPHB4*) (21), and Ras p21 protein activator (*RASA1*) (22)—was completed by a laboratory and returned with negative results for these genes. There was no family history consistent with HHT. Mother and father denied any history of recurrent nosebleeds. While most cases of HHT are autosomal dominant as mentioned above, there are de novo cases reported (23). There is also an increasing amount of data supporting that the first patient with HHT in a family may be mosaic, and therefore show a negative result on blood testing for HHT (24,25). The patient did ultimately meet clinical diagnostic criteria for possible HHT under the Curaçao Criteria (6), the recommended criteria used for clinical diagnosis of HHT (Table 2).

The patient is currently asymptomatic and being monitored in the Cincinnati HHT Center of Excellence every 6 months. This Center is a multidisciplinary clinic, with the core specialties involved including a hematologist, a genetic counselor, interventional radiology, and an ear, nose, and throat surgeon. Our Center also includes a pulmonologist, a cardiologist, a gastroenterologist, a neurologist, a neuroradiologist, a neurosurgeon, and a plastic surgeon (26). As part of her monitoring, she will require several clinical tests per international guidelines (27). She will need a repeat brain MRI upon entering adulthood, as brain AVMs can grow with puberty. She will require repeat bubble echocardiograms every 2–3 years to monitor the state of her existing pulmonary AVMs and evaluate for anticipated complications, such as the risk of embolic and/or infectious events. She will require testing of annual hemoglobin levels starting at age 35 years due to risk for internal bleeding from AVMs (especially GI AVMs), as well as annual liver screening with liver function testing. She will also need routine colonoscopy at age 50 years, or sooner if she has evidence of GI bleed (27). Finally, she will also require antibiotic prophylaxis due to the risk for infectious emboli secondary to her existing pulmonary AVMs.

DISCUSSION

Hereditary hemorrhagic telangiectasia is an autosomal dominant disease characterized by vascular dysplasia (28). Mutations in several genes—*ENG, ACVRL1, MADH4, SMAD, GDF2, RASA1* (20), *EPHB4* (21), *RASA1* (22)—have been associated with HHT, all of which affect proteins in the transforming growth factor β (TGFβ) superfamily (20,28). Unfortunately, a subset of patients and families (15–20%) with HHT do not have an identifiable mutation known to cause HHT (28). The majority of patients who do have a known genetic mutation will have one in the *ENG* or *ACVRL1* genes (28).

The vascular abnormalities that characterize HHT can occur in any organ, with large vessel involvement affecting the liver, lung, and brain (in descending order) and small vessel involvement affecting the nasal mucosa, GI tract, and skin (29). Patients classically present with a triad of epistaxis, telangiectasias, and a family history of HHT (6). However, in reality, patients can present in a myriad of ways, none of which are pathognomonic for the disease (29). Additionally, diagnosis in childhood is complicated because the Curaçao Criteria (Table 2) have been shown to have poor sensitivity and specificity in children who are less than 15 years old (30). More concerning, patients with vascular malformations may remain asymptomatic until the development of a catastrophic event urges them to receive medical attention (29).

Our patient only had a few scattered telangiectasias. These were noted only after very careful examination and imaging suggestive of telangiectasias had been obtained. The most important diagnostic clue in evaluating this patient was her presentation with digital clubbing, which was due to her pulmonary AVMs (29).

Pulmonary AVMs are quite common in HHT, affecting up to 50% of patients throughout their lifetime (28). Moreover, of all cases of pulmonary AVMs, 80–90% are due to an underlying diagnosis of HHT (28). The data are sparse on sporadic pulmonary AVMs, and according to our experts, occurrence of pulmonary AVMs outside of HHT is rare unless it occurs with severe pulmonary, liver, or cardiac disease, which our patient lacked.

The diagnosis of pulmonary AVMs is of particular importance. The direct connection between arterial and venous blood supply leaves the patient vulnerable to paradoxical emboli that can cause cerebral abscess or stroke (28). Treatment is aimed at embolization of the malformation (28). Other

risks associated with pulmonary AVMs include rupture, leading to massive blood loss (28). Therefore, any patient suspected of having HHT should be screened with a bubble contrast echocardiogram (31).

Hepatic AVMs are even more common in HHT, with up to 75% of patients being affected, though not all patients experience symptoms (31). When patients do experience symptoms, most present with heart failure, portal hypertension, or biliary disease. These symptoms can be managed medically (31), though liver transplantation is favored in Europe and has demonstrated good outcomes (32). Until recently, initial or routine screening for hepatic AVMs if asymptomatic (lacking a bruit and with normal transaminases) was not routinely recommended in HHT patients (31). Our institution recommends checking liver function tests, including gamma glutamyl transferase and brain natriuretic peptide, annually.

For rheumatologists, HHT is important as it can mimic the skin findings of limited systemic sclerosis (SSc) (formerly the CREST variant of scleroderma). Limited SSc is characterized by calcinosis, Raynaud's phenomenon, esophageal involvement, sclerodactyly, and telangiectasia. This was not in our initial diagnosis but is worth mentioning here as skin findings can precede other findings by years (33). It has been suggested that anticentromere antibodies can distinguish limited SSc from HHT (33). However, there are case reports where this is not the circumstance and distinguishing between limited SSc and HHT may present a diagnostic dilemma (34).

There have been reports of other diseases associated with HHT. An example is juvenile polyposis HHT (JP-HHT) (31,35), an autosomal dominant disorder characterized by gastrointestinal polyps and increased risk for GI cancer (35). These patients require further screening with colonoscopy and endoscopy every 1–2 years beginning with first symptom or at age 12 years, whichever comes first (31).

Interestingly, patients with JP-HHT have a known genetic variation in the *SMAD4* gene (35,36). *SMAD4* encodes a protein in the TGFβ pathway (36). This is relevant, as there is also a case report of a patient with systemic JIA and JP-HHT with a known variation in the *SMAD4* gene (36). Though our patient ultimately did not have persistent fevers or blood work supporting a diagnosis of systemic JIA, this association may be important for rheumatologists who have patients presenting with fevers, clubbing, and findings of telangiectasias.

FINAL DIAGNOSIS

Hereditary hemorrhagic telangiectasia.

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. Schultz 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. Schultz, Divanovic, Towe, Miethke, Wusik, Hammill, Brunner.

Acquisition of data. Schultz, Divanovic, Towe, Miethke, Wusik, Hammill, Brunner.

Analysis and interpretation of data. Schultz, Divanovic, Towe, Miethke, Wusik, Hammill, Brunner.

REFERENCES

1. Spicknall KE, Zirwas MJ, English JC III. Clubbing: an update on diagnosis, differential diagnosis, pathophysiology, and clinical relevance. J Am Acad Dermatol 2005;52:1020–8.
2. Jyotsna M, Tharakan J. Clinical Sign: clubbing. Indian J Cardiovasc Dis Women WINCARS 2017;02:e1–9.
3. Zhang Z, Zhang C, Zhang Z. Primary hypertrophic osteoarthropathy: an update. Front Med China 2013;7:60–4.
4. Vece TJ, Fan LL. Diagnosis and management of diffuse lung disease in children. Paediatr Respir Rev 2011;12:238–42.
5. Khurshid I, Downie GH. Pulmonary arteriovenous malformation. Postgrad Med J 2002;78:191–7.
6. Shovlin CL, Guttmacher AE, Buscarini E, Faughnan ME, Hyland RH, Westermann CJ, et al. Diagnostic criteria for hereditary hemorrhagic telangiectasia (Rendu-Osler-Weber syndrome). Am J Med Genet 2000;91:66–7.
7. Dorney ER, Fowler NO, Mannix EP. Unilateral clubbing of the fingers due to absence of the aortic arch. Am J Med 1955;18:150–4.
8. Levin SE, Harrisberg JR, Govendrageloo K. Familial primary hypertrophic osteoarthropathy in association with congenital cardiac disease. Cardiol Young 2002;12:304–7.
9. Van Manen MJ, Vermeer LC, Moor CC, Vrijenhoeff R, Grutters JC, Veltkamp M, et al. Clubbing in patients with fibrotic interstitial lung diseases. Respir Med 2017;132:226–31.
10. Schulert GS, Yasin S, Carey B, Chalk C, Do T, Schapiro AH, et al. Systemic juvenile idiopathic arthritis–associated lung disease: characterization and risk factors. Arthritis Rheumatol 2019;71:1943–54.
11. Kimura Y, Weiss JE, Haroldson KL, Lee T, Punaro M, Oliveira S, et al. Pulmonary hypertension and other potentially fatal pulmonary complications in systemic juvenile idiopathic arthritis. Arthritis Care Res (Hoboken) 2013;65:745–52.
12. Indian Pediatrics. CINCA syndrome. URL: http://www.indianpediatrics.net/dec2007/dec-933-936.htm.
13. Vece TJ, Watkin LB, Nicholas SK, Canter D, Braun MC, Guillerman RP, et al. Copa syndrome: a novel autosomal dominant immune dysregulatory disease. J Clin Immunol 2016;36:377–87.
14. Noorelahi R, Perez G, Otero HJ. Imaging findings of Copa syndrome in a 12-year-old boy. Pediatr Radiol 2018;48:279–82.
15. Brennan M, McDougall C, Walsh J, Crow Y, Davidson J. Copa mutation—a new condition to consider with polyarthritis and interstitial lung disease [case report G426]. Arc Dis Child 2017;102: A167–8.
16. Herrera CN, Tomala-Haz JE. Portal hypertension: an uncommon clinical manifestation of Takayasu arteritis in a 9-year-old child. Open Access Rheumatol 2016;8:115–8.
17. Yancey J, Luxford W, Sharma OP. Clubbing of the fingers in sarcoidosis. JAMA 1972;222:582.
18. McPhee SJ. Clubbing. In: Clinical methods: the history, physical, and laboratory examinations. Boston: Butterworths; 1990. URL: http://www.ncbi.nlm.nih.gov/pubmed/21250207.
19. Craig JW. Hypertrophic pulmonary osteo-arthropathy as first symptom of neoplasm. Br Med J 1937;1:750–2.

20. McDonald J, Wooderchak-Donahue W, VanSant Webb C, Whitehead K, Stevenson DA, Bayrak-Toydemir P. Hereditary hemorrhagic telangiectasia: genetics and molecular diagnostics in a new era. Front Genet 2015;6:1.
21. Wooderchak-Donahue WL, Akay G, Whitehead K, Briggs E, Stevenson DA, O'Fallon B, et al. Phenotype of CM-AVM2 caused by variants in EPHB4: how much overlap with hereditary hemorrhagic telangiectasia (HHT)? Genet Med 2019;21:2007–14.
22. Wooderchak-Donahue W, Stevenson DA, McDonald J, Grimmer JF, Gedge F, Bayrak-Toydemir P. RASA1 analysis: Clinical and molecular findings in a series of consecutive cases. Eur J Med Genet 2012;55:91–5.
23. Gedge F, McDonald J, Phansalkar A, Chou LS, Calderon F, Mao R, Lyon E, et al. Clinical and analytical sensitivities in hereditary hemorrhagic telangiectasia testing and a report of de novo mutations. J Mol Diagn 2007;9:258–65.
24. Tørring PM, Kjeldsen AD, Ousager LB, Brusgaard K. ENG mutational mosaicism in a family with hereditary hemorrhagic telangiectasia. Mol Genet Genomic Med 2018;6:121–5.
25. McDonald J, Wooderchak-Donahue WL, Henderson K, Paul E, Morris A, Bayrak-Toydemir P. Tissue-specific mosaicism in hereditary hemorrhagic telangiectasia: implications for genetic testing in families. Am J Med Genet A 2018;176:1618–21.
26. Cincinnati Children's Hospital Medical Center. HHT Center of Excellence. URL: https://directory.curehht.org/hospital/cincinnati-childrens-hospital-medical-center.
27. CureHHT. HHT International Clinical Guidelines. URL: https://curehht.org/resource/hht-international-clinical-guidelines/.
28. Kühnel T, Wirsching K, Wohlgemuth W, Chavan A, Evert K, Vielsmeier V. Hereditary hemorrhagic telangiectasia. Otolaryngol Clin North Am 2018;51:237–54.
29. Sabba C, Gallitelli M, Pasculli G, Suppressa P, Resta F, Tafaro E. HHT: A rare disease with a broad spectrum of clinical aspects. Curr Pharm Des 2006;12:1217–20.
30. Pahl KS, Choudhury A, Wusik K, et al. Applicability of the Curaçao Criteria for the diagnosis of hereditary hemorrhagic telangiectasia in the pediatric population. J Pediatr 2018;197:207–13.
31. Kroon S, Snijder RJ, Faughnan ME, Mager HJ. Systematic screening in hereditary hemorrhagic telangiectasia. Curr Opin Pulm Med 2018;24:260–8.
32. Lerut J, Orlando G, Adam R, et al. Liver transplantation for hereditary hemorrhagic telangiectasia: report of the European liver transplant registry. Ann Surg 2006;244:854–64.
33. Fritzler MJ, Arlette JP, Behm AR, Kinsella TD. Hereditary hemorrhagic telangiectasia versus CREST syndrome: can serology aid diagnosis? J Am Acad Dermatol 1984;10:192–6.
34. Lee JB, Ben-Aviv D, Covello SP. The diagnostic quandary of hereditary haemorrhagic telangiectasia vs. CREST syndrome. Br J Dermatol 2001;145:646–9.
35. Gallione C, Aylsworth AS, Beis J, Berk T, Bernhardt B, Clark RD, et al. Overlapping spectra of SMAD4 mutations in juvenile polyposis (JP) and JP-HHT syndrome. Am J Med Genet 2010;152A:333–9.
36. Bishop J, Britton J, Murphy A, et al. Juvenile idiopathic arthritis associated with combined JP–HHT Syndrome: a novel phenotype associated with a novel variant in SMAD4. J Pediatr Genet 2018;07:78–82.

Arthritis Care & Research
Vol. 73, No. 10, October 2021, pp 1387–1395
DOI 10.1002/acr.24635

American College of Rheumatology
Empowering Rheumatology Professionals

Using Critical Race Theory to Understand Trial Participation Among Black Individuals With Systemic Lupus Erythematosus: A Qualitative Study of Patients and Caregivers

Rodlescia S. Sneed,[1] Maryann Mason,[2] Jessica N. Williams,[3] Corine Sinnette,[3] Kreager Taber,[3] Karen Mancera-Cuevas,[4] Gina Curry,[5] Patricia Canessa,[6] Rosalind Ramsey-Goldman,[2] and Candace H. Feldman[3]

Objective. Black patients with systemic lupus erythematous (SLE) experience greater disease incidence and severity than White patients, and yet they are underrepresented in SLE clinical trials. We applied Critical Race Theory to qualitatively explore the influence of racism on the underrepresentation of Black patients in SLE clinical trials and to develop a framework for future intervention.

Methods. We conducted focus group sessions in Chicago and Boston with Black adults (ages ≥18 years) with SLE and their caregivers. We queried the participants about their knowledge regarding clinical trials, factors that might motivate or hinder trial participation, and how race and experiences of racism might impact clinical trial participation. Focus group responses were transcribed verbatim and analyzed thematically.

Results. We held 4 focus groups (n = 31 participants); 20 participants had SLE, and 11 were caregivers. All participants were Black, 90% were women, and the mean age was 54 years. Qualitative analyses revealed several themes that negatively impact trial participation, including mistrust related to racism, concerns about assignment to placebo groups, strict study exclusion criteria, and SLE-related concerns. Factors that motivated trial participation included recommendations from physicians and reputable institutions, a desire to help the greater good, and culturally sensitive marketing of trials.

Conclusion. Actions to improve clinical trial participation among Black individuals should focus on reframing how trial information is presented and disseminated and on reevaluating barriers that may restrict trial participation. Additionally, researchers must acknowledge and respond to the presence of racial bias in health care. Community–academic partnerships may help build trust and reduce fears of mistreatment among Black individuals with SLE.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a complex, multisystem autoimmune disease with significant racial disparities in incidence, prevalence, and severity. Black individuals are 2–4 times more likely to develop SLE than White individuals (1) and have earlier and more abrupt disease onset (2), greater risk of lupus nephritis (2,3), and higher disease-related mortality (4) than White individuals.

Despite high disease burden (making up 43% of SLE cases), Black individuals are underrepresented in SLE clinical trials, comprising only 14% of trial participants (5). Failure to represent diverse populations in clinical research violates ethical principles of distributive justice, which require that burdens and benefits of research be equitably distributed across racial, ethnic, sex, and social groups (6). Additionally, there may be differences in the safety and efficacy of new medications based on genetic background (7). Although race is a social construct, there is evidence of genetic

The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the Department of Health and Human Services.

Supported by the Department of Health and Human Services (Office for Minority Health grant 1 CPIMP181168-01-00).

[1]Rodlescia S. Sneed, PhD, MPH: Michigan State University, Flint; [2]Maryann Mason, PhD, Rosalind Ramsey-Goldman, MD, DrPH: Feinberg School of Medicine, Northwestern University, Chicago, Illinois; [3]Jessica N. Williams, MD, MPH, Corine Sinnette, MA, MPH, Kreager Taber, BA, Candace H. Feldman, MD, MPH, ScD: Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; [4]Karen Mancera-Cuevas, DrPH, MS, MPH, CHES: Illinois Department of Public Health, Springfield, and Feinberg School of Medicine, Northwestern University, Chicago, Illinois; [5]Gina Curry, MPH, MBA: University of Chicago, Chicago, Illinois; [6]Patricia Canessa, PhD: Illinois Public Health Association, Springfield.

Drs. Ramsey-Goldman and Feldman contributed equally to this work.

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

Address correspondence to Rodlescia S. Sneed, PhD, MPH, Division of Public Health, College of Human Medicine, Michigan State University, 200 East 1st Street, Flint, MI 48502. Email: sneedrod@msu.edu.

Submitted for publication August 27, 2020; accepted in revised form April 22, 2021.

SIGNIFICANCE & INNOVATIONS

- We applied the Public Health Critical Race Praxis, an adaptation of Critical Race Theory, to explicitly explore the role of both historical injustices and ongoing structural and interpersonal racism on underrepresentation of Black individuals in systemic lupus erythematosus (SLE) clinical trials.
- Through focus groups in 2 US cities with Black patients with SLE and their caregivers, we documented the pervasive role that racism and distrust play in the decision to enroll in trials.
- We identified factors that can reduce barriers to enrollment, including community engagement, strong provider–patient relationships and trust building, and culturally sensitive study documents.

variability both among Black individuals and between Black and White individuals (8). Without the inclusion of Black patients in clinical trials, research findings related to treatment safety and efficacy will not be generalizable.

We conducted a systematic review to identify factors impacting the participation of underrepresented groups in rheumatic disease research (9). We found that no studies in the rheumatic disease population discussed the role of racism, despite historical exploitation of Black individuals in research studies and ongoing structural, institutional, and interpersonal racism in the US. To fill this gap, we used Critical Race Theory as a framework to understand clinical trial decision-making among Black SLE patients and their caregivers (10,11). We applied the Public Health Critical Race Praxis (PHCRP) (11), which adapts Critical Race Theory to public health to understand how exposure to racism may explain, in part, the underrepresentation of Black individuals in SLE clinical trials. The 4 main components of the PHCRP are as follows: 1) understanding racism as it currently exists within society ("contemporary racial relations"), 2) understanding how preexisting beliefs about racial groups may shape a project or how the project may reinforce those beliefs ("production knowledge"), 3) recognizing race as a social construct that may intersect with other sources of systemic oppression ("conceptualization and measurement"), and 4) use of the knowledge obtained from research to help stop root causes of inequity ("action") (12-15). We used this framework to identify motivating factors, barriers, and mediators to trial participation.

PATIENTS AND METHODS

Application of the PHCRP. We integrated the 4 components of the PHCRP directly into the research process. We framed our research questions by operationalizing race as a social rather than biologic construct and acknowledged that clinical trials have often functioned as a structural mechanism where racism/racial hierarchies have operated and been perpetuated. A racially diverse, multidisciplinary study team developed the focus group moderator guide, drawing on existing literature and the PHCRP framework. To incorporate diverse perspectives among participants, we held focus groups in 2 racially, ethnically, and socioeconomically diverse cities.

Participants and data collection. We conducted focus groups in Chicago and Boston with Black individuals with SLE ages ≥18 years and their caregivers, who participated in medical decision-making. We included both caregivers and SLE caregivers in our focus groups, as several studies describe the value of including caregiver–patient dyadic units in research designed to improve patient outcomes (16,17). Participants were identified through community-based networks and organizations in primarily Black neighborhoods, SLE support groups, local and national associations, and hospital clinics. This study was approved by the institutional review boards of Northwestern Feinberg School of Medicine (Chicago) and Mass General Brigham (Boston).

Focus groups in both cities were led by experienced Black moderators, who followed the same written guide (see Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24635/abstract). The sessions were ~90 minutes, were recorded and transcribed verbatim, and detailed notes were taken. Consistent with the PHCRP, participants were asked about how race and experiences of racial discrimination might impact trial participation, ways to increase Black participation in clinical trials, and their attitudes about clinical trials designed to recruit Black patients.

Analyses. Focus group transcripts were analyzed using the constant comparative method (18), an inductive data-coding process used for categorizing and comparing qualitative data. After transcription, 3 researchers (JNW, CS, CHF) reviewed the moderator guide and transcripts to mutually develop a coding system that reflected the overarching themes, subthemes, and proposed actions gleaned from the focus groups. Initial codes were determined, after reviewing the transcripts, based on the research question of interest and guided by the PHCRP. Specific attention was paid to terms relevant to experiences of discrimination, stigma, historical factors, distrust and structural racism, and the ways in which these factors permeated identified themes and subthemes. One researcher (MM) additionally analyzed transcripts with a specific focus on motivators, barriers, and mediators of trial participation and developed corresponding codes, which were then reviewed, adjudicated, and incorporated by the full study team (JNW, CS, RRG, CHF). We used an iterative approach to develop a standard coding framework, which was applied to all text from the focus group transcripts. Two coders (MM, CHF) independently coded the focus group transcripts; codes were entered into ATLAS.ti (MM). Coding results were then compared for consistency. The team met multiple times to review themes

and codes in order to ensure consistency and agreement. Focus groups were conducted until we determined that thematic saturation had been achieved (19). Based on our previous work among Black patients with SLE in Boston and Chicago (20), our a priori hypothesis was that there would be differences in attitude, beliefs, and experiences related to clinical trials between participants in the 2 cities. Thus, we stratified our analyses by city.

RESULTS

We conducted 4 focus group sessions in Boston and Chicago in 2019. The 2 focus groups in Boston included 6 and 9 participants, respectively. The 2 focus groups in Chicago each included 8 participants. All participants identified as Black. Sixty-five percent of participants were diagnosed with SLE (n = 20); 11 were caregivers. Twelve percent of participants had a high school education, 23% had attended some college/technical school, and 65% had at least a college degree. Fifty-seven percent of participants were employed. The mean age of all participants was 54 years (range 29–75 years), and 90% of participants were female. The mean age of SLE patients was 51.7 years (range 29–71 years), and the mean age of SLE caregivers was 59.6 years (range 34–75 years). There were notable demographic differences between participants from the 2 cities. Seventy-three percent of Boston participants had previously participated in research, compared to only 43% of Chicago participants. Boston participants had longer disease duration (mean 24.4 years) than Chicago participants (mean 15.4 years), and 93% of Chicago participants were active in a church/faith community compared to 33% of Boston participants. Additionally, 87% of Chicago participants were college educated, compared to 40% of Boston participants (see Supplementary Tables 1 and 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24635/abstract). Thematic analysis of transcripts resulted in 6 key themes, including skepticism of clinical trials/research (subthemes of trust, racism and historical context, and placebo arms), the greater good, SLE-specific factors, physician/institutional influences, influence of social network members, and trial/research-specific factors (subthemes of exclusion criteria, information presentation/marketing of studies).

Skepticism of clinical trials and research. *Trust, racism, and historical context.* A key component of critical race theory is characterizing how racism currently operates within society. We wanted to understand how racism might be perpetuated within health care and research settings; therefore, we explored whether racism impacted the way in which our focus group participants experience research and clinical trial participation.

Some participants described general mistrust of clinical trials and medical practice due to historical racism in health care and research. They expressed mistrust of research stemming from

Table 1. Quotations from participants expressing concern about trust, racism, historical context, and placebo arms

Trust, racism, and historical context
But there is a deep mistrust with the African American community. First thing a lot of us think of is the Tuskegee Experiment…So there's a deep distrust between the medical community and the African American community. And that's definitely a barrier, even if you are trusting and see the benefits. It still lingers in your thought process. (Chicago focus group 1, quotation A)
We have had so much racism,… We're thinking of all the things that have happened to us throughout the years. …we're gonna say, oh, no. You all want to do all that disease stuff and testing. No. We know what our ancestors went through. (Boston focus group 1, quotation B)
There's that trust factor because maybe people of my generation, they're familiar with that syphilis study… where they were supposedly treating these people and they were not. (Chicago focus group 2, quotation C)
People, the higher levels of education, the more open they are to research, new ideas, participating in studies, helping with that… They can understand. Whereas people with lower levels, they tend to be more involved with the old belief system. (Chicago focus group 1, quotation D)
One of the things we didn't really talk about is, there's a divide in the socioeconomic groups. I work in a relatively upper middle class, and you see they take full advantage of all the medical benefits and everything… I come home, I talk to my people, and they have no idea. (Chicago focus Group 1, quotation E)
A lot of minorities, especially black families, 'cause we don't go to the hospital, we wait until the very last minute, so when we do go to the hospital, we die. "Don't go to the hospital. They kill you. They're killing you." I think we're not educating ourselves to learn more about your symptoms, when to go to the hospital… we are afraid to go to the hospital. (Boston focus group 1, quotation F)
Sometimes when you have an African American woman like myself come in and complain of pain, it's not always taken to the level of severity if my Caucasian counterpart came in and complained of pain. They don't make that diagnosis as quick. That's scientifically proven. (Chicago focus group 2, quotation G)
I kept saying, "I thought this was anonymous"… it's not anonymous 'cause I notice every time a person of color comes in there, they go and mark that paper a certain way, so don't tell me this is anonymous….If you're differentiating the Blacks and the Whites, just tell me that. I'm still gonna do it, but be honest with me. Don't say, "Oh, no. It's anonymous." No, it isn't. You're putting a number down there every time you give out one of these things to somebody of color, and I saw the whole thing go down. I said, "This is not right. Not right." (Boston focus Group 1, quotation H)
A lot of times, when you hear that statement, "We're looking for Black people,".....our antennas go up. Is this racist or something like that? (Boston focus group 1, quotation I)
Placebo arms
It wouldn't bother me because that's the only way you're gonna be able to evaluate. (Boston focus group 1, quotation J)
I'm always afraid to participate because the drug interactions. If you don't know whether you're gonna get a placebo. Placebo's fine, but if it's one of the drugs, it may interact negatively with some of the medicines because I take a lot of medications. (Boston focus group 1, quotation K)
It would bother me if I had to participate for months or years and found out afterwards that I was taking a placebo. (Boston focus group 1, quotation L)
If I finally in my head say, I'll go ahead and give it a shot….I would be teed off if I was in there, and I stayed there 6 months in your little study or trial thing, and I was getting the placebo for 6 months. (Boston focus group 1, quotation M)

(Continued)

Table 1. (*Cont'd*)

That's kind of one of those things where I'm just like, I don't want to mess around and not get treatment. That's probably why my family is very adamant against clinical trials, because they don't trust the process. (Chicago focus group 1, quotation N)
Yes. I would participate in a study. I might feel a little sad if I found out I was the one that got the placebo, but I would still participate because I do want to help others. (Boston focus group 1, quotation O)
It comforts me to know that it wouldn't just be a placebo, it would be the current medication that is being used. (Chicago focus group 2, quotation P)

their knowledge of wrongdoings that occurred during the Tuskegee Syphilis Study (21). For these participants, the legacy of Tuskegee persisted despite the perceived benefits of clinical research (Table 1, quotations A, B, and C).

Two Chicago focus group participants expressed that research skepticism would be most common for older people, who had more familiarity with the Tuskegee Syphilis Study and more negative past experiences with racism (Table 1, quotation C). One participant suggested that more educated Black individuals might be less skeptical about research, implying that they may have more knowledge of the health care system and would be more likely to take advantage of system benefits (Table 1, quotation D).

Several participants highlighted other historical instances of racism in medicine, including forced sterilization of Black women (22) and the story of Henrietta Lacks (23). Both were described as having long-term impact on health care utilization by Black individuals and as reasons to be skeptical of research. Participants were divided about whether these historical situations could occur again. Some thought that there were now significant safeguards to protect research participants from exploitation; others were still skeptical and were therefore hesitant to participate. An overall cultural mistrust of health care institutions, leading to less utilization of health care services and poorer health outcomes, was highlighted as a consequence of racism in research (Table 1, quotation F).

Participants also described their current experiences with racism. One participant noted a lack of cultural competency by physicians, while another reported a history of racial bias in assessment and management of her chronic pain (Table 1, quotation G). One participant described an experience where people of color were inappropriately tracked in a research study that was supposed to be anonymous (Table 1, quotation H). Another participant noted skepticism of trials that were only open to Black participants, expressing concern that researchers might have racist motives (Table 1, quotation I). Others, despite being skeptical, felt that new therapies needed to be tested among Black individuals to demonstrate their usefulness for the Black population.

Placebo arms. Focus group moderators explained randomized trials, specifically random assignment to treatment or placebo (control) groups. Reactions to placebo groups were mixed. Some participants recognized that placebo arms were necessary for evaluating treatment effectiveness (Table 1, quotation J). Others preferred to get a placebo over a new treatment, as they thought that active treatments would interfere with their current medication regimen (Table 1, quotation K). Several participants said they would be disappointed to get assigned to a placebo group, as they wanted the potential benefits of active treatment (Table 1, quotations L–N). One participant even stated that the risk of getting a placebo would discourage her from participating altogether. Despite possibly being assigned to a placebo group, some participants said that they would still participate in order to help others (Table 1, quotation O). There was less participant resistance to clinical trials when they compared the current standard treatment to a new treatment (Table 1, quotation P).

The greater good. Several participants felt that clinical trial participation was of little personal benefit to them. They thought participation would advance science (Table 2, quotation A) or help others with SLE (Table 2, quotations B–F), including younger people and family members. A few participants specifically wanted to help other Black patients with SLE (Table 2, quotation G). Several participants thought trial participation was crucial for addressing SLE racial disparities (Table 2, quotation H).

SLE-specific factors. Some participants were hesitant to participate in trials because of concerns that trial drugs would interact negatively with their current medications or that new medications might be metabolized in organ systems already impacted by disease (e.g., liver, kidneys). Others were afraid of adding another drug to their already complicated medication regimen (Table 3, quotation A). Some participants described great difficulty in receiving their initial SLE diagnosis and in identifying successful treatments; consequently, they were not interested in receiving experimental medications that might negatively impact their currently stable disease status (Table 3, quotations B and C).

A few participants feared being treated like a guinea pig as a result of trial participation, or that living with SLE felt like being a part of a large experiment due to frequent clinical visits and specimen collections (Table 3, quotations D and E). Several participants felt that clinical trials added an extra burden to their already burdensome clinical care experience.

Physician and institutional influences. Most participants were reluctant to enroll in trials without getting approval from both their primary care physicians and the doctors who managed their SLE (e.g., rheumatologists or nephrologists) (Table 4, quotation A). Most participants did not care about being approached for a clinical trial by a physician of the same race or sex (Table 4, quotation B); participants were most concerned about being approached by physicians who were credible, trustworthy, good listeners, and culturally competent (Table 4, quotations C–E). Two people had participated in trials before based on their rheumatologist's

Table 2. Quotations from participants expressing ideas about helping the greater good

I know...that to advance in medicine, to advance, that scientifically, research must occur. And I know that there's limits on what researchers can do on nonhumans. We're only going to get so far. (Chicago focus group 2, quotation A)
I've got to keep in mind, this is not for you. This is for somebody else... So you have to go into it with the right spirit, that being, I'm willing to participate so the information can be gathered that could be helpful for somebody else." (Chicago focus group 2, quotation B)
I had a really good doctor who found out exactly what I had and continued to work with me for probably the last 7 years and a good rheumatologist, so I just make sure that I take care of myself. I would do whatever I could to help the next person. (Boston focus group 1, quotation C)
Seeing someone sick and not being able to help them made me want to learn more about research and participating in any study that I can. (Boston focus group 1, quotation D)
I know any trial that I might be involved in is really not gonna help me, so—but seeing these very, very young children that are being diagnosed with severe disease.....They need a chance. In that regard, yeah, I'm all for it. Take me. Use me. Take any part of me that you want. (Boston focus group 1, quotation E)
And it's like, okay, but I'm trying to live. I'm trying to maybe see them grandbabies. While at the same time I'm for advancement. And if something's not going to help me immediately, I want to make it better for the future. It may help my daughters, it may help my grandkids. Who knows? Who knows what's in this gene pool? (Chicago focus group 2, quotation F)
I'm positive for it, because we're so severely underrepresented as far as research is concerned. And we need to give an extra push to get somebody behind us, because sometimes we're just lagging behind. We just may get left behind completely. (Chicago focus group 1, quotation G)
Because with lupus, that's the higher percentage of people who get lupus.... I also think that if research is not done on African Americans, I don't think the cure will ever be found...Because that's where it lies. (Boston focus group 2, quotation H)

recommendation. They found comfort in the fact that their physician was often the investigator leading lupus research studies (Table 4, quotation F).

Beyond the physician, participants cared about institutional credibility and reputation. They also wanted institutional presence within the local community beyond conducting research studies (Table 4, quotations G and H).

Influence of social network members. Most participants relied on social support from family members, friends, church members, or lupus support group members. Two participants said that their friends/family discouraged research participation due to knowledge of historical racism in research (Table 4, quotation I). One participant felt that family members were concerned that participation might cause disease-related setbacks (Table 4, quotation J).

Participants usually discussed health-related plans with family members, since family members often bear the responsibility of providing illness-related support (Table 4, quotation K). While some friends and family discouraged trial participation, some participants stated that their friends and family would be the best people to get them to consider clinical trials. One participant stated that her sister had told her about research studies and offered to accompany her to visits. (Table 4, quotation L).

Trial/research-specific factors. *Exclusion criteria.* Several people identified trial exclusion criteria as a barrier to trial participation among Black patients (Table 5, quotation A). They felt that researchers failed to explain why they were ineligible for studies. One participant said that the lack of adequate communication made her less interested in future trials and less likely to refer her friends and family to trials (Table 5, quotation B). One participant was concerned that she might be excluded from trials because of racial discrimination (Table 5, quotation C).

Information presentation and "marketing" of studies. Some participants did not know how to get involved in trials. Those familiar with trials emphasized the importance of having detailed trial information, including written descriptions of study drugs, information about prior trial results, and cost. Participants wanted to consult their physicians about trial medications (Table 5, quotation D) and to know in advance what role their own physicians would play in the clinical trial process. Several participants worried about whether clinical trial participation would require them to change primary care doctors (Table 5, quotation E).

When asked about ideal recruitment tools, participants mentioned YouTube videos and infomercials (Table 5, quotation F). A few participants wanted researchers to train trusted community leaders and liaisons to talk about SLE clinical trials instead of the researchers themselves (Table 5, quotations G–I). They found this to be crucial in overcoming issues of trust within the Black community. One participant suggested that people who had SLE would be the most credible recruiters for clinical trials.

Several participants stressed the importance of in-person meetings before, during, and after clinical trial enrollment (Table 5, quotation J). They wanted to meet with other SLE patients, their

Table 3. Quotations from participants expressing systemic lupus erythematosus–specific concerns

It all depends if there's a medication involved 'cause I'm telling you, I got 17 in the morning, 12, and then 8. They'd have to be able to work with those meds. If it worked with those, then I could fly with you a little bit. (Boston focus group 1, quotation A)
If I'm going to try something new, I want to make sure it doesn't interact with what I'm currently taking. (Chicago focus group 2, quotation B)
If I'm stable, why would I want to do a trial? Why would I want to rock the boat? (Chicago focus group 2, quotation C)
Right, I'm already functioning, you know, to what the lupus normal level is. I don't want to be worse. (Chicago focus group 2, quotation D)
I felt like a guinea pig, it's because I was seeing my doctor—I got to watch my words—weekly. It wasn't even monthly. Weekly. I felt like a guinea pig 'cause during my normal blood and urine, she was taking 13 tubes of blood from me. (Boston focus group 1, quotation E)
When I said guinea pig.....They're still learning about lupus. They'll forever be learning about lupus. It's not a set answer. (Boston focus group 1, quotation F)

Table 4. Quotations from participants discussing the influence of physicians, institutions, and social network members

Physician/institutional influences
I have a team of 6 doctors, so any one thing, I have to talk to my kidney doctor before I can take whatever my rheumatologist wants. So I just can't go blindly into some study. My kidney doctor would have a fit. (Chicago focus group 2, quotation A)
It's not who's saying it. It's just how you say it, period. It could be if you're a black lady and you're saying, "I want you for this research 'cause you're a black woman," I'm still gonna be like, "Why do you want me for the research?" (Boston focus group 1, quotation B)
I don't really care about the ethnicity of the doctor, but if he cannot listen, and he does not include me at the center of the care plan….you're off my case. Because I don't allow that. I need to be involved in the decision making. (Chicago focus group 2, quotation C)
For me, that (race) doesn't matter. I just want to know that the person is being truthful. (Boston focus group 2, quotation D)
The best doctor is going to say, look, I don't know everything about lupus, but you tell me what goes on with you. Now, that's the doctor I want, because he's listening. The one that walks in the room thinking he knows everything, he's not for me. (Chicago focus group 2, quotation E)
What made me comfortable was the fact that it was my physician, and he is well known. He's done a lot of research on lupus. And so it made me a lot more comfortable to move forward with it. (Chicago focus group 2, quotation F)
Is this place in the community outside of this research? Are they visible? Are they present? Do we know of them in another capacity other than wanting to come and do a clinical study or clinical trial? (Chicago focus group 2, quotation G)
Another consideration would probably be who's doing this research…….sometimes reputable names kind of draw more attention, patients may feel more comfortable with a reputable name like that versus some research facility that they haven't heard of, they know nothing about, have no connections. (Chicago focus group 2, quotation H)
Influence of social network members
My friends…..they always say, "Don't do it. Don't forget your history." (Boston focus group 1, quotation I)
It's a topic with my family and myself for quite a few things. I would like to be part of clinical trials, but the family is pretty adamant to say no. Because they don't know if the treatment will work and if it will make my condition worse. I have had times of really bad flares or hospitalizations and things like that, and they just don't want to see that again. (Chicago focus group 1, quotation J)
I'm very considerate about their feelings, because they are the ones that help me when I'm not well…. that decision does impact my family, because they're not around the corner, and they're not up the street….I do have to take that into consideration. (Chicago focus group 1, quotation K)
My sister, she gets a study, she'll tell me of them…As soon as they hear about something, they tell me, "Why don't you go and do this?" My sister will be the one to say, "I'll go with you." She'll take that fear off me…She'll say, "I'll go with you." That makes it much easier. (Boston focus group 1, quotation L)

caregivers, and their health care teams in diverse locations (e.g., schools, churches, lupus support group meetings, hospitals) to talk about their clinical trial experiences.

Identifying fears, motivating factors, structural barriers, and mediators of clinical trial participation. We identified specific motivating factors, fears, structural barriers and mediators of clinical trial participation among Black patients with SLE. A desire to help the greater good was an important facilitator of research participation in our sample. Many participants expressed motivation to participate in order to help others, including their family members, children with SLE, the Black community, and science in general. Participants were also motivated to participate if they could see some personal benefit, such as being assigned to an active treatment that improved their health. Participant fears included potentially being treated like a guinea pig during a trial, reliving previous historical mistreatment of Black individuals

Table 5. Quotations from participants discussing trial/research-specific factors

Exclusion criteria
A lot of the clinical trials, we would like to do them, but a lot of times, we don't qualify because we have lupus or other ailments. Why do they make it so hard? Why do you have to be so healthy? We may feel healthy and everything, but why do you have to have all these standards? You should have some unhealthy people in the trial too to test it out on, and most of the time, with African American women, we don't qualify for a lot of the studies unless you lie. (Boston focus group 1, quotation A)
So, if I have a question, you must be willing. Not just…cut me off. Or, "You're not eligible." But actually, talk to me and bring me into what you're doing…Cuz even if I can't participate…I may tell my sister to participate because I believe in your research because you've explained it better to me. (Boston focus group 2, quotation B)
If you all are doing a clinical trial and, say…The researcher calls me and asks me, do I have any other health issues besides lupus. And I say to him, "I have menopause." They say, "Oh, well. I'm sorry. You…can't do it." Is that discrimination? Or is that because of the research? (Boston focus group 2, quotation C)
Do we have permission to consult our physicians about these medications? (Boston focus group 2, quotation D)
Would we have to change doctors to try this clinical-trial stuff? (Boston focus group 2, quotation E)
Information presentation/marketing of studies
Something you can even watch. Sometimes that influences more than even reading. Where something you can watch that talks about research being done. How it's done. What benefits are. How the procedure went through. There's nothing that anybody watches like that. I haven't seen anything. (Boston focus group 2, quotation F)
Leaders in the communities. People that are influencers. People that have a heart for people. Those kinds of things. People that have the time or want to do it. And it's open for whoever wants to, so that people can come in and do that. (Boston focus group 2, quotation G)
The community may not trust the research, where it's coming from. But if there's steps and there are levels, and you get to the people they do trust, then I think that works. (Boston focus group 2, quotation H)
Different places, for example, like in a church where they have health information or whatever. Places that people trust. Maybe not just a person, but communities. Or people, things, communities that are trusted in the community? (Boston focus group 2, quotation I)
I tried to get my 2 cents in. I would love to do a clinical trial, but—and I'd also like to have meetings like this with the people that are on the clinical trial, whether it's the placebo or not, and for us to sit and talk about what has happened to our bodies since we started taking this particular medication. If these people are gonna be in the trial, you need to test them. You need to talk to them about their trial before they take the medication or the whatever, samples, then while they're taking it and then afterwards. (Boston focus group 1, quotation J)

in research, and being misled during research participation. Fears regarding placebos were mixed. Some participants feared receiving a placebo rather than active treatment, while at least 1 person expressed a preference for placebos due to fears of exacerbating existing health issues via new treatments. Structural barriers that impacted clinical trial participation included racism (which traversed many of the key themes identified), researcher biases, and study exclusion criteria, which often alienated Black patients with SLE from clinical trial participation. With respect to mediators, individuals were more likely to participate in trials if they were recommended by their physicians or if they were conducted by community-engaged health care institutions. Some individuals required the full support of their friends and family members in order to participate in research, especially since their loved ones were involved in their ongoing care and disease management. Appropriate presentation of clinical trials to potential participants also served as a mediator. Participants wanted transparency on why Black individuals were being targeted for clinical trials as well as clear, ongoing information about the trials, from the consent process, through study completion, and beyond.

DISCUSSION

Guided by critical race theory and the PHCRP, focus group participants explored contemporary racialization as it applies to clinical trial participation. Our moderator guide was designed to allow participants to address the role of race and discrimination, past and present, in research. Most participants wanted open discussion and felt that it was a critical consideration for trial participation. While there were differences in some sociodemographic factors of participants between cities, themes and quotes paralleled each other. The explicit goal of this work was to give a voice to underrepresented populations through focus group participation. Additionally, we included a racially and ethnically diverse team of researchers at every stage of the study. In terms of the PHCRP focus on conceptualization and measurement, we considered the number of individuals included within each focus group, keeping the size small enough to allow for comfort discussing difficult topics. We also aimed for participation diversity with respect to geography to explore its interaction with race. Finally, consistent with the PHCRP framework, focus group participants discussed actionable ways to reduce barriers and improve trial representation. Through direct discussions with these groups, we can move closer to increasing clinical trial participation among Black patients.

Racism and discrimination are frequently discussed in the context of health care and clinical trials, as racism-related barriers to research participation have been well-documented (24,25). The historical exploitation of Black individuals in research studies has embedded mistrust of researchers and the research process in Black communities (26–29). Modern-day structural, institutional, and interpersonal racism toward Black individuals is an ongoing concern.

Our findings build on prior work in this area. In our previous systematic review (9), we observed that community engagement and ongoing discussions with social network members were crucial for promoting clinical trial participation among patients from underrepresented groups. Trust in physicians was a key theme in most studies included in the review, as several studies demonstrated that physician trust impacted willingness to participate in clinical trials.

Strengths of the present study include deliberately devoted attention to the role of racism exposure in willingness to participate in clinical research using a theoretical framework specifically designed to detect and evaluate racism as a contributor to racial health disparities. While prior studies have investigated the role of race in underenrollment (30–32), we aimed to explicitly address the role of racism as a key factor. The degree to which issues related to racism would have been raised as a prominent concern without our explicitly asking is unknown; however, the degree and depth of discussions suggest that by explicitly addressing racism, we provided a forum that made it acceptable for racism to be discussed. We used PHCRP to frame our study design, data collection, and analyses in order to robustly describe the role of racism in trial participation.

This study was not without limitations. Many of our participants were older and had been living with lupus for years; thus, we did not capture the perspective of young, newly diagnosed individuals. Ninety percent of the focus group participants were female; therefore, the perspectives of Black men were limited. Although SLE is more common in women than men, there still may be intersectionality between race and sex that we could not explore. We relied on convenience samples of Black patients who primarily attended urban SLE clinics that were run by study researcher clinicians. Thus, their attitudes and experiences may be different from those of patients less connected to academic research centers. Additionally, Chicago participants were primarily middle-class Black patients whose perspectives may not represent those of Black patients from other socioeconomic status groups. We did not clearly differentiate between patients with SLE and their caregivers in our analyses. It is possible that we did have some increased affirmation of perspectives because of the caregivers' presence. Because of the small sample size and the sensitive nature of the topics discussed, we did not present separate data for patients and caregivers or attribute quotes directly to specific individuals. We did this to preserve complete anonymity, which was part of our commitment to our participants.

Additionally, while our study included Black participants from 2 diverse metropolitan areas (Boston and Chicago), we did not specifically design the study to address diverse perspectives among Black participants. Because this distinction was not incorporated into the study design, we were not able to make specific inferences about group differences.

Finally, our findings may not be generalizable to Black patients living in other geographic areas, including the American South, Europe, or rural settings within the US. Although we cannot

generalize to all Black patients, we did achieve considerable diversity in our sample through the recruitment of patients from the 2 cities, as patients across cities were different with respect to age, disease duration, education, research experience, and church/faith community involvement. Despite these differences, discussions and themes were consistent across cities.

In considering Black participation in SLE clinical trials, researchers should recognize the role that disease burden may play in both study eligibility and desire for participation. Black patients tend to have more severe disease manifestation than White patients, which may be either a barrier or a facilitator of trial participation. Sicker patients may be more overwhelmed with their disease and its impact on their quality of life, potentially making them less likely to participate. Alternatively, if their current treatment regimen is not working or causing side effects, they may want to try something else. Likewise, differences in disease severity may contribute to concerns about receiving placebo. Finally, given that SLE trials may have strict criteria that exclude sicker patients, researchers should re-examine how study exclusion criteria impact potential Black participants.

Consistent with the PHCRP framework, we identified how researchers can promote SLE trial participation in Black communities. Researchers must acknowledge and respond to historical and current issues related to racial bias in health care. Engagement with stakeholders (e.g., trusted physicians, friends, family members) is crucial for establishing trust and reducing fears of mistreatment and discrimination among Black patients with SLE. Our participants desired clear and open communication about all aspects of SLE clinical trials before, during, and after trial participation. To gain trust, researchers must address their own implicit biases and invest in Black communities beyond just achieving their research goals. They should form authentic community–academic partnerships that actively engage community leaders in the research process. Such partnerships can foster clinical trial competency within community spaces and provide an important step toward increasing trial participation.

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. Sneed 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. Mancera-Cuevas, Canessa, Ramsey-Goldman, Feldman.

Acquisition of data. Williams, Sinnette, Taber, Curry, Ramsey-Goldman, Feldman.

Analysis and interpretation of data. Sneed, Mason, Williams, Sinnette, Ramsey-Goldman, Feldman.

REFERENCES

1. McCarty DJ, Manzi S, Medsger TA Jr, Ramsey-Goldman R, Laporte RE, Kwoh CK. Incidence of systemic lupus erythematosus race and gender differences. Arthritis Rheum 1995;38:1260–70.
2. Somers EC, Marder W, Cagnoli P, Lewis EE, DeGuire P, Gordon C, et al. Population-based incidence and prevalence of systemic lupus erythematosus: the Michigan Lupus Epidemiology and Surveillance program. Arthritis Rheumatol 2014;66:369–78.
3. Bastian HM, Roseman JM, McGwin G Jr, Alarcon GS, Friedman AW, Fessler BJ, et al. Systemic lupus erythematosus in three ethnic groups. XII. Risk factors for lupus nephritis after diagnosis. Lupus 2002;11:152–60.
4. Lim SS, Helmick CG, Bao G, Hootman J, Bayakly R, Gordon C, et al. Racial disparities in mortality associated with systemic lupus erythematosus—Fulton and DeKalb Counties, Georgia, 2002–2016. MMWR Morb Mortal Wkly Rep 2019;68:419.
5. Falasinnu T, Chaichian Y, Bass MB, Simard JF. The representation of gender and race/ethnic groups in randomized clinical trials of individuals with systemic lupus erythematosus. Curr Rheumatol Rep 2018;20:20.
6. The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1979. URL: https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html.
7. Ortega VE, Meyers DA. Pharmacogenetics: implications of race and ethnicity on defining genetic profiles for personalized medicine. J Allergy Clin Immunol 2014;133:16–26.
8. Burchard EG, Ziv E, Coyle N, Gomez SL, Tang H, Karter AJ, et al. The importance of race and ethnic background in biomedical research and clinical practice. N Engl J Med 2003;348:1170–5.
9. Lima K, Phillip CR, Williams J, Peterson J, Feldman CH, Ramsey-Goldman R. Factors associated with participation in rheumatic disease-related research among underrepresented populations: a qualitative systematic review. Arthritis Care Res (Hoboken) 2020;72:1481–9.
10. Bell D, Solórzano DG. Critical race theory's intellectual roots. In: Lynn M, Dixson AD, editors. Handbook of critical race theory in education.1st ed. New York: Routledge; 2013. p. 48–68.
11. Ford CL, Airhihenbuwa CO. The public health critical race methodology: praxis for antiracism research. Soc Sci Med 2010;71:1390–8.
12. Williams JN, Ford CL, Morse M, Feldman CH. Racial disparities in rheumatology through the lens of critical race theory. Rheum Dis Clin North Am 2020 Nov;46:605–12.
13. Ford CL, Airhihenbuwa CO. Critical race theory, race equity and public health: toward antiracism praxis. Am J Public Health 2010;100 Suppl 1:S30–5.
14. Ford CL, Airhihenbuwa CO. Commentary: just what is critical race theory and what's it doing in a progressive field like public health? Ethn Dis 2018;28 Suppl 1:223–30.
15. Ford CL, Airhihenbuwa CO. The public health critical race methodology: praxis for antiracism research. Soc Sci Med 2010;71:1390–8.
16. Jolly M, Thakkar A, Mikolaitis RA, Block JA. Caregiving, dyadic quality of life and dyadic relationships in lupus. Lupus 2015;24:918–26.
17. Patient and caregiver perceptions of lymphoma care and research opportunities: a qualitative study. Cancer 2019;125:4096–104.
18. Glaser BG. The constant comparative method of qualitative analysis. In: Social problems. 1965;12:436–45.
19. Urquhart C. Grounded theory for qualitative research: a practical guide. 1st ed. Thousand Oaks: Sage Publications; 2012.
20. Phillip CR, Mancera-Cuevas K, Leatherwood C, Chmiel JS, Erickson DL, Freeman E, et al. Implementation and dissemination of an African American popular opinion model to improve lupus awareness: an academic–community partnership. Lupus 2019;28:1441–51.
21. Brandt AM. Racism and research: the case of the Tuskegee Syphilis Study. The Hastings Center Report 1978;8:21–9.

22. Rodriguez-Trias H. Sterilization abuse. Women Health 1978;3: 10–5.
23. Corbie-Smith G, Thomas SB, George DM. Distrust, race, and research. Arch Int Med 2002;162:2458–63.
24. Scharff DP, Mathews KJ, Jackson P, Hoffsuemmer J, Martin E, Edwards D. More than Tuskegee: understanding mistrust about research participation. J Health Care Poor Underserved 2010;21:879.
25. Hughes TB, Varma VR, Pettigrew C, Albert MS. African Americans and clinical research: evidence concerning barriers and facilitators to participation and recruitment recommendations. Gerontologist 2017;57:348–58.
26. Luebbert R, Perez A. Barriers to clinical research participation among African Americans. J Transcult Nurs 2016;27:456–63.
27. Alsan M, Wanamaker M. Tuskegee and the health of black men. Q J Econ 2018;133:407–55.
28. Corbie-Smith G, Thomas SB, Williams MV, Moody-Ayers S. Attitudes and beliefs of African Americans toward participation in medical research. J Gen Intern Med 1999;14:537–46.
29. Brandon DT, Isaac LA, LaVeist TA. The legacy of Tuskegee and trust in medical care: is Tuskegee responsible for race differences in mistrust of medical care? J Natl Med Assoc 2005;97:951–6.
30. Anjorin A, Lipsky P. Engaging African ancestry participants in SLE clinical trials. Lupus Sci Med 2018;5:e000297.
31. Arriens C, Aberle T, Carthen F, Kamp S, Thanou A, Chakravarty E, et al. Lupus patient decisions about clinical trial participation: a qualitative evaluation of perceptions, facilitators and barriers. Lupus Sci Medicine 2020;7:e000360.
32. 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. 73, No. 10, October 2021, pp 1396–1404
DOI 10.1002/acr.24346

Racial Differences in Contraception Encounters and Dispensing Among Female Medicaid Beneficiaries With Systemic Lupus Erythematosus

Jessica N. Williams,[1] Chang Xu,[1] Karen H. Costenbader,[1] Bonnie L. Bermas,[2] Lydia E. Pace,[1] and Candace H. Feldman[1]

Objective. African American and Hispanic women with systemic lupus erythematosus (SLE) have the highest rates of potentially avoidable pregnancy complications, yet racial disparities in family planning among reproductive-age women with SLE have not been well-studied. Our objective was to examine whether there are racial differences in contraception encounters and dispensing among US Medicaid-insured women with SLE.

Methods. Using Medicaid claims data from 2000–2010, we identified women ages 18–50 years with SLE. We examined contraception encounters and uptake over 24 months. We used multivariable logistic regression to estimate the odds ratio and 95% confidence interval by race/ethnicity of contraception encounters, any contraception dispensing, and highly effective contraception (HEC) use, adjusted for age, region, year, SLE severity, and contraindication to estrogen. We also compared contraception encounters and dispensing among women with SLE to the general population and women with diabetes mellitus.

Results. We identified 24,693 reproductive-age women with SLE; 43% were African American, 35% White, 15% Hispanic, 4% Asian, 2% other race, and 1% American Indian/Alaska Native. Nine percent had a contraceptive visit, 10% received any contraception, and 2% received HEC. Compared to White women, African American and Asian women had lower odds of contraception dispensing, and African American women had lower odds of HEC use. Women with SLE were more likely to receive HEC than the general population and women with diabetes mellitus.

Conclusion. In this study of reproductive-age women with SLE, African American and Asian women had lower odds of contraception dispensing and African American women had lower odds of HEC use. Further study is needed to understand the factors driving these racial disparities among this population.

INTRODUCTION

Systemic lupus erythematosus (SLE) is an autoimmune condition that is most common in women of reproductive age (15–50 years) (1). Many medications used to treat SLE and its complications are teratogenic and require concurrent use of effective contraception for women at risk for pregnancy. Active SLE, including lupus nephritis, is associated with adverse pregnancy outcomes that are potentially avoidable if patients are able to time pregnancy for when their SLE is well-controlled (2–4). Furthermore, there are known racial/ethnic disparities in pregnancy outcomes among women with SLE. As compared to White women with SLE, African American and Hispanic women with SLE have higher rates of pregnancy complications, including preterm labor, preeclampsia, and fetal growth restriction, even after controlling for baseline medical conditions and insurance status (5). These complications may result in acute care use and poorer outcomes. Thus, comprehensive contraceptive counseling and care in the outpatient setting for women with SLE, in particular African American and Hispanic women, is important to minimize pregnancy complications.

Highly effective contraception (HEC) includes long-acting reversible contraception (LARC) methods (intrauterine devices [IUDs] and implants), as well as sterilization. LARC methods are safe for women with SLE (6–8) and have a <1% failure rate for preventing pregnancy (9). LARC methods also do not contain estrogen, which is contraindicated in women with SLE who have

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Dr. Feldman's work was supported by the NIH/National Institute of Arthritis and Musculoskeletal and Skin Diseases (K23-AR-071500).

[1]Jessica N. Williams, MD, MPH, Chang Xu, MS, Karen H. Costenbader, MD, MPH, Lydia E. Pace, MD, MPH, Candace H. Feldman, MD, ScD: Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts;

[2]Bonnie L. Bermas, MD: University of Texas Southwestern Medical Center, Dallas.

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

Address correspondence to Candace H. Feldman, MD, ScD, 60 Fenwood Road, Boston, MA 02115. Email: cfeldman@bwh.harvard.edu.

Submitted for publication February 6, 2020; accepted in revised form June 2, 2020.

SIGNIFICANCE & INNOVATIONS

- Within a large population of reproductive-age women with systemic lupus erythematosus (SLE), rates of contraceptive visits, contraception dispensing, and highly effective contraception (HEC) use were low, even among women with lupus nephritis or those using teratogens, who are at higher risk of potentially avoidable adverse pregnancy outcomes.
- Despite more contraceptive visits compared to White women, African American women had lower odds of receiving contraception and HEC. Asian women had fewer contraceptive visits and lower odds of receiving contraception.
- Compared to the general population, reproductive-age women with SLE were less likely to receive any form of contraception. Those patients with SLE who received contraception were more likely to receive HEC or long-acting reversible contraception, potentially due to avoidance of estrogen-based methods in women with high lupus disease activity or positive antiphospholipid antibodies. Across the SLE and general populations, African American women were more likely to have a contraceptive visit but were less likely to receive HEC.
- These racial disparities in contraception care identified among reproductive-age women with SLE and lupus nephritis are likely multifactorial in origin. Further research is necessary to ensure that patients who are at the highest risk for avoidable, adverse pregnancy outcomes are receiving appropriate reproductive care.

positive antiphospholipid antibodies or high disease activity (10). Thus, LARC is an ideal form of contraception for many women with SLE. Additionally, sterilization is another HEC method for women who desire permanent contraception. Prior studies have revealed variable rates of contraception use among reproductive-age women with SLE (11–21), and rates of HEC use among this population have been shown to be low, in the range of 6–9%, though slightly higher than the 5% HEC use reported in the general population of reproductive-age women (19,22).

Despite evidence that contraceptive care for reproductive-age women with SLE is suboptimal and that African American and Hispanic women with SLE have worse pregnancy outcomes, racial disparities in contraception care among reproductive-age women with SLE have not been well-studied. However, prior qualitative studies have found that mistrust is more likely to complicate the relationships of African American and Hispanic female patients with their contraception providers, which may ultimately lead to disparities in contraception care. For example, a thematic analysis of coded interview data from 27 African American and Hispanic reproductive-age women found that the majority felt implicit pressure to accept their providers' preferred method of birth control, despite concerns that risks were not appropriately emphasized. Consequences of patient-provider interactions included discontinuation of recommended contraceptive methods and disengagement from the health care system (23). A separate study of 135 women in New York found that in survey data, African American and Hispanic women were much more likely to report birth-control related mistrust of the medical system and government than White women ($P < 0.001$) (24).

Using US Medicaid data, we examined racial/ethnic differences in outpatient contraceptive encounters, contraception dispensing, and specifically HEC use, among reproductive-age women with SLE and among subpopulations of women with SLE using teratogens and those with lupus nephritis. We also compared contraceptive encounters and dispensing in the SLE population with the general and diabetes mellitus (DM) populations. We chose to examine DM because this disease is another chronic medical condition that may adversely impact pregnancy outcomes, and patients with DM may struggle with similar factors impacting contraception use among women with SLE such as pill burden and competing priorities. We hypothesized that contraceptive encounters, contraceptive dispensing, and HEC use would be more likely in patients who are at the highest risk for potentially avoidable adverse pregnancy outcomes, including African American patients, Hispanic patients, patients with lupus nephritis, and teratogen users.

PATIENTS AND METHODS

Patient population. We used Medicaid claims data (Medicaid Analytic Extract [MAX]) from 47 US states between 2000 and 2006 and from the 29 most populated US states between 2007 and 2010 to identify reproductive-age women (age 18–50 years) with prevalent SLE (≥3 International Classification of Diseases, Ninth Revision [ICD-9] codes 710.0 separated by ≥30 days; the index date was the date of the third SLE code) (25). Only 29 states were analyzed during the period of 2007–2010 because our data-use agreement was limited to these areas, which included the majority of the US and SLE population during this time period. MAX includes demographics information, health care encounters, and drug prescription and dispensing information. Additionally, we examined the subpopulation who had ≥1 teratogenic medication prescription during the baseline period or on the index date (methotrexate, mycophenolate mofetil, cyclophosphamide, leflunomide, angiotensin-converting enzyme inhibitor, or warfarin), as well as the subpopulation with lupus nephritis (defined as ≥2 ICD-9 codes for nephritis, proteinuria, and/or renal failure on or after the SLE diagnosis date and ≥30 days apart; the index date was the date of the second lupus nephritis–related code) (26). Last, we age-matched female Medicaid patients in the general population (4:1) and DM population (2:1) to patients with SLE enrolled between 2007 and 2010. General population patients did not have any codes for SLE and had hospital discharge diagnoses or physician visit claims on the same index date as each patient with prevalent SLE. DM patients did not have any codes for SLE and had ≥3 ICD-9 codes

for DM from hospital discharge diagnoses or physician visit claims separated by ≥30 days on the same index date as each patient with prevalent SLE (see Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24346/abstract, for DM codes).

For the nonmatched analyses, we required ≥6 months of continuous enrollment prior to the index date (baseline period) and ≥24 months of continuous enrollment after (follow-up period). For the matched analyses, we required ≥6 months of continuous enrollment prior to the index date (baseline period) and ≥12 months of continuous enrollment after (follow-up period); the matched analyses had a shorter follow-up period because data were available for the general and DM populations from 2007 to 2010 only. Patients were excluded if they were ineligible for new contraception use (baseline period codes for HEC, hormone replacement therapy, hysterectomy, premature ovarian failure, or menopause; see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24346/abstract). We also excluded patients with baseline or follow-up period codes for pregnancy (see Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24346/abstract). Patients were censored at death, end of the study database, or the end of Medicaid enrollment (whichever came first). This study was approved by the Institutional Review Board at Brigham and Women's Hospital, which provided a waiver of written informed consent. The study was conducted in compliance with the Declaration of Helsinki.

Exposure and outcomes. The exposure of interest was race/ethnicity, which was determined by patient self-report. Possible categories included White Non-Hispanic, Black Non-Hispanic or African American, Hispanic, Asian or Pacific Islander, American Indian or Alaska Native, or other race. Outcomes of interest included outpatient encounters for contraception management, receipt of any form of contraception (including combined hormonal pills, progestin-only pills, IUDs, implants, estrogen-based intravaginal rings, estrogen-based transdermal patches, medroxyprogesterone injections, or sterilization), receipt of HEC (IUDs, implants, or sterilization), and receipt of LARC (IUDs or implants). These 4 outcomes were assessed using encounter, pharmacy, and procedure codes (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24346/abstract) during the follow-up period, as well as drug dispensing data. We were not able to assess the use of barrier methods in this data set.

Baseline covariates. In the multivariable logistic regression models, we adjusted for age group at the index date (age 18–24, 25–31, 32–38, 39–45, and 46–50 years), calendar year of the index date, geographic region (Northeast, Midwest, South, West), the Ward SLE risk adjustment index (proxy for SLE severity) (27), and a composite variable for contraindication to estrogen (baseline period codes for ischemic heart disease, hypertension, stroke, venous thromboembolism, smoking, migraine with aura, or breast cancer; see Supplementary Appendix A, available on

Table 1. Baseline characteristics of female Medicaid beneficiaries with systemic lupus erythematosus (SLE) ages 18–50 years, 2000–2010*

Characteristics	SLE (n = 24,693)	Teratogen use (n = 5,754)	Lupus nephritis (n = 5,229)
Age, years			
18–24	2,685 (11)	780 (14)	1,042 (20)
25–31	4,153 (17)	921 (16)	1,199 (23)
32–38	5,918 (24)	1,243 (22)	1,201 (23)
39–45	7,122 (29)	1,643 (29)	1,104 (21)
46–50	4,815 (20)	1,167 (20)	683 (13)
Race/ethnicity			
White	8,585 (35)	1,630 (28)	1,085 (21)
African American	10,687 (43)	2,760 (48)	2,805 (54)
Hispanic	3,648 (15)	894 (16)	884 (17)
Asian	945 (4)	260 (5)	308 (6)
American Indian/Alaska Native	272 (1)	56 (1)	49 (1)
Other	556 (2)	154 (3)	98 (2)
Geographic region			
Northeast	5,102 (21)	1,190 (21)	980 (19)
Midwest	5,004 (20)	1,227 (21)	1,074 (21)
South	9,512 (39)	2,168 (38)	2,115 (41)
West	5,075 (21)	1,169 (20)	1,060 (20)
Contraindication to estrogen	9,553 (39)	3,200 (56)	2,490 (48)
Ward SLE risk adjustment index ≥1	8,870 (36)	2,918 (51)	3,581 (69)
Calendar year of index date			
2000–2004	13,125 (53)	3,355 (58)	2,514 (48)
2005–2006	3,742 (15)	900 (16)	793 (15)
2007–2008	7,826 (32)	1,499 (26)	1,922 (37)

* Values are the number (%).

the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24346/abstract). We also examined zip code median household income, but this covariate was removed from the model because it did not contribute significantly. We did not have data on sexual activity, practice setting (public versus private), provider specialty, or census tract, and thus these were not included as covariates.

Statistical analysis. For the nonmatched analyses, we employed multivariable logistic regression models adjusted for the aforementioned covariates to estimate the odds ratio (OR) and 95% confidence interval (95% CI) by race/ethnicity of: 1) any outpatient encounter for contraceptive management (versus none), 2) any contraception (versus none), 3) HEC (versus no HEC), 4) LARC (versus no LARC), 5) HEC (versus no HEC) in teratogen user subgroup analyses, and 6) HEC (versus no HEC) in lupus nephritis subgroup analyses. For HEC use, we tested for an interaction between age and calendar year of the index date, as well as age and race, and the terms were not statistically significant. For the matched analyses, we employed conditional logistic regression models adjusted for the aforementioned covariates (except for the calendar year of the index date, given a shortened study period of 2007–2010), to estimate the OR (95% CI) of any outpatient encounter for contraceptive management (versus none), any contraception (versus none), HEC (versus no HEC), and LARC (versus no LARC). In these models, we also examined the OR (95% CI) by race/ethnicity of any outpatient encounter for contraceptive management (versus none), any contraception (versus none), HEC (versus no HEC), and LARC (versus no LARC).

Additionally, we estimated multivariable logistic regression models adjusted for age group, race/ethnicity, and state of residence (reference: Massachusetts), to examine state-level trends in any outpatient encounter for contraceptive management (versus none), any contraception (versus none), and HEC (versus no HEC), among reproductive-age women with SLE in the 29 most-populated US states between 2007 and 2010.

We set $\alpha = 0.05$ to determine statistical significance, and all *P* values were 2-sided. Data were analyzed using SAS software, version 9.4. Medicaid data were obtained through a data-use agreement with the Centers for Medicare and Medicaid Services and are presented in accordance with their policies (cell sizes <11 are suppressed). We also conducted sensitivity analyses including

Table 2. Multivariable logistic regression analyses examining factors associated with contraception encounters and dispensing among female Medicaid beneficiaries with systemic lupus erythematosus (SLE) ages 18–50 years, 2000–2010 (n = 24,693)*

Characteristic	Encounter for contraception management	Any contraception dispensing†	HEC use	LARC use
Age, years				
46–50	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)
18–24	18.03 (14.25–22.82)‡	21.72 (17.06–27.66)‡	12.36 (7.12–21.44)‡	11.91 (6.25–22.69)‡
25–31	10.61 (8.41–13.40)‡	13.74 (10.83–17.44)‡	11.65 (6.81–19.91)‡	8.15 (4.31–15.43)‡
32–38	5.73 (4.53–7.24)‡	6.78 (5.34–8.62)‡	7.14 (4.17–12.24)‡	5.46 (2.88–10.35)‡
39–45	2.53 (1.98–3.24)‡	3.29 (2.57–4.21)‡	3.62 (2.08–6.31)‡	2.41 (1.23–4.74)‡
Race				
White	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)
African American	1.40 (1.25–1.55)‡	0.89 (0.81–0.98)‡	0.71 (0.56–0.89)‡	0.98 (0.72–1.32)
Hispanic	1.38 (1.20–1.59)‡	0.89 (0.77–1.01)	1.12 (0.85–1.47)	1.36 (0.94–1.98)
Asian	0.66 (0.50–0.88)‡	0.72 (0.57–0.92)‡	0.69 (0.40–1.22)	1.34 (0.73–2.47)
American Indian/Alaska Native	1.45 (0.95–2.23)	0.94 (0.61–1.46)	1.63 (0.79–3.38)	2.16 (0.86–5.44)
Other	1.24 (0.89–1.75)	1.00 (0.73–1.36)	1.27 (0.68–2.37)	1.57 (0.67–3.67)
Region				
Northeast	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)
South	1.14 (1.01–1.30)‡	1.00 (0.88–1.12)	0.89 (0.69–1.16)	0.87 (0.61–1.25)
West	1.31 (1.13–1.51)‡	1.06 (0.93–1.22)	0.98 (0.74–1.31)	1.18 (0.81–1.74)
Midwest	1.14 (0.99–1.32)	1.10 (0.96–1.26)	0.97 (0.72–1.30)	1.25 (0.85–1.84)
Ward SLE risk adjustment index	0.88 (0.80–0.98)‡	0.71 (0.64–0.78)‡	0.94 (0.76–1.16)	1.07 (0.81–1.41)
Contraindication to estrogen	0.88 (0.80–0.98)‡	0.86 (0.78–0.95)‡	0.99 (0.80–1.23)	0.95 (0.72–1.27)
Calendar year of index date				
2000–2004	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)
2005–2006	1.12 (0.99–1.27)	0.90 (0.80–1.01)	1.37 (1.04–1.80)‡	1.50 (1.01–2.22)‡
2007–2008	0.90 (0.81–1.00)	0.46 (0.41–0.51)‡	1.91 (1.55–2.35)‡	3.12 (2.36–4.12)‡

* Values are the odds ratio (95% confidence interval). Regression model adjusted for age group, geographic region, calendar year of the index date, the SLE risk adjustment index, and composite variable for contraindication to estrogen. HEC = highly effective contraception; LARC = long-acting reversible contraception; Ref. = reference.
† Any contraception dispensing includes sterilization, intrauterine devices, implants, oral contraceptives, estrogen patch, estrogen ring, and medroxyprogesterone injections. Data on the use of barrier methods were not available.
‡ Statistically significant.

women with pregnancy codes, as well as examining incident SLE (≥3 ICD-9 codes 710.0 separated by ≥30 days, with 24 months without prior SLE codes) (23).

RESULTS

We identified 24,693 female Medicaid beneficiaries with SLE between 2000 and 2010 (Table 1). The mean ± SD age was 38 ± 9 years; 43% were African American , 35% White, 15% Hispanic, 4% Asian, 2% other race, and 1% American Indian/Alaska Native. In all, 39% of patients resided in the South, 21% in the Northeast, 21% in the West, and 20% in the Midwest. Nine percent had an encounter for contraceptive management, 10% received any form of contraception, and 2% received HEC. Among the subpopulation of 5,754 women with SLE using a teratogenic medication and the subpopulation of 5,229 women with prevalent lupus nephritis, there were no differences in frequency of contraceptive visits, contraception dispensing, or HEC use as compared to the overall SLE population.

SLE cohort. The results of the multivariable logistic regression analyses are shown in Table 2. Compared to White women with SLE, African American women with SLE had 1.40 times higher odds (95% CI 1.25–1.55) of a contraceptive visit but 0.89 times lower odds (95% CI 0.81–0.98) of any contraception dispensing and 0.71 times lower odds (95% CI 0.56–0.89) of HEC use. Compared to White women with SLE, Hispanic women with SLE had 1.38 times higher odds (95% CI 1.20–1.59) of a contraceptive visit. Compared to White women with SLE, Asian women with SLE had significantly lower odds of both a contraceptive visit (OR 0.66 [95% CI 0.50–0.88]) and any contraception dispensing (OR 0.72 [95% CI 0.57–0.92]).

Younger age (<46 years) was strongly associated with contraceptive visits, contraceptive dispensing, HEC use, and LARC use. Living in the South (OR 1.14 [95% CI 1.01–1.30]) and living in the West (OR 1.31 [95% CI 1.13–1.51]) was associated with increased odds of a contraceptive visit. More severe SLE was associated with lower odds of a contraceptive visit (OR 0.88 [95% CI 0.80–0.98]) and lower odds of any contraception dispensing (OR 0.71 [95% CI 0.64–0.78]). Having a contraindication to estrogen was also associated with lower odds of a contraceptive visit (OR 0.88 [95% CI 0.80–0.98]) and lower odds of any contraception dispensing (OR 0.86 [95% CI 0.78–0.95]). Having an index date during the later years of 2005–2008 was associated with increased odds of HEC use (OR 1.37 [95% CI 1.04–1.80] for 2005–2006; OR 1.91 [95% CI 1.55–2.35] for 2007–2008) and LARC use (OR 1.50 [95% CI 1.01–2.22] for 2005–2006; OR 3.12 [95% CI 2.36–4.12] for 2007–2008).

SLE subpopulation using teratogens. Among the 5,754 women with SLE using teratogens, there were no significant racial differences in HEC use (Table 3). Younger age was strongly associated with HEC use among this subpopulation, as was index date year of 2007–2008 (OR 2.39 [95% CI 1.57–3.64]).

SLE subpopulation with lupus nephritis. Among the subpopulation of patients with lupus nephritis (n = 5,229), American Indian/Alaska Native race was associated with increased odds of HEC use (OR 7.95 [95% CI 2.48–25.50]) (Table 3). There were no other significant racial differences in HEC use among this subpopulation. Younger age, SLE severity (OR 2.42 [95% CI 1.26–4.66]), and index date year of 2007–2008 (OR 1.67 [95% CI 1.03–2.72]) were strongly associated with HEC use among this subpopulation.

Table 3. Multivariable logistic regression analyses examining factors associated with highly effective contraception (HEC) use among female Medicaid beneficiaries with systemic lupus erythematosus (SLE) and teratogen use and lupus nephritis, ages 18–50 years, 2000–2010*

Characteristic	HEC use on teratogen (n = 5,754)†	HEC use with lupus nephritis (n = 5,229)
Age, years		
46–50	1.00 (Ref.)	1.00 (Ref.)
18–24	5.36 (2.29–12.53)‡	6.18 (2.89–13.20)‡
25–31	5.57 (2.42–12.82)‡	3.83 (1.75–8.41)‡
32–38	3.44 (1.48–8.03)‡	3.01 (1.33–6.82)‡
39–45	2.37 (1.01–5.59)‡	1.00 (Ref.)§
Race		
White	1.00 (Ref.)	1.00 (Ref.)
African American	0.84 (0.52–1.36)	0.82 (0.45–1.48)
Hispanic	1.20 (0.67–2.17)	0.98 (0.48–2.02)
Asian	0.75 (0.25–2.21)	0.84 (0.30–2.40)
American Indian/ Alaska Native	1.77 (0.40–7.74)	7.95 (2.48–25.50)‡
Other	2.08 (0.78–5.55)	<0.001 (<0.001 to >999.99)
Region		
Northeast	1.00 (Ref.)	1.00 (Ref.)
South	1.37 (0.79–2.39)	0.98 (0.50–1.95)
West	1.22 (0.65–2.27)	1.33 (0.64–2.77)
Midwest	1.20 (0.63–2.26)	1.29 (0.63–2.67)
Ward SLE risk adjustment index	1.04 (0.69–1.56)	2.42 (1.26–4.66)‡
Contraindication to estrogen	0.76 (0.51–1.14)	0.93 (0.57–1.51)
Calendar year of index date		
2000–2004	1.00 (Ref.)	1.00 (Ref.)
2005–2006	1.56 (0.90–2.69)	1.15 (0.57–2.30)
2007–2008	2.39 (1.57–3.64)‡	1.67 (1.03–2.72)‡

* Values are the odds ratio (95% confidence interval). Regression model adjusted for age group, geographic region, calendar year of the index date, the SLE risk adjustment index, and composite variable for contraindication to estrogen. Ref. = reference.
† Teratogens included methotrexate, mycophenolate mofetil, mycophenolic acid, cyclophosphamide, leflunomide, angiotensin-converting enzyme inhibitor, or warfarin.
‡ Statistically significant.
§ Age 39–50 years was used as the reference group because there were no patients with HEC use in the 46–50 years age group.

We also examined contraception use by age group for the overall SLE population and for the lupus nephritis subpopulation. In the 18–24 years age group, 29% of patients with SLE and 19% of patients with lupus nephritis received any contraception. In the 25–31 years age group, 21% of patients with SLE and 10% of patients with lupus nephritis received any contraception. In the 32–38 years age group, 11% of patients with SLE received any contraception versus 7% of patients with lupus nephritis, and in the oldest 2 age categories (39–50 years, combined given the smaller size of the lupus nephritis subpopulation), 8% of patients with SLE received any contraception versus 3% of patients with lupus nephritis.

State-level analyses. We identified 7,826 reproductive-age women with SLE residing in the 29 most populated US states between 2007 and 2010 for the state-level analyses. Alabama residence (OR 0.14 [95% CI 0.03–0.66]) and Ohio residence (OR 0.29 [95% CI 0.10–0.85]) were associated with lower odds of any contraception dispensing. Texas residence (OR 2.08 [95% CI 1.03–4.18]) and Wisconsin residence (OR 2.33 [95% CI 1.09–4.99]) were associated with higher odds of a contraceptive management visit. There were no significant state-level differences in HEC use. These results did not change when race was excluded from the model.

Age-matched SLE, DM, and general populations. For the age-matched results, we identified 6,946 women with SLE, 13,718 women with DM, and 29,698 women in the general population (Table 4). There were more African American patients in the SLE population (46%) as compared to the DM (31%) and general populations (22%). There were more Hispanic patients in the general population (27%) as compared to the SLE (13%) and DM populations (15%). Table 5 shows results of the conditional logistic regression analyses comparing contraception encounters and dispensing among the age-matched SLE, DM, and general populations. As compared to the DM population, the SLE population had significantly lower odds of contraceptive visits (OR 0.82 [95% CI 0.73–0.92]) and any contraception dispensing (OR 0.42 [95% CI 0.37–0.47]) but significantly higher odds of HEC use (OR 1.39 [95% CI 1.12–1.72]) and LARC use (OR 1.36 [95% CI 1.05–1.75]). As compared to the general population, the SLE population had significantly lower odds of contraceptive visits (OR 0.64 [95% CI 0.55–0.75]) and any contraception dispensing (OR 0.45 [95% CI 0.38–0.52]) but significantly higher odds of HEC use (OR 1.39 [95% CI 1.03–1.89]) and LARC use (OR 1.52 [95% CI 1.06–2.19]). Across all 3 populations, African American women had higher odds of a contraceptive visit but lower odds of HEC use (Table 6). Among the general population, American Indian/Alaska Native race and other race were associated with higher odds of a contraceptive visit, and Hispanic race was associated with lower odds of a contraceptive visit but with higher odds of any contraception dispensing, HEC use, and LARC use.

Table 4. Baseline characteristics of age-matched female Medicaid beneficiaries in the systemic lupus erythematosus (SLE) population, diabetes mellitus (DM) population, and general population (GP), 2007–2010*

Characteristic	SLE	DM	GP
Total number	6,946	13,718	29,698
Age, mean ± SD years	38 ± 9	38 ± 9	37 ± 9
Race/ethnicity			
White	2,472 (36)	6,121 (45)	12,466 (42)
African American	3,187 (46)	4,254 (31)	6,385 (22)
Hispanic	917 (13)	2,079 (15)	8,147 (27)
Asian	213 (3)	321 (2)	790 (3)
American Indian/Alaska Native	67 (1)	192 (1)	312 (1)
Other	90 (1)	751 (6)	1,598 (5)
Geographic region			
Northeast	1,352 (20)	2,693 (20)	5,605 (19)
Midwest	1,555 (22)	3,112 (23)	5,939 (20)
South	2,856 (41)	5,380 (39)	8,135 (27)
West	1,183 (17)	2,533 (19)	10,019 (34)

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

Sensitivity analyses. For our sensitivity analyses including women with pregnancy codes, we identified 29,206 reproductive-age women with SLE between 2000 and 2010. A total of 15% had an encounter for contraceptive management, 15% received any form of contraception, and 5% received HEC. Similar to our primary results, we found that as compared to White women, despite 1.24 times higher odds (95% CI 1.15–1.35) of a contraceptive visit, African American women had 0.87 times lower odds (95% CI 0.80–0.94) of any contraception dispensing and 0.71 times lower odds (95% CI 0.62–0.81) of HEC use. Asian women had significantly lower odds of contraceptive visits (OR 0.62 [95% CI 0.49–0.77]), any contraception dispensing (OR 0.59 [95% CI 0.48–0.73]), and HEC use (OR 0.47 [95% CI 0.32–0.70]).

For our sensitivity analyses examining women with incident SLE, we identified 9,521 reproductive-age women with incident SLE between 2000 and 2010. A total of 10% had an encounter for contraceptive management, 11% received any form of contraception, and 2% received HEC. We found that African American patients (OR 1.46 [95% CI 1.24–1.73]) and Hispanic patients (OR 1.30 [95% CI 1.04–1.62]) with incident SLE were more likely to have a visit for contraception management, but there were no significant racial differences in receipt of contraception dispensing or HEC use.

DISCUSSION

In this nationwide study of reproductive-age women with SLE, we found that contraception dispensing and HEC uptake were very low. Low contraception uptake remained true among women at the highest risk for potentially avoidable adverse pregnancy outcomes: teratogen users or women with lupus nephritis. Despite the lupus nephritis subpopulation being younger, we did not find that they were more likely than the overall

Table 5. Conditional logistic regression results examining contraception encounters and dispensing among female Medicaid beneficiaries in the systemic lupus erythematosus (SLE) population, as compared to the age-matched diabetes mellitus (DM) population and general population (GP), 2007–2010*

Comparison group	Encounter for contraception management	Any contraception dispensing†	HEC use	LARC use
SLE vs. DM	0.82 (0.73–0.92)	0.42 (0.37–0.47)	1.39 (1.12–1.72)	1.36 (1.05–1.75)
SLE vs. GP	0.64 (0.55–0.75)	0.45 (0.38–0.52)	1.39 (1.03–1.89)	1.52 (1.06–2.19)

* Values are the odds ratio (95% confidence interval). All values are statistically significant. HEC = highly effective contraception; LARC = long-acting reversible contraception.
† Any contraception dispensing includes sterilization, intrauterine devices, implants, oral contraceptives, estrogen patch, estrogen ring, and medroxyprogesterone injections. Data on use of barrier methods were not available.

SLE population to receive contraception, even when stratified by age group. We also identified racial disparities in contraception encounters and dispensing. Despite more contraceptive visits compared to White women, African American women had lower odds of receiving contraception and HEC, while Asian women had fewer contraceptive visits and lower odds of receiving contraception. As compared to the DM and general populations, reproductive-age women with SLE were less likely to receive any form of contraception but more likely to receive HEC or LARC. Across the SLE, DM, and general populations, African American women were more likely to have a contraceptive visit but less likely to receive HEC. We also found notable state-based differences in contraception encounters and dispensing in exploratory analyses. Compared to Massachusetts residence, Alabama and Ohio residence was associated with decreased odds of any contraception dispensing, and Texas and Wisconsin residence was associated with increased odds of a contraceptive visit.

A listing of the studies to date examining contraception care among women with SLE is included in Supplementary Table 1, available on the *Arthritis Care & Research* website at

Table 6. Conditional logistic regression results examining contraception encounters and dispensing by race/ethnicity among age-matched female Medicaid beneficiaries in the systemic lupus erythematosus (SLE) population, diabetes mellitus (DM) population, and general population (GP), 2007–2010*

Group, race/ethnicity	Encounter for contraception management	Any contraception dispensing†	HEC use	LARC use
SLE				
White	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)
African American	1.27 (1.03–1.57)‡	0.89 (0.71–1.12)	0.68 (0.48–0.98)‡	1.00 (0.65–1.53)
Hispanic	1.29 (0.95–1.73)	0.93 (0.67–1.28)	1.07 (0.67–1.71)	1.18 (0.66–2.11)
Asian	0.98 (0.55–1.73)	0.72 (0.38–1.34)	1.29 (0.60–2.77)	1.88 (0.81–4.36)
American Indian/ Alaska Native	1.15 (0.44–3.02)	0.93 (0.32–2.68)	0.57 (0.08–4.20)	0.96 (0.13–7.21)
Other	1.55 (0.76–3.17)	1.44 (0.69–3.03)	2.00 (0.77–5.18)	2.68 (0.92–7.82)
DM				
White	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)
African American	1.20 (1.04–1.38)‡	0.91 (0.80–1.02)	0.66 (0.48–0.92)‡	0.88 (0.61–1.27)
Hispanic	1.13 (0.94–1.37)	0.92 (0.78–1.08)	1.04 (0.72–1.51)	0.95 (0.60–1.49)
Asian	1.06 (0.68–1.66)	1.24 (0.88–1.75)	1.53 (0.75–3.09)	1.93 (0.90–4.12)
American Indian/ Alaska Native	0.95 (0.54–1.66)	1.05 (0.68–1.63)	0.81 (0.25–2.59)	0.83 (0.20–3.44)
Other	0.99 (0.74–1.32)	0.89 (0.69–1.13)	0.67 (0.35–1.29)	0.97 (0.49–1.90)
GP				
White	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)
African American	1.16 (1.05–1.27)‡	1.06 (0.97–1.15)	0.77 (0.63–0.95)‡	0.87 (0.69–1.10)
Hispanic	0.82 (0.72–0.92)‡	1.77 (1.63–1.92)‡	1.55 (1.31–1.84)‡	1.65 (1.37–2.00)‡
Asian	0.76 (0.56–1.02)	1.10 (0.91–1.32)	0.89 (0.59–1.35)	1.01 (0.66–1.55)
American Indian/ Alaska Native	1.80 (1.28–2.54)‡	0.96 (0.71–1.30)	0.97 (0.51–1.85)	1.09 (0.55–2.16)
Other	1.29 (1.09–1.52)‡	0.91 (0.78–1.06)	0.90 (0.64–1.27)	0.94 (0.64–1.37)

* Values are the odds ratio (95% confidence interval). HEC = highly effective contraception; LARC = long-acting reversible contraception; Ref. = reference.
† Any contraception dispensing includes sterilization, intrauterine devices, implants, oral contraceptives, estrogen patch, estrogen ring, and medroxyprogesterone injections. Data on the use of barrier methods were not available.
‡ Statistically significant.

http://onlinelibrary.wiley.com/doi/10.1002/acr.24346/abstract (8,11–22,28–33). These prior studies have revealed variable rates of contraception use among reproductive-age women with SLE (22–85%) (11–21), and rates of HEC use among this population are even lower, in the range of 6–9% (19,22). Additionally, studies examining contraceptive counseling in the outpatient setting among women with SLE have shown that 28–59% of patients at risk for pregnancy are not receiving counseling (13,18–20,28), including up to 46% of women using teratogens (29). Unplanned pregnancies and lack of contraception use have been found to be significantly more common among women with SLE as compared to women with rheumatoid arthritis or healthy controls (17). Moreover, combined hormonal contraception use has been found to be common in women with SLE who have contraindications to its use, including the presence of antiphospholipid antibodies (30).

Despite known racial disparities in pregnancy outcomes among women with SLE, only 1 of these studies examined racial differences in contraception care among this population (28), and the majority of the studies did not provide any racial data. Although racial differences in contraception care among women with SLE have not been well-studied, prior qualitative studies have found that African American and Hispanic women perceive that their providers are pressuring them about their contraception choices (23) and report higher levels of birth control–related mistrust as compared to White women (24), which may adversely affect their reproductive choices and lead to racial disparities.

In this study, we found that as compared to patients with DM or the general population, patients with SLE were less likely to receive contraception in general but more likely to receive HEC or LARC. Higher rates of HEC use among women with SLE could reflect enhanced concerns about estrogen exacerbating disease activity or leading to thromboembolic disease. These findings also suggest that the presence of a chronic disease is not the only factor leading to decreased contraception dispensing among women with SLE. As far as our state of residence findings, state-level variation in contraceptive policies and access for Medicaid-insured women could contribute to these differences. For example, a prior study of state-level differences in postpartum contraception use found that women in Alabama and Ohio had very low rates of LARC use (2%) (34). We also found that during the study period, Texas had a state-funded program to cover family planning services for adult women living at ≤185% of the Federal poverty level, and Wisconsin has a state plan amendment to provide family planning services for those living at ≤306% of the Federal poverty level, which is the highest threshold among all states (35).

Strengths of our study include the use of a large, nationwide, and racially diverse study population with detailed medication data. We had detailed information on a range of contraceptive services, including encounters for contraception management, and pharmacy and medical claims for combined oral contraceptive pills, progestin-only pills, IUDs, implants, estrogen-based intravaginal rings, estrogen-based transdermal patches, medroxyprogesterone injections, and sterilization.

The limitations of this study include lack of data on the use of nonprescription methods of contraception or contraceptives that were not reimbursed by Medicaid. We were also not able to ascertain practice setting (public versus private), provider specialty, patient census tract, male partner sterilization, whether patients desired pregnancy, whether patients declined contraception, and whether patients were sexually active with men during the study period. We were limited in the number of covariates we could include in our models given the small number of outcomes, and thus there was a possibility of residual confounding. Last, due to variability in enrollment in Medicaid, we were unable to identify LARC insertions prior to the baseline period for our population. However only 0.6% of women in our cohort had LARC removals during the follow-up period without prior claims for insertion during the baseline and follow-up periods. Given this small percentage, we feel it is unlikely that we are missing a significant number of LARC users who had placement prior to the baseline period.

In conclusion, racial disparities in contraceptive encounters and dispensing among African American and Asian reproductive-age women with SLE in the US Medicaid population is likely multifactorial. Contributing factors may include patient preference, cultural factors, ineffective communication and/or mistrust between health care providers and patients, racial bias of providers (implicit or explicit), state-level policies affecting reproductive care, and the legacy of historical injustices (for example, the history of forced sterilization of African American women in the US may account for lower odds of HEC use but not LARC use among African American women in our study). In particular, ineffective communication and mistrust in the patient-provider relationship may account for our observation of increased contraception management visits but decreased receipt of contraception or HEC among African American women with SLE. With known disparities in pregnancy outcomes by race/ethnicity among women with SLE, qualitative studies are needed to further understand the factors that are driving these racial differences. We feel that every effort should be made to mitigate these racial disparities and to ensure that patients who are at the highest risk for adverse pregnancy outcomes are receiving appropriate reproductive care.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. 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. Williams, Costenbader, Bermas, Pace, Feldman.

Acquisition of data. Costenbader, Feldman.

Analysis and interpretation of data. Williams, Xu, Costenbader, Bermas, Pace, Feldman.

REFERENCES

1. Pons-Estel JG, Alarcón GS, Scofield L, Reinlib L, Cooper GS. Understanding the epidemiology and progression of systemic lupus erythematosus. Semin Arthritis Rheum 2010;39:257–68.
2. Feldman CH, Speyer C, Ashby R, Bermas BL, Bhattacharyya S, Chakravarty E, et al. Development of a set of lupus-specific ambulatory care sensitive, potentially preventable adverse conditions: a Delphi consensus study. Arthritis Care Res (Hoboken) 2019;73:146–57.
3. Clowse ME, Jamison M, Myers E, James AH. A national study of the complications of lupus in pregnancy. Am J Obstet Gynecol 2008;199:127.e1–6.
4. Yasmeen S, Wilkins EE, Field NT, Sheikh RA, Gilbert WM. Pregnancy outcomes in women with systemic lupus erythematosus. J Matern Fetal Med 2001;10:91–6.
5. Clowse ME, Grotegut C. Racial and ethnic disparities in the pregnancies of women with systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2016;68:1567–72.
6. Rebelo RC, Pignaton E, Valeria Bahamondes M, Costallat LTL, Appenzeller S, Bahamondes L, et al. Disease activity and thromboembolic events in women with systemic lupus erythematosus with and without anti-phospholipid syndrome: users of the 52-mg levonorgestrel-releasing intrauterine system. Arch Gynecol Obstet 2019;299:1597–605.
7. Sammaritano LR. Contraception in patients with systemic lupus erythematosus and antiphospholipid syndrome. Lupus 2014;23:1242–5.
8. Sánchez-Guerrero J, Uribe AG, Jiménez-Santana L, Mestanza-Peralta M, Lara-Reyes P, Seuc AH, et al. A trial of contraceptive methods in women with systemic lupus erythematosus. N Engl J Med 2005;353:2539–49.
9. Winner B, Peipert JF, Zhao Q, Buckel C, Madden T, Allsworth JE, et al. Effectiveness of long-acting reversible contraception. N Engl J Med 2012;366:1998–2007.
10. Curtis KM, Tepper NK, Jatlaoui TC, Berry-Bibee E, Horton LG, Zapata LB, et al. U.S. Medical eligibility criteria for contraceptive use, 2016. MMWR Recomm Rep 2016;65:1–103.
11. Birru Talabi M, Clowse ME, Blalock SJ, Moreland L, Siripong N, Borrero S. Contraception use among reproductive-age women with rheumatic diseases. Arthritis Care Res (Hoboken) 2019;71:1132–40.
12. Schwarz EB, Manzi S. Risk of unintended pregnancy among women with systemic lupus erythematosus. Arthritis Rheum 2008;59:863–6.
13. Yazdany J, Trupin L, Kaiser R, Schmajuk G, Gillis JZ, Chakravarty E, et al. Contraceptive counseling and use among women with systemic lupus erythematosus: a gap in health care quality? Arthritis Care Res (Hoboken) 2011;63:358–65.
14. DeNoble AE, Hall KS, Xu X, Zochowski MK, Piehl K, Dalton VK. Receipt of prescription contraception by commercially insured women with chronic medical conditions. Obstet Gynecol 2014;123:1213–20.
15. Champaloux SW, Tepper NK, Curtis KM, Zapata LB, Whiteman MK, Marchbanks PA, et al. Contraceptive use among women with medical conditions in a nationwide privately insured population. Obstet Gynecol 2015;126:1151–9.
16. Brito MB, Casqueiro JS, Alves FS, Lopes JB, Alves RD, Santiago M. Low prevalence of contraceptive use among Brazilian women of reproductive age with systemic lupus erythematosus. J Obstet Gynaecol 2018;38:975–8.
17. Galappatthy P, Jayasinghe JD, Paththinige SC, Sheriff RM, Wijayaratne LS. Pregnancy outcomes and contraceptive use in patients with systemic lupus erythematosus, rheumatoid arthritis and women without a chronic illness: a comparative study. Int J Rheum Dis 2017;20:746–54.
18. Dalkilic E, Tufan AN, Oksuz MF, Sahbazlar M, Coskun BN, Seniz N, et al. Comparing female-based contraceptive methods in patients with systemic lupus erythematosus, rheumatoid arthritis and a healthy population. Int J Rheum Dis 2014;17:653–7.
19. Kittisiam T, Werawatakul Y, Nanagara R, Wantha O. Low prevalence of contraceptive counseling at Srinagarind hospital, Thailand among women of reproductive age with systemic lupus erythematosus. Reprod Health 2013;10:21.
20. Lakasing L, Khamashta M. Contraceptive practices in women with systemic lupus erythematosus and/or antiphospholipid syndrome: what advice should we be giving? J Fam Plann Reprod Health Care 2001;27:7–12.
21. Julkunen HA, Kaaja R, Friman C. Contraceptive practice in women with systemic lupus erythematosus. Br J Rheumatol 1993;32:227–30.
22. Pujol TA, Serban N, Swann J, Kottke M. Medicaid claims for contraception among women with medical conditions after release of the US medical eligibility criteria for contraceptive use. Prev Chronic Dis 2019;16:E03.
23. Gomez AM, Wapman M. Under (implicit) pressure: young Black and Latina women's perceptions of contraceptive care. Contraception 2017;96:221–6.
24. Rosenthal L, Lobel M. Gendered racism and the sexual and reproductive health of Black and Latina women. Ethn Health 2020;25:367–92.
25. Feldman CH, Hiraki LT, Liu J, Fischer MA, Solomon DH, Alarcón GS, et al. Epidemiology and sociodemographics of systemic lupus erythematosus and lupus nephritis among US adults with Medicaid coverage, 2000–2004. Arthritis Rheum 2013;65:753–63.
26. Chibnik LB, Massarotti EM, Costenbader KH. Identification and validation of lupus nephritis cases using administrative data. Lupus 2010;19:741–3.
27. Ward MM. Development and testing of a systemic lupus-specific risk adjustment index for in-hospital mortality. J Rheumatol 2000;27:1408–13.
28. Ferguson S, Trupin L, Yazdany J, Yelin E, Barton J, Katz P. Who receives contraception counseling when starting new lupus medications? The potential roles of race, ethnicity, disease activity, and quality of communication. Lupus 2016;25:12–7.
29. Quinzanos I, Davis L, Keniston A, Nash A, Yazdany J, Fransen R, et al. Application and feasibility of systemic lupus erythematosus reproductive health care quality indicators at a public urban rheumatology clinic. Lupus 2015;24:203–9.
30. Mendel A, Bernatsky S, Pineau CA, St-Pierre Y, Hanly JG, Urowitz MB, et al. Use of combined hormonal contraceptives among women with systemic lupus erythematosus with and without medical contraindications to oestrogen. Rheumatology (Oxford) 2019;58:1259–67.
31. Cravioto MD, Jiménez-Santana L, Mayorga J, Seuc AH. Side effects unrelated to disease activity and acceptability of highly effective contraceptive methods in women with systemic lupus erythematosus: a randomized, clinical trial. Contraception 2014;90:147–53.
32. Ekblom-Kullberg S, Kautiainen H, Alha P, Helve T, Leirisalo-Repo M, Julkunen H. Reproductive health in women with systemic lupus erythematosus compared to population controls. Scand J Rheumatol 2009;38:375–80.
33. Petri M, Kim MY, Kalunian KC, Grossman J, Hahn BH, Sammaritano LR, et al. Combined oral contraceptives in women with systemic lupus erythematosus. N Engl J Med 2005;353:2550–8.
34. White K, Potter JE, Hopkins K, Grossman D. Variation in postpartum contraceptive method use: results from the Pregnancy Risk Assessment Monitoring System (PRAMS). Contraception 2014;89:57–62.
35. Kaiser Family Foundation. States that have expanded eligibility for coverage of family planning services under Medicaid. URL: https://www.kff.org/medicaid/state-indicator/family-planning-services-waivers/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D.

Arthritis Care & Research
Vol. 73, No. 10, October 2021, pp 1405–1412
DOI 10.1002/acr.24341

American College of Rheumatology
Empowering Rheumatology Professionals

Obesity and the Risk of Incident Chronic Opioid Use in Rheumatoid Arthritis

Joshua F. Baker,[1] Andrew Stokes,[2] Sofia Pedro,[3] Ted R. Mikuls,[4] Michael George,[5] Bryant R. England,[4] Harlan Sayles,[4] Fred Wolfe,[3] and Kaleb Michaud[6]

Objective. The present study was undertaken to evaluate whether the rate of incident chronic opioid use is higher in obese patients with rheumatoid arthritis (RA).

Methods. Participants with RA in the FORWARD databank were asked about their use of weak and strong opioid medications on semiannual surveys. Incident chronic opioid use was defined as new reported use extending over 2 contiguous surveys (~7–12 months). Cox proportional hazards models were used to evaluate associations between body mass index (BMI) at enrollment and incident chronic opioid use (overall use and strong opioid use). Models adjusted for demographics, smoking, disease duration, RA treatments, household income, and education level. The predicted 5-year cumulative incidence was calculated from Cox models.

Results. Among 19,794 participants, 2,802 experienced an incident episode of chronic opioid use over 93,254 person-years of follow-up. Higher BMI was associated with higher risk of chronic opioid use. Severe obesity (BMI >35 kg/m^2) was associated with a higher risk of overall use (adjusted hazard ratio [HR_{adj}] 1.74 [95% confidence interval (95% CI) 1.72–2.04], $P < 0.0001$) and strong opioid use (HR_{adj} 2.11 [95% CI 1.64–2.71], $P < 0.001$) compared to normal BMI. This association was partially explained by greater comorbidity, pain, and disability in obese groups. The attributable risk for obesity was 15% of overall opioid use and 24% of strong opioid use.

Conclusion. Obesity is associated with a substantially higher risk of incident chronic opioid use. Approximately 1 in 4 cases of incident use of strong opioids may be attributable to obesity, suggesting a major public health impact. Interventions to prevent or reduce obesity could have an important impact on the use of opioids.

INTRODUCTION

Chronic use of prescription opioids increases the risk of prescription opioid overdose and death and is a major ongoing public health problem (1,2). Arthritis is a major contributor to chronic opioid use as the result of long-term joint pain (3). Recent evidence suggests that the prevalence of chronic prescription opioid use among patients with rheumatoid arthritis (RA) is ~17–40% and has increased over the last decade (4–6).

A recent study identified predictors of opioid use in patients with RA (6). These included chronic pain, antidepressant use, higher disease activity, and greater disability. However, this study did not evaluate the impact of obesity. Some studies have identified social determinants and other factors associated with opioid use in individuals with arthritis (4,7). However, we are not aware of studies quantifying the impact of obesity on this public health problem in patients with arthritis. This is despite a number of studies suggesting greater symptoms and more severe disability in patients with RA who are obese (8–11).

The contents herein do not represent the views of the Department of Veterans Affairs or the US Government.

Dr. Baker's work was supported by the Department of Veterans Affairs (Veterans Affairs Clinical Science Research and Development Merit award I01-CX-001703). Dr. Stokes' work was supported by the Robert Wood Johnson Foundation. Dr. Mikuls' work was supported by the Department of Veterans Affairs (Merit award CX000896), the Rheumatology Research Foundation, and the NIH (National Institute of General Medical Sciences grant U54-GM-115458). Dr. England's work was supported by the Rheumatology Research Foundation and the NIH (National Institute of General Medical Sciences grant U54-GM-115458). Dr. Michaud's work was supported by the Rheumatology Research Foundation.

[1]Joshua F. Baker, MD, MSCE: Philadelphia VA Medical Center and University of Pennsylvania, Philadelphia; [2]Andrew Stokes, PhD: Boston University School of Public Health, Boston, Massachusetts; [3]Sofia Pedro, PhD, Fred Wolfe, MD: FORWARD, The National Databank for Rheumatic Diseases, Wichita, Kansas; [4]Ted R. Mikuls, MD, MSPH, Bryant R. England, MD, Harlan Sayles, MS: VA Nebraska–Western Iowa Health Care System and University of Nebraska Medical Center, Omaha; [5]Michael George, MD, MSCE: University of Pennsylvania, Philadelphia; [6]Kaleb Michaud, PhD: FORWARD, The National Databank for Rheumatic Diseases, Wichita, Kansas, and VA Nebraska–Western Iowa Health Care System and University of Nebraska Medical Center, Omaha, Nebraska.

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

Address correspondence to Joshua F. Baker, MD, MSCE, Division of Rheumatology, Department of Medicine, 5 White Building, 3400 Spruce Street, Hospital of the University of Pennsylvania, Philadelphia, PA 19104. Email: bakerjo@uphs.upenn.edu.

Submitted for publication December 19, 2019; accepted in revised form May 22, 2020.

SIGNIFICANCE & INNOVATIONS

- Obesity is a major public health problem with a far-reaching impact on health.
- Obesity is associated with a substantially higher risk of incident chronic opioid use, with 1 in 4 cases of all incident use of strong opioids possibly being attributable to obesity.
- Greater opioid use in obesity is partially explained by greater pain, pain centralization, joint dysfunction, and disability in obese patients.

Obesity is well described as an important contributor to the burden of chronic pain in the US (12). It is associated with numerous conditions that increase the risk of chronic pain, such as degenerative arthritis and lower back pain, among others (13–15). In addition, obesity is strongly associated with depression (16), central pain sensitization (17), and prevalent and progressive disability (18). Thus, there are a number of reasons why patients with RA who are obese may experience worse pain and joint dysfunction and may be at particular risk for chronic opioid use. To date, the impact of obesity on the opioid crisis in patients with arthritis has been overlooked and is likely to be significant.

Thus, patients with inflammatory arthritis experience chronic pain and are at risk for chronic opioid use, and obesity is an important comorbidity that may directly lead to a greater utilization of chronic opioids in this high-risk population. In this study, we aimed to determine if obesity is associated with a higher risk of incident chronic opioid use in RA and to estimate the public health impact of obesity on incident chronic opioid use in this population (both overall use and strong opioid use). We hypothesized a substantially higher rate of incident chronic opioid use over time in obese patients with RA independent of demographics, social determinants, and disease-related factors.

PATIENTS AND METHODS

Study setting. The study was conducted in FORWARD, The National Databank for Rheumatic Diseases. FORWARD is a patient-based multi-disease, multipurpose rheumatic disease registry with patients enrolled from community-based rheumatology practices across the US and followed up with regular biannual questionnaires. All patients with RA have received a diagnosis from a rheumatologist. Key patient data are validated regularly using medical records. The registry has been described in detail elsewhere (19,20). We utilized data from January 1, 1999 to February 28, 2019. The study was approved by Via Christi Hospitals Wichita Institutional Review Board (IRB00001674). All patients sign informed consent prior to participating.

Body mass index (BMI). All patients reported their weight and height on each biannual questionnaire. BMI was calculated at enrollment (weight [kg]/modal height [m]^2) and categorized according to published World Health Organization categories as underweight (<18.5 kg/m^2), normal weight (≥18.5–25 kg/m^2), overweight (≥25–30 kg/m^2), obese (≥30–35 kg/m^2), and severely obese (≥35 kg/m^2) (21). Patients with a BMI <14 kg/m^2 were excluded.

Chronic opioid use. Incident chronic opioid use was defined as reporting of any opioid use on 2 consecutive questionnaires (typically 6 months apart) among a cohort of participants who did not report use at enrollment. This classification was based on definitions used in prior studies (6) and resulted in at least 7 months of consistent opioid use. In secondary analyses, we separated strong and weak opioids. Weak opioids included codeine, tramadol, hydrocodone, and dextropropoxyphene. Strong opioids included morphine, fentanyl, methadone, hydromorphone, oxycodone, and oxymorphone (22). We also explored the use of non-opioid analgesics including acetaminophen and nonsteroidal antiinflammatory drugs.

Covariables. Covariates of interest included factors considered likely to be associated with incident opioid use and were evaluated as confounders and potential mediators based on a priori hypotheses. Demographics, smoking, disease duration, household income, and education level were assessed at enrollment and considered potential confounders. Other obesity-related factors were considered to likely fall on the causal pathway between obesity and opioid use and included physical functioning, disability, comorbidity, and chronic pain. Physical functioning was quantified using the Health Assessment Questionnaire (HAQ), a validated patient-reported outcome measure (23). Participants were also asked whether they were unable to work due to disability. Comorbidity was assessed using the Rheumatic Disease Comorbidity Index, a validated quantitative measure of comorbid illness (24). Specific individual comorbidities (depression, hypertension, heart disease, cancer, diabetes mellitus) were also assessed separately. Patient assessment of overall physical and mental health was derived from the Short Form 36 health survey (25). Pain scores were assessed using a visual analog scale, and pain sensitization was assessed using the Fibromyalgia Symptom Severity Scale and the Widespread Pain Index (26).

Statistical analysis. Enrollment characteristics of the study population were described at baseline across BMI categories. Cox proportional hazards models were used to evaluate associations between enrollment BMI category and incident chronic opioid use among patients who were not using opioids at baseline. Multivariable models included potential confounders such as demographics, smoking, disease duration, baseline treatments (including the number of prior biologics), household income, and highest level of education achieved. In order to estimate the potential public health impact of obesity on chronic opioid use, we determined the cumulative incidence of chronic

Table 1. Characteristics of non-opioid users at enrollment by body mass index (BMI) category*

Characteristic	Underweight, <18.5 kg/m^2 (n = 387)	Normal, 18.5–25 kg/m^2 (n = 7,181)	Overweight, 25–30 kg/m^2 (n = 6,456)	Obese, 30–35 kg/m^2 (n = 3,431)	Severely obese, >35 kg/m^2 (n = 2,339)	*P*
Age, mean ± SD years	60.0 ± 16.9	59.5 ± 14.5	60.6 ± 12.6	59.1 ± 11.9	56.1 ± 11.6	<0.001
Female	350 (90)	5,939 (82.7)	4,520 (70.0)	2,549 (74.3)	2,002 (85.6)	<0.001
White	364 (94)	6,828 (95)	6,079 (94)	3,180 (93)	2,229 (91)	<0.001
Current smoking	32 (8)	384 (5.4)	292 (4.5)	177 (5.2)	102 (4.4)	<0.01
Education, mean ± SD years	13.9 ± 2.4	13.8 ± 2.3	13.6 ± 2.3	13.4 ± 2.4	13.5 ± 2.4	<0.001
Income (per $1,000), mean ± SD	47.4 ± 33.8	53.7 ± 32.9	51.9 ± 31.7	47.7 ± 30.6	47.5 ± 30.5	<0.001
RA factors						
Disease duration, mean ± SD years	15.9 ± 13.0	14.0 ± 12.4	13.2 ± 12.1	12.6 ± 12.0	12.0 ± 11.3	<0.001
Methotrexate use	210 (54)	4,045 (56)	3,729 (58)	1,890 (55.09)	1,322 (57)	0.099
Biologic use	121 (31)	2,612 (36)	2,358 (37)	1,352 (39)	997 (43)	<0.001
Prednisone use	171 (44)	2,730 (38)	2,424 (38)	1,259 (37)	778 (33)	<0.001
NSAID use	229 (59)	4,234 (59)	3,879 (60)	2,061 (60)	1,400 (60)	0.69
Acetaminophen use	85 (22)	1,590 (22)	1,400 (22)	699 (20)	535 (23)	0.19
Comorbidities						
Comorbidity score, mean ± SD	1.5 ± 1.5	1.3 ± 1.4	1.4 ± 1.4	1.7 ± 1.5	2.1 ± 1.6	<0.001
Kidney disease	8 (2.3)	104 (1.5)	119 (2.0)	69 (2.2)	47 (2.2)	0.11
Depression	36 (9.9)	675 (10.0)	630 (10.4)	453 (14.3)	500 (23)	<0.001
Diabetes mellitus	10 (2.7)	234 (3.5)	404 (6.7)	367 (27)	351 (26)	<0.001
Hypertension	66 (18)	1,418 (21)	1,847 (31)	1,263 (23)	1,030 (18)	<0.001
Heart disease	34 (9.3)	462 (6.8)	406 (6.7)	235 (7.4)	172 (8.0)	0.09
Pain and disability						
Work disabled	36 (9.1)	577 (8.0)	567 (8.8)	428 (13)	405 (20)	<0.001
HAQ score, mean ± SD (range 0–3)	1.04 ± 0.76	0.92 ± 0.70	0.93 ± 0.70	1.05 ± 0.70	1.25 ± 0.70	<0.001
Pain VAS score, mean ± SD (range 0–10)	3.5 ± 2.6	3.1 ± 2.6	3.3 ± 2.6	3.8 ± 2.7	4.4 ± 2.7	<0.001
FMSS score, mean ± SD (n = 8,891)	3.8 ± 2.3	3.5 ± 2.2	3.6 ± 2.2	4.0 ± 2.4	4.8 ± 2.5	<0.001
WPI score, mean ± SD (n = 8,970)	4.8 ± 5.1	4.0 ± 4.2	4.4 ± 4.5	5.1 ± 4.7	6.0 ± 5.2	<0.001
SF-36 MCS score, mean ± SD	50.6 ± 10.5	50.6 ± 10.4	50.8 ± 10.3	49.4 ± 11.0	47.3 ± 11.9	<0.001
SF-36 PCS score, mean ± SD	37.8 ± 11.2	40.4 ± 10.6	39.3 ± 10.5	37.0 ± 10.2	34.2 ± 9.9	<0.001

* Values are the number (%) unless indicated otherwise. FMSS = Fibromyalgia Symptom Severity Scale; HAQ = Health Assessment Questionnaire; MCS = mental component summary; NSAID = nonsteroidal antiinflammatory drugs; PCS = physical component summary; RA = rheumatoid arthritis; SF-36 = Short Form 36; VAS = visual analog scale; WPI = Widespread Pain Index.

opioid use at 5 years based on Cox models. This was performed by defining the baseline hazard and adjusting groups based on categorical risk. The population attributable fraction (PAF) and the number needed to treat (NNT) were calculated based on the absolute risk differences at 5 years. These metrics are used to quantify the public health impact of an exposure. The PAF can be interpreted as the proportion of the population that experienced the outcome that would be avoided were they not to have been exposed. In this study, the NNT represents the number of people for whom the exposure would need have been prevented in order to avoid a single outcome. We also explored adjusting models for obesity-associated risk factors that we considered to likely lie on the causal pathway between obesity and opioid use, including time-varying pain scores, comorbidity scores, HAQ scores, and the effect of work disability on obesity, to assess their contribution to the long-term risks associated with obesity. Finally, in order to consider the possibility of reverse causality (either due to pain-related weight gain or disease-related weight loss) affecting estimates of risk, we performed a sensitivity analysis assessing associations between BMI reported at age 30 years and incident chronic opioid use in a subset of participants with available data. We also performed 2 exploratory analyses to determine 1) whether pain and disability at enrollment were associated with significant weight gain, and 2) to determine if greater BMI at baseline was associated with worsening pain and disability. Analyses were performed using Stata software, version 14.2.

RESULTS

A total of 37,868 patients with RA were included, with a mean ± SD age of 64.4 ± 12.7 years and a mean ± SD BMI of 28.4 ± 6.9 kg/m^2. Obesity was observed in 34% of the population and severe obesity in 15%. The prevalence of any opioid use at enrollment was 27%, with only 5% reporting the use of strong opioids. The prevalence of opioid use at enrollment was greater in higher BMI categories (normal weight: 2,883 [23%]; overweight: 3,023 [25%]; obese: 2,032 [29%]; and severely obese: 2,010 [35%]; $P < 0.001$). This was also true for strong opioids (normal weight: 520 [4.2%]; overweight: 541 [4.5%]; obese: 385 [5.5%]; and severely obese: 458 [8.0%]; $P < 0.001$).

There were 19,794 total participants who did not report the use of opioids at baseline and had sufficient data to be included in further analyses examining incident use. The characteristics of this study population are shown in Table 1. Obese participants were younger and more likely to be female, non-White, and to report greater comorbidity, disability (including work disability), worse quality of life, lower income, greater pain, greater fibromyalgia symptoms, and higher regional pain scores. Obese participants were more likely to report the use of biologic therapies and less likely to report the use of glucocorticoids at baseline. There was no significant relationship between obesity and the use of non-opioid analgesics at baseline.

In models adjusting for age, sex, race, RA disease duration, smoking, methotrexate use, prednisone use, tumor necrosis factor inhibitor (TNFi) use, non-TNFi biologic use, prior biologic use, level of education, and household income, obesity was associated with a greater risk of incident chronic opioid use in a dose-dependent manner, with severe obesity being associated with a substantially higher risk of incident chronic use (adjusted hazard ratio [HR_{adj}] 1.74 [95% confidence interval (95% CI) 1.54–1.95], $P < 0.001$) (Table 2). Based on this model, the overall predicted incidence of any chronic opioid use in patients with RA was 15% at 5 years (Table 3). The predicted cumulative incidence at 5 years was 21% among severely obese patients and 13% among individuals of normal weight (Table 3), with an absolute risk difference of 8%. This absolute risk difference corresponds to an NNT of 13, suggesting that the prevention or reversal of severe obesity in 13 patients might prevent 1 occurrence of chronic opioid use at 5 years.

The overall population attributable fraction for overweight and obesity was 15% (Table 3). Notably, among severely obese patients, there were more cases of chronic opioid use at 5 years than would have been expected based on the rate observed in the individuals of normal weight (487 versus 304). Thus, among severely obese patients, 38% of opioid use was potentially attributable to the obesity. In sensitivity analyses, similar associations were observed between BMI at age 30 years and incident chronic opioid use. For example, patients who reported a severely obese BMI at age 30 years were at significantly higher risk of incident chronic opioid use (HR_{adj} 1.88 [95% CI 1.47–2.40], $P < 0.001$) compared to participants with normal BMI at age 30 years (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24341/abstract).

The overall incidence of strong opioid use was lower, with a predicted 5-year incidence of 1.8%. The relationship between obesity and opioid use, however, was somewhat stronger. For example, the risk of chronic strong opioid use was >2-fold higher for severe obesity compared to normal weight (HR_{adj} 2.11 [95% CI 1.64–1.71], $P < 0.001$) (Table 2). The absolute risk difference between severe obesity and normal weight was 1.4% at 5 years (NNT = 71). The overall population attributable risk fraction for obesity and being overweight was 26%. Among those with severe obesity, there were more cases of incident strong opioid use at 5 years than would have been expected at the rate observed in the normal weight category (199 versus 94). Thus, 52% of strong opioid use among severely obese patients was potentially attributable to obesity. In sensitivity analyses, patients who reported a severely obese BMI at age 30 years were also at a much higher risk of incident chronic strong opioid use (HR_{adj} 2.84 [95% CI 1.58–5.11]), $P < 0.001$) (see Supplementary Table 1, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24341/abstract). Other factors observed to be associated with chronic opioid use included smoking, female sex, longer disease duration, use of biologics, a greater number of total biologics used, prednisone use, lower income, and lower education level (see Supplementary

Table 2. Associations between body mass index (BMI) category and incident use of opioids among nonusers*

	Any opioid use among non-opioid users (n = 19,794)†		Strong opioid use among non-opioid users (n = 20,155)‡	
	HR_{adj} (95% CI)	*P*	HR_{adj} (95% CI)	*P*
BMI category				
Underweight	1.16 (0.88–1.52)	0.29	0.98 (0.39–2.41)	0.97
Normal	1 (ref.)	–	1 (ref.)	0.04
Overweight	1.22 (1.11–1.34)	<0.001	1.38 (1.51–1.81)	0.02
Obese	1.42 (1.27–1.58)	<0.001	1.54 (1.12–2.10)	0.007
Severely obese	1.73 (1.54–1.95)	<0.001	2.11 (1.64–2.71)	<0.001

* Adjusted for age, sex, race, rheumatoid arthritis disease duration, smoking, methotrexate use, prednisone use, tumor necrosis factor inhibitor (TNFi) use, non-TNFi biologic use, prior biologic use, level of education, and household income. 95% CI = 95% confidence interval; HR_{adj} = adjusted hazard ratio; ref. = reference.
† Person-years = 93,254; no. cases = 2,802.
‡ Person-years = 99,352; no. cases = 347.

Table 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24341/abstract).

To evaluate to what degree obesity-associated symptoms and comorbidities might explain the estimated associations between obesity and incident opioid use, we constructed models with additional adjustment for time-varying pain scores, comorbidity index, diabetes mellitus, hypertension, heart disease, and work disability. In these models, the associations between obesity and opioid use were partially attenuated for any opioid use (obese HR_{adj} 1.22 [95% CI 1.09–1.36]; severely obese HR_{adj} 1.30 [95% CI 1.14–1.48]) and for strong opioid use (obese HR_{adj} 1.18 [95% CI 0.85–1.66]; severely obese HR_{adj} 1.42 [95% CI 1.00–2.01]) (Table 4). The coefficients for obesity and severe obesity were reduced by 46% and 53%, respectively, with this adjustment. Higher pain scores, work disability, and higher comorbidity index were also each associated with incident chronic opioid use (see Supplementary Table 2, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24341/abstract).

Exploratory analyses suggested that greater pain and HAQ scores at baseline were not associated with significant changes in BMI over time. However, greater BMI at baseline was strongly associated with worsening of pain and HAQ scores (see Supplementary Table 3 and Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24341/abstract).

DISCUSSION

Results from this study demonstrated a substantially higher rate of incident chronic opioid use among obese patients with

Table 3. Estimated incidence, number of total cases, and the cases of chronic opioid use attributable to obesity in the specific group and in the entire population based on models from Table 2*

	No. of cases	Predicted 5-year incidence	Cases attributable to excess weight	PAF, %†
Any opioid use				
Underweight	387	57 (15)	NA	NA
Normal	7,181	905 (13)	0 (0)	0
Overweight	6,456	975 (15)	129 (13)	4.3
Obese	3,431	597 (17)	137 (23)	4.5
Severely obese	2,339	489 (21)	187 (38)	6.2
All	19,794	3,023 (15)	446	15
Strong opioid use				
Underweight	394	5 (1.3)	NA	NA
Normal	7,292	95 (1.3)	0 (0)	0
Overweight	6,556	118 (1.8)	33 (28)	9.3
Obese	3,509	70 (2.0)	25 (36)	7.1
Severely Obese	2,404	65 (2.7)	34 (52)	9.6
All	20,155	353 (1.8)	92	26

* Values are the number (%) unless indicated otherwise. NA = not applicable; PAF = population attributable fraction.
† The PAF represents the proportion of cases in the whole population that are attributable to this category of obesity. For example, 9.6% of all cases of chronic strong opioid use in the entire population are attributable to severe obesity.

Table 4. Predictors of overall incident chronic opioid and chronic strong opioid use before and after adjusting for additional factors considered likely to be in the causal pathway*

	Model 1 (n = 19,794)†		Model 2 (n = 18,511)‡	
	HR_{adj} (95% CI)	*P*	HR_{adj} (95% CI)	*P*
Any opioid use among nonusers				
BMI category				
Underweight	1.16 (0.88–1.52)	0.29	1.12 (0.85–1.48)	0.42
Normal	1 (ref.)	–	1 (ref.)	–
Overweight	1.22 (1.11–1.34)	<0.001	1.14 (1.04–1.26)	0.007
Obese	1.42 (1.27–1.58)	<0.001	1.22 (1.09–1.36)	0.001
Severely obese	1.73 (1.54–1.95)	<0.001	1.30 (1.15–1.48)	<0.001
Strong opioid use among nonusers				
BMI category				
Underweight	0.99 (0.40–2.44)	0.99	0.75 (0.27–2.04)	0.57
Normal	1 (ref.)	–	1 (ref.)	–
Overweight	1.38 (1.05–1.81)	0.02	1.24 (0.94–1.65)	0.13
Obese	1.53 (1.12–2.09)	0.008	1.18 (0.85–1.65)	0.32
Severely obese	2.11 (1.52–2.92)	<0.001	1.42 (1.00–2.01)	0.05

* 95% CI = 95% confidence interval; BMI = body mass index; HR_{adj} = adjusted hazard ratio; ref. = reference.
† Model 1: person-years = 93,254; cases = 2,802. Adjusted for age, sex, race, rheumatoid arthritis disease duration, smoking, methotrexate use, prednisone use, tumor necrosis factor inhibitor (TNFi) use, non-TNFi biologic use, prior biologic use, level of education, and household income.
‡ Model 2: person-years = 85,570; cases = 2,631. Model 1 plus adjustment for the rheumatic disease comorbidity index, diabetes mellitus, hypertension, heart disease, depression, work disability, pain score on visual analog scale, and Health Assessment Questionnaire score.

RA. The overall rate of incident chronic opioid use at 5 years was relatively high in this population (15%) and consistent with prior reports (6). In our study population, excess weight could potentially account for 15% of all incident chronic use and 26% of all incident strong opioid use in the population at 5 years. Strikingly, 52% of cases of strong opioid use at 5 years among severely obese patients were potentially attributable to the obesity itself, suggesting that prevention or reversal of severe obesity could have a substantial impact on reducing opioid use in this group.

These findings have important public health implications. Within this population at risk of chronic opioid use due to inflammatory arthritis, obesity was a critical risk factor. Our study also suggests that the risk of opioid use in obese patients may be related to the observation of greater pain, centralization of pain, disability, comorbidity, and poor quality of life in these participants over time. These factors have been strongly linked to obesity in other populations (13–18). We have previously shown that obese patients with RA had more rapid progression of disability over time than patients of normal weight, suggesting a direct contribution of obesity to the development of disability (18). In this study, when considered in regression models, factors such as pain, disability, and comorbidity explained approximately one-half of the measured effect. It is likely that other factors, not assessed in the current study, may be important, such as characteristics or location of pain, other comorbidities, and other behavioral or social factors. Because comorbidities such as pain and disability are likely caused by obesity, we considered it inappropriate to consider them as confounders in these analyses. It is, however, difficult to fully disentangle the complex relationship between obesity and these related comorbidities.

An alternative hypothesis might be that pain and disability lead to both obesity and opioid use. However, several analyses presented here suggest that this is less likely. For example, the observation that BMI at age 30 years has a comparable association with incident use suggests that reverse causality is not a substantial contributor to these observed relationships because BMI in early life will often temporally precede other risk factors such as pain and disability from inflammatory arthritis. Furthermore, in this population, greater HAQ score and pain were not significantly associated with subsequent changes in BMI. Overall, these data support the hypothesis that public health approaches to prevent or treat obesity could substantially reduce the use of chronic opioids among patients with RA through a reduction in pain and pain centralization, joint dysfunction and disability, and improvements in quality of life.

We are aware of 1 prior study characterizing the incident use of chronic opioids in patients with RA (6). This study found that chronic pain, antidepressant use, higher disease activity, and greater disability were all associated with a higher rate of chronic opioid use. Our study confirms these prior observations and builds on these data by evaluating obesity as a risk factor for chronic opioid use. Our study illustrates how obesity might act as a common etiologic factor that may lead to poor health status, disability, chronic pain, and ultimately opioid use.

While this study was not aimed at defining all risk factors for chronic opioid use in this population, we did identify a number of associations that deserve further study. In our population, evidence of longer disease duration, prior use of biologic drugs, and the use of prednisone were associated with the use of opioids, suggesting greater use in refractory and longstanding disease. An association with more severe disease is intuitive and confirms results from

prior studies (6). Current use of biologics was protective, perhaps related to better disease control or perhaps related to other health or social factors that may be different in individuals receiving these therapies. Differences in sex and associations with income and education level suggest that there are socioeconomic determinants of opioid use. Prior studies in the general population have suggested that chronic prescription opioid use and prescription opioid overdose are each associated with socioeconomic determinants (27,28). The effect of obesity in this study was independent of these socioeconomic determinants. Further study is necessary to evaluate each of these factors and their individual roles.

A limitation of these data is the self-report of the exposure and the lack of information regarding dosing of opioid medications. Self-report of BMI may lead to an underestimation of BMI, particularly among more obese individuals (29). The impact of this bias might be hypothesized to result in a further underestimation of the risks of obesity. In addition, we did not have access to objective measures of disease activity (i.e., swollen joint counts, inflammatory markers). Thus, one might speculate that the associations observed for obese patients with RA are driven by more severe inflammatory arthritis in this group. However, numerous reports suggest that obese patients with RA have more modest inflammatory disease measured by advanced imaging (30,31) and show reduced rates of radiographic damage progression, suggesting a milder phenotype of the inflammatory arthritis itself (32–35). This study may not be entirely generalizable to all populations, as risk factors for chronic opioid use and the prevalence of obesity will vary by population. However, this registry is a large and national sample of the US population of patients with RA. Since pain and poor health status might be hypothesized to lead to obesity, it is difficult to entirely rule out the possibility of an effect of reverse causality occurring (prior to study enrollment). However, a strength of the current study is the use of long-term longitudinal data and evaluation of incident chronic opioid use among nonusers at baseline. In addition, the observed associations with BMI at age 30 years suggest that excess weight likely preceded obesity-related factors in most cases. Finally, because obese patients more commonly use opioids at enrollment (and were therefore ineligible for the primary analysis), we may have underestimated the risk of obesity due to the exclusion of these prevalent users.

Strengths of the study include the longitudinal study design, long-term follow-up, robust clinical variables measured over time, the quantification of the potential public health impact, the assessment of potential mediators such as pain and disability, and the confirmation of results in sensitivity analyses using early life weight.

In conclusion, obesity is strongly linked to incident chronic opioid use among patients with RA. As much as 15% of all incident opioid use in this population, and 25% of strong opioid use, might be attributable to excess weight and its associated comorbidities. These data support a focus on the prevention and management of obesity in order to stem the chronic use of opioids in patients with arthritis.

ACKNOWLEDGMENTS

The authors thank the participants in the FORWARD databank.

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. Baker 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. Baker, Stokes, Michaud.

Acquisition of data. Baker, Michaud.

Analysis and interpretation of data. Baker, Stokes, Pedro, Mikuls, George, England, Sayles, Wolfe, Michaud.

REFERENCES

1. Quinn PD, Hur K, Chang Z, Krebs EE, Bair MJ, Scott EL, et al. Incident and long-term opioid therapy among patients with psychiatric conditions and medications: a national study of commercial health care claims. Pain 2017;158:140–8.
2. Katz C, El-Gabalawy R, Keyes KM, Martins SS, Sareen J. Risk factors for incident nonmedical prescription opioid use and abuse and dependence: results from a longitudinal nationally representative sample. Drug Alcohol Depend 2013;132:107–13.
3. Janakiram C, Fontelo P, Huser V, Chalmers NI, Lopez Mitnik G, Brow AR, et al. Opioid prescriptions for acute and chronic pain management among Medicaid beneficiaries. Am J Prev Med 2019;57:365–73.
4. Zamora-Legoff JA, Achenbach SJ, Crowson CS, Krause ML, Davis III JM, Matteson EL. Opioid use in patients with rheumatoid arthritis 2005–2014: a population-based comparative study. Clin Rheumatol 2016;35:1137–44.
5. Curtis JR, Xie F, Smith C, Saag KG, Chen L, Beukelman T, et al. Changing trends in opioid use among patients with rheumatoid arthritis in the United States. Arthritis Rheumatol 2017;69:1733–40.
6. Lee YC, Kremer J, Guan H, Greenberg J, Solomon DH. Chronic opioid use in rheumatoid arthritis: prevalence and predictors. Arthritis Rheumatol 2019;71:670–7.
7. Power JD, Perruccio AV, Gandhi R, Veillette C, Davey JR, Lewis SJ, et al. Factors associated with opioid use in presurgical knee, hip, and spine osteoarthritis patients. Arthritis Care Res (Hoboken) 2019;71:1178–85.
8. Ajeganova S, Andersson ML, Hafström I, for the BARFOT Study Group. Association of obesity with worse disease severity in rheumatoid arthritis as well as with comorbidities: a long-term followup from disease onset. Arthritis Care Res (Hoboken) 2013;65:78–87.
9. Katz P, Margaretten M, Trupin L, Schmajuk G, Yazdany J, Yelin E. Role of sleep disturbance, depression, obesity, and physical inactivity in fatigue in rheumatoid arthritis. Arthritis Care Res (Hoboken) 2016;68:81–90.
10. Humphreys JH, Verstappen SM, Mirjafari H, Bunn D, Lunt M, Bruce IN, et al. Association of morbid obesity with disability in early inflammatory polyarthritis: results from the Norfolk Arthritis Register. Arthritis Care Res (Hoboken) 2013;65:122–6.
11. George MD, Baker JF. The obesity epidemic and consequences for rheumatoid arthritis care. Curr Rheumatol Rep 2016;18:6.
12. Relieving pain in America: a blueprint for transforming prevention, care, education, and research. Mil Med 2016;181:397–9.
13. Shiri R, Karppinen J, Leino-Arjas P, Solovieva S, Viikari-Juntura E. The association between obesity and low back pain: a meta-analysis. Am J Epidemiol 2010;171:135–54.
14. Okifuji A, Hare BD. The association between chronic pain and obesity. J Pain Res 2015;8:399–408.

15. Kulkarni K, Karssiens T, Kumar V, Pandit H. Obesity and osteoarthritis. Maturitas 2016;89:22–8.
16. Jantaratnotai N, Mosikanon K, Lee Y, McIntyre RS. The interface of depression and obesity. Obes Res Clin Pract 2017;11:1–10.
17. Bigal ME, Lipton RB, Holland PR, Goadsby PJ. Obesity, migraine, and chronic migraine: possible mechanisms of interaction. Neurology 2007;68:1851–61.
18. Baker JF, England BR, Mikuls TR, Sayles H, Cannon GW, Sauer BC, et al. Obesity, weight loss, and progression of disability in rheumatoid arthritis. Arthritis Care Res (Hoboken) 2018;70:1740–7.
19. Michaud K. The National Data Bank for Rheumatic Diseases (NDB). Clin Exp Rheumatol 2016;34 Suppl 101:S100–1.
20. Wolfe F, Michaud K. A brief introduction to the National Data Bank for Rheumatic Diseases. Clin Exp Rheumatol 2005;23 Suppl 39:S168–71.
21. Obesity: preventing and managing the global epidemic: report of a WHO consultation. World Health Organ Tech Rep Ser 2000;894:i–xii, 1–253.
22. Wolfe F, Walitt BT, Katz RS, Lee YC, Michaud KD, Hauser W. Longitudinal patterns of analgesic and central acting drug use and associated effectiveness in fibromyalgia. Eur J Pain 2013;17:581–6.
23. Wolfe F, Kleinheksel SM, Cathey MA, Hawley DJ, Spitz PW, Fries JF. The clinical value of the Stanford Health Assessment Questionnaire Functional Disability Index in patients with rheumatoid arthritis. J Rheumatol 1988;15:1480–8.
24. England BR, Sayles H, Mikuls TR, Johnson DS, Michaud K. Validation of the Rheumatic Disease Comorbidity Index. Arthritis Care Res (Hoboken) 2015;67:865–72.
25. Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. BMJ 1992;305:160–4.
26. Wolfe F, Clauw DJ, Fitzcharles MA, Goldenberg DL, Katz RS, Mease P, et al. The American College of Rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity. Arthritis Care Res (Hoboken) 2010;62:600–10.
27. Pear VA, Ponicki WR, Gaidus A, Keyes KM, Martins SS, Fink DS, et al. Urban-rural variation in the socioeconomic determinants of opioid overdose. Drug Alcohol Depend 2019;195:66–73.
28. Svendsen K, Fredheim OM, Romundstad P, Borchgrevink PC, Skurtveit S. Persistent opioid use and socio-economic factors: a population-based study in Norway. Acta Anaesthesiol Scand 2014;58:437–45.
29. Kovalchik S. Validity of adult lifetime self-reported body weight. Public Health Nutr 2009;12:1072–7.
30. Mangnus L, Nieuwenhuis WP, van Steenbergen HW, Huizinga TW, Reijnierse M, van der Helm-van Mil AH. Body mass index and extent of MRI-detected inflammation: opposite effects in rheumatoid arthritis versus other arthritides and asymptomatic persons. Arthritis Res Ther 2016;18:245.
31. Baker JF, Ostergaard M, George M, Shults J, Emery P, Baker DG, et al. Greater body mass independently predicts less radiographic progression on X-ray and MRI over 1–2 years. Ann Rheum Dis 2014;73:1923–8.
32. Kaufmann J, Kielstein V, Kilian S, Stein G, Hein G. Relation between body mass index and radiological progression in patients with rheumatoid arthritis. J Rheumatol 2003;30:2350–5.
33. Van der Helm-van Mil AH, van der Kooij SM, Allaart CF, Toes RE, Huizinga TW. A high body mass index has a protective effect on the amount of joint destruction in small joints in early rheumatoid arthritis. Ann Rheum Dis 2008;67:769–74.
34. Westhoff G, Rau R, Zink A. Radiographic joint damage in early rheumatoid arthritis is highly dependent on body mass index. Arthritis Rheum 2007;56:3575–82.
35. Baker JF, George M, Baker DG, Toedter G, Feldt JM, Leonard MB. Associations between body mass, radiographic joint damage, adipokines, and risk factors for bone loss in rheumatoid arthritis. Rheumatology (Oxford) 2011;50:2100–7.

Arthritis Care & Research
Vol. 73, No. 10, October 2021, pp 1413–1424
DOI 10.1002/acr.24363

American College of Rheumatology
Empowering Rheumatology Professionals

Is There Any Role for Opioids in the Management of Knee and Hip Osteoarthritis? A Systematic Review and Meta-Analysis

Mikala C. Osani,[1] L. Stefan Lohmander,[2] and Raveendhara R. Bannuru[1]

Objective. Opioids have long been prescribed for chronic pain conditions, including osteoarthritis (OA). However, there is little information about their temporal efficacy, or differences in efficacy and safety between opioids with strong versus weak/intermediate μ opioid receptor–binding affinity. To explore these research questions, we conducted a systematic review and meta-analyses of randomized controlled trials (RCTs) conducted in patients with knee and/or hip OA.

Methods. We searched Medline, Embase, PubMed Central, and the Cochrane Central Register of Controlled Trials from inception to December 2019 and sought unpublished data. Placebo-controlled RCTs of oral opioids in patients with knee and/or hip OA were included. Standardized mean differences (SMDs) were calculated for pain and function at 2, 4, 8, and 12 weeks. Subgroup analyses for strong and weak/intermediate opioids were conducted. Meta-regression was performed to assess the impact of dosage (morphine equivalency) on pain relief. Risk ratios were calculated for safety at the final follow-up.

Results. A total of 18 RCTs (9,283 participants) were included. Opioids demonstrated small benefits on pain at each time point, with SMDs ranging from –0.28 (95% confidence interval [95% CI] –0.38, –0.17) to –0.19 (95% CI –0.29, –0.08); similar effects were observed for function. Strong opioids demonstrated consistently inferior efficacy and overall worse safety than weak/intermediate opioids. Meta-regression revealed that incremental pain relief achieved beyond 20–50-mg doses was not substantial in the context of increased safety risks.

Conclusion. Opioids provide minimal relief of OA symptoms within a 12-week period, and they are known to cause discomfort in a majority of patients. Clinicians and policy makers should reconsider the utility of opioids in the management of OA.

INTRODUCTION

The use of opioids in chronic, noncancer pain has long been controversial. The clinical use of these drugs increased considerably in the mid-1990s, contemporaneous to the widespread reframing of pain as "the fifth vital sign" that was adopted by numerous clinical and research societies in an effort to improve patient care and satisfaction (1–4). Indeed, some research suggested that opioids exert similar or even more profound benefits on pain in individuals with chronic conditions than nonsteroidal antiinflammatory drugs (NSAIDs) or acetaminophen (5–7). Opioids have also been considered favorable to conventional analgesics because they do not result in major organ toxicity, despite long-term use (1–3). For this reason, they have been recommended for patients experiencing chronic pain for whom NSAIDs or acetaminophen are contraindicated (4). All these reasons coupled with aggressive marketing campaigns contributed to the dramatic increases in opioid prescriptions worldwide (2,8–18). However, concern about the devastating potential for opioid misuse among chronic pain patients has come to the forefront in the past decade (13,19–21). More recently, higher-quality evidence syntheses have also reported a lack of benefit of opioid medications on pain in individuals with chronic conditions, with notably low rates of tolerability of the drugs (22–28).

Dr. Bannuru's work was supported by the National Center for Complementary and Integrative Health (K23-AT-009374).

[1]Mikala C. Osani, BA, Raveendhara R. Bannuru, MD, PhD: Tufts Medical Center, Boston, Massachusetts; [2]L. Stefan Lohmander, MD, PhD: Lund University, Lund, Sweden.

Dr. Lohmander has received consulting fees from Arthro Therapeutics AB, GlaxoSmithKline, Janssen Research and Development, Pfizer, and Regeneron (less than $10,000 each). No other disclosures relevant to this article were reported.

Address correspondence to Raveendhara R. Bannuru, MD, PhD, Center for Treatment Comparison and Integrative Analysis, Division of Rheumatology, Box 406, Tufts Medical Center, 800 Washington Street, Boston, MA 02111. Email: rbannuru@tuftsmedicalcenter.org.

Submitted for publication December 19, 2019; accepted in revised form June 16, 2020.

SIGNIFICANCE & INNOVATIONS

- Opioids demonstrated only small benefits on pain and function at 2, 4, 8, and 12 weeks, which decreased with time.
- Strong opioids consistently demonstrated inferior effects to weak/intermediate opioids on pain and function.
- Participants who received strong opioids were more likely to experience more adverse events than those who received weak/intermediate opioids.

Osteoarthritis (OA) is one of the most common conditions contributing to the prevalence and global burden of chronic musculoskeletal pain worldwide (29,30). More recent clinical practice guidelines for OA have generally taken a cautious approach in recommending opioid use (24,25,31), with statements ranging from absolute avoidance of the drugs in any patient subgroup (24) to favorable recommendations for patients whose disease has failed medical therapy, in whom surgical intervention is contraindicated (31).

Recent meta-analyses assessing the efficacy and safety of opioids versus placebo in patients with OA showed little evidence that the drugs are beneficial on pain or function, and reported high rates of adverse events (AEs) (22,23,26–28). However, all reviews that assessed efficacy did so at the final follow-up time, masking potential temporal changes in the action of opioids on pain and/or function. Temporal assessments can reveal peak periods of efficacy and can provide clinicians with a blueprint for optimal regimen lengths for particular pharmacologic treatments (32,33). Furthermore, the impact of μ receptor–binding affinity (strong versus weak or intermediate opioids) on the observed efficacy of opioids was not adequately assessed, with 1 review conducting analyses specific to single drug formulations, rather than by binding affinity class (23). Likewise, none of these reviews assessed the impact of opioid medications on sleep quality or depression.

In light of the apparent lack of benefit of opioids on pain and function, additional information on these outcomes may shed light on other potential benefits opioids may have specific to their psychoactive and/or sedative effects. Some researchers have recommended closely monitored use of opioids in chronic pain for individuals who are experiencing sleep deprivation due to pain (34), although there is a possibility that chronic opioid use can result in long-term disruptions to sleep architecture, including the inhibition of the rapid eye movement (REM) stage of sleep, which is essential to adequate rest and recovery (35,36). Assessing mood and sleep quality variables is also important in understanding the relationship between analgesics with psychoactive and/or sedative effects, because diagnosed depression and self-reported sleep deprivation have both been associated with higher levels of pain (37,38). The relationship between these variables may provide a clue as to the muted efficacy response observed in patients with OA receiving opioids.

We aimed to characterize temporal patterns in pain relief and functional improvement among participants receiving oral opioids for knee or hip OA and to assess the safety profile of the drugs. Meta-analyses of opioid effects on pain and function were conducted for 2-, 4-, 8-, and 12-week time points. Other relevant patient-reported outcomes were assessed, including quality of life, sleep quality, and depression. Differences in efficacy and safety between strong and weak/intermediate opioids were also evaluated.

MATERIALS AND METHODS

This study was performed according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (see Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract) (39). We were unable to register our protocol at the PROSPERO registry (40), because a portion of the quality assessment was previously completed by our team as part of a different project (24). The analyses and results described within this article have not been published elsewhere.

Data sources/searches. A systematic search strategy was implemented of Medline, Embase, PubMed Central, and the Cochrane Central Register of Controlled Trials from inception to April 2019 (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract). Unpublished data were sought through ClinicalTrials.gov, and we reviewed the reference lists of relevant systematic reviews, meta-analyses, and supplements of conference proceedings that had been published up to April 2019 by hand. The electronic search was updated and reviewed again in December 2019. Our search criteria included randomized controlled trials (RCTs) that tested the effects of any US Food and Drug Administration (FDA)–approved oral opioid medication on participants diagnosed with musculoskeletal disorders. No restrictions were placed on publication date, status, or language.

Study selection. Placebo-controlled randomized trials assessing the efficacy and/or safety of FDA-approved oral opioid drugs in participants with knee and/or hip OA were included. Data from all dosage groups were collected from dose-ranging studies and pooled for the final analyses. RCTs incorporating combination treatment with NSAIDs in the study protocol were excluded. Studies that focused on peri-operative outcomes were not eligible. Studies in which the location of OA was undefined were excluded. Studies using enriched enrollment or inappropriate placebo controls (e.g., double-dummy designs incorporating non-oral routes of administration) were also excluded.

References gathered by the systematic search underwent an initial abstract screening in which 2 independent reviewers assessed them for potential eligibility according to preestablished

inclusion and exclusion criteria (MCO and RRB). The full manuscripts of references that were included in the initial round of screening were reviewed more thoroughly for suitability in the analysis (MCO and RRB). Discordant results in inclusion or exclusion that resulted during either screening stage were discussed and resolved; in the event that a disagreement in eligibility could not be resolved by consensus, we consulted a third reviewer (LSL).

Data extraction and quality assessment. A data extraction form was created to collect information on study and population characteristics, opioid classification, dosage and frequency, rescue medication protocol, and relevant efficacy and safety outcomes. Data were independently extracted from each included RCT by 2 reviewers (MCO and RRB). Opioid classes were defined by the strength of μ opioid receptor–binding affinity, as either strong (hydromorphone, morphine, oxycodone, oxymorphone) or weak/intermediate (codeine, hydrocodone, tapentadol, tramadol) (41,42). Data were collected at all reported time points and grouped into the following time point categories: 2 weeks (0–2 weeks), 4 weeks (3–6 weeks), 8 weeks (7–10 weeks), and 12 weeks (11–16 weeks) (32). We collected outcome data that were reported by any validated scale; in the event that >1 scale was reported, results for all scales were collected. Data that were presented in manuscript text or tables were prioritized over graphical data; data that were available only via figures or graphs were recovered using Engauge Digitizer and double-checked by a second reviewer (43). Wherever possible, intent-to-treat analyses were preferred.

Quality was independently assessed at the study level by 2 reviewers (MCO and RRB) using the Cochrane Risk of Bias Tool (44). To assess overall study quality, we developed criteria for very low, low, moderate, and high quality a priori; details on quality designations can be found in Supplementary Appendix B, page 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract. Quality was assessed at the outcome level using GRADE (Grading of Recommendations Assessment, Development and Evaluation) methodology (45). GRADE evidence profiles were constructed for the final follow-up time. GRADE quality assessment was undertaken by 2 independent reviewers (MCO and RRB), who resolved conflicts through discussion and consensus. A GRADE quality assessment rubric was designed a priori and used as a reference during the assessment (see Supplementary Table 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract).

Outcome definitions. The primary outcome was pain. Secondary outcomes included functional status, quality of life, sleep quality, depression, opioid withdrawal symptoms, discontinuation due to lack of efficacy, discontinuation due to AEs, rates of AEs and serious AEs, rates of gastrointestinal AEs, and rates of somnolence. Continuous outcomes were reported as the mean change from baseline to follow-up. We established an a priori extraction hierarchy for pain and functional outcome scales using the Cochrane Musculoskeletal Research Group's list of proposed outcomes (46). In analyses of pain and function, Western Ontario and McMaster Universities Osteoarthritis Index scales were prioritized (47); in analyses of quality of life, component summary scores of the Medical Outcomes Survey Short Form 36-item questionnaire (48) took precedence. If no other scales were available, nonstandard Likert scales were included in analyses of pain and function. Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract, provides additional details of the methods used to collect and assess different outcomes.

Statistical analysis. For continuous outcomes, we calculated standardized mean differences (SMDs) and 95% confidence intervals (95% CIs) as the bias-corrected Hedges' g statistic, using the mean change from baseline to follow-up (49). Meta-analyses were conducted using the inverse variance method and employed random effects models, as described by DerSimonian and Laird, to account for methodologic and clinical heterogeneity (50). Dichotomous outcomes were analyzed using the Mantel-Haenszel method and reported the effects as risk ratios and 95% CIs (51). Heterogeneity was assessed using the I^2 statistic (52). Subgroup analyses of broad groupings of strong and weak/intermediate opioids were conducted for pain at every time point, and for safety outcomes. Subgroup analyses of pain isolating RCTs (or study arms of relevant RCTs) involving tramadol were also planned for every time point a priori, to investigate potential differences in pain relief that may result from its dual mechanism of action (53). Meta-regression was performed to assess the impact of dosage (morphine equivalency) on pain relief. Morphine equivalency was calculated from the total daily dose of opioid using an online conversion calculator that was created using Centers for Disease Control and Prevention benchmarks (https://www.oregonpainguidance.org/opioidmedcalculator). For studies involving variable dosing of opioids by individual participant tolerance, the mean daily dose was used to calculate morphine equivalency. Subgroup analyses based on dose were performed on 4 dose groups <20 mg, 20–50 mg, 51–100 mg, and >100 mg. The potential for publication bias was evaluated using Egger's test and visual inspection of funnel plots (54). All analyses were conducted using R, version 3.6.1 with the "meta" package (55).

RESULTS

The systematic search returned 3,421 potentially relevant abstracts. Of these, 173 were eligible for full text review, and 18 RCTs involving 9,283 participants were included in final analyses (Figure 1). Supplementary Table 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract, provides the reasons for late-stage exclusion for studies that originally passed full-text review but were ultimately

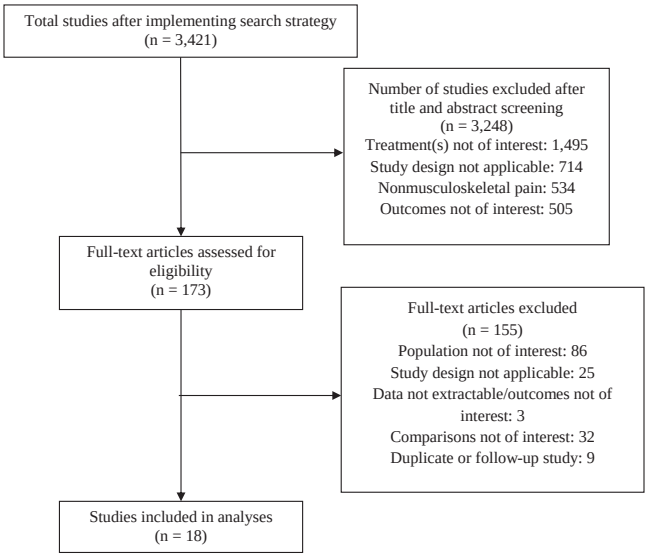


Figure 1. Study flow diagram.

deemed ineligible. Included studies were published between the years 2000 and 2017; data from 2 unpublished studies were gathered from ClinicalTrials.gov.

Study characteristics and demographics of the included RCTs are shown in Supplementary Table 4, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract. Seven RCTs involved patients with knee OA only, and 11 RCTs involved mixed populations of patients with knee and/or hip OA. No eligible RCTs involved patients with hip OA alone. Eleven trials compared weak/intermediate opioids to placebo: codeine (n = 1) (56), tapentadol (n = 3) (57–59), and tramadol (n = 7) (60–66); 10 trials compared strong opioids to placebo: morphine (n = 1) (67), hydromorphone (n = 2) (68,69), oxycodone (n = 5) (57–59,70,71), and oxymorphone (n = 2) (72,73); three 3-armed trials involved both strong and weak opioids (57–59); and one 3-armed trial compared 2 different strong opioids against placebo (72). The follow-up duration of the included RCTs ranged from 10 days to 16 weeks (median 12 weeks); 61% of RCTs had a follow-up of at least 12 weeks.

Demographic characteristics among study participants were relatively consistent across studies and were comparable to those expected among patients with OA. The proportion of female participants in the included RCTs ranged from 49% to 73% (median 62%). The mean age ranged from 54 to 67 years (median 61 years), and mean body mass index (BMI) ranged from 27.9 to 34.3 kg/m^2 (median 33.5 kg/m^2); BMI was not reported in 44% of studies. Four studies (22%) reported that the majority of participants' previous experience with opioids classified them as nonopioid users (58,67,71,73). Eleven RCTs (61%) reported patients' previous experience with opioids at baseline; however, those studies did not report the exact percentages of individuals who were opioid experienced at baseline. Limited doses of acetaminophen were the most common type of rescue medication (56% of RCTs) allowed during the course of the follow-up; 2 studies did not permit rescue medication usage.

Quality assessment of the included trials showed that the majority of trials were of low to very low quality, with the majority of high risk-of-bias ratings being related to attrition bias and potential reporting bias (Figure 2 and Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract). Aside from attrition bias concerns, the methodologic quality of the trials was moderate overall. Of 18 included RCTs, only 3 (17%) did not report funding sources in adequate detail (56,64,67), and the remaining 15 studies reported industry sponsorship and/or direct industry involvement of ≥1 investigator. Visual inspection of funnel plots did not reveal evidence of publication bias (see Supplementary Figure 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract). The overall quality of evidence at the outcome level as assessed by GRADE was low (see Supplementary Table 5, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract). Downgrades were most commonly attributed to risk-of-bias concerns and heterogeneity.

Temporal effects of opioids on pain and function. Opioids demonstrated small benefits on pain over placebo at every follow-up time through 12 weeks (see Supplementary Figure 3A and Supplementary Table 6, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract). The largest effects were observed at 2 and 4 weeks, with SMDs of –0.27 (95% CI –0.37, –0.17; n = 11 RCTs) and –0.28 (95% CI –0.38, –0.17; n = 9 RCTs), respectively (see Supplementary Figures 4 and 5, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract). Benefits on pain decreased notably (32%) from 4 weeks to 8 weeks and remained very small until the 12-week follow-up, with SMDs of –0.19 (95% CI –0.29, –0.08; n = 6 RCTs) and –0.21 (95% CI –0.30, –0.11; n = 10 RCTs), respectively (see Supplementary Tables 5–7, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract). All analyses demonstrated moderate heterogeneity, with I^2 ranging from 59% to 67%. Analyses isolating RCTs comparing tramadol to placebo demonstrated nearly identical effects on pain to those including all RCTs at every time point, with lower heterogeneity (see Supplementary Table 6, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract).

The effects of opioids on functional outcomes were similarly small (see Supplementary Figure 3B, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract). The functional benefits of opioids were most pronounced at 4 weeks (–0.26 [95% CI –0.37, –0.15]; n = 5 RCTs). Heterogeneity among functional analyses was more variable, with values ranging from 0% at 2 weeks to 77% at 8 weeks.

The overall quality of evidence for both pain and functional outcomes at the final follow-up time, as assessed by GRADE, was low due to risk-of-bias concerns and inconsistency ratings that resulted from moderate-to-high heterogeneity (see Supplementary Table 5, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract).

Overall effects of opioids on quality of life, sleep quality, and depression. Small benefits in sleep quality were observed in participants receiving opioids at the last follow-up (0.20 [95% CI 0.14, 0.27]; n = 9 RCTs; I^2 = 0%); effects did not differ between strong and weak/intermediate opioids (see Supplementary Table 6, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract). The quality of evidence for sleep was assessed to be moderate. Opioids demonstrated no benefits on quality of life (SMD 0.04 [95% CI –0.10, 0.17]; n = 7 RCTs; I^2 = 77%) at the final follow-up. The quality of evidence for quality of life was assessed to be very low due to severe heterogeneity and risk-of-bias concerns. Participants receiving opioids reported slightly worse depression scores compared to those who received placebo at the 12-week follow-up in the 1 RCT that reported the outcome (–0.15 [95% CI –0.28, –0.02]); the GRADE quality of evidence for this outcome was moderate, but inconsistency could not be assessed due to the fact that only 1 RCT contributed to this analysis.

Strong versus weak/intermediate opioids for pain. With respect to pain relief, strong opioids consistently underperformed compared to weak/intermediate opioids versus placebo (Figure 3 and Supplementary Figures 4–7, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract). Weak/intermediate opioids demonstrated effect sizes that were 19%, 32%, and 130% greater than the effect sizes for strong opioids at 2, 4, and 8 weeks, respectively. The difference between the effects of weak/intermediate opioids versus strong opioids was not as pronounced at 12 weeks, with weak/intermediate groups demonstrating effects that were 16% larger than those observed for strong opioids versus placebo. To further examine this relationship, post hoc meta-analyses assessing direct comparisons of strong versus weak/intermediate opioids were conducted for 2-, 4-, 8-, and 12-week time points using data from three 3-armed RCTs; the only drug comparison addressed by these 3 studies was tapentadol versus oxycodone (57–59). The results of these analyses reflected similar trends observed in our indirect comparisons of strong versus weak/intermediate opioids, with tapentadol demonstrating superiority over oxycodone at 2-, 4-, and 8-week time points, and with effects of the 2 groups converging at 12 weeks; tapentadol demonstrated statistically significant superiority at weeks 4 and 8 (see Supplementary Table 7, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract). Meta-regression

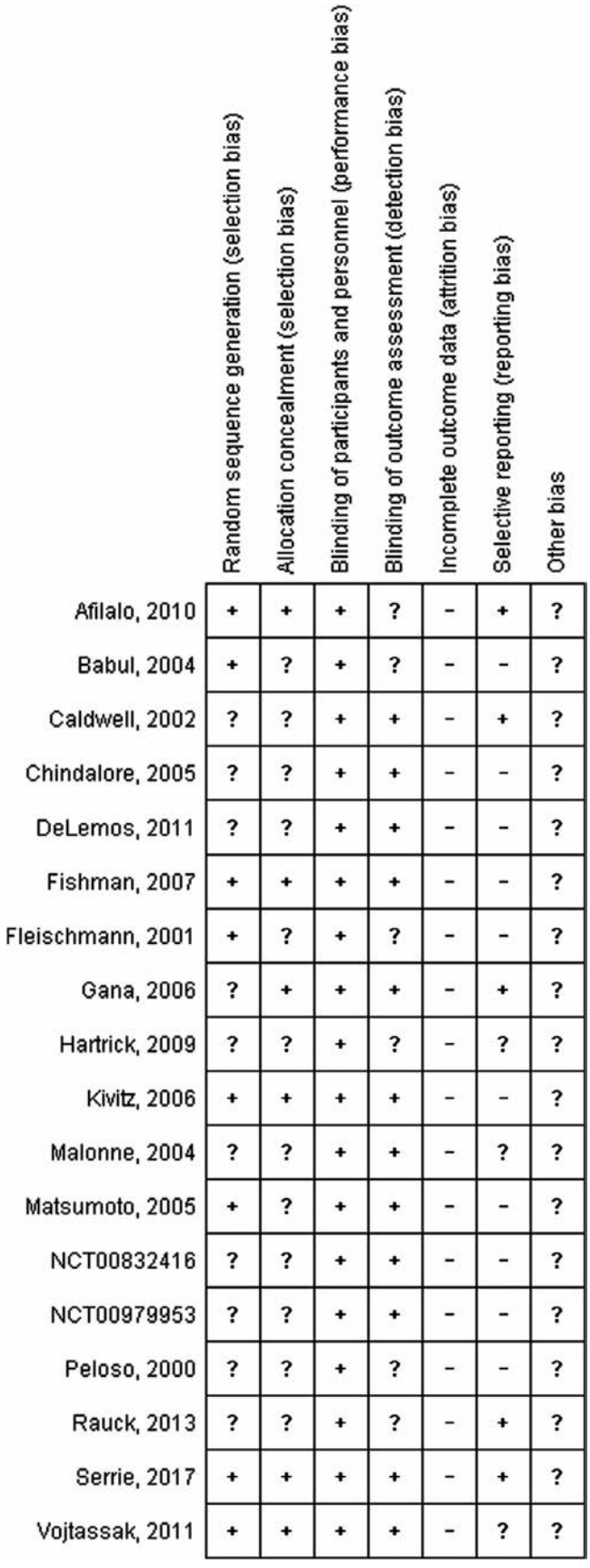


Figure 2. Risk-of-bias summary of included studies. + = low risk of bias; ? = moderate risk of bias; – = high risk of bias.

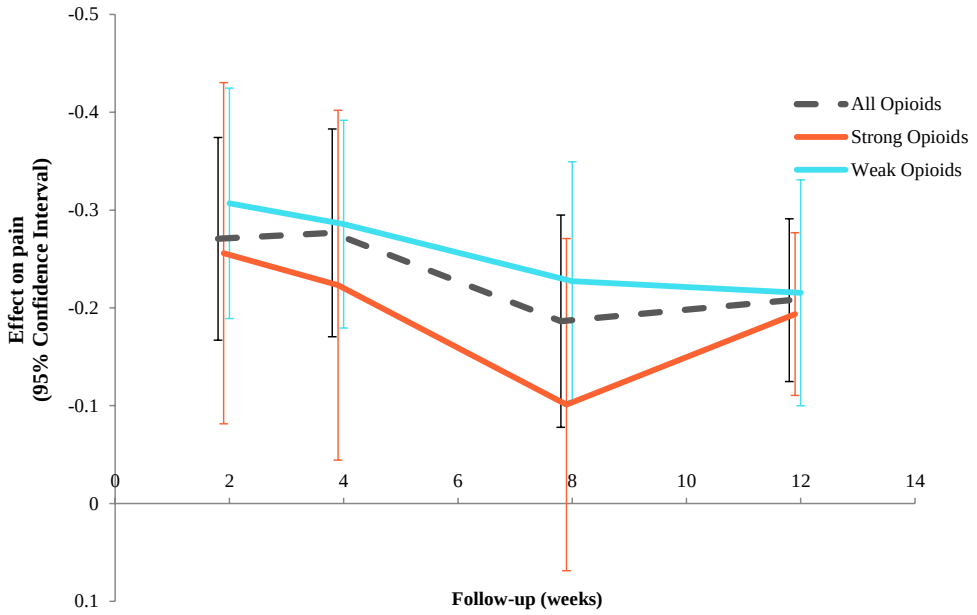


Figure 3. Pain trajectories based on opioid strength.

exploring dose effects revealed that opioid dosage (measured as morphine equivalency) was related to the magnitude of pain relief observed at the final follow-up time (P = 0.03), and that dose effects accounted for approximately 18% of the variability in standardized effect sizes for pain (R^2 = 18.05%) (Figure 4). In the subgroup analyses based on morphine dose equivalency ranges (<20 mg/day, 20–50 mg/day, 51–100 mg/day, or >100 mg/day), the effect sizes for pain tended to increase with an increase in dose, ranging from –0.11 (95% CI –0.33, 0.12) for <20 mg/day to –0.40 (95% CI –0.57, –0.23) for >100 mg/day (see Supplementary Table 8, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract). However, the clinical relevance of the possible increases in pain relief achieved beyond a 20–50 mg dose was questionable in light of the additional risks encountered at these doses (see Supplementary Table 8).

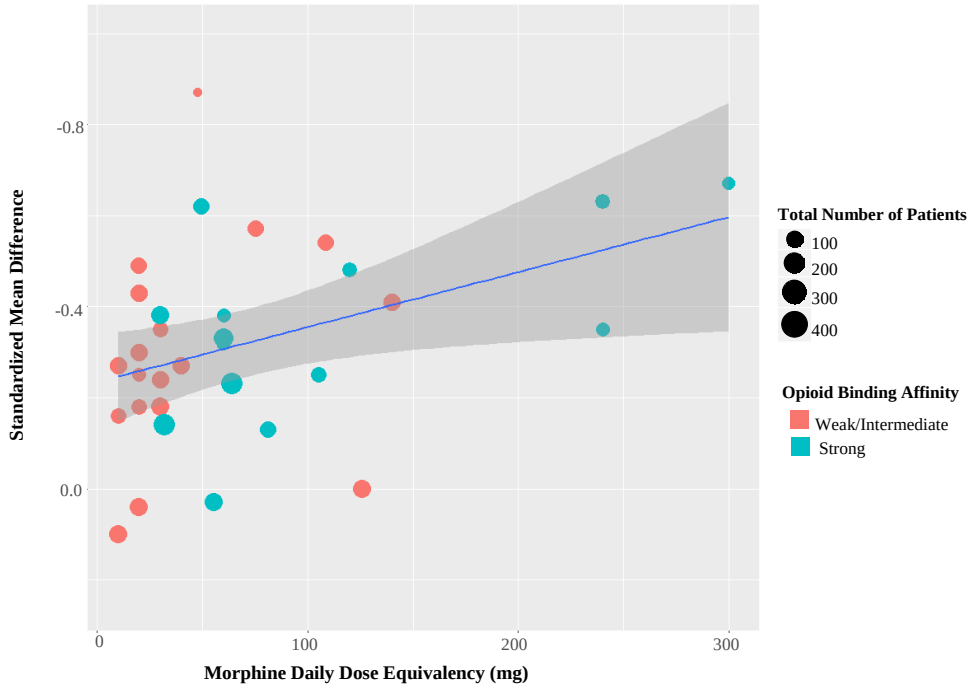


Figure 4. Relationship between opioid dose and pain relief at final follow-up. Different dosage groups within the same study were assessed as separate observations; therefore, 32 comparisons versus placebo were assessed.

Strong versus weak/intermediate opioids for function. With respect to physical function outcomes, strong opioids underperformed compared to weak/intermediate opioids versus placebo at 4-week and 12-week follow-up times, with effect sizes that were 29% and 5% smaller, respectively (see Supplementary Table 6, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract). Comparison of strong versus weak/intermediate opioids at 8 weeks was not possible, since no RCT involving strong opioids reported functional outcomes at 8 weeks.

Overall safety of opioids. Participants receiving opioids were more likely to discontinue treatment during the study period due to AEs (risk ratio 3.88 [95% CI 3.18, 4.74]) (Table 1). Opioid recipients were 1.5 times more likely to experience any AE or serious AE. The risks of experiencing a gastrointestinal AE or somnolence were 3.5 times and 4.1 times higher, respectively, in participants receiving opioids than placebo. The most common AEs reported included nausea, constipation, diarrhea, and vomiting. Interpretation of the results of analyses of total AEs and gastrointestinal AEs was limited by extremely high heterogeneity (Table 1).

The risk of opioid withdrawal symptoms was not markedly different from placebo, but data on this outcome were reported in only 3 trials. GRADE quality assessment of safety outcomes ranged from very low to moderate, primarily on the basis of heterogeneity (see Supplementary Table 5, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract). Heterogeneity was specifically noted in analyses of total AEs and gastrointestinal AEs, most likely due to a lack of standardization in reporting of these measures (e.g., reporting a sum total versus individual events).

Comparison of the safety of strong versus weak opioids. The relative risk of discontinuation due to AEs was higher among participants receiving strong opioids versus those receiving weak/intermediate opioids (relative risk 5.44 [95% CI 4.50, 6.57] versus relative risk 2.92 [95% CI 2.32, 3.66]) (Table 1). The relative risks of experiencing any AE, experiencing a gastrointestinal AE, and experiencing somnolence were similar for the 2

Table 1. Safety results*

Class	Follow-up, median (range) weeks	No. RCTs (no. patients)†	Risk ratio vs. placebo (95% CI)‡	I^2, %
Opioid withdrawal symptoms				
All opioids	12 (12–13)	3 (2,211)	1.96 (0.87, 4.41)	46
Strong opioids	12 (NA)	2 (811)	2.78 (1.41, 5.49)§	43
Weak opioids	12 (12–13)	3 (1,796)	1.06 (0.19, 5.80)	64
Discontinuation due to adverse events				
All opioids	12 (1.43–16)	18 (9,314)	3.88 (3.18, 4.74)§	41
Strong opioids	12 (1.43–16)	10 (4,686)	5.44 (4.50, 6.57)§	0
Weak opioids	12 (1.43–13)	11 (5,474)	2.92 (2.32, 3.66)§	36
Discontinuation due to lack of efficacy				
All opioids	12 (1.43–16)	16 (9,314)	0.48 (0.42, 0.55)§	29
Strong opioids	12 (1.43–16)	9 (4,686)	0.37 (0.29, 0.49)§	38
Weak opioids	12 (1.43–13)	10 (5,244)	0.55 (0.48, 0.62)§	0
Total adverse events				
All opioids	12 (1.43–16)	14 (8,263)	1.50 (1.34, 1.69)§	89
Strong opioids	12 (1.43–16)	7 (3,661)	1.63 (1.36, 1.95)§	93
Weak opioids	12 (1.43–13)	10 (5,452)	1.41 (1.22, 1.64)§	89
Serious adverse events				
All opioids	12 (1.43–16)	12 (7,439)	1.51 (1.08, 2.11)§	0
Strong opioids	12 (1.43–16)	6 (3,165)	1.66 (0.96, 2.88)	9
Weak opioids	12 (1.43–13)	8 (5,117)	1.37 (0.88, 2.12)	3
Gastrointestinal adverse events				
All opioids	12 (1.43–16)	16 (7,777)	3.49 (2.85, 4.27)§	79
Strong opioids	12 (1.43–16)	7 (3,479)	3.42 (2.64, 4.43)§	79
Weak opioids	12 (1.43–13)	10 (4,972)	3.12 (2.28, 4.26)§	87
Somnolence				
All opioids	12 (1.43–16)	15 (8,320)	4.07 (3.25, 5.09)§	0
Strong opioids	12 (1.43–16)	9 (4,391)	4.06 (3.15, 5.24)§	0
Weak opioids	12 (1.43–13)	9 (4,772)	4.27 (2.85, 6.41)§	29

* 95% CI = 95% confidence interval; NA = not applicable; RCT = randomized controlled trial.
† The number of RCTs overlaps in some cases due to 3-arm trials including both strong and weak opioids.
‡ Risk ratios >1 indicate greater risk of an event in the opioid group.
§ Statistically significant.

groups, compared to placebo. Participants receiving strong opioids were more likely to report symptoms of opioid withdrawal versus placebo.

DISCUSSION

We aimed to update and enhance the evidence base for opioid use in knee and hip OA to provide a holistic assessment of what role, if any, opioids may play in a contemporary OA treatment regimen. The results of our study showed that opioids overall demonstrated only small benefits on pain and function, compared with placebo, in participants with OA from 2 to 12 weeks of treatment. The use of opioids contributed no measurable benefit to quality of life versus placebo. The risk of experiencing any adverse safety outcome of interest, with the exception of opioid withdrawal symptoms, was higher among participants receiving any type of opioid. Interestingly, strong opioids consistently underperformed compared to weak/intermediate opioids for pain and function outcomes. Participants receiving strong opioids were also at a greater risk of experiencing any safety outcome than those who received weak/intermediate opioids.

A recent study that evaluated the therapeutic trajectory of NSAIDs found that they exerted moderate effects on pain, peaking within the first 2 to 8 weeks of treatment (SMDs ranging from –0.43 to –0.36) and plateauing by 26 weeks (SMD –0.21) (33). In comparison, the magnitudes of the effect sizes for pain demonstrated by opioids over time in the current study were consistently small, ranging from –0.27 at 2 weeks to –0.21 by 12 weeks. Put into context, the largest effect sizes observed for opioids versus placebo at 2 weeks did not exceed the treatment effect (–0.29) of intraarticular placebos versus oral placebos observed in a network meta-analysis of common treatments for OA (74). Thus, these results call into question the clinical relevance and utility of opioid drugs in both short- and long-term OA pain management regimens.

Our results are consistent with recent studies assessing the efficacy and safety of opioids versus placebo in participants with OA, despite methodologic differences (22,23,26–28). The review by Megale et al focused on broad musculoskeletal pain and incorporated data from both transdermal and oral opioids; likewise, the review by Fuggle et al included OA at any location (22,27). To constrain heterogeneity, our inclusion criteria were limited to RCTs involving oral opioid use in patients with knee and/or hip OA. The findings of our study are also in agreement with the results of a review by Da Costa et al (23). Although these authors analyzed individual opioid formulations separately, there was a general trend for weak/intermediate opioids to demonstrate larger effects on pain than strong opioids at the final follow-up time. Our study differed from these 2 studies in that it included some additional studies, including 1 newly published RCT and 2 unpublished studies (58,63,70).

Most reviews included enriched enrollment with standard study designs in their analyses (75) and also included RCTs that did not meet our inclusion standards (see Supplementary Table 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract) (76–85). We thoroughly evaluated the therapeutic trajectory of oral opioids for knee and hip OA and attempted to tease out differences between strong and weak/intermediate opioids, in an effort to offer better clinical insight. We also provided specific data on the pain relief trajectory of tramadol and found that it offered comparable effects to other opioids. This study also reports data on sleep quality and depression, as well as clinically relevant AEs.

An interesting finding from our study is the underperformance of strong opioids. The most likely explanation for this finding could be the relationship between pain relief and the tolerability of opioids based on dose. We observed a relevant relationship between morphine dose equivalency and the magnitude of pain relief at the final follow-up (Figure 4). However, the relative risk of discontinuation due to AEs among participants receiving strong opioids was nearly twice that of participants receiving weak/intermediate opioids (Table 1). Many participants who received strong opioids may have been unable to achieve the optimal therapeutic dose as a result of attrition related to a lack of tolerability. Participants who might have received the optimal therapeutic dosage and started experiencing superior pain relief would have withdrawn from the trials due to AEs, leaving only the participants with suboptimal therapeutic doses, ultimately biasing the results toward the null.

Participants who received opioids reported experiencing better sleep quality by the final follow-up time; however, the benefits of opioids on sleep quality did not differ based on weak/intermediate or strong opioid-binding affinity (SMD 0.21 [95% CI 0.11, 0.31] versus 0.20 [95% CI 0.10, 0.30], respectively). Opioid users were also more than 4 times more likely to report somnolence, or daytime drowsiness. Since scales that measure sleep quality rely on self-report, we are unable to clarify what aspects of sleep were improved by opioid use from patients' perspectives. The benefits to patients' sleep that were reported may have been more focused on ease of falling asleep, and the impacts of opioids on REM sleep may have contributed to the incidence of somnolence. However, the simple act of taking an opioid during the day may have increased the likelihood of experiencing drowsiness at inappropriate times.

The results of this study are subject to some limitations. As in any meta-analysis, the quality and credibility of our results are limited by the quality and quantity of the contributing RCTs. Overall, the quality of the RCTs that contributed to our analyses was assessed to be low to very low, with the majority of methodologic concerns pertaining to attrition bias as well as reporting bias related to inadequate reporting of baseline data. There was specifically a trend for differential rates in discontinuation due to AEs in the opioid groups and lack of treatment efficacy in the placebo groups. Participants in the opioid groups were approximately 4 times more likely to withdraw from the study due to AEs

compared to those in the placebo group. Possibly the efficacy data from patients who were responders, or potential responders, to opioid treatment who experienced issues with tolerance were not captured, resulting in a bias toward the null (see Supplementary Figure 8, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract). We also observed lower rates of discontinuation due to lack of efficacy in the opioid groups (risk ratio 0.48 [95% CI 0.42, 0.55]; n = 16 RCTs) (Table 1). Differential rates of discontinuation due to lack of efficacy could result in inflated placebo effects by excluding participants in the placebo group who experienced the least benefit from the full study follow-up (see Supplementary Figure 9, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24363/abstract), resulting in a bias toward the null. These factors might have resulted in underestimating the true effect of opioids, but the fact that many participants were unable to tolerate the therapeutic doses would also mean that the effects we report here may be more pragmatic and likely to be seen in clinical practice.

Potential reporting bias was documented in over half of the included RCTs, most commonly related to insufficient documentation of baseline characteristics or baseline values for efficacy variables. Unfortunately, the potential impact that the omission of baseline data for efficacy outcomes may have had on efficacy results cannot be ascertained. The interpretability of the results of analyses of total AEs and gastrointestinal AEs was limited by extremely high heterogeneity for all opioids and for strong and weak/intermediate subgroups. Likewise, heterogeneity in our efficacy analyses was concerning. Although we attempted to constrain heterogeneity through strict inclusion criteria and subgroup analysis, we tried to further explore and explain various sources of this heterogeneity, including BMI, severity of OA, baseline pain, and study quality.

Our sensitivity analyses excluding very low-quality studies yielded similar effect sizes, with a 22% decrease in heterogeneity at 2 weeks but a 4% increase at 12 weeks. BMI was reported in only 10 studies (5 studies reporting a BMI of 34, 2 reporting 33, 2 reporting 31). Although the lack of variability in these values would normally have precluded us from conducting a meaningful meta-regression analysis, post hoc meta-regression of pain relief values on BMI were not significant ($P = 0.37$). Baseline pain values were reported in only 10 studies, with minimum variability among those values (SD of 6.58 points on a standardized 0–100 scale); post hoc meta-regression of pain relief on baseline pain (standardized to a 0–100 scale) did not reveal a significant relationship ($P = 0.88$). The data for disease severity were not reported in a way that is amenable for any kind of analysis or further exploration.

Another major limitation of the current study is the lack of longer-term follow-up data. None of the RCTs extended beyond 16 weeks of follow-up. Since OA is a chronic disorder, and since many opioid users are prescribed these drugs to manage OA and other painful chronic conditions, the lack of longer-term data may mean that the results from any of the currently available RCT data for opioids in this population are not generalizable to a large portion of individuals with OA who need long-term pain medications. Although a recent meta-analysis of data from open-label extension trials concluded that opioids may be effective and safe for long-term use, only 4 of the RCTs specifically involved patients with OA, 3 of which extended to only 26 weeks (86). Furthermore, the proportion of participants retained from the original trials ranged from 11% to 29% and likely consists mainly of participants who tolerated and responded to opioids. Thus, the conclusions drawn in the current study must be considered in light of the aforementioned limitations, as well as the limitations inherent to open-label, noncomparative study designs. Finally, our study was limited in its ability to capture patient attitudes and perspectives on opioid treatment for OA. Future research should focus on the impact of patient education on attitudes and beliefs about the use of opioids for treating OA symptoms.

Opioids, when compared to placebo, show only small benefits on pain and function from 2 to 12 weeks of treatment, contribute no measurable benefit to quality of life, and show an increased risk of harms. Strong opioids demonstrated consistently inferior efficacy and overall worse safety than weak/intermediate opioids. In light of this evidence and in the context of a rapidly shifting paradigm for pain management, clinicians and policy makers should reconsider the utility of opioids in the management of OA.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Campbell JN. APS 1995 presidential address. J Pain 1996;5:85–8.
2. Manchikanti L, Fellows B, Ailinani H, Pampati V. Therapeutic use, abuse, and nonmedical use of opioids: a ten-year perspective. Pain Physician 2010;13:401–35.
3. Levy N, Sturgess J, Mills P. "Pain as the fifth vital sign" and dependence on the "numerical pain scale" is being abandoned in the US: Why? Br J Anaesth 2018;120:435–8.
4. Rummans TA, Burton MC, Dawson NL. How good intentions contributed to bad outcomes: the opioid crisis. Mayo Clin Proc 2018;93:344–50.
5. Furlan AD, Sandoval JA, Mailis-Gagnon A, Tunks E. Opioids for chronic noncancer pain: a meta-analysis of effectiveness and side effects. CMAJ 2006;174:1589–94.
6. Morone NE, Weiner DK. Pain as the fifth vital sign: exposing the vital need for pain education. Clin Ther 2013;35:1728–32.
7. Smith SR, Deshpande BR, Collins JE, Katz JN, Losina E. Comparative pain reduction of oral non-steroidal anti-inflammatory drugs and opioids for knee osteoarthritis: systematic analytic review. Osteoarthritis Cartilage 2016;24:962–72.

8. Eriksen J, Sjogren P, Bruera E, Ekholm O, Rasmussen NK. Critical issues on opioids in chronic non-cancer pain: an epidemiological study. Pain 2006;125:172–9.
9. Jeffery MM, Hooten WM, Henk HJ, Bellolio MF, Hess EP, Meara E, et al. Trends in opioid use in commercially insured and Medicare Advantage populations in 2007–16: retrospective cohort study. BMJ 2018;362:k2833.
10. Huxtable CA, Roberts LJ, Somogyi AA, MacIntyre PE. Acute pain management in opioid-tolerant patients: a growing challenge. Anaesth Intensive Care 2011;39:804–23.
11. Birke H, Kurita GP, Sjogren P, Simonsen MK, Juel K, Ekholm O. Chronic non-cancer pain and the epidemic prescription of opioids in the Danish population: trends from 2000 to 2013. Acta Anaesthesiol Scand 2016;60:623–33.
12. Atluri S, Sudarshan G, Manchikanti L. Assessment of the trends in medical use and misuse of opioid analgesics from 2004 to 2011. Pain Physician 2014;17:E119–28.
13. Volkow ND, McLellan AT. Opioid abuse in chronic pain: misconceptions and mitigation strategies. N Engl J Med 2016;374:1253–63.
14. Boulanger A, Clark AJ, Squire P, Cui E, Horbay GL. Chronic pain in Canada: have we improved our management of chronic noncancer pain? Pain Res Manag 2007;12:39–47.
15. Weir RJ. The insidious relationship between pharmaceutical marketing to physicians, opioid prescriptions, and overdose deaths. Chicago Policy Review. 2019. URL: https://chicagopolicyreview.org/2019/05/16/the-insidious-relationship-between-pharmaceutical-marketing-to-physicians-opioid-prescriptions-and-overdose-deaths/.
16. Lee AJ, Bandari J, Macleod LC, Davies BJ, Jacobs BL. Concentration of opioid-related industry payments in opioid crisis areas. J Gen Intern Med 2019;34:187–9.
17. Hadland SE, Rivera-Aguirre A, Marshall BDL, Cerda M. Association of pharmaceutical industry marketing of opioid products with mortality from opioid-related overdoses. JAMA Netw Open 2019;2:e186007.
18. Chidgey BA, McGinigle KL, McNaull PP. When a vital sign leads a country astray: the opioid epidemic. JAMA Surg 2019;154:987–8.
19. Deveza LA, Hunter DJ, Van Spil WE. Too much opioid, too much harm. Osteoarthritis Cartilage 2018;26:293–5.
20. Vowles KE, McEntee ML, Julnes PS, Frohe T, Ney JP, van der Goes DN. Rates of opioid misuse, abuse, and addiction in chronic pain: a systematic review and data synthesis. Pain 2015;156:569–76.
21. Thorlund JB, Turkiewicz A, Prieto-Alhambra D, Englund M. High inappropriate use of prescribed opioids in patients with incident knee or hip osteoarthritis. Osteoarthritis Cartilage 2019;27:S49.
22. Megale RZ, Deveza LA, Blyth FM, Naganathan V, Ferreira PH, McLachlan AJ, et al. Efficacy and safety of oral and transdermal opioid analgesics for musculoskeletal pain in older adults: a systematic review of randomized, placebo-controlled trials. J Pain 2018;19:475.e471–5.
23. Da Costa BR, Nüesch E, Kasteler R, Husni E, Welch V, Rutjes AW, et al. Oral or transdermal opioids for osteoarthritis of the knee or hip. Cochrane Database Syst Rev 2014;9:CD003115.
24. Bannuru RR, Osani MC, Vaysbrot EE, Arden NK, Bennell K, Bierma-Zeinstra SM, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthritis Cartilage 2019;27:1578–89.
25. National Clinical Guideline. National Institute for Health and Clinical Excellence: guidance. In: Osteoarthritis: care and management in adults. London: National Institute for Health and Care Excellence; 2014.
26. Myers J, Wielage RC, Han B, Price K, Gahn J, Paget MA, et al. The efficacy of duloxetine, non-steroidal anti-inflammatory drugs, and opioids in osteoarthritis: a systematic literature review and meta-analysis. BMC Musculoskelet Disord 2014;15:76.
27. Fuggle N, Curtis E, Shaw S, Spooner L, Bruyere O, Ntani G, et al. Safety of opioids in osteoarthritis: outcomes of a systematic review and meta-analysis. Drugs Aging 2019;36:129–43.
28. Stewart M, Cibere J, Sayre EC, Kopec JA. Efficacy of commonly prescribed analgesics in the management of osteoarthritis: a systematic review and meta-analysis. Rheumatol Int 2018;38:1985–97.
29. Neogi T. The epidemiology and impact of pain in osteoarthritis. Osteoarthritis Cartilage 2013;21:1145–53.
30. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017;390:1211–59.
31. Kolasinski SL, Neogi T, Hochberg MC, Oatis C, Guyatt G, Block J, et al. 2019 American College of Rheumatology/Arthritis Foundation guideline for the management of osteoarthritis of the hand, hip, and knee. Arthritis Care Res (Hoboken) 2020;72:149–62.
32. Bannuru RR, Natov NS, Obadan IE, Price LL, Schmid CH, McAlindon TE. Therapeutic trajectory of hyaluronic acid versus corticosteroids in the treatment of knee osteoarthritis: a systematic review and meta-analysis. Arthritis Rheum 2009;61:1704–11.
33. Osani MC, Vaysbrot EE, Zhou M, McAlindon TE, Bannuru RR. Duration of symptom relief and early trajectory of adverse events for oral nonsteroidal antiinflammatory drugs in knee osteoarthritis: a systematic review and meta-analysis. Arthritis Care Res (Hoboken) 2020;72:641–51.
34. Brennan MJ, Lieberman JA III. Sleep disturbances in patients with chronic pain: effectively managing opioid analgesia to improve outcomes. Curr Med Res Opin 2009;25:1045–55.
35. Cronin A, Keifer JC, Baghdoyan HA, Lydic R. Opioid inhibition of rapid eye movement sleep by a specific mu receptor agonist. Br J Anaesth 1995;74:188–92.
36. Dimsdale JE, Norman D, DeJardin D, Wallace MS. The effect of opioids on sleep architecture. J Clin Sleep Med 2007;3:33–6.
37. Agarwal P, Sambamoorthi U. Healthcare expenditures associated with depression among individuals with osteoarthritis: post-regression linear decomposition approach. J Gen Intern Med 2015;30:1803–11.
38. Sale JE, Gignac M, Hawker G. The relationship between disease symptoms, life events, coping and treatment, and depression among older adults with osteoarthritis. J Rheumatol 2008;35:335–42.
39. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009;151:264–9.
40. Davies S. The importance of PROSPERO to the National Institute for Health Research. Syst Rev 2012;1:5.
41. Trescot AM, Datta S, Lee M, Hansen H. Opioid pharmacology. Pain Physician 2008;11 Suppl:S133–53.
42. Vadivelu N, Timchenko A, Huang Y, Sinatra R. Tapentadol extended-release for treatment of chronic pain: a review. J Pain Res 2011;4:211.
43. Markum Mitchell. Engauge digitizer. URL: https://github.com/markummitchell/engauge-digitizer.
44. Cochrane Collaboration. The Cochrane risk of bias tool. URL: http://handbook.cochrane.org.
45. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924–26.
46. Cochrane Musculoskeletal. Proposed outcomes. URL: http://musculoskeletal.cochrane.org/proposed-outcomes.
47. 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.

48. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–83.
49. Hedges LV, Olkin I. Statistical methods for meta-analysis. New York: Academic Press; 1985.
50. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177–88.
51. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959;22:719–48.
52. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21:1539–58.
53. Cepeda MS, Camargo F, Zea C, Valencia L. Tramadol for osteoarthritis: a systematic review and metaanalysis. J Rheumatol 2007;34:543–55.
54. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315:629–34.
55. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid Based Ment Health 2019;22:153–60.
56. Peloso PM, Bellamy N, Bensen W, Thomson GT, Harsanyi Z, Babul N, et al. Double blind randomized placebo control trial of controlled release codeine in the treatment of osteoarthritis of the hip or knee. J Rheumatol 2000;27:764–71.
57. Afilalo M, Etropolski MS, Kuperwasser B, Kelly K, Okamoto A, Van Hove I, et al. Efficacy and safety of tapentadol extended release compared with oxycodone controlled release for the management of moderate to severe chronic pain related to osteoarthritis of the knee: a randomized, double-blind, placebo- and active-controlled phase III study. Clin Drug Investig 2010;30:489–505.
58. Serrie A, Lange B, Steup A. Tapentadol prolonged-release for moderate-to-severe chronic osteoarthritis knee pain: a double-blind, randomized, placebo-and oxycodone controlled release-controlled study. Curr Med Res Opin 2017;33:1423–32.
59. Hartrick C, Van Hove I, Stegmann JU, Oh C, Upmalis D. Efficacy and tolerability of tapentadol immediate release and oxycodone HCl immediate release in patients awaiting primary joint replacement surgery for end-stage joint disease: a 10-day, phase III, randomized, double-blind, active-and placebo-controlled study. Clin Ther 2009;31:260–271.
60. Fleischmann RM, Caldwell JR, Roth SH, Tesser JRP, Olson W, Kamin M. Tramadol for the treatment of joint pain associated with osteoarthritis: a randomized, double-blind, placebo-controlled trial. Curr Ther Res 2001;62:113–28.
61. Babul N, Noveck R, Chipman H, Roth SH, Gana T, Albert K. Efficacy and safety of extended-release, once-daily tramadol in chronic pain: a randomized 12-week clinical trial in osteoarthritis of the knee. J Pain Symptom Manage 2004;28:59–71.
62. Fishman RL, Kistler CJ, Ellerbusch MT, Aparicio RT, Swami SS, Shirley ME, et al. Efficacy and safety of 12 weeks of osteoarthritic pain therapy with once-daily tramadol (Tramadol Contramid OAD). J Opioid Manag 2007;3:273–80.
63. Labopharm. NCT00832416. A tour-arm study comparing the analgesic efficacy and safety of tramadol once a day 100, 200 and 300 mg versus placebo for the treatment of pain due to osteoarthritis of the knee. 2009. URL: https://clinicaltrials.gov/ct2/show/results/NCT00832416.
64. Malonne H, Coffiner M, Sonet B, Sereno A, Vanderbist F. Efficacy and tolerability of sustained-release tramadol in the treatment of symptomatic osteoarthritis of the hip or knee: a multicenter, randomized, double-blind, placebo-controlled study. Clin Ther 2004;26:1774–82.
65. Gana TJ, Pascual ML, Fleming RR, Schein JR, Janagap CC, Xiang J, et al. Extended-release tramadol in the treatment of osteoarthritis: a multicenter, randomized, double-blind, placebo-controlled clinical trial. Curr Med Res Opin 2006;22:1391–401.
66. DeLemos BP, Xiang J, Benson C, Gana TJ, Pascual ML, Rosanna R, et al. Tramadol hydrochloride extended-release once-daily in the treatment of osteoarthritis of the knee and/or hip: a double-blind, randomized, dose-ranging trial. Am J Ther 2011;18:216–26.
67. Caldwell JR, Rapoport RJ, Davis JC, Offenberg HL, Marker HW, Roth SH, et al. Efficacy and safety of a once-daily morphine formulation in chronic, moderate-to-severe osteoarthritis pain: results from a randomized, placebo-controlled, double-blind trial and an open-label extension trial. J Pain Symptom Manage 2002;23:278–91.
68. Vojtassak J, Jacobs A, Rynn L, Waechter S, Richarz U. A phase IIIb, multicentre, randomised, parallel-group, placebo-controlled, double-blind study to investigate the efficacy and safety of OROS hydromorphone in subjects with moderate-to-severe chronic pain induced by osteoarthritis of the hip or the knee. Pain Res Treat 2011;2011:239501.
69. Rauck R, Rapoport R, Thipphawong J. Results of a double-blind, placebo-controlled, fixed-dose assessment of once-daily OROS(R) hydromorphone ER in patients with moderate to severe pain associated with chronic osteoarthritis. Pain Pract 2013;13:18–29.
70. Cubist. NCT00979953. Efficacy and safety study evaluating ADL5859 and ADL5747 in participants with pain due to osteoarthritis of the knee. 2015. URL: https://clinicaltrials.gov/ct2/show/results/NCT00979953.
71. Chindalore VL, Craven RA, Yu KP, Butera PG, Burns LH, Friedmann N. Adding ultralow-dose naltrexone to oxycodone enhances and prolongs analgesia: a randomized, controlled trial of oxytrex. J Pain 2005;6:392–9.
72. Matsumoto AK, Babul N, Ahdieh H. Oxymorphone extended-release tablets relieve moderate to severe pain and improve physical function in osteoarthritis: results of a randomized, double-blind, placebo- and active-controlled phase III trial. Pain Med 2005;6:357–66.
73. Kivitz A, Ma C, Ahdieh H, Galer BS. A 2-week, multicenter, randomized, double-blind, placebo-controlled, dose-ranging, phase III trial comparing the efficacy of oxymorphone extended release and placebo in adults with pain associated with osteoarthritis of the hip or knee. Clin Ther 2006;28:352–64.
74. Bannuru RR, McAlindon TE, Sullivan MC, Wong JB, Kent DM, Schmid CH. Effectiveness and implications of alternative placebo treatments: a systematic review and network meta-analysis of osteoarthritis trials. Ann Intern Med 2015;163:365–72.
75. Burch F, Fishman R, Messina N, Corser B, Radulescu F, Sarbu A, et al. A comparison of the analgesic efficacy of tramadol contramid OAD versus placebo in patients with pain due to osteoarthritis. J Pain Symptom Manage 2007;34:328–38.
76. Zautra AJ, Smith BW. Impact of controlled-release oxycodone on efficacy beliefs and coping efforts among osteoarthritis patients with moderate to severe pain. Clin J Pain 2005;21:471–7.
77. Schnitzer TJ, Kamin M, Olson WH. Tramadol allows reduction of naproxen dose among patients with naproxen-responsive osteoarthritis pain: a randomized, double-blind, placebo-controlled study. Arthritis Rheum 1999;42:1370–7.
78. Silverfield JC, Kamin M, Wu SC, Rosenthal N. Tramadol/acetaminophen combination tablets for the treatment of osteoarthritis flare pain: a multicenter, outpatient, randomized, double-blind, placebo-controlled, parallel-group, add-on study. Clin Ther 2002;24:282–97.
79. Emkey R, Rosenthal N, Wu SC, Jordan D, Kamin M. Efficacy and safety of tramadol/acetaminophen tablets (Ultracet) as add-on therapy for osteoarthritis pain in subjects receiving a COX-2 nonsteroidal antiinflammatory drug: a multicenter, randomized, double-blind, placebo-controlled trial. J Rheumatol 2004;31:150–6.
80. Etropolski M, Kelly K, Okamoto A, Rauschkolb C. Comparable efficacy and superior gastrointestinal tolerability (nausea, vomiting, constipation) of tapentadol compared with oxycodone hydrochloride. Adv Ther 2011;28:401–17.

81. Kjærsgaard-Andersen P, Nafei A, Skov O, Madsen F, Andersen HM, Kroner K, et al. Codeine plus paracetamol versus paracetamol in longer-term treatment of chronic pain due to osteoarthritis of the hip: a randomised, double-blind, multi-centre study. Pain 1990;43:309–18.
82. Quiding H, Grimstad J, Rusten K, Stubhaug A, Bremnes J, Breivik H. Ibuprofen plus codeine, ibuprofen, and placebo in a single- and multidose cross-over comparison for coxarthrosis pain. Pain 1992;50:303–7.
83. Spierings EL, Fidelholtz J, Wolfram G, Smith MD, Brown MT, West CR. A phase III placebo- and oxycodone-controlled study of tanezumab in adults with osteoarthritis pain of the hip or knee. Pain 2013;154:1603–12.
84. Markenson JA, Croft J, Zhang PG, Richards P. Treatment of persistent pain associated with osteoarthritis with controlled-release oxycodone tablets in a randomized controlled clinical trial. Clin J Pain 2005;21:524–35.
85. Roth SH, Fleischmann RM, Burch FX, Dietz F, Bockow B, Rapoport RJ, et al. Around-the-clock, controlled-release oxycodone therapy for osteoarthritis-related pain: placebo-controlled trial and long-term evaluation. Arch Intern Med 2000;160:853–60.
86. Bialas P, Maier C, Klose P, Hauser W. Efficacy and harms of long-term opioid therapy in chronic non-cancer pain: systematic review and meta-analysis of open-label extension trials with a study duration ≥26 weeks. Eur J Pain 2020;24:265–78.

Arthritis Care & Research
Vol. 73, No. 10, October 2021, pp 1425–1429
DOI 10.1002/acr.24354

American College of Rheumatology
Empowering Rheumatology Professionals

BRIEF REPORT

Targeted Program in an Academic Rheumatology Practice to Improve Compliance With Opioid Prescribing Guidelines for the Treatment of Chronic Pain

Eric J. Wang, Rebecca Helgesen, Chadwick R. Johr, Hannah S. Lacko, Michael A. Ashburn, and Peter A. Merkel

Objective. The Centers for Disease Control and Prevention (CDC) and many state governments have issued guidelines for opioid prescribing for the treatment of chronic noncancer-associated pain. We sought to decrease practice variation and increase compliance with these guidelines in a tertiary academic rheumatology practice by developing an interdisciplinary opioid working group and using electronic health record (EHR)–integrated data feedback.

Methods. Division leadership and providers established shared goals at interdisciplinary meetings involving rheumatology, pain medicine, nursing, and pharmacy. Interventions included educational sessions on opioid prescribing guidelines and the sharing of individual de-identified prescribing patterns. An opioid dashboard page within the EHR allowed every provider to see individualized and division-wide data that tracked process measures based on CDC and state-specific guidelines. Baseline data from June to August 2017 were compared with monthly data through December 2018.

Results. At baseline, 40% of patients had an active opioid agreement (a Pennsylvania guideline and a New Jersey law), 25% had a urine drug screen result within 12 months of their most recent opioid prescription, and 24% had a concurrent benzodiazepine prescription. After 16 months, these percentages improved to 88%, 66%, and 16%, respectively. The average number of opioid tablets prescribed per month decreased from 59,733 to 48,966 (–18%; $P = 0.02$).

Conclusion. Shared goals developed through interdisciplinary input and readily accessible data feedback can markedly increase provider compliance with national and state-specific guidelines for opioid prescribing for the treatment of chronic noncancer-associated pain in rheumatology.

INTRODUCTION

Opioids are well-documented as having a limited role for the treatment of chronic noncancer-associated pain, with a potential modest benefit matched or exceeded by harms (1,2), including death from causes other than overdose (3). Despite this limitation, opioid prescriptions for chronic noncancer-associated pain increased 4-fold between 1999 and 2010 (4).

The epidemic of opioid-related deaths in the US has not ended (4), and the continued overprescribing of opioids is a likely contributor. Over 70,000 people died from drug overdose in the US in 2017, and two-thirds of these deaths involved opioids (5). Approximately one-third of the opioid-related deaths involved a prescription opioid (5), and the overprescribing of opioids may also facilitate their misuse, abuse, and diversion (6). Of the nearly 12 million people age ≥12 years who engaged in opioid misuse in 2016, 92% (10.9 million) misused solely prescription opioids (7).

Reducing clinically unwarranted opioid prescriptions has become a key strategy taken by Federal and state institutions to address the opioid epidemic (4). The Centers for Disease Control and Prevention (CDC) Guideline for Prescribing Opioids for Chronic Pain recommends a thorough risk-benefit evaluation prior to starting or maintaining opioid therapy, and at the lowest effective dosage for the least amount of time possible (8). The majority of state governments in the nation have also established limits defining either the maximum days' supply or the maximum morphine milligram equivalents (MME) of first-time opioid prescriptions (9).

The Division of Rheumatology at the University of Pennsylvania examined its own practices to minimize the risk of inadvertent

Eric J. Wang, MD, Rebecca Helgesen, MBA, Chadwick R. Johr, MD, Hannah S. Lacko, MA, CSSBB, Michael A. Ashburn, MD, MPH, Peter A. Merkel, MD, MPH: Hospital of the University of Pennsylvania, Philadelphia.

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

Address correspondence to Eric J. Wang, MD, Tuttleman Building, 2nd Floor, 1840 South Street, Philadelphia, PA 19146. Email: ewang29@jhmi.edu.

Submitted for publication January 30, 2020; accepted in revised form June 9, 2020.

SIGNIFICANCE & INNOVATIONS

- This study demonstrates how interdisciplinary goal-setting, prescriber education, and individualized data feedback integrated within an electronic health record can markedly reduce variations in opioid prescribing practices and increase compliance with national and state-specific guidelines for the treatment of chronic noncancer-associated pain in a rheumatology practice.
- Prescribing patterns for opioids can be changed relatively quickly by establishing shared goals through interdisciplinary input and by allowing providers in real-time to readily access, track, and compare data describing their own performance.
- This study may have implications for improving adherence to process measures in other clinical settings.

opioid overprescribing. This academic practice included 18 providers, 5 clinical sites, and a patient volume of approximately 25,000 visits per year. An internal review in 2016 found that the group was the fifth-largest prescriber of opioids by number of tablets within the clinical practices of the University of Pennsylvania. There was also evidence of significant practice variation within the division regarding compliance with CDC and state-specific opioid prescribing guidelines for chronic noncancer-associated pain.

To address these issues regarding opioid prescribing, the Division of Rheumatology at the University of Pennsylvania created a division-specific opioid working group. By collecting baseline data, planning and implementing an integrated set of interventions, and assessing their impact to guide future actions, we used a plan-do-study-act cycle. We report on this ongoing quality improvement initiative to decrease practice variability and increase compliance with national and state-specific guidelines regarding opioid prescribing for the treatment of chronic noncancer-associated pain in an academic rheumatology practice.

MATERIALS AND METHODS

Opioid working group. The Division of Rheumatology created an opioid working group to address its goals of optimizing treatment of pain while lowering the risk of harm. This group was informed by a broader opioid task force of the University of Pennsylvania Health System (UPHS). The working group consisted of a rheumatology physician lead, the rheumatology division administrator, the musculoskeletal and rheumatology service line director of quality, a rheumatology nurse, a pain medicine physician, and a clinical pharmacist. Process measures were selected based on the CDC Guideline for Prescribing Opioids for Chronic Pain, Commonwealth of Pennsylvania guidelines, and New Jersey state law. These measures included the percentage of patients receiving chronic opioid prescriptions (≥3 prescriptions in a 12-month period) through the Division of Rheumatology who: 1) had an active opioid agreement within 12 months of their most recent opioid prescription (a Pennsylvania-specific guideline and a New Jersey law); 2) had a visit with the provider within 3 months of their most recent opioid prescription; 3) had a urine drug screen within 12 months of their most recent opioid prescription; and 4) had an active concurrent benzodiazepine prescription (8,10,11). In addition, the total number of opioid tablets prescribed per month was documented, with the expectation that increased compliance with these process measures would lead to a decrease in this number. We focused on Pennsylvania opioid prescribing guidelines and a New Jersey opioid prescribing law because the vast majority of our patients taking opioids fill their prescriptions in 1 of these 2 states.

Opioid dashboard. UPHS created an opioid dashboard page that was integrated into the electronic health record (EHR) system (Epic Systems Corporation), which recorded the process measures above for each provider. This dashboard allowed collection and presentation of data from the institutional level, to the practice location, down to the individual provider level. Every provider was able to look up his or her own data and those of anyone else within UPHS.

Data feedback, provider education, and EHR optimization. A division-wide meeting was held that included physicians (including trainees), nurse practitioners, nurses, practice managers, and pharmacists. Baseline data on the above process measures were shared, and a consensus was reached regarding the necessity of decreasing practice variability and increasing compliance with national and state guidelines regarding opioid prescriptions. Specific goals and a long-term action plan to improve compliance with these process measures were established. A division-wide educational session regarding these measures was led by a pain medicine physician. Throughout this process, de-identified physician-level performance data were openly shared and discussed to facilitate an understanding of goals and existing challenges, rather than placing blame on individual providers. Through these educational and iterative efforts, a standardized process for the prescription of opioids for treating chronic noncancer-associated pain was developed and implemented. A financial incentive to the division for programmatic initiatives was administered through the UPHS musculoskeletal and rheumatology service line, based on the division meeting selected goals for these process measures. Individual clinicians did not receive direct financial benefit.

To integrate into the clinical workflow the tasks necessary to accomplish the agreed-upon process goals, the UPHS opioid task force identified opportunities to optimize the EHR. These opportunities included providing integrated access to the Pennsylvania and New Jersey Prescription Drug Monitoring Program (PDMP)

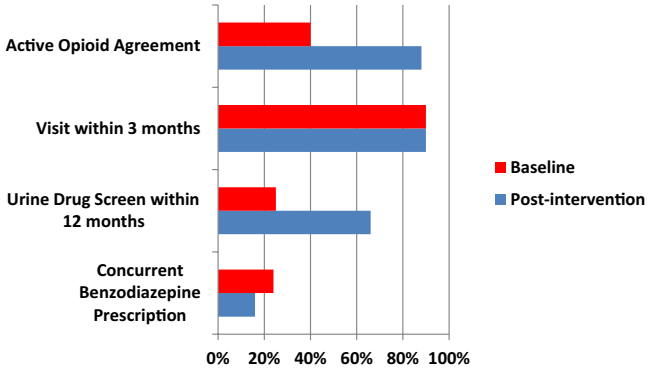


Figure 1. Baseline and postintervention characteristics of opioid prescribing process measures.

databases within each patient's electronic chart. This access was accomplished in collaboration with the Department of Health of the Commonwealth of Pennsylvania. Previously, providers had to exit their patient's electronic chart, access the database's website, sign in, and then enter the patient's identifying information into the website to access PDMP data. This integration allowed providers instead to click on 1 link within the patient's record to immediately access the PDMP database.

The UPHS opioid task force also identified the fact that the existing workflow for completing opioid agreements presented a barrier to compliance. Individual practices within the health system were expected to create their own versions of an opioid agreement. These documents were in paper form, and after they were signed, they had to be individually scanned into each patient's electronic record. However, there was no standardized location for where to store these agreements within the EHR, and there was no standardized name for these files. In response, the UPHS opioid task force created a single electronic opioid agreement that conformed to clinical practice guidelines and state laws. This electronic agreement replaced the previous paper version, and all patients who had not signed an opioid agreement within 1 year had their opioid agreement renewed using the new electronic signature process. This new workflow established an institutional standard for the opioid agreement, and because the document was electronic, it was automatically uploaded into the patient's chart into a standardized location immediately following signature.

Individual prescriber performance with the process measures was monitored on an ongoing basis through the opioid dashboard. Divisional leadership ensured that all prescribers were aware of their performance through regular communication of individual and division-wide performance. Divisional leadership also used the opioid dashboard to identify prescribers who appeared not to be improving, facilitating individualized education and coaching. In addition, the opioid dashboard allowed all providers to create lists of those patients out of adherence with specific process measures (e.g., not having an active opioid agreement on file). Providers and support staff could then easily identify these patients and bring their care into compliance.

RESULTS

Baseline data (June 2017 to August 2017) were obtained in September 2017 and monthly data were obtained through December 2018. A summary of the data is shown in Figure 1. At baseline, 40% of patients had an active opioid agreement, all signed on paper. This rate improved to 88% (all electronically signed) by December 2018. The percentage of patients with a documented provider visit within 3 months of receiving any opioid prescription was 90% at baseline and remained above 90% throughout the study period. Only 25% of patients had undergone a urine drug screen within 12 months of receiving an opioid prescription at baseline; this rate increased to 74% by June 2018 and remained above 66% through December 2018. The percentage of patients having active prescriptions of both opioids and benzodiazepines (not necessarily from the same prescriber or institution) was 24% at baseline; this percentage dropped to 20% by June 2018 and further decreased to 16% by December 2018.

Total opioid tablets prescribed per month. At baseline, the mean number of opioid tablets prescribed per month was 59,733, with a median of 62,404 tablets. From September 2017 through December 2018, the mean number of opioid tablets prescribed decreased to 48,966 per month, with a median of 48,901 tablets, a reduction of the mean by 18% ($P = 0.02$). A summary of the data is shown in Figure 2.

Number of opioid tablets per prescription per month. At baseline, there was a mean ± SD of 115.4 ± 3.18 tablets per opioid prescription per month (n = 3 months of baseline data, median = 116.9). From September 2017 through December 2018, the mean ± SD number of tablets per opioid prescription per month decreased to 112.5 ± 3.08. This reduction was not

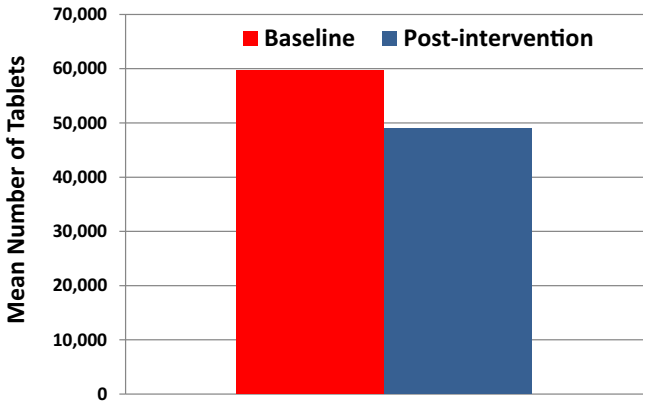


Figure 2. Mean total opioid tablets prescribed per month.

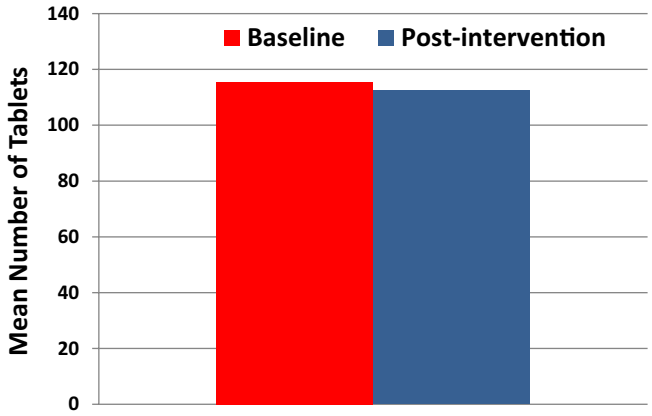


Figure 3. Mean number of opioid tablets per prescription per month.

statistically significant (P = 0.14; n = 16 months of postintervention data; median = 112.4). A summary of the data is shown in Figure 3.

DISCUSSION

In this ongoing initiative within an academic rheumatology practice, an interdisciplinary working group that included all opioid prescribing stakeholders established clearly defined process measures consistent with current CDC and state opioid prescribing guidelines and law. Directly through the EHR, all prescribers could access real-time data tracking their own performance in meeting these process measures and also compare their own data with those of the division and their practice location.

This program was quite successful. Over the first 16 months of this initiative, practice variation decreased and compliance increased regarding 3 key process measures: 1) the percentage of patients with an active opioid agreement within 12 months of their most recent opioid prescription; 2) the percentage of patients with a urine drug screen within 12 months of their most recent opioid prescription; and 3) the percentage of patients with an active concurrent benzodiazepine prescription from any provider or institution. The improved compliance with these process measures is consistent with the concept that motivated providers, leadership buy-in, and timely data feedback can together change practice behavior (12). Compliance with the fourth process measure (the percentage of patients with a provider visit within 3 months of their most recent opioid prescription) also remained high.

The total number of opioid tablets prescribed per month decreased even though the number of opioid tablets per prescription per month did not significantly change. One plausible explanation is that when a provider has assessed a patient to meet the indications for receiving a new or renewed opioid prescription, the decision on how many tablets to prescribe per month is based largely on objective clinical findings. However, with increased compliance with national and state-specific guidelines for opioid prescribing, the number of total prescriptions written may have decreased, ideally reflecting a reduction in opioid overprescribing, and leading to an overall decrease in the total number of tablets prescribed.

Notably, this initiative was associated with a decrease in the percentage of patients receiving a concurrent benzodiazepine prescription, because the risk of harm (e.g., respiratory depression) in patients receiving both opioids and benzodiazepines is significantly increased compared to patients taking opioids alone (8). Nationally, approximately 20–25% of patients who have an opioid or benzodiazepine prescription in the outpatient setting have also been co-prescribed the other (13). Our baseline percentage of patients with concurrent opioid and benzodiazepine prescriptions was similar at 24%, but decreased to 16%. Since nearly all of the benzodiazepine prescriptions were from providers outside the practice, this decrease may be attributable to fewer patients receiving opioid prescriptions from the division.

This project has a number of strengths. The initiative had strong provider buy-in and satisfaction, and all relevant opioid-prescribing stakeholders were involved. Many providers expressed gratitude with the rheumatology opioid working group for helping them achieve a better understanding of current best practices for opioid prescribing. This initiative also resulted in improvements in 3 key process measures, even though all providers were under significant time pressure in busy outpatient practices, and well-established patients had to agree to perform additional tasks on routine follow-up visits (e.g., signing an opioid agreement). The optimizations the UPHS opioid task force contributed to the EHR, namely the creation of a standardized electronic opioid agreement and integrating access to PDMP data, helped make these tasks achievable.

An additional strength of this work is that it facilitates adaptability to future iterations of guidelines and laws regarding opioid prescribing. As national opioid prescribing guidelines continue to be refined and state governments pass new legislation, process measures can be modified, and providers will have the infrastructure in place to understand new requirements, adjust their practice patterns accordingly, and track their compliance.

Our study also has limitations. This initiative only addressed chronic noncancer-associated pain, though such pain is the likely reason why patients with rheumatic diseases receive chronic opioid prescriptions (14). Also, although our EHR tracks the number of tablets per opioid prescription per month, it does not track the number of opioid prescriptions written per provider per month. Therefore, we are unable to directly measure whether the number of opioid prescriptions per provider decreased, even though the total number of opioid tablets prescribed per month in the division had decreased. In addition, although our EHR tracks the MME of initial opioid prescriptions, it does not track the average MME of subsequent opioid prescriptions. We are thus unable to state whether the decrease over time in total opioid tablets prescribed per month was accompanied by a concomitant decrease in MME

prescribed per month. Providers can plausibly reduce the number of opioid tablets they prescribe while increasing the dosage of the tablets, leading to an equivalent or potentially larger MME. However, such actions are unlikely because they would contradict our divisional goals, which emphasize using the lowest MME necessary per prescription, consistent with national and state guidelines (8,10,11). Subsequent revisions to our EHR will aim to track the number of opioid prescriptions written per provider and the average MME per prescription. A factor limiting the implications of our work is that we were only able to track process measures. Information feedback initiatives may have a greater impact on process measures than patient outcomes (15).

Overall, this work demonstrates the fact that interdisciplinary goal-setting, provider education, and real-time data feedback can markedly reduce practice variation and increase compliance with national and state-specific guidelines for the prescribing of opioids for the treatment of chronic noncancer-associated pain. Further studies are needed to determine whether increased compliance with these guidelines is sufficient to improve patient safety.

ACKNOWLEDGMENTS

The authors thank the faculty and nursing staff of the Division of Rheumatology at the University of Pennsylvania for their enthusiastic support and participation in this study.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. Wang 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. Helgesen, Johr, Ashburn, Merkel.
Acquisition of data. Helgesen, Lacko.
Analysis and interpretation of data. Wang, Ashburn, Merkel.

REFERENCES

1. Busse JW, Wang L, Kamaleldin M, Craigie S, Riva JJ, Montoya L, et al. Opioids for chronic noncancer pain: a systematic review and meta-analysis. JAMA 2018;320:2448–60.
2. Chou R, Turner JA, Devine EB, Hansen RN, Sullivan SD, Blazina I, et al. The effectiveness and risks of long-term opioid therapy for chronic pain: a systematic review for a National Institutes of Health Pathways to Prevention Workshop. Ann Intern Med 2015;162:276–86.
3. Ray WA, Chung CP, Murray KT, Hall K, Stein CM. Prescription of long-acting opioids and mortality in patients with chronic noncancer pain. JAMA 2016;315:2415–23.
4. Bonnie RJ, Schumacher MA, Clark JD, Kesselheim AS. Pain management and opioid regulation: continuing public health challenges. Am J Public Health 2019;109:31–4.
5. Scholl L, Seth P, Kariisa M, Wilson N, Baldwin G. Drug and opioid-involved overdose deaths: United States, 2013–2017. MMWR Morb Mortal Wkly Rep 2019;67:1419–27.
6. Mundkur ML, Franklin JM, Abdia Y, Huybrechts KF, Patorno E, Gagne JJ, et al. Days' supply of initial opioid analgesic prescriptions and additional fills for acute pain conditions treated in the primary care setting: United States, 2014. MMWR Morb Mortal Wkly Rep 2019;68:140–3.
7. Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the United States: results from the 2016 National Survey on Drug Use and Health (HHS Publication No. SMA 17-5044, NSDUH Series H-52). Rockville (MD): Center for Behavioral Health Statistics and Quality; 2017. URL: https://www.samhsa.gov/data/.
8. Dowell D, Haegerich TM, Chou R. CDC guideline for prescribing opioids for chronic pain: United States, 2016. JAMA 2016;315:1624–45.
9. National Conference of State Legislators. Prescribing policies: states confront opioid overdose epidemic. URL: http://www.ncsl.org/research/health/prescribing-policies-states-confront-opioid-overdose-epidemic.aspx.
10. Ashburn MA, Boateng S, Boll J Jr, Christopher T, Cleaver M, Consuelos M, et al. Treating chronic non-cancer pain: safe prescribing of opioids in chronic non-cancer pain. Commonwealth of Pennsylvania. Revised June 11, 2018. URL: https://www.health.pa.gov/topics/Documents/Opioids/Non-cancer%20Pain%20Guidelines%20Final.pdf.
11. S3 [1R] Vitale, Sweeney. State of New Jersey 217th Legislature. URL: https://www.njleg.state.nj.us/2016/Bills/S0500/3_R1.PDF.
12. Bradley EH, Holmboe ES, Mattera JA, Roumanis SA, Radford MJ, Krumholz HM. Data feedback efforts in quality improvement: lessons learned from US hospitals. Qual Saf Health Care 2004;13:26–31.
13. Agarwal SD, Landon BE. Patterns in outpatient benzodiazepine prescribing in the United States. JAMA Netw Open 2019;2:e187399.
14. Chen SK, Feldman CH, Brill G, Lee YC, Desai RJ, Kim SC. Use of prescription opioids among patients with rheumatic diseases compared to patients with hypertension in the USA: a retrospective cohort study. BMJ Open 2019;9:e027495.
15. Van der Veer S, de Keizer NF, Ravelli AC, Tenkink S, Jager KJ. Improving quality of care: a systematic review on how medical registries provide information feedback to health care providers. Int J Med Inform 2010;79:305–23.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 73, No. 10, October 2021, pp 1430–1435
DOI 10.1002/acr.24447

BRIEF REPORT

Trends in Office Visits During Which Opioids Were Prescribed for Adults With Arthritis in the US, 2006–2015

Loredana Santo,[1] Susan M. Schappert,[1] Jennifer M. Hootman,[2] and Charles G. Helmick[2]

Objective. To analyze trends for visits to office-based physicians at which opioids were prescribed among adults with arthritis in the US, from 2006 to 2015.

Methods. We analyzed nationally representative data on patient visits to office-based physicians from 2006 to 2015 from the National Ambulatory Medical Care Survey (NAMCS). Visit percentages for first- and any-listed diagnosis of arthritis by age groups and sex were reported. Time points were grouped into 2-year intervals to increase the reliability of estimates. Annual percentage point change and 95% confidence intervals (95% CIs) were reported from linear regression models.

Results. From 2006 to 2015, the percentage of visits to office-based physicians by adults with a first-listed diagnosis of arthritis increased from 4.1% (95% CI 3.5%, 4.7%) in 2006–2007 to 5.1% (95% CI 3.9%, 6.6%) in 2014–2015 ($P = 0.033$). Among these visits, the percentage of visits with opioids prescribed increased from 16.5% (95% CI 13.1%, 20.5%) in 2006–2007 to 25.6% (95% CI 17.9%, 34.6%) in 2014–2015 ($P = 0.017$). The percentage of visits with any-listed diagnosis of arthritis increased from 6.6% (95% CI 5.9%, 7.4%) in 2006–2007 to 8.4% (95% CI 7.0%, 10.0%) in 2014–2015 ($P = 0.001$). Among these visits, the percentage of visits with opioids prescribed increased from 17.4% (95% CI 14.6%, 20.4%) in 2006–2007 to 25.0% (95% CI 19.7%, 30.8%) in 2014–2015 ($P = 0.004$).

Conclusion. From 2006 to 2015, the percentage of visits to office-based physicians by adults with arthritis increased and the percentage of opioids prescribed at these visits also increased. NAMCS data will allow continued monitoring of these trends after the implementation of the 2016 Centers for Disease Control and Prevention Guideline for prescribing opioids for chronic pain.

INTRODUCTION

In the US, arthritis affected an estimated 54.4 million (22.7%) adults in 2013–2015 and is projected to affect 78.4 million by 2040 (1,2). Arthritis is characterized by chronic pain that can be managed by a combination of nonpharmacologic interventions, such as physical activity and disease-management education, and pharmacologic therapy (3). The most common pharmacologic therapy includes nonsteroidal antiinflammatory drugs and new disease-modifying antirheumatic drugs. However, most patients with dispensed opioid prescriptions have arthritis, suggesting that arthritis pain is often treated with opioids (4). This is a controversial issue for this population because of the potential adverse effects, especially among older adults with arthritis. The 2016 Centers for Disease Control and Prevention (CDC) Guideline for Prescribing Opioids for Chronic Pain recommends nonpharmacologic therapy and non-opioid therapy for chronic pain and consideration of opioid therapy for chronic pain–related conditions, including osteoarthritis, only if expected benefits for pain and function are anticipated to outweigh the risk due to the possible harms of opioids (5). Moreover, usage of opioids to treat moderate-to-severe chronic back, hip, or knee osteoarthritis pain has not been deemed superior to using nonopioid medications (6). There are gaps in the literature regarding the use of prescription opioids for adults with arthritis in the ambulatory care setting, where most arthritis is diagnosed and managed. To begin addressing these research gaps, we characterized opioid prescriptions among adults with

The findings and conclusions contained herein are those of the authors and do not necessarily represent those of the CDC.

[1]Loredana Santo, MD, MPH, Susan M. Schappert, MA: National Center for Health Statistics, US Centers for Disease Control and Prevention, Hyattsville, Maryland; [2]Jennifer M. Hootman, PhD, Charles G. Helmick, MD: National Center for Chronic Disease Prevention and Health Promotion, US Centers for Disease Control and Prevention, Atlanta, Georgia.

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

Address correspondence to Charles G. Helmick, MD, Division of Population Health, Centers for Disease Control and Prevention, 4770 Buford Hwy NE, Atlanta, GA 30341. Email: chelmick@cdc.gov.

Submitted for publication May 11, 2020; accepted in revised form September 8, 2020.

SIGNIFICANCE & INNOVATIONS

- Arthritis and opioid use represent significant public health and clinical problems in the US. Arthritis patients taking opioid medications are at risk for the potentially adverse effects of these drugs. There are gaps in the literature regarding the prescribing of opioids for adults with arthritis in the ambulatory care setting.
- We used 2006–2015 National Ambulatory Medical Care Survey data to determine trends.
- We used the percentage of visits by adult patients to office-based physicians at which opioids were prescribed.
- During 2006–2015, the percentage of visits to office-based physicians by adult patients with a diagnosis of arthritis increased significantly. Among these visits, the percentage of visits at which opioids were prescribed increased as well.

arthritis by analyzing trends for visits to office-based physicians at which opioids were prescribed between 2006 and 2015.

PATIENTS AND METHODS

The National Ambulatory Medical Care Survey (NAMCS) is an annual, nationally representative survey of visits to nonfederal, office-based physicians in the US conducted by the National Center for Health Statistics (7). The NAMCS currently uses a stratified 2-stage probability sampling design, with physicians selected at the first stage and visits selected at the second stage. Survey data are weighted to produce national estimates. Detailed information on the NAMCS methodology is available on the CDC website (https://www.cdc.gov/nchs/ahcd/). This analysis covers the period from 2006 to 2015, preceding the 2016 CDC Guideline for Prescribing Opioids for Chronic Pain (5). We began with the 2006 NAMCS, the first year NAMCS adopted a new drug coding system and started coding drugs in terms of their generic components and therapeutic classifications, using the Lexicon Plus proprietary database (Cerner Multum). We ended with the 2015 NAMCS, the last year using the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes needed to capture arthritis, in order to maintain a consistent definition of arthritis; comparable arthritis codes using the new ICD-10-CM codes for 2016 and later were not available. In 2016, the NAMCS switched from coding diagnoses using the ICD-9-CM to the ICD-10-CM. Response rates (the percentage of in-scope physicians for whom at least one-half of their expected number of visit records were completed) ranged from 58.9% in 2006 to 35.4% in 2015.

The unit of analysis used was ambulatory care visit (the number of visits rather than the number of people). Visits by adults ages ≥18 years with a first-listed or first 3–listed (hereafter called any-listed) diagnoses of arthritis were analyzed. Arthritis was defined using the ICD-9-CM codes 274, 710, 712-716, 719, 729.1 (8), based on a subset of arthritis codes originally defined by the National Arthritis Data Workgroup (9). Arthritis visits with prescribed opioids were defined as arthritis visits where at least 1 opioid was prescribed (i.e., provided, new, or continued prescription). Opioids were defined using Cerner Multum's Lexicon third level therapeutic category codes for narcotic analgesics (code 60) and narcotic analgesic combinations (code 191). Records with prescriptions of buprenorphine only were excluded. By 2014, the NAMCS was collecting data on as many as 30 medications prescribed (i.e., provided or a new or continued prescription) at office visits. However, in 2005–2009, only 8 medications were collected, which increased to 10, and later 12, medications. For comparability, only data for the first 8 drugs were used across all years. The 2 analysis groups included in this study were visits to office-based physicians made by adults with a first-listed diagnosis of arthritis and visits to office-based physicians made by adults with any-listed diagnosis of arthritis. First-listed diagnoses generally represent the primary diagnosis for the visit. In 2006–2013, up to 3 diagnoses could be reported at each visit. In 2014 and 2015, as many as 5 diagnoses could be reported, but only the first 3 diagnoses were reviewed in this analysis, to be comparable across all study years.

Arthritis visits with prescribed opioids were defined as visits with a diagnosis of arthritis where at least 1 opioid was found among the first 8 drugs reported at the sampled visit. An opioid can be prescribed at the sampled visit, provided at the sampled visit, or continued if it was prescribed prior to the sampled visit. The NAMCS instrument does not determine which drugs were prescribed for which diagnosis, so it is possible that among visits made by adults with arthritis as well as another diagnosis, opioids were prescribed for pain related to other conditions. The NAMCS collects data on medications that are prescribed during a patient visit, but it does not measure whether the patient actually took the medication; consequently, medication adherence was not examined in this analysis.

Visit percentages for first- or any-listed diagnosis with 95% confidence intervals (95% CIs) were reported. Time points were grouped into 2-year intervals to increase the reliability of estimates. Orthogonal polynomial contrasts were used to assess linear and quadratic trends (10). Annual percentage point changes and 95% CIs were reported from linear regression models. Statistical analyses accounted for the complex survey design and were conducted using SAS, version 9.4, and SUDAAN (RTI International), version 11.0.

RESULTS

Osteoarthritis and allied disorders (ICD-9-CM code 715) was the most frequent form of arthritis in our sampled first-listed arthritis visits and any-listed arthritis visits (31% and 29%, respectively).

Table 1. Trends in percentage of visits to office-based physicians by adults with a first-listed or any-listed diagnosis of arthritis: United States, 2006–2015*

	Overall		Ages 18–44 years		Ages 45–64 years		Ages ≥65 years		Women		Men	
	No. of visits	% (95% CI)	No. of visits	% (95% CI)	No. of visits	% (95% CI)	No. of visits	% (95% CI)	No. of visits	% (95% CI)	No. of visits	% (95% CI)
First-listed diagnosis												
2006–2007†	1,395	4.1 (3.5, 4.7)	261	2.4 (2.0, 2.8)	583	4.9 (4.0, 5.9)	551	4.9 (4.1, 5.7)	865	4.2 (3.5, 5.0)	530	3.9 (3.4, 4.4)
2008–2009†	1,356	3.5 (3.1, 3.9)	242	1.9 (1.6, 2.3)	569	4.0 (3.5, 4.5)	545	4.5 (3.8, 5.2)	854	3.6 (3.1, 4.1)	502	3.4 (3.0, 3.9)
2010–2011†	1,556	4.8 (4.1, 5.7)	318	3.3 (2.6, 4.2)	649	5.4 (4.6, 6.3)	589	5.7 (4.7, 6.8)	968	5.1 (4.2, 6.2)	588	4.5 (3.8, 5.2)
2012–2013‡	4,564	4.5 (4.1, 5.0)	792	2.6 (2.3, 3.0)	2,062	5.6 (5.1, 6.2)	1,710	5.1 (4.6, 5.7)	2,866	4.8 (4.3, 5.3)	1,698	4.2 (3.8, 4.7)
2014–2015‡	2,621	5.1 (3.9, 6.6)	436	2.8 (2.2, 3.5)	1,107	5.9 (4.5, 7.6)	1,078	6.1 (4.3, 8.4)	1,622	5.0 (3.8, 6.6)	999	5.2 (4.0, 6.7)
Year of trend change	–	None	–	None	–	None	–	None	–	None	–	None
Annual % point change (95% CI)§	–	0.31 (0.02, 0.61)	–	0.08 (0.002, 0.31)	–	0.38 (–0.01, 0.74)	–	0.32 (–0.14, 0.77)	–	0.29 (–0.03, 0.61)	–	0.36 (0.06, 0.65)
P for trend	–	0.033	–	0.047	–	0.044	–	0.17	–	0.07	–	0.019
Any-listed diagnosis												
2006–2007	2,322	6.6 (5.9, 7.4)	427	3.8 (3.2, 4.4)	940	7.7 (6.7, 8.9)	955	8.2 (7.3, 9.2)	1,471	6.9 (6.1, 7.9)	851	6.1 (5.4, 6.9)
2008–2009	2,206	5.9 (5.4, 6.4)	403	3.1 (2.7, 3.6)	925	6.7 (6.0, 7.5)	878	7.5 (6.7, 8.4)	1,391	6.0 (5.4, 6.6)	815	5.8 (5.2, 6.4)
2010–2011	2,464	7.3 (6.5, 8.2)	475	4.6 (3.8, 5.5)	1,054	8.4 (7.4, 9.4)	935	8.7 (7.5, 9.9)	1,536	7.6 (6.6, 8.7)	928	6.8 (6.1, 7.6)
2012–2013	7,612	7.7 (7.2, 8.3)	1,379	4.7 (4.2, 5.2)	3,363	9.3 (8.6, 10.0)	2,870	8.7 (8.0, 9.4)	4,799	8.0 (7.4, 8.7)	2,833	7.2 (6.6, 7.8)
2014–2015	4,361	8.4 (7.0, 10.0)	715	4.5 (3.7, 5.3)	1,827	9.6 (7.8, 11.6)	1,819	10.3 (8.1, 12.9)	2,707	8.5 (7.0, 10.2)	1,654	8.3 (6.9, 10.0)
Year of trend change	–	None	–	None	–	None	–	None	–	None	–	None
Annual % point change (95% CI)**	–	0.55 (0.21, 0.90)	–	0.30 (0.10, 0.50)	–	0.64 (0.19, 1.1)	–	0.56 (0.02, 1.1)	–	0.35 (0.15, 0.90)	–	0.59 (0.24, 0.95)
P for trend	–	0.001	–	0.003	–	0.005	–	0.04	–	0.006	–	0.001

* Values are the percentage of visits for first- or any-listed diagnosis (95% confidence interval [95% CI]) unless indicated otherwise. Unweighted number of visits represents the number of sampled visits to office-based physicians made by adults with first- or any-listed diagnosis of arthritis. Percentages reflect sample data that were weighted to produce national estimates of visits to non–federally employed, office-based physicians by adults ages ≥18 years with first- or any-listed diagnosis of arthritis. For source of data, see ref. 7.

† For 2006–2011, visits to physicians working in community health centers were excluded from the analysis to be consistent with the 2012–2015 National Ambulatory Medical Care Survey sampling design.

‡ For 2012–2015, sample sizes were higher than previous years due to additional funding. For 2014–2015, as many as 5 diagnoses could be reported, but only the first 3 diagnoses were reviewed, to be comparable across all study years.

§ Annual percentage point change, 95% CI, and *P* values were calculated using linear regression models.

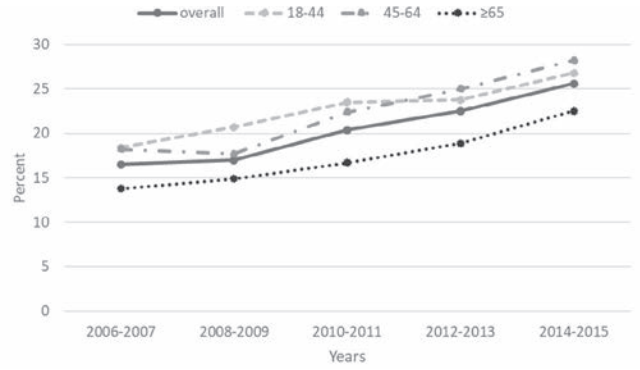


Figure 1. Trends in percentage of visits to office-based physicians made by adults with a first-listed diagnosis of arthritis in which at least 1 opioid was prescribed, by age group in the US from 2006 to 2015.

First-listed arthritis visits. Overall, the percentage of visits to office-based physicians by adults with a first-listed diagnosis of arthritis increased from 4.1% (95% CI 3.5%, 4.7%) in 2006–2007 to 5.1% (95% CI 3.9%, 6.6%) in 2014–2015 (P = 0.033), with similar trends in the age 45–64 years group. The increase in the age ≥65 years group was not statistically significant (P = 0.17). During the study period, the percentage of visits significantly increased for men (P = 0.019) but not for women (P = 0.07).

Any-listed arthritis visits. Overall, the percentage of visits with any-listed diagnosis of arthritis increased from 6.6% (95% CI 5.9%, 7.4%) in 2006–2007 to 8.4% (95% CI 7.0%, 10.0%) in 2014–2015 (P = 0.001). Similar trends were observed among all 3 age groups and for women and men (Table 1).

First-listed arthritis visits with opioids prescribed. Overall, among visits with a first-listed diagnosis of arthritis, the percentage of visits with opioids prescribed increased from 16.5% (95% CI 13.1%, 20.5%) in 2006–2007 to 25.6% (95% CI 17.9%, 34.6%) in 2014–2015 (P = 0.017). Among adults ages 45–64 years, this percentage increased from 18.2% (95% CI 13.5%, 23.7%) in 2006–2007 to 28.2% (95% CI 18.4%, 39.8%) in 2014–2015 (P = 0.028). The increases in the age 18–44 years and age ≥65 years groups were not statistically significant (P = 0.13 and P = 0.052, respectively) (Figure 1 and Table 2). Among women, the percentage of visits with opioids prescribed increased from 18.0% (95% CI 14.0%, 22.6%) in 2006–2007 to 24.1% (95% CI 18.3%, 30.7%) in 2014–2015 (P = 0.029). Similar trends were found among men (P = 0.034; intervening biennial data details not shown).

Any-listed arthritis visits with opioids prescribed. Overall, among visits with any-listed diagnosis of arthritis, the percentage of visits with opioids prescribed increased from 17.4% (95% CI 14.6%, 20.4%) in 2006–2007 to 25.0% (95% CI 19.7%, 30.8%) in 2014–2015 (P = 0.004). Among adults ages 45–64 years, this percentage increased from 19.2% (95% CI 15.6%, 23.3%) in 2006–2007 to 27.9% (95% CI 21.4%, 35.3%) in 2014–2015 (P = 0.009). Among adults ages ≥65 years, the percentage of visits with opioids prescribed increased from 13.7% (95% CI 10.8%, 17.0%) in 2006–2007 to 22.8% (95% CI 17.1%, 29.3%) in 2014–2015 (P = 0.005). No statistically significant trends among adults ages 18–44 years were noted (Figure 2 and Table 3). Among women, the percentage of visits with opioids prescribed increased from 19.5% (95% CI 16.3%, 22.9%) in 2006–2007 to 24.5% (95% CI 20.7%, 28.7%) in 2014–2015 (P = 0.008). Similar trends were found among men (P = 0.016; intervening biennial data details not shown).

DISCUSSION

From 2006 to 2015, the percentage of visits to office-based physicians by adults with arthritis, whether first-listed or any-listed, significantly increased overall for most age groups and for men and often women. This growth in ambulatory care is consistent with the growing estimates and projections of arthritis prevalence and suggests 2 related problems. The first problem is the growth in the cost of arthritis, which was $304 billion in 2013 (8). The second problem is the growth of arthritis as a common comorbid condition that can reduce physical activity and thereby interfere with the management of diabetes mellitus, heart disease, obesity, and other chronic conditions (11). During this same study period, the overall percentage of arthritis visits with opioids prescribed for this increasing ambulatory population, whether first-listed or any-listed, significantly increased as well. In part, this increase may reflect greater opioid prescriptions during this time period. It also suggests, along with a report that those with arthritis comprise the majority of those prescribed opioids (4), a growing ambulatory care problem of arthritis pain management.

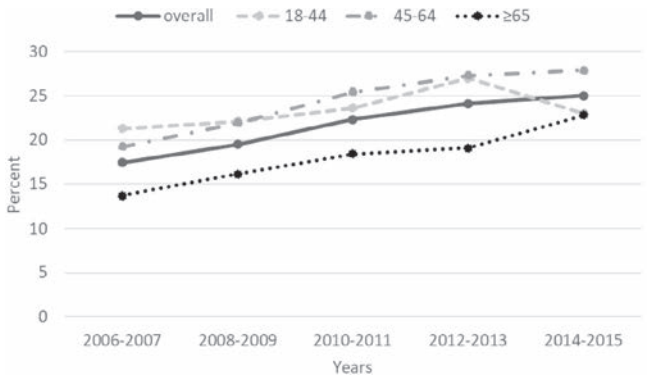


Figure 2. Trends in percentage of visits to office-based physicians made by adults with any-listed diagnosis of arthritis in which at least 1 opioid was prescribed, by age group in the US from 2006 to 2015.

Table 2. Trends in percentage of visits to office-based physicians made by adults with a first-listed diagnosis of arthritis in which at least 1 opioid was prescribed, by age group in the US, 2006–2015*

	2006–2007		2008–2009		2010–2011		2012–2013		2014–2015		
Age range	No. of visits (%)	95% CI	No. of visits (%)	95% CI	No. of visits (%)	95% CI	No. of visits (%)	95% CI	No. of visits (%)	95% CI	*P* for trend
Overall	216 (16.5)	13.1–20.5	242 (17.0)	(14.1–20.2)	317 (20.4)	17.4–23.7	992 (22.5)	20.2–24.9	536 (25.6)	17.9–34.6	0.017
18–44 years	47 (18.4)	12.3–25.9	53 (20.7)	(14.8–27.6)	71 (23.5)	17.2–30.7	177 (23.8)	19.2–28.9	96 (26.8)	17.2–38.4	0.13
45–64 years	101 (18.2)	13.5–23.7	111 (17.7)	(14.0–21.9)	146 (22.4)	18.3–27.0	490 (25.0)	21.8–28.4	257 (28.2)	18.4–39.8	0.028
≥65 years	68 (13.8)	10.1–18.4	78 (14.9)	(11.1–19.4)	100 (16.7)	12.7–21.3	325 (18.9)	16.2–21.8	183 (22.5)	14.2–32.7	0.052

* 95% CI = 95% confidence interval.

There is, however, limited evidence that opioids help with chronic arthritis pain (6).

Some limitations of this study include the inability to ascertain medication adherence, the inability to directly link prescribed opioids to an arthritis diagnosis (although, the 3 most common other diagnostic categories were endocrine, nutritional, metabolic diseases, and immunity disorders [17%], diseases of the circulatory system [8%], and diseases of the nervous system [5%], for which opioids are not typically prescribed), and the need to combine 2-year periods to enable reliable estimation. Even with 2 years of data combined, statistical power was limited to detect significant trends for some groups, and stratified analysis by specific diagnoses within the arthritis definition could not be conducted. Strengths include using a large, standard, nationally representative survey with multiple years of data and the ability to address diagnoses, medications, and demographic variables concurrently. The NAMCS data from 2016 and later may help monitor these trends.

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. Helmick 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. Santo, Hootman, Helmick.

Acquisition of data. Santo.

Analysis and interpretation of data. Santo, Schappert.

REFERENCES

1. 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.
2. Hootman JM, Helmick CG, Barbour KE, Theis KA, Boring MA. Updated projected prevalence of self-reported doctor-diagnosed arthritis and arthritis-attributable activity limitation among US adults, 2015–2040. Arthritis Rheumatol 2016;68:1582–7.
3. Arthritis in America. URL: https://www.cdc.gov/vitalsigns/arthritis/index.html.
4. Murphy LB, Cisternas MG, Theis KA, Brady TJ, Bohm MK, Guglielmo D, et al. All-cause opioid prescriptions dispensed: the outsized role of adults with arthritis. Am J Prev Medicine 2020;59:355–66.
5. Dowell D, Haegerich TM, Chou R. CDC guideline for prescribing opioids for chronic pain: United States, 2016. MMWR Recomm Rep 2016;65:1–49.
6. Krebs EE, Gravely A, Nugent S, Jensen AC, DeRonne B, Goldsmith ES, et al. Effect of opioid vs nonopioid medications on pain-related function in patients with chronic back pain or hip or knee osteoarthritis pain: the SPACE randomized clinical trial. JAMA 2018;319:872–82.
7. National Center for Health Statistics. National Ambulatory Medical Care Survey. URL: https://www.cdc.gov/nchs/ahcd/.
8. Murphy LB, Cisternas MG, Pasta DJ, Helmick CG, Yelin EH. Medical expenditures and earnings losses among US adults with arthritis in 2013. Arthritis Care Res (Hoboken) 2018;70:869–76.

Table 3. Trends in percentage of visits to office-based physicians made by adults with any-listed diagnosis of arthritis in which at least one opioid was prescribed, by age group in the US, 2006–2015*

	2006–2007		2008–2009		2010–2011		2012–2013		2014–2015		
Age range	No. of visits (%)	95% CI	No. of visits (%)	95% CI	No. of visits (%)	95% CI	No. of visits (%)	95% CI	No. of visits (%)	95% CI	*P* value for trends
Overall	388 (17.4)	14.6–20.4	444 (19.5)	17.0–22.2	547 (22.3)	19.9–24.9	1,784 (24.1)	21.9–26.5	984 (25.0)	19.7–30.8	0.004
18–44	91 (21.3)	15.7–27.9	95 (22.1)	17.5–27.3	114 (23.6)	18.5–29.3	351 (27.0)	23.1–31.2	162 (23.0)	16.2–31.1	0.398
45–64	180 (19.2)	15.6–23.3	214 (21.9)	18.3–25.9	265 (25.4)	21.6–29.5	873 (27.3)	24.1–30.6	469 (27.9)	21.4–35.3	0.009
≥65	117 (13.7)	10.8–17.0	135 (16.1)	12.9–19.8	168 (18.4)	15.5–21.6	560 (19.1)	17.0–21.5	353 (22.8)	17.1–29.3	0.005

* 95% CI = 95% confidence interval.

9. Murphy LB, Cisternas MG, Greenlund KJ, Giles W, Hannan C, Helmick CG. Defining arthritis for public health surveillance: methods and estimates in four US population health surveys. Arthritis Care Res (Hoboken) 2017;69:356–67.
10. Ingram DD, Malec DJ, Makuc DM, Kruszon-Moran D, Gindi RM, Albert M, et al. National Center for Health Statistics guidelines for analysis of trends. Vital Health Stat 2 2018;179: 1–71.
11. Barbour KE, Moss S, Croft JB, Helmick CG, Theis KA, Brady TJ, et al. Geographic variations in arthritis prevalence, health-related characteristics, and management: United States, 2015. MMWR Surveill Summ 2018;67:1–28.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 73, No. 10, October 2021, pp 1436–1443
DOI 10.1002/acr.24349

Cognitive Function Trajectories in Association With the Depressive Symptoms Trajectories in Systemic Lupus Erythematosus Over Time

Zahi Touma,[1] Bahar Moghaddam,[1] Jiandong Su,[1] and Patricia Katz[2]

Objective. Cognitive function may change over time in patients with systemic lupus erythematosus (SLE), and cognitive function trajectories have not been well studied. We aimed to identify cognitive function trajectories in SLE and describe them with depressive symptoms trajectories, and we also aimed to identify baseline factors associated with class membership in the dual trajectories.

Methods. Longitudinal data from the University of California San Francisco Lupus Outcomes Study were analyzed. Two outcome trajectories were studied jointly, the Hopkins Verbal Learning Test–Revised and the Center for Epidemiologic Studies Depression Scale (CES-D) (administered annually). Univariate/multivariable logistic regression analyses examined baseline factors associated with class memberships.

Results. A total of 755 patients were studied, and 4 latent classes were identified: 1) low CES-D scores and low cognitive scores (no depression plus cognitive impairment; 20%), 2) lowest CES-D scores and highest normal cognitive scores (no depression plus normal cognition; 48%), 3) highest CES-D scores and lowest cognitive scores (depression plus cognitive impairment; 9%), and 4) high CES-D scores and cognitive score at borderline (depression plus borderline cognition; 23%).

Conclusion. In all, 4 distinct classes of dual cognitive function and depressive symptoms were identified. Persistently low cognitive performance in 28% of patients (classes 1 and 3) did not significantly improve over 7 years. Cognitive impairment was associated with depression status in 9% of patients (class 3). Other factors also predicted latent class membership: ethnicity, education, disease activity, physical functioning, and bodily pain. These results highlight the importance of periodic assessment of cognitive function and of different aspects relevant for assessing and managing cognitive function over time in SLE.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a chronic autoimmune disorder with multisystemic manifestations that include neuropsychiatric SLE (1). The American College of Rheumatology (ACR) has proposed 19 discrete central and peripheral nervous system syndromes as neuropsychiatric SLE (2). This proposal includes both cognitive dysfunction and mood disorder. Cognitive impairment is among the most commonly reported neuropsychiatric symptoms among patients with SLE, with a prevalence of 33–43% (1–3). Cognitive impairment can involve any of the following functions as defined by ACR nomenclature: "memory (learning and recall), complex attention, simple attention, executive skills (planning, organizing, and sequencing), visual-spatial processing, language (e.g., verbal, fluency), reasoning/problem solving, and psychomotor speed" (2). Previous studies have shown that patients with SLE perform poorly compared to controls on standardized neuropsychiatric testing, with decreased attention, impairment in working memory, and decreased executive function (4,5).

Impairment in cognitive function can be subtle and can fluctuate over time. Several studies have addressed the longitudinal course of cognitive impairment, but their results are limited by

The Lupus Outcomes Study was supported by the NIH/National Institute of Arthritis and Musculoskeletal and Skin Diseases (grant AR053308). Dr. Touma's work was supported by the Arthritis Society Salary Support and by the Canadian Rheumatology Association Arthritis Society Clinician Investigator Salary Award.

[1]Zahi Touma, MD, PhD, Bahar Moghaddam, MD, Jiandong Su, MSc, BMSc, BSc: Toronto Western Hospital and University of Toronto, Toronto, Ontario, Canada; [2]Patricia Katz, PhD: University of California, San Francisco.

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

Address correspondence to Zahi Touma, MD, PhD, Division of Rheumatology, Department of Medicine, Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada (email: Zahi.Touma@uhn.ca); or to Patricia Katz, PhD, Professor of Medicine and Health Policy, Division of Rheumatology, Department of Medicine, Philip R. Lee Institute for Health Policy Studies, University of California San Francisco, San Francisco, CA 94143 (email: Patti.Katz@ucsf.edu).

Submitted for publication February 21, 2020; accepted in revised form June 2, 2020.

SIGNIFICANCE & INNOVATIONS

- Persistently low cognitive performance in 28% of patients (classes 1 and 3) did not significantly improve over 7 years.
- Normal cognitive performance in 72% (classes 2 and 4) followed a stable trajectory.
- Four latent classes of dual cognitive function and depressive symptoms were identified.
- Factors associated with latent class membership include ethnicity, education, disease activity, physical functioning, and bodily pain.

small sample size and short duration of follow-up. The sample size in these studies ranges between 28 and 188 patients, and the duration of follow-up did not surpass 3–5 years (6–8). Furthermore, while studies have shown that cognitive function may fluctuate over time in SLE patients (3), cognitive function trajectories have not been previously studied in SLE.

Comorbid depression is also common in SLE patients (9,10), with a prevalence of 30–40% based on the results of a recent systematic review (11). Depression has been shown to be associated with cognitive impairment in cross-sectional SLE studies (3,10,12). These studies have shown that the association of depression and cognitive impairment in SLE can be independent of baseline demographic factors and disease activity (13). Despite this, the longitudinal relationship between cognition and depression in SLE patients has not been studied previously.

Our study is the first to identify joint cognitive function and depressive symptom trajectories in SLE. We used group-based trajectory modeling to determine whether cognitive function and depressive symptoms in patients with SLE can be clustered in discrete latent classes (14). We have used previously validated tools, including the Hopkins Verbal Learning Test–Revised (HVLT-R) and the Center for Epidemiologic Studies Depression Scale (CES-D) in generating trajectories (10–12). In addition, we aimed to identify baseline factors associated with trajectory membership.

PATIENTS AND METHODS

Patients and data. Data were obtained from the University of California San Francisco (UCSF) Lupus Outcomes Study (15), in which patients are followed longitudinally since 2002 via annual telephone structured surveys conducted by trained interviewers. Patients with physician-diagnosed SLE were confirmed to meet the 1997 ACR revised SLE criteria (16). The annual survey encompasses the following domains: demographic characteristics and socioeconomic status, SLE disease activity by Systemic Lupus Activity Questionnaire (SLAQ) (17), bodily pain and physical functioning as reported on the Short Form 36 (SF-36) health questionnaire (18), education and employment, cognitive function, and medications use (glucocorticoids, antimalarials, and immunosuppressants). Income was coded as above or below 125% of the US Federal poverty threshold, based on income and household size. All participants provided informed consent for the data collection and the study protocol was approved by the UCSF Committee on Human Research.

Measures. The HVLT-R, which measures verbal memory, was administered annually in years 2–7, providing up to 6 waves of observation (wave 2 is the baseline assessment for cognitive function). Age- and education-adjusted Z scores were derived for HVLT-R delayed recall. The HVLT-R is a word-list learning and memory test that encompasses recall and recognition. This tool provides an assessment of verbal learning efficiency, ability to access newly learned information, and retention (19). The CES-D (range 0–60; a score ≥24 represents depression in SLE) was administered yearly (20). Each CES-D item includes 4 response categories, with possible scores between 0 and 3. The CES-D score is a sum of the 20 items, which can total between 0 and 60. The Systemic Lupus Erythematosus Activity Questionnaire (SLAQ) is available from the second interview year and then annually. The percentage of missing data for different variables is shown in Table 1.

Statistical analysis. Demographic characteristics were described by mean ± SD for continuous variables and frequency (percentage) for categorical variables. Patients' HVLT-R and CES-D scores over follow-up were modeled using dual trajectory analysis (21). For modeling the trajectory analysis, the study period starts from the first administration of the cognitive tests and extends over 6 waves of observations performed annually. Dual trajectory analysis allows the investigation of the dynamic interrelationship between 2 outcomes over time. In our study, dual trajectory modeling was used to assess the extent of association between latent patterns of cognitive function (measured by the HVLT-R) and latent patterns of depressive symptoms (measured by the CES-D). Dual latent class trajectory analysis for the HVLT-R and the CES-D was performed using a group-based trajectory model on 755 patients in SAS software, procedure PROC TRAJ, version 9.3 (22). The dual models for the HVLT-R and the CES-D with up to 2–6 classes were assessed, and the best models were determined by a combination of clinical plausibility (we excluded models in which individual groups were too small [<10%] to have clinical significance, where the differences between cognitive trajectories would not be clinically significant) and statistical criteria (planning to look for the minimum points of the Bayesian information criterion and the Akaike information criterion [AIC]). The model outputs included the identification of the appropriate number of dual trajectories and their shapes, the percentage of patients of each trajectory, and the estimated combined and conditional probabilities of group membership of each dual trajectory.

Posterior probability is a parameter of model adequacy when grouping individuals into a particular trajectory. Using only members with a dual latent class posterior probability of

Table 1. Characteristics at the baseline cognitive assessment for all study participants*

Characteristic	All study participants (n = 755)	Class 1 (n = 151)	Class 2 (n = 360)	Class 3 (n = 71)	Class 4 (n = 173)
Female	698 (92.5)	134 (88.7)	332 (92.2)	68 (95.8)	164 (94.8)
Age at SLE diagnosis, mean ± SD years	34.3 ± 13.4	35.5 ± 14.2	33.2 ± 13.3	35.3 ± 12.5	35.6 ± 11.9
Disease duration, mean ± SD years	15.5 ± 8.5	15.9 ± 8.7	15.5 ± 8.7	15.6 ± 8.4	15.1 ± 7.8
Ethnicity					
Missing	34 (4.5)	7 (4.6)	19 (5.3)	3 (4.2)	5 (2.9)
White	519 (68.7)	95 (62.9)	254 (70.6)	41 (57.7)	129 (74.6)
Hispanic	54 (7.2)	16 (10.)	18 (5.0)	6 (8.5)	14 (8.1)
African American	47 (6.2)	11 (7.3)	15 (4.2)	10 (14.1)	11 (6.4)
Asian	61 (8.1)	15 (9.9)	32 (8.9)	5 (7.0)	9 (5.2)
Other	40 (5.3)	7 (4.6)	22 (6.1)	6 (8.5)	5 (2.9)
Education					
Less than high school	15 (2.0)	3 (2.0)	2 (0.6)	8 (11.3)	2 (1.2)
High school graduate	81 (10.7)	23 (15.2)	21 (5.8)	18 (25.4)	19 (11.0)
Some college	214 (28.3)	56 (37.1)	85 (23.6)	18 (25.4)	55 (31.8)
Trade or vocational school	136 (18.0)	23 (15.2)	55 (15.3)	20 (28.2)	38 (22.0)
College graduate	175 (23.2)	35 (23.2)	100 (27.8)	4 (5.6)	36 (20.8)
Postgraduate degree	134 (17.7)	11 (7.3)	97 (26.9)	3 (4.2)	23 (13.3)
Employment status					
Unemployed	397 (52.6)	82 (54.3)	147 (40.8)	59 (83.1)	109 (63.0)
Employed	358 (47.4)	69 (45.7)	213 (59.2)	12 (16.9)	64 (37.0)
Below poverty threshold	83 (11.0)	19 (12.6)	19 (5.3)	23 (32.4)	22 (12.7)
Smoking status					
Missing	4 (0.5)	0 (0.0)	2 (0.6)	0 (0.0)	2 (1.2)
Current smoker	72 (9.5)	11 (7.3)	27 (7.5)	9 (12.7)	25 (14.5)
Former smoker	242 (32.1)	55 (36.4)	107 (29.7)	23 (32.4)	57 (32.9)
Never smoker	437 (57.9)	85 (56.3)	224 (62.2)	39 (54.9)	89 (51.4)
Glucocorticoid exposure					
Oral	320 (42.4)	69 (45.7)	144 (40.0)	30 (42.3)	77 (44.5)
Intravenous	47 (6.2)	13 (8.6)	17 (4.7)	10 (14.1)	7 (4.0)
DMARD therapy					
Hydroxychloroquine	384 (50.9)	86 (57.0)	179 (49.7)	40 (56.3)	79 (45.7)
Azathioprine	55 (7.3)	13 (8.6)	26 (7.2)	3 (4.2)	13 (7.5)
Methotrexate, oral	51 (6.8)	10 (6.6)	21 (5.8)	6 (8.5)	14 (8.1)
Methotrexate, subcutaneous	23 (3.0)	3 (2.0)	7 (1.9)	2 (2.8)	11 (6.4)
Mycophenolate mofetil	73 (9.7)	12 (7.9)	34 (9.4)	8 (11.3)	19 (11.0)
Cyclophosphamide	6 (0.8)	2 (1.3)	0 (0.0)	2 (2.8)	2 (1.2)
SLE disease activity (SLAQ)					
Mean ± SD	4.1 ± 2.7	4.0 ± 2.8	3.1 ± 2.4	6.2 ± 2.3	5.3 ± 2.5
Median (IQR)	4 (2–6)	4 (2–6)	3 (1–5)	7 (5–8)	5 (4–7)
SF-36 physical function					
Mean ± SD	58.9 ± 29.7	56.5 ± 30.2	71.0 ± 26.7	31.4 ± 19.4	47.3 ± 25.8
Median (IQR)	60 (35–85)	60 (30–85)	80 (50–95)	30 (20–45)	50 (25–65)
SF-36 bodily pain					
Mean ± SD	41.8 ± 11.0	42.4 ± 10.1	46.4 ± 10.3	31.5 ± 8.3	36.0 ± 8.3
Median (IQR)	41 (33–50)	42 (33–50)	46 (37–55)	31 (25–37)	37 (29–42)

* Values are the number (%) unless indicated otherwise. DMARD = disease-modifying antirheumatic drug; IQR = interquartile range; SF-36 = Short Form 36 health survey; SLAQ = Systemic Lupus Activity Questionnaire; SLE = systemic lupus erythematosus.

>0.80 (n = 655), we further conducted univariate/multivariable logistic regression analyses to examine baseline factors associated with latent classes memberships, including sex, ethnicity, education, disease duration, treatments, physical functioning, bodily pain, and self-reported disease activity. We first tested for proportionality of clinical variables, including education, medication use (glucocorticoids and immunosuppressants), SLE disease activity, physical functioning, and bodily pain in between the 4 latent classes. Under the circumstances of lacking proportionality in predictors, we analyzed the association in 2 groups with the best and worst trajectories. Logistic regression analysis comparing class 2 and class 3 was performed, using the latter as a reference. This approach allowed us to determine factors that are associated with normal cognitive function and the absence of depression in the studied cohort. A step-down variable selection method was used in variable selection in multivariable analysis. Variables with the highest *P* values were dropped 1 by 1 until the lowest AIC value was reached. The demographic characteristics of patients with missing data were compared with the rest of the cohort. All patients had visits or were followed up in the first 2

waves (2 and 3), and 31 patients had no more follow-up data after wave 4 (the third year after wave 2).

RESULTS

Patient characteristics. Of the 815 participants in the study, 755 had at least 2 scores recorded for both the HVLT-R and the CES-D in the follow-up period and were included in the analysis. Demographic and clinical characteristics of the cohort are shown in Table 1. The mean ± SD age of the analytic sample was 50.1 ± 12.6 years, and 92.5% (n = 698) were female. The mean ± SD age at SLE diagnosis and disease duration at first visit in the study were 34.3 ± 13.4 and 15.5 ± 8.5 years, respectively. The mean ± SD SLE disease activity as measured by the SLAQ was 4.1 ± 2.7 at the baseline assessment. Over 40% of participants reported exposure to oral glucocorticoids in the past 12 months (at baseline visit). In total, 31 patients had missing data (loss to follow-up). There was no statistically significant difference between patients with loss to follow-up and patients in the first 2 waves and patients who had more follow-up after wave 2 (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24349/abstract).

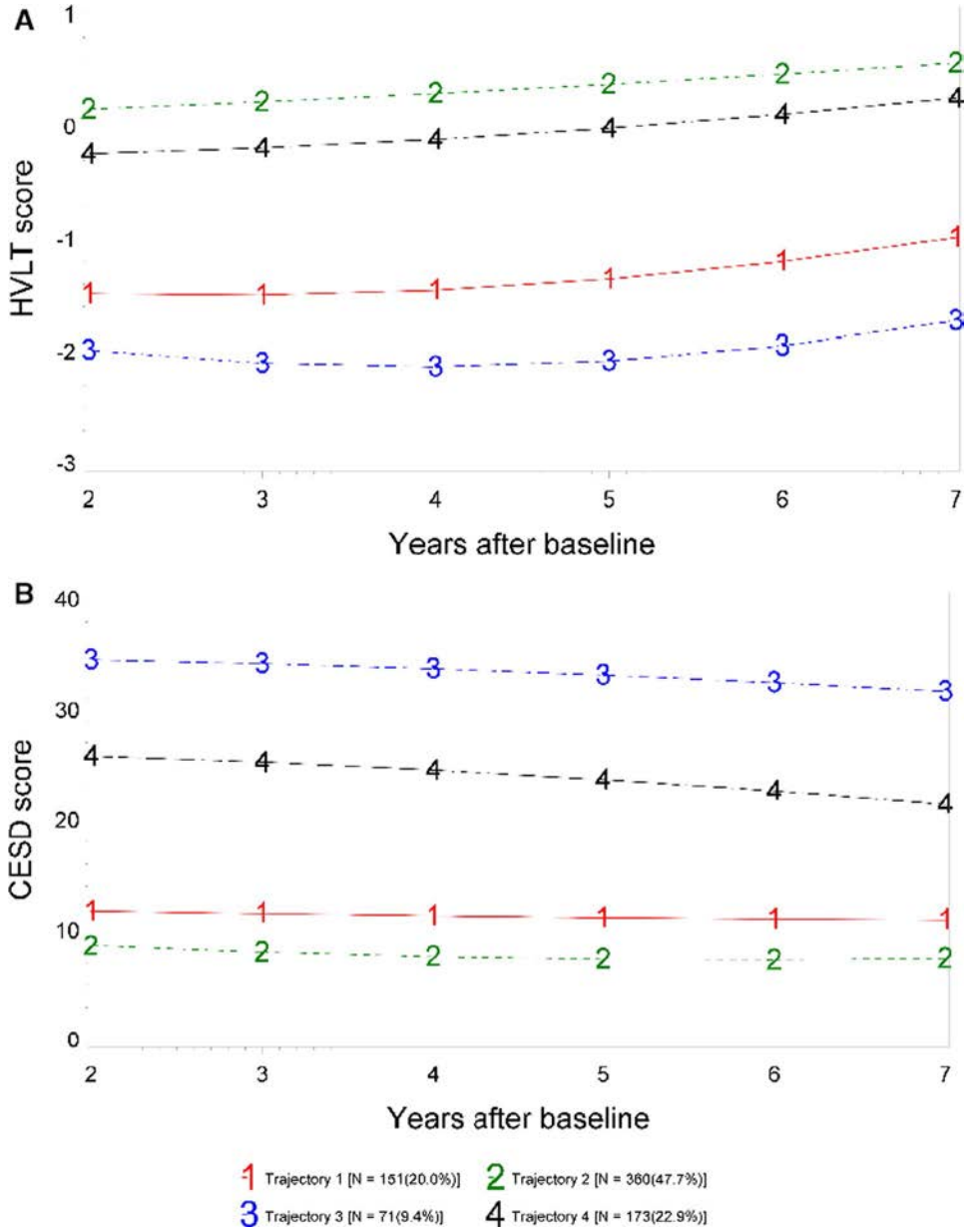


Figure 1. Dual trajectory model of the Hopkins Verbal Learning Test–Revised (HVLT-R) and the Center for Epidemiologic Studies Depression Scale (CES-D) over 7 years. **A**, Predicted HVLT-R score in 4 trajectories: HVLT-R scores presented as Z scores compared to controls. Z scores less than or equal to −1 indicate cognitive impairment. **B**, Predicted CES-D score in 4 trajectories: a score ≥24 represents depression in systemic lupus erythematosus. Members in each class C1 (red), C2 (green), C3 (blue), and C4 (black) are the same across each figure.

Dual latent classes trajectories for cognitive function and depressive symptoms. We identified the best model as having 4 latent classes, with 4 trajectories for cognitive function and 4 trajectories for depressive symptoms over the 6-year follow-up period. Class 1 was defined as low CES-D score and low cognitive scores (no depression plus cognitive impairment; n = 119, 18.6%). Class 2 comprised those with the lowest CES-D score and highest normal cognitive scores (no depression plus normal cognition; n = 334, 51.0%). Class 3 included those with the highest CES-D scores and lowest cognitive scores (depression plus cognitive impairment; n = 61, 9.3%). Class 4 consisted of those with high CES-D scores and normal cognitive scores (depression plus normal cognition; n = 141, 21.5%) (Figure 1).

Classes with persistent HVLT-R Z scores of –1 and below reflect persistent low cognition, and classes with CES-D scores of ≥24 reflect persistent depression. Two classes, 1 and 3, comprising 18.6% and 9.3% of patients, respectively, displayed persistently low cognition. Two classes, 3 and 4, comprising 9.3% and 21.5% of patients, respectively, displayed persistent depression (Figure 2 and Figure 3).

Demographic information and clinical characteristics of each class are shown in Table 1. There were no significant differences in age or disease duration between the 4 latent classes. Ethnic composition varied between latent classes, with a higher percentage of White patients in classes 2 and 4, where cognitive function was graded as normal (70.6% and 74.6%, respectively), and a higher percentage of African American patients (14.1%) in class 3, where both depressive symptoms and cognitive impairment are observed. The highest education levels were observed in class 2, with 56.8% having a college degree or higher education. The highest frequency of lower income levels was observed in class 3 and the lowest in class 2 (31.1% and 4.8%, respectively). Patients in class 2 also reported higher quality of life measures, with the highest values of SF-36 physical functioning and bodily pain among all 4 classes. Additionally, this group had the lowest disease activity as determined by the SLAQ.

Univariate/multivariable logistic regression analyses (modeling factors at baseline associated with normal cognition and no depression). The cohort size after removing patients with low posterior probability (<0.80) was 655 (patient characteristics are in Supplementary Table 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24349/abstract). White ethnicity and higher levels of education were associated with normal cognitive function and the absence of depression (Table 2). White patients were found to have an odds ratio (OR) of 2.1 (95% confidence interval [95% CI] 1.19–3.8) and OR 4.3 (95% CI 1.69–10.86) times the odds of normal cognitive function and low CES-D score, respectively, in univariate and multivariate regression analyses. Similarly, higher levels of education were associated with normal cognitive performance and lower depressive symptoms with OR 2.22 (95% CI 1.75–2.81) and OR 2.52 (95% CI 1.73–3.66), respectively, in univariate and multivariate analyses. Higher SF-36 scores in physical function were associated with membership in class 2, with OR 1.06 (95% CI 1.05–1.07) and OR 1.04 (95% CI 1.02–1.06) in univariate and multivariate regression, respectively. Similarly, higher SF-36 scores in bodily pain were associated with membership in class 2, with OR 1.17 (95% CI 1.12–1.21) and OR 1.12 (95% CI 1.06–1.18) in univariate and multivariate regression, respectively. Higher disease activity at baseline and the use of disease-modifying antirheumatic drugs (DMARDs) were not associated with class 2 membership (normal cognition and the absence of depression).

DISCUSSION

Neuropsychiatric SLE manifestations, including cognitive impairment and depression, are commonly reported in literature (3). Our study is the first to describe joint cognitive function and depressive symptom trajectories in patients with SLE. Using group-based trajectory modeling, we were able to identify 4 distinct dual trajectory patterns for depression and cognitive function in SLE patients over a 7-year period.

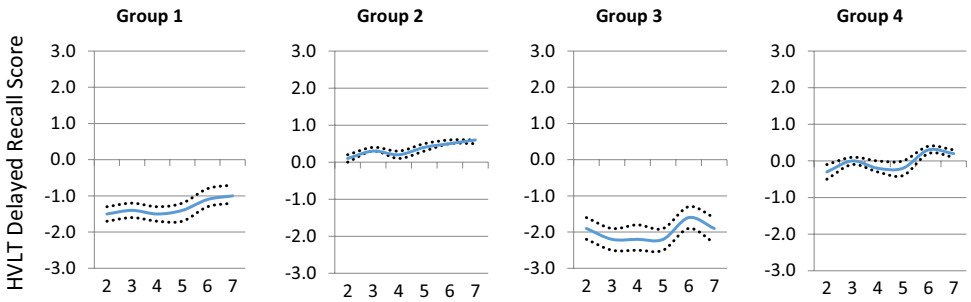


Figure 2. Individual trajectories for Hopkins Verbal Learning Test–Revised (HVLT-R) scores in latent classes. HVLT-R scores presented as Z scores compared to controls. Z scores less than or equal to –1 indicate cognitive impairment. Group 1 is no depression and with cognitive impairment (n = 119); group 2 is no depression and with normal cognition (n = 334); group 3 is depression and with cognitive impairment (n = 61); group 4 is depression and normal cognition (n = 141). Blue lines indicate individual trajectories. Broken black lines indicate 95% confidence intervals. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24349/abstract.

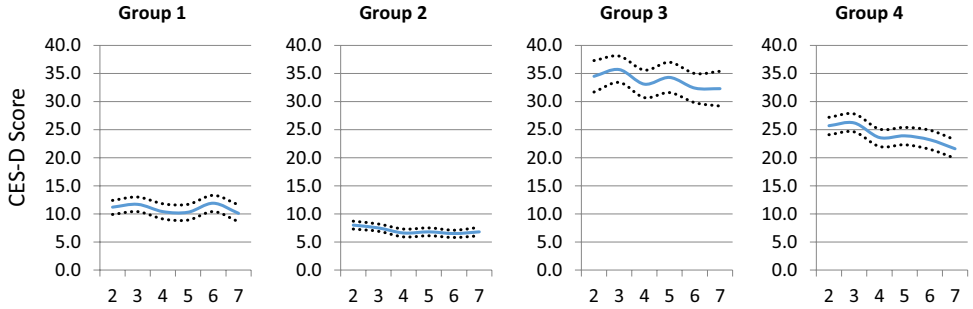


Figure 3. Individual trajectories for Center for Epidemiologic Studies Depression Scale (CES-D) scores in latent classes. A score of ≥24 represents depression in systemic lupus erythematosus. Group 1 is no depression and with cognitive impairment (n = 119); group 2 is no depression and with normal cognition (n = 334); group 3 is depression and with cognitive impairment (n = 61); group 4 is depression and normal cognition (n = 141). Blue lines indicate individual trajectories. Broken black lines indicate 95% confidence intervals. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24349/abstract.

Overall, we observed persistently low cognitive scores as determined by poor performance in the HVLT-R in 27.9% of our cohort (classes 1 and 3). In both latent classes, the cognitive function trajectories did not fluctuate significantly throughout the follow-up period, demonstrating a stable trajectory without further decline or improvement. Similarly, cognitive function trajectories did not fluctuate with normal cognition (classes 2 and 4), again following a stable trajectory without progression to cognitive impairment.

Our study provides a long follow-up period for monitoring cognitive function in SLE patients. Previous longitudinal studies on the course and outcome of cognitive impairment in SLE patients have reported variable findings and have been limited by small sample size and short duration of follow-up. Hanly et al reported cognitive impairment in 21% of a small prospective cohort (n = 70) of SLE patients and observed resolution of cognitive impairment in the majority of their cohort after 1 year of follow-up, with only 12% exhibiting persistent cognitive impairment (23). In a 5-year follow-up study of the same cohort, persistent cognitive impairment was only present in 4% (2 patients) and resolution of cognitive impairment in 19% of their cohort (8). In another small prospective cohort (28 patients), Waterloo et al reported stable cognitive function in a 5-year follow-up period, with resolution of cognitive impairment in a small subset of patients who had improvement of underlying psychiatric disorders (24). In another small prospective cohort (n = 43), Ceccarelli et al reported a prevalence of 20.9% of cognitive impairment at baseline and 13.9% at follow-up in 10 years (25).

Contrary to the aforementioned studies, we did not observe resolution of cognitive impairment over time in class 1 (20%) and class 3 (9%). This difference may be accounted for by the differences in follow-up period and sample sizes or because we used a research cohort rather than a clinical sample, and individuals with significant levels of cognitive impairment may have been screened out of the study due to inability to provide informed consent. Furthermore, each study used different neuropsychiatric assessments of cognitive dysfunction. While all of these assessment tools have been previously validated in studying cognitive impairment in patients with SLE, they vary in specificity and sensitivity, and a direct comparison between the results of these studies may not be possible (26).

Depression, as determined by high scores on the CES-D, was observed in 31% of the study cohort (classes 3 and 4). Similar to our finding of stable cognitive function, trajectories of depressive

Table 2. Baseline factors associated with normal cognitive function and absence of depression*

Variable	Univariate	*P*	Multivariate	*P*
Female	0.39 (0.09–1.677)	0.20	–	–
White	2.13 (1.19–3.80)	0.01	4.28 (1.69–10.86)	0.002
Disease duration	0.99 (0.97–1.03)	0.96	–	–
Education	2.22 (1.75–2.81)	<0.001	2.52 (1.73–3.66)	<0.0001
Income below poverty threshold	0.113 (0.05–0.24)	<0.0001	–	–
Smoking status	0.71 (0.41–1.24)	0.23	–	NS
SLE disease activity	0.60 (0.52–0.68)	<0.001	0.77 (0.62–0.94)	0.01
Glucocorticoid, IV	0.27 (0.12–0.63)	0.002	–	–
Glucocorticoid, oral	0.83 (0.48–1.44)	0.52	–	–
Any DMARDs†	1.13 (0.58–2.18)	0.73	–	–
Hydroxychloroquine	0.76 (0.44–1.32)	0.33	0.23 (0.09–0.60)	0.002
SF-36 physical function	1.06 (1.05–1.07)	<0.0001	1.04 (1.02–1.06)	0.0004
SF-36 bodily pain	1.17 (1.12–1.21)	<0.0001	1.12 (1.06–1.18)	<0.0001

* Values are the odds ratio (95% confidence interval), unless indicated otherwise. The cohort size after removing patients with low posterior probability (<0.80) was 655. Class 3 is the reference group as compared to class 2. DMARDs = disease-modifying antirheumatic drugs; IV = intravenous; SF-36 = Short Form 36 health survey; SLE = systemic lupus erythematosus.
† Any immunosuppressant includes any use of methotrexate, oral or subcutaneous, azathioprine, mycophenolate mofetil, and cyclophosphamide.

symptoms did not fluctuate significantly throughout the follow-up period. Previous studies have reported a wide range in prevalence of depression in SLE, ranging from 30% to 40% (11,26). Despite this high prevalence, trends of depression in persons with SLE over time have not been consistently reported in literature. Huang et al (27) reported an incidence of 29.7 per 1,000 person-years in a cohort followed over 26 years. However, incidence of depression was determined based on recorded chart data, diagnostic codes, and antidepressant use, as opposed to validated screening tools.

We compared 2 latent classes to determine baseline factors associated with membership in each trajectory. We chose class 2 (high cognitive and low depression scores) and class 3 (low cognitive and high depression scores) for this comparison because they represent the most differing classes. In our univariate regression analysis, we found White ethnicity (OR 2.13 [95% CI 1.19–3.80]) and higher education levels (OR 2.22 [95% CI 1.75–2,81]) to be associated with normal cognitive function and the absence of depression. This association remained significant in the multivariate regression analysis. High baseline SLE disease activity and use of DMARDs were not associated with normal cognitive function and the absence of depression. Demographic differences in cognitive impairment in SLE, including ethnicity and education levels, have been previously explored in several studies (28–31). While some did not find any association between ethnicity and cognitive performance in SLE patients, others have reported a higher prevalence of cognitive impairment among African American patients and a lower prevalence among Asian patients. These differences may be related to background socioeconomic factors (28), which may influence access to resources, including education. These demographic differences in turn can alter cognitive reserve, which has been shown to be increased in those with higher education levels (32). Indeed, having a higher education level has been linked to higher cognitive reserve and larger hippocampal volumes, allowing individuals to capitalize on greater structural integrity of the brain (33). Sociodemographic factors, specifically lower income, financial strain, and low education have also been associated with depression in those with SLE (9).

There are limitations and strengths to our study. The first limitation is the use of the HVLT-R as opposed to the ACR comprehensive neuropsychologic battery of tests. While the HVLT-R is both time- and cost-efficient, it only identifies impairment in episodic verbal learning memory. Several previous studies have shown that cognitive impairment in SLE can occur in any cognitive domain, including decreased attention, impaired working memory, executive function, and overall cognitive slowing (4). However, the HVLT-R has been previously validated against the ACR SLE neuropsychologic battery and has been shown to have a sensitivity of 81% in identifying cognitive impairment in patients with SLE (10), and it is therefore a reasonable assessment tool for our study. A possibility exists that those with the cognitive impairment had fewer years of follow-up or were more likely to drop out. In our study, 31 patients were lost to follow-up after wave 2. The demographic characteristics of patients lost to follow-up did not differ from those who remained in the study. We conducted annual telephone surveys to collect our data as opposed to in-person interviews.

While face-to-face assessments are preferred, especially for evaluation of depression, our current model allowed the collection of a larger pool of data within the limits of resources. Last, in interpreting our results, an important consideration to keep in mind is the complex interplay between depression and cognitive impairment. For instance, cognitive impairment can reduce occupational productivity and interfere with several domains of social functioning. These emotional and social factors may lead to depression. Conversely, symptoms of depression, including poor concentration or psychomotor slowing, may impact performance on cognitive tasks (34,35).

Our study has the advantage of a long-term follow-up period with annual reassessment, providing continuous monitored data points. However, individuals with severe illness, including those with cognitive impairment, may be underrepresented. Additionally, our results are based on outcomes from a large cohort and are more likely to be reflective of a general lupus patient population. Finally, our study is the first to identify distinct trajectories in combined cognitive function and depressive symptoms in SLE.

We have shown that persistently low cognitive performance in 28% of patients (classes 1 and 3) did not significantly improve over 7 years. In addition, normal cognitive performance in 72% (classes 2 and 4) followed a stable trajectory. This highlights the importance of periodic, yearly assessment of cognitive function in SLE. We have identified 4 distinct classes of combined cognitive function and depressive symptoms in patients with SLE. Overall, low cognitive function is associated with persistent depression in 9.3% of patients. Several factors may be associated with membership in each latent class. These include ethnicity, education level, disease activity, physical functioning, and bodily pain. Further studies are needed to determine how to best integrate our understanding of factors related to depression and cognitive impairment, in assessment and management of SLE patients. In addition, determining whether the management of depression would alter cognitive trajectories is important, particularly in patients with low cognitive function and persistent depression (class 3).

Our results have confirmed the complexity of the underlying elements of sociodemographic factors, disease activity, and comorbidities associated with persistent cognitive impairment, which require further research. There is still, however, an unmet need to identify patients at high risk of developing cognitive impairment, to distinguish between different trajectories of cognitive impairment (persistent cognitive impairment and normal cognitive function), and to monitor cognitive impairment progression over time. These results highlight different aspects relevant for assessing and managing cognitive function over time in SLE. Improving the assessment of cognitive function in patients with SLE will help facilitate the development of treatments or interventions in the future.

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. Touma 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. Touma, Moghaddam, Su, Katz.
Acquisition of data. Touma, Moghaddam, Su, Katz.
Analysis and interpretation of data. Touma, Moghaddam, Su, Katz.

REFERENCES

1. Hanly JG. Diagnosis and management of neuropsychiatric SLE. Nat Rev Rheumatol 2014;10:338–47.
2. ACR Ad Hoc Committee on Neuropsychiatric Lupus Nomenclature. The American College of Rheumatology nomenclature and case definitions for neuropsychiatric lupus syndromes. Arthritis Rheum 1999;42:599–608.
3. Rayes HA, Tani C, Kwan A, Marzouk S, Colosimo K, Medina-Rosas J, et al. What is the prevalence of cognitive impairment in lupus and which instruments are used to measure it? A systematic review and meta-analysis. Semin Arthritis Rheum 2018;48:240–55.
4. Hanly JG, Su L, Omisade A, Farewell VT, Fisk JD. Screening for cognitive impairment in systemic lupus erythematosus. J Rheumatol 2012;39:1371–7.
5. Hanly JG, Omisade A, Su L, Farewell V, Fisk JD. Assessment of cognitive function in systemic lupus erythematosus, rheumatoid arthritis, and multiple sclerosis by computerized neuropsychological tests. Arthritis Rheum 2010;62:1478–86.
6. Denburg SD, Carbotte RM, Denburg JA. Psychological aspects of systemic lupus erythematosus: cognitive function, mood, and self-report. J Rheumatol 1997;24:998–1003.
7. Appenzeller S, Cendes F, Costallat LT. Cognitive impairment and employment status in systemic lupus erythematosus: a prospective longitudinal study. Arthritis Rheum 2009;61:680–7.
8. Hanly JG, Cassell K, Fisk JD. Cognitive function in systemic lupus erythematosus: results of a 5-year prospective study [letter]. Arthritis Rheum 1997;40:1542–3.
9. McCormick N, Trupin L, Yelin EH, Katz PP. Socioeconomic predictors of incident depression in systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2018;70:104–13.
10. Julian LJ, Yazdany J, Trupin L, Criswell LA, Yelin E, Katz PP. Validity of brief screening tools for cognitive impairment in rheumatoid arthritis and systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2012;64:448–54.
11. Moustafa AT, Moazzami M, Engel L, Bangert E, Hassanein M, Marzouk S, et al. Prevalence and metric of depression and anxiety in systemic lupus erythematosus: a systematic review and meta-analysis. Semin Arthritis Rheum 2020;50:84–94.
12. Nantes SG, Su J, Dhaliwal A, Colosimo K, Touma Z. Performance of screening tests for cognitive impairment in systemic lupus erythematosus. J Rheumatol 2017;44:1583–9.
13. Petri M, Naqibuddin M, Carson KA, Wallace DJ, Weisman MH, Holliday SL, et al. Depression and cognitive impairment in newly diagnosed systemic lupus erythematosus. J Rheumatol 2010;37:2032–8.
14. Liu Z, Han L, Gahbauer EA, Allore HG, Gill TM. Joint trajectories of cognition and frailty and associated burden of patient-reported outcomes. J Am Med Dir Assoc 2018;19:304–309.
15. 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.
16. 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.
17. Karlson EW, Daltroy LH, Rivest C, Ramsey-Goldman R, Wright EA, Partridge AJ, et al. Validation of a systemic lupus activity questionnaire (SLAQ) for population studies. Lupus 2003;12:280–6.
18. McHorney CA, Ware JE Jr, Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36). III: tests of data quality, scaling assumptions, and reliability across diverse patient groups. Med Care 1994;32:40–66.
19. Gaines JJ, Shapiro A, Alt M, Benedict RH. Semantic clustering indexes for the Hopkins Verbal Learning Test-Revised: initial exploration in elder control and dementia groups. Appl Neuropsychol 2006;13:213–22.
20. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977;1:385–401.
21. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. Annu Rev Clin Psychol 2010;6:109–38.
22. Jones BL, Nagin DS. Advances in group-based trajectory modeling and an SAS procedure for estimating them. Sociol Methods Res 2007;35:542–71.
23. Hanly JG, Fisk JD, Sherwood G, Eastwood B. Clinical course of cognitive dysfunction in systemic lupus erythematosus. J Rheumatol 1994;21:1825–31.
24. Waterloo K, Omdal R, Husby G, Mellgren SI. Neuropsychological function in systemic lupus erythematosus: a five-year longitudinal study. Rheumatology (Oxford) 2002;41:411–5.
25. Ceccarelli F, Perricone C, Pirone C, Massaro L, Alessandri C, Mina C, et al. Cognitive dysfunction improves in systemic lupus erythematosus: results of a 10 years prospective study. PLoS One 2018;13:e0196103.
26. Kwan A, Katz P, Touma Z. The assessment of anxiety and depression and its associated factors in SLE. Curr Rheumatol Rev 2019;15:90–8.
27. Huang X, Magder LS, Petri M. Predictors of incident depression in systemic lupus erythematosus. J Rheumatol 2014;41:1823–33.
28. Knight AM, Trupin L, Katz P, Yelin E, Lawson EF. Depression risk in young adults with juvenile- and adult-onset lupus: twelve years of followup. Arthritis Care Res (Hoboken) 2018;70:475–80.
29. Meara A, Davidson N, Steigelman H, Zhao S, Brock G, Jarjour WN, et al. Screening for cognitive impairment in SLE using the Self-Administered Gerocognitive Exam. Lupus 2018;27:1363–7.
30. Doninger NA, Fink JW, Utset TO. Neuropsychologic functioning and health status in systemic lupus erythematosus: does ethnicity matter? J Clin Rheumatol 2005;11:250–6.
31. Borowoy AM, Pope JE, Silverman E, Fortin PR, Pineau C, Smith CD, et al. Neuropsychiatric lupus: the prevalence and autoantibody associations depend on the definition: results from the 1000 faces of lupus cohort. Semin Arthritis Rheum 2012;42:179–85.
32. Stern Y. Cognitive reserve: implications for assessment and intervention. Folia Phoniatr Logop 2013;65:49–54.
33. O'Shea DM, Langer K, Woods AJ, Porges EC, Williamson JB, O'Shea A, et al. Educational attainment moderates the association between hippocampal volumes and memory performances in healthy older adults. Front Aging Neurosci 2018;10:361.
34. Knight MJ, Baune BT. Cognitive dysfunction in major depressive disorder. Curr Opin Psychiatry 2018;31:26–31.
35. Nishiguchi Y, Takano K, Tanno Y. The need for cognition mediates and moderates the association between depressive symptoms and impaired effortful control. Psychiatry Res 2016;241:8–13.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 73, No. 10, October 2021, pp 1444–1450
DOI 10.1002/acr.24356

Mortality Among Hospitalized Individuals With Systemic Lupus Erythematosus in the US Between 2006 and 2016

Christine Anastasiou,[1] Laura Trupin,[1] David V. Glidden,[1] Jing Li,[1] Milena Gianfrancesco,[1] Stephen Shiboski,[1] Gabriela Schmajuk,[2] and Jinoos Yazdany[1]

Objective. To evaluate time trends in mortality for hospitalized adults with systemic lupus erythematosus (SLE) compared to the general hospitalized population (GHP), and to identify factors associated with increased risk of death among hospitalized SLE patients.

Methods. We used the National (Nationwide) Inpatient Sample to estimate all-cause mortality for adults discharged from community hospitals in the US between 2006 and 2016. Poisson regression models were used to estimate the risk of in-hospital death among all patients, including demographic characteristics, socioeconomic factors, comorbidity score, hospital region, SLE diagnosis, and race/ethnicity as covariates.

Results. Among 340,467,049 hospitalizations analyzed, 1,903,279 had a discharge diagnosis of SLE. In adjusted analysis, the risk of inpatient death decreased among hospitalizations for patients with SLE from 2.2% to 1.5% ($P < 0.001$) between 2006 and 2016. All of the decrease in SLE mortality occurred between 2006 and 2008; after 2008, mortality stabilized at a rate statistically similar to the GHP. Hospitalizations for Black, Hispanic, and Asian/Pacific Islander patients with SLE were more likely to end in death compared to hospitalizations for either White patients with SLE or individuals of the same non-White race/ethnicity without SLE.

Conclusion. In the largest study of in-hospital SLE mortality published to date, we found significant improvements in mortality for hospitalized patients with SLE in the US from 2006 until 2008, after which mortality stabilized at a level similar to that of the GHP. Our results also demonstrate a persistently high mortality burden among Black and Hispanic patients with SLE in the US and contribute new data revealing high mortality among Asian/Pacific Islander patients with SLE.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a multisystem autoimmune disease that can result in organ damage, frequent hospitalization, and premature death (1–4). Advances in therapy in recent decades have altered rates of damage accrual, hospitalizations, and mortality for patients with SLE, with observational studies suggesting that overall SLE survival has improved from <50% 5-year survival in the 1950s (5) to >90% 10-year survival by 2000 (6). A recent analysis based on death certificate data in the US noted a decrease in the SLE age-standardized mortality rate between 1968 and 2013 but reported that the ratio of SLE to non-SLE mortality had increased 30% since 1968 (7), which raises the question of whether SLE advances have lagged behind medical improvements for other conditions. Most other studies evaluating SLE mortality have been based on patient cohorts, which may not represent the general population, and capture a relatively small number of total deaths (1,6,8–10).

The content herein is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Dr. Anastasiou's work was supported by the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases grant 5T32-AR-007304-40 and National Institute on Aging grant 5T32-AG-049663-04). Dr. Gianfrancesco's work was supported by NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases grant K01-AR-075085). Dr. Schmajuk's work was supported by the Russell/Engleman Medical Research Center for Arthritis. Dr. Yazdany's work was supported by the Robert L. Kroc Chair in Rheumatic and Connective Tissue Diseases, the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases grants P30-AR-070155 and K24-AR-074534), and the Russell/Engleman Medical Research Center for Arthritis.

[1]Christine Anastasiou, MD, Laura Trupin, MPH, David V. Glidden, PhD, Jing Li, MPH, Milena Gianfrancesco, PhD, Stephen Shiboski, PhD, Jinoos Yazdany, MD, MPH: University of California, San Francisco; [2]Gabriela Schmajuk, MD, MSc: University of California, San Francisco and San Francisco VA Medical Center, San Francisco, California.

Dr. Glidden has received consulting fees from Gilead Sciences and Merck & Company (less than $10,000 each). Dr. Yazdany has received consulting fees from Astra Zeneca and Eli Lilly and Company (less than $10,000 each) and research support from Pfizer. No other disclosures relevant to this article were reported.

Address correspondence to Christine Anastasiou, MD, 400 Parnassus Avenue, Floor B1, San Francisco, CA 94143. Email: Christine.Anastasiou@ucsf.edu.

Submitted for publication January 8, 2020; accepted in revised form June 9, 2020.

SIGNIFICANCE & INNOVATIONS

- To our knowledge, this is the largest ever performed primary analysis of hospitalizations for individuals with systemic lupus erythematosus (SLE) and the first large-scale population-based study in >20 years to evaluate in-hospital mortality among Asian/Pacific Islander patients with SLE in the US.
- In-hospital survival for individuals with SLE improved from 2006 through 2008 and then plateaued.
- Hospitalizations for Asian/Pacific Islander, Hispanic, and Black patients with SLE had a higher risk of ending in death compared to hospitalizations for White patients.

The most commonly cited reasons for SLE hospitalization both in the US and internationally are SLE flare and infection (11–15). SLE and connective tissue disorders had a 27% 30-day hospital readmission rate in the US in 2010, ranking it as the sixth most likely principal diagnosis associated with readmission (4). It is known that the majority of deaths in the general population occur in the hospital and that patients are more likely to be admitted to the hospital in the 6 months prior to their death (16). Nevertheless, studies evaluating SLE in-hospital mortality are limited. The most recent multiyear national evaluation of mortality for hospitalized SLE patients in the US is from data spanning 1998–2002 (17). The few studies describing SLE in-hospital mortality in the US over the past 15 years suggest that mortality may be decreasing over time (12,18). To our knowledge, this research conducted over the past 2 decades also did not investigate in-hospital mortality risk differences in the US by race or ethnicity.

Previous cohort and population-based studies of SLE mortality have consistently shown increased damage accrual and mortality among Black patients compared to White patients (1,7,10,19–21). Although studies of Hispanic mortality rates have sometimes had conflicting results (1,21–23), most studies suggest that Hispanic ethnicity is associated with more active SLE and higher mortality. Population-based studies conducted in the US in the 1970s and 1980s described mortality rates 3–5 times greater among Asian/Pacific Islander patients compared to White patients (24,25). A study in the early 1990s evaluating 10,000 hospitalizations similarly showed a higher odds of mortality among hospitalized Asian patients compared to White patients (odds ratio 1.77 [95% confidence interval (95% CI) 1.26–2.47]) (26). Other studies of damage accrual and all-cause mortality among Asian/Pacific Islander patients in the US and internationally that did not focus specifically on in-hospital mortality have had inconsistent results (1,7,23–28).

In this study, we present results from the largest investigation of SLE hospital mortality to date. We evaluated mortality trends over the time span 2006–2016 for admissions with and without a diagnosis of SLE and assessed risk factors we hypothesized would be associated with death. We also analyzed discrepancies in mortality risk by race/ethnicity, performing the first large-scale study in the past 20 years to evaluate in-hospital mortality for Asian/Pacific Islander patients with SLE in the US.

PATIENTS AND METHODS

Data source and population. We analyzed data from 2006 to 2016 from the National Inpatient Sample and the Nationwide Inpatient Sample (NIS) of the Healthcare Cost and Utilization Project (HCUP), Agency for Healthcare Research and Quality (29). (The Nationwide Inpatient Sample was renamed the National Inpatient Sample in 2012 at the time of the sampling redesign.) In 2016, the NIS yielded annual national estimates for >35 million hospitalizations after weighting, representing over 97% of US community hospital discharges (non-federal, general, and specialty hospitals, including public hospitals and academic medical centers). Details regarding the NIS sampling design and available data elements have been published extensively elsewhere (29–31). In 2012, the NIS was redesigned from a sample of hospitals from which all discharges were retained to a sample of discharges from all hospitals participating in HCUP (32). The NIS is ideal for longitudinal analyses, and estimates over study years can be reliably calculated for the entire time of interest using discharge trend weights provided by HCUP for data prior to 2012 (32,33). The primary unit of analysis in the NIS is discharge record, and individual patients are not identifiable. Discharge diagnoses used included 1 primary and up to 30 secondary diagnosis codes as well as up to 15 procedure codes. International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes were used prior to October 2015, and ICD-10-CM codes were used thereafter. Given that this analysis uses publicly available data and contains no individual patient identifiers, it was exempt from approval by the University of California, San Francisco Institutional Review Board.

Measures. The primary outcome in this study was all-cause mortality during hospitalization. SLE was captured in ICD-9-CM code 710.0 and ICD-10-CM codes M32.1x, M32.8, and M32.9. Demographic characteristics for hospitalized individuals included age, sex, and race/ethnicity (White, Black, Asian/Pacific Islander, Hispanic, Native American, or other). Socioeconomic status was categorized based on quartiles of median household income for the patient's ZIP code. Primary expected payer was categorized as Medicare, Medicaid, private insurance, self-pay, or other (no charge, workers compensation, Title 5 and other government programs). Analyses were performed using the primary payer for each admission. Hospital census region was as defined by the US Census Bureau, localizing hospitals to the Northeast, Midwest, South, or West. A Charlson comorbidity index score (34) was calculated with discharge diagnoses for each encounter using coding algorithms previously published (35),

Table 1. Demographic and clinical characteristics of adults with and without systemic lupus erythematosus (SLE) discharged from community hospitals in the US between 2006 and 2016*

Characteristic	SLE (n = 1,903,279)†	No SLE (n = 338,563,769)†	*P*‡
Age, mean ± SE years	51.4 ± 0.07	57.2 ± 0.08	<0.001
Female	89.0	59.4	<0.001
Race			<0.001
White	53.0	68.9	–
Black	29.7	14.4	–
Hispanic	12.0	10.7	–
Asian/Pacific Islander	2.1	2.4	–
Native American	0.7	0.7	–
Other	2.6	3.0	–
Primary payer			<0.001
Medicare	47.0	45.5	–
Private	28.2	29.7	–
Medicaid	18.1	15.9	–
Self-pay	3.7	5.0	–
Other	3.0	3.8	–
Income quartile§			<0.001
Quartile 1 (low income)	33.7	29.6	–
Quartile 2	24.8	26.0	–
Quartile 3	22.6	23.7	–
Quartile 4	18.9	20.7	–
Hospital region¶			<0.001
Northeast	17.7	19.5	–
Midwest	20.2	22.9	–
South	43.7	38.6	–
West	18.4	18.9	–
Comorbidity index, mean ± SE	2.61 ± 0.005	1.54 ± 0.005	<0.001

* Values are the % unless indicated otherwise.
† National estimates were generated from raw counts using discharge and trend weights provided by the Healthcare Cost and Utilization Project. There were 69,526,462 observations for adults without SLE, and 390,360 observations for adults with SLE.
‡ *P* value by *t*-test for continuous variables and Pearson's chi-square test for categorical measures.
§ Household median income quartile in the ZIP code of residence.
¶ Hospital location by census region defined by US Census Bureau.

with minor modifications to include a small number of additional billing codes to ensure accurate index score calculations for discharges from 2006 to 2016 in the US (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24356/abstract).

Statistical analysis. All analyses were designed and performed in adherence with the guidelines and recommendations outlined by HCUP for appropriate use and interpretation of NIS data (29). We applied discharge weights and trend weights provided by HCUP to account for the stratified sampling design of the NIS and to generate nationally representative estimates over multiple years (33). The proportion of hospitalizations with SLE was compared between years using chi-square tests. We compared the characteristics of hospitalized individuals with and without SLE using *t*-tests for continuous measures and chi-square tests for categorical measures. We used Poisson regression with a log link and robust variance estimate to calculate the relative risk (RR) of in-hospital death associated with the following variables: demographic characteristics (age, sex, race/ethnicity), hospital region, comorbidity index score, socioeconomic factors (health insurance and income quartile for ZIP code), and diagnosis of SLE. We fit a model to investigate possible interaction between SLE diagnosis (yes/no) and year of admission to evaluate trends in these hospitalizations over time. We also fit an expanded model to investigate possible interaction between SLE diagnosis (yes/no) and race/ethnicity. Finally, we restricted the original model to include only hospitalized SLE patients. Marginal predictions with 95% CIs were calculated. The race/ethnicity variable had 13% missing values, while vital status, age, sex, primary expected payer, and income quartile by ZIP code had between 0.06% and 2.3% missing values. In order to assess potential bias due to missing observations for these variables, we estimated regression results with and without imputation. Missing values were imputed using multiple imputation by chained equations (MICE) (36) using NIS hospital number, hospital stratum, discharge/trend weights, hospital region, SLE diagnosis, comorbidity index score, and year as predictors. The MICE method adheres to the recommendations outlined by HCUP for handling missing values (37). We present results using imputation for missing observations. Restricting the analysis to complete case data resulted in no substantial change in study conclusions (results not shown). Data analysis was performed using Stata/MP,

version 16.0 (38), and survey data analysis commands were used to account for the complex sampling design of the NIS. Multiple imputation was performed using the Stata mi package (39), and marginal estimates were obtained using the mimargins program (40).

RESULTS

After weighting, there were 1,903,279 adult hospitalizations with a discharge diagnosis of SLE during the studied time frame. The percentage of hospitalizations containing a diagnosis code for SLE increased slightly between 2006 and 2016 from 0.5% (n = 153,645) to 0.6% (n = 173,749) (P <0.001). A description of the demographic and clinical characteristics for all hospitalized adults stratified by SLE diagnosis is provided in Table 1. Compared with the general hospitalized population, individuals with SLE were more likely to be young, female, Black or Hispanic race/ethnicity, reside in a lower income ZIP code quartile, live in the South, and have a higher mean comorbidity index score.

Next, we performed an analysis designed to assess time trends in hospital mortality. In unadjusted analysis, the proportion of hospitalizations for patients without SLE that ended in death slightly decreased from 2.35% (95% CI 2.29–2.42%) in 2006 to 2.19% (95% CI 2.16–2.23%) in 2016 (Figure 1). Among hospitalizations with a diagnosis code of SLE, in-hospital mortality decreased without statistical significance from 2.23% (95% CI 2.05–2.40%) in 2006 to 1.95% (95% CI 1.80–2.10%) in 2016. After including age, sex, race/ethnicity, income quartile for ZIP code, and comorbidity index score as covariates in the regression, all-cause yearly inpatient mortality decreased among admissions for individuals without SLE from 1.92% (95% CI 1.87–1.98%) to 1.43% (95% CI 1.40–1.46%) between 2006 and 2016. During the same time period, the risk of inpatient death decreased by 30% among hospitalizations for individuals with SLE, from 2.20% (95% CI 2.03–2.37%) to 1.54% (95% CI 1.42–1.66%). Nationally, there were an estimated 2,672 inpatient SLE deaths in community hospitals in the US in 2016. This amounts to an estimated 1,147 fewer deaths in 2016 compared to what would have been expected under 2006 rates. All of the decrease in in-hospital mortality for SLE occurred between 2006 and 2008, after which the mortality rate was statistically similar to that of the general hospitalized population and remained stable throughout the rest of the years analyzed.

In an analysis designed to assess potential interactions between SLE diagnosis and race/ethnicity (Figure 2), we found that hospitalized Asian patients with SLE had a 43% higher risk of death compared to those without SLE (P < 0.001) and a 79% higher risk of death compared to hospitalizations for White patients with SLE (P < 0.001). Hospitalizations for Black and Hispanic individuals with SLE were more likely to end in death compared to hospitalizations for individuals of the same racial/ethnic group without SLE (P < 0.001) and compared to hospitalizations for White patients with SLE (P < 0.001). Conversely, hospitalizations among White patients with SLE were less likely to end in death compared to those without SLE (P < 0.001).

Finally, we performed an analysis designed to assess factors associated with death among hospitalizations with an SLE diagnosis. Asian/Pacific Islander race/ethnicity was associated with a 1.65 (95% CI 1.42–1.92) higher risk of death compared to White patients (Table 2; all results from Poisson regression including covariates described above). The RR of in-hospital mortality was lower in the Midwest (0.84 [95% CI 0.77–0.92]) compared to the Northeast. In comparison to SLE hospitalizations covered by Medicare, those not covered by insurance (self-pay) were associated with a higher risk of death (RR 1.36 [95% CI 1.18–1.57]). Older patients, men, and those with a higher comorbidity index score also had a higher risk of death during hospitalization.

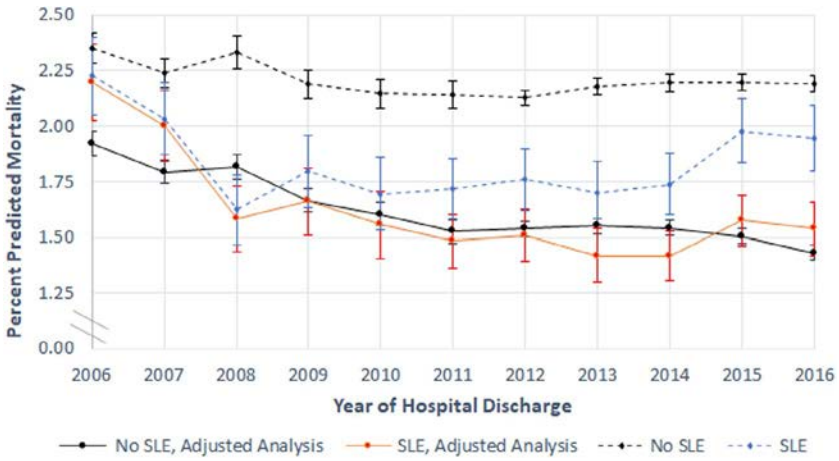


Figure 1. Predicted mortality during hospitalization based on diagnosis of systemic lupus erythematosus (SLE) by year of hospitalization. Adjusted analysis results show marginal predictions from a Poisson regression model including age, sex, race/ethnicity, residence in a low income ZIP code, comorbidity index score, primary payer, and an interaction term for SLE and time. Vertical error bars represent 95% confidence intervals.

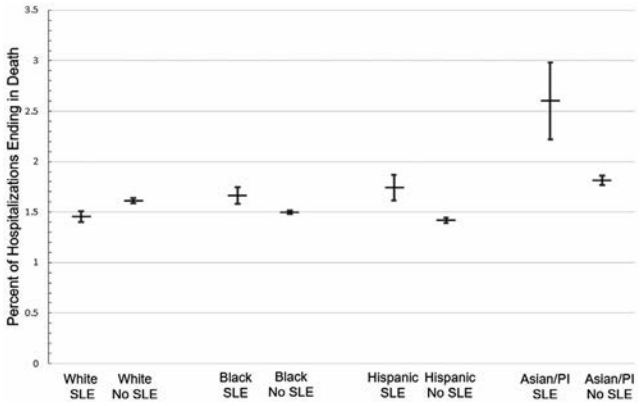


Figure 2. Predicted in-hospital mortality (%) by race and diagnosis of systemic lupus erythematosus (SLE), 2006–2016, with marginal predictions and 95% confidence intervals from a Poisson regression model including age, sex, race, income quartile by patient ZIP code, comorbidity index score, primary payer, and an interaction term for race and SLE. Native American and other race/ethnicity are not included in the figure. Asian/PI = Asian/Pacific Islander.

DISCUSSION

We conducted the largest primary data analysis of SLE hospitalizations in the US to date. Our results showed that in adjusted analysis, SLE hospitalizations were more likely than non-SLE hospitalizations to end in death in 2006 and 2007; the mortality rate for SLE hospitalizations decreased through 2008, after which mortality was statistically similar between the 2 groups. Our findings also demonstrate the continued high mortality burden of SLE among admitted Black patients, Hispanic patients, and men and contribute data highlighting high mortality among Asian/Pacific Islander patients.

Overall, our study and others conducted in prior years suggest improved in-hospital survival. An earlier publication describing a Californian cohort from 1991 to 1994 estimated that 5.1% of patients died during their admission (26). In-hospital mortality was estimated at 3.1% using the NIS from 1998 to 2002 (17). This decreasing mortality rate parallels that found in Washington state, where in-hospital mortality was estimated to decrease from 3.1% in 2003 to 1.3% by 2011 (18). Our results suggest that interventions to reduce all-cause in-hospital mortality, such as programs to lessen hospital-acquired infections (41,42), and interventions primarily targeting patients with SLE, such as a broader array of immunosuppressive drugs, may have successfully improved in-hospital survival over 1 decade. The percentage of SLE admissions only marginally increased in our study, suggesting that the findings cannot be explained by a change in admission thresholds. Furthermore, differences in coding practices over time would likely affect all hospitalized patients, highlighting once again the important strength of our study comparing SLE mortality to that of the general hospitalized population.

Despite significant advances in survival among individuals hospitalized with SLE, some racial/ethnic groups continue to have a disproportionately high risk of death. Our results demonstrate that Asian, Hispanic, and Black individuals with SLE all have a higher adjusted risk of in-hospital death compared to hospitalized individuals of the same racial/ethnic group without SLE. Each of these racial/ethnic groups with SLE also had a higher risk of in-hospital death compared to White patients with SLE. We note the strikingly high mortality rate for Asian/Pacific Islander patients discharged with SLE from community hospitals in the US; these hospitalizations were associated with a 79% higher risk of death compared to hospitalizations for White patients and a 43% higher risk of death compared to hospitalizations for Asians/Pacific Islander patients without SLE. Since our unit of analysis is hospital discharges, it could be possible that Asian/Pacific Islander patients have unmeasured factors that make them less likely than other racial/ethnic groups to present to the hospital except when severely ill. This explanation however would not account for the clinically important difference in mortality that we report between hospitalized SLE and non-SLE patients of Asian/Pacific Islander race/ethnicity, nor the results of smaller prior studies that used individual patients as the unit of analysis and similarly reported high Asian mortality rates compared to White patients (24–26). Future studies should further investigate potential reasons for persistent inequalities in outcomes among these racial and ethnic groups.

Table 2. Risk of in-hospital death for adults with systemic lupus erythematosus (SLE) discharged from community hospitals in the US between 2006 and 2016*

Characteristic	RR for death (95% CI)	*P*
Age, per decade	1.31 (1.28–1.33)	<0.001
Female	0.81 (0.76–0.86)	<0.001
Race		<0.001
White	Ref.	–
Black	1.03 (0.97–1.09)	–
Hispanic	1.10 (1.02–1.19)	–
Asian/Pacific Islander	1.65 (1.42–1.92)	–
Native American	0.97 (0.71–1.33)	–
Other	1.36 (1.16–1.60)	–
Primary payer		<0.001
Medicare	Ref.	–
Private	0.99 (0.93–1.06)	–
Medicaid	1.06 (0.98–1.15)	–
Self-pay	1.36 (1.18–1.57)	–
Other	1.26 (1.08–1.46)	–
Low income†	1.02 (0.97–1.08)	0.42
Hospital region‡		<0.001
Northeast	Ref.	–
Midwest	0.84 (0.77–0.92)	–
South	1.04 (0.96–1.13)	–
West	1.05 (0.96–1.15)	–
Comorbidity index	1.32 (1.30–1.33)	<0.001

* Results from a Poisson regression model controlling for survey year and including all covariates shown. Analyses were weighted to reflect the sampling design. 95% CI = 95% confidence interval; Ref. = reference; RR = relative risk.
† Based on residence in a ZIP code in the lowest quartile of median household income.
‡ Hospital located in census region defined by US Census Bureau.

Male sex was also associated in our study with a higher risk of in-hospital mortality for SLE. These findings are similar to prior analyses in hospitalized patients showing worse outcomes in men (17–18,26). Although SLE is more common in female patients, male patients may have faster disease progression and damage accrual (1,43,44). As expected, increasing age and higher comorbidity index score were associated with higher mortality among patients hospitalized with SLE. Self-pay status, which reflects a lack of insurance, was also associated with higher mortality. Future work could investigate whether self-pay SLE patients receive a different quality of inpatient care as a result of monetary constraints, whether likely limited primary preventive care may be driving differences in outcomes, or whether higher inpatient mortality per admission is a reflection of a higher threshold of illness severity to present for medical evaluation compared to individuals with insurance.

One potential limitation of this study is that we rely on ICD-9-CM and ICD-10-CM codes for SLE to be included in the discharge summary from the treating physician. Prior studies evaluating the accuracy of administrative data diagnoses using ICD-9-CM coding for SLE have suggested sensitivities of 67–98% and specificities of 72–90% (45,46). The Charlson comorbidity index using the Quan coding algorithms has been validated as an acceptable predictor of in-hospital mortality (area under the curve 0.76–0.87) (35,47). Another potential limitation of the study is the 13% missing race values in the NIS, which are known not to be missing completely at random. In order to decrease the risk for bias, we performed multiple imputation by chained equations according to the recommendations from HCUP (37). Additionally, the NIS sampling scheme was redesigned in 2012. We have fully addressed this change in design using trend weights specifically developed by the NIS to allow for analysis over time (32,33). There was no significant change in mortality among either SLE patients or the general adult population in our study between 2011 and 2012 when a change in mortality could potentially be expected to have been due to changes in survey design.

In conclusion, our results demonstrate a decrease in adjusted all-cause mortality among SLE hospitalizations in the US to levels similar to that of the general hospitalized population. We show that mortality was relatively stable for both SLE and non-SLE admissions from 2008 to 2016, reflecting a shift from previous decades when inpatient mortality levels were decreasing over time. Despite advances in overall all-cause mortality among all patients with SLE, hospitalized Asian/Pacific Islander, Hispanic, and Black patients have a higher risk of inpatient death compared to both patients without SLE and to White patients. Our study results demonstrate the importance of considering Asian/Pacific Islander patients with SLE in the US as at potentially heightened risk for poor outcomes and therefore warranting special clinical attention and inclusion in future research studies. Comprehensive efforts addressing differences in disease severity, access to health care, and social determinants of health are likely necessary to narrow disparities in hospital mortality among men and various racial/ethnic groups with 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. Anastasiou 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. Anastasiou, Shiboski, Yazdany.

Acquisition of data. Anastasiou, Yazdany.

Analysis and interpretation of data. Anastasiou, Trupin, Glidden, Li, Gianfrancesco, Shiboski, Schmajuk, Yazdany.

REFERENCES

1. Bruce IN, O'Keeffe AG, Farewell V, Hanly JG, Manzi S, Su L, et al. Factors associated with damage accrual in patients with systemic lupus erythematosus: results from the Systemic Lupus International Collaborating Clinics (SLICC) Inception Cohort. Ann Rheum Dis 2015;74:1706–13.
2. Alarcón GS, McGwin G Jr, Bastian HM, Roseman J, Lisse J, Fessler BJ, et al. Systemic lupus erythematosus in three ethnic groups. VIII. Predictors of early mortality in the LUMINA cohort. Arthritis Rheum 2001;45:191–202.
3. Yazdany J, Marafino BJ, Dean ML, Bardach NS, Duseja R, Ward MM, et al. Thirty-day hospital readmissions in systemic lupus erythematosus: predictors and hospital- and state-level variation. Arthritis Rheumatol 2014;66:2828–36.
4. Elixhauser A, Steiner C, for the Agency for Healthcare Policy and Research. Readmissions to U.S. hospitals by diagnosis, 2010. Healthcare Cost and Utilization Project (HCUP) statistical brief 153. URL: https://www.hcup-us.ahrq.gov/reports/statbriefs/sb153.pdf.
5. Merrell M, Shulman LE. Determination of prognosis in chronic disease, illustrated by systemic lupus erythematosus. J Chron Dis 1955;1:12–32.
6. Cervera R, Khamashta MA, Font J, Sebastiani GD, Gil A, Lavilla P, et al. Morbidity and mortality in systemic lupus erythematosus during a 10-year period: a comparison of early and late manifestations in a cohort of 1,000 patients. Medicine (Baltimore) 2003;82:299–308.
7. Yen EY, Shaheen M, Woo JM, Mercer N, Li N, McCurdy DK, et al. 46-year trends in systemic lupus erythematosus mortality in the United States, 1968 to 2013: a nationwide population-based study. Ann Intern Med 2017;167:777–85.
8. Uramoto KM, Michet CJ Jr, Thumboo J, Sunku J, O'Fallon WM, Gabriel SE. Trends in the incidence and mortality of systemic lupus erythematosus, 1950–1992. Arthritis Rheum 1999;42:46–50.
9. Urowitz MB, Gladman DD, Tom BD, Ibañez D, Farewell VT. Changing patterns in mortality and disease outcomes for patients with systemic lupus erythematosus. J Rheumatol 2008;35:2152–8.
10. Bernatsky S, Boivin JF, Joseph L, Manzi S, Ginzler E, Gladman DD, et al. Mortality in systemic lupus erythematosus. Arthritis Rheum 2006;54:2550–7.
11. Lee J, Dhillon N, Pope J. All-cause hospitalizations in systemic lupus erythematosus from a large Canadian referral centre. Rheumatology (Oxford) 2013;52:905–9.
12. Anastasiou C, Dulai O, Baskaran A, Proudfoot J, Verhaegen S, Kalunian K. Immunosuppressant use and hospitalisations in adult patients with systemic lupus erythematosus admitted to a tertiary academic medical centre. Lupus Sci Med 2018;5:e000249.
13. Chan K, Dekis A, Clarke AE, Pineau CA, Vinet E, Nashi E, et al. Hospitalizations in patients with systemic lupus erythematosus:

updated analyses from 2006 to 2011. Arthritis Res Ther 2012;14 Suppl 3:A59.

14. Rosa GP, Ortega MF, Teixeira A, Espinosa G, Cervera R. Causes and factors related to hospitalizations in patients with systemic lupus erythematosus: analysis of a 20-year period (1995–2015) from a single referral centre in Catalonia. Lupus 2019;28:1158–66.
15. Gu K, Gladman DD, Su J, Urowitz MB. Hospitalizations in patients with systemic lupus erythematosus in an academic health science center. J Rheumatol 2017;44:1173–8.
16. Dartmouth Atlas Project. End of life care. URL: https://www.dartmouthatlas.org/interactive-apps/end-of-life-care/.
17. Krishnan E. Hospitalization and mortality of patients with systemic lupus erythematosus. J Rheumatol 2006;33:1770–4.
18. Goss LB, Ortiz JR, Okamura DM, Hayward K, Goss CH. Significant reductions in mortality in hospitalized patients with systemic lupus erythematosus in Washington state from 2003 to 2011. PLoS One 2015;10:e0128920.
19. Kasitanon N, Magder LS, Petri M. Predictors of survival in systemic lupus erythematosus. Medicine (Baltimore) 2006;85:147–56.
20. Lim SS, Helmick CG, Bao G, Hootman J, Bayakly R, Gordon C, et al. Racial disparities in mortality associated with systemic lupus erythematosus: Fulton and DeKalb counties, Georgia, 2002–2016. MMWR Morb Mortal Wkly Rep 2019;68:419–22.
21. Alarcón GS, Friedman AW, Straaton KV, Moulds JM, Lisse J, Bastian HM, et al. Systemic lupus erythematosus in three ethnic groups: III. A comparison of characteristics early in the natural history of the LUMINA cohort. LUpus in MInority populations: NAture vs. Nurture. Lupus 1999;8:197–209.
22. Walsh SJ, DeChello LM. Geographical variation in mortality from systemic lupus erythematosus in the United States. Lupus 2001;10:637–46.
23. Gómez-Puerta JA, Barbhaiya M, Guan H, Feldman CH, Alarcón GS, Costenbader KH. Racial/ethnic variation in all-cause mortality among United States Medicaid recipients with systemic lupus erythematosus: a Hispanic and Asian paradox. Arthritis Rheumatol 2015;67:752–60.
24. Kaslow RA. High rate of death caused by systemic lupus erythematosus among U. S. residents of Asian descent. Arthritis Rheum 1982;25:414–8.
25. Serdula MK, Rhoads GG. Frequency of systemic lupus erythematosus in different ethnic groups in Hawaii. Arthritis Rheum 1979;22:328–33.
26. Ward MM. Hospital experience and mortality in patients with systemic lupus erythematosus. Arthritis Rheum 1999;42:891–8.
27. Jakes RW, Bae SC, Louthrenoo W, Mok CC, Navarra SV, Kwon N. Systematic review of the epidemiology of systemic lupus erythematosus in the Asia-Pacific region: prevalence, incidence, clinical features, and mortality. Arthritis Care Res (Hoboken) 2012;64:159–68.
28. Lee YH, Choi SJ, Ji JD, Song GG. Overall and cause-specific mortality in systemic lupus erythematosus: an updated meta-analysis. Lupus 2016;25:727–34.
29. Agency for Healthcare Research and Quality. Healthcare Cost and Utilization Project (HCUP). HCUP databases. URL: www.hcup-us.ahrq.gov/nisoverview.jsp.
30. Rubin DS, Matsumoto MM, Moss HE, Joslin CE, Tung A, Roth S. Ischemic optic neuropathy in cardiac surgery: incidence and risk factors in the United States from the National Inpatient Sample 1998 to 2013. Anesthesiology 2017;126:810–21.
31. Burton BN, Jafari A, Asmerom B, Swisher MW, Gabriel RA, DeConde A. Inpatient mortality after endoscopic sinus surgery for invasive fungal rhinosinusitis. Ann Otol Rhinol Laryngol 2019;128:300–8.
32. Houchens RL, Ross DN, Elixhauser A, Jiang J, for the Agency for Healthcare Research and Quality. Nationwide Inpatient Sample redesign final report. 2014. HCUP NIS related reports. April 4, 2014. URL: https://www.hcup-us.ahrq.gov/db/nation/nis/reports/NISRedesignFinalReport040914.pdf.
33. Agency for Healthcare Research and Quality. Trend weights for HCUP NIS data. 2015. URL: https://www.hcup-us.ahrq.gov/db/nation/nis/trendwghts.jsp.
34. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40:373–83.
35. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 2005;43:1130–9.
36. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med 2011;30:377–99.
37. Houchens R, for the Agency for Healthcare Research and Quality. Missing data methods for the NIS and the SID. 2015. HCUP methods series report 2015-01. January 22, 2015. URL: https://www.hcup-us.ahrq.gov/reports/methods/2015_01.pdf.
38. Stata statistical software: release 16.0. College Station (TX): StataCorp; 2019.
39. StataCorp. Stata multiple imputation reference manual: release 16. College Station (TX): 2019.
40. UCLA Institute for Digital Research & Education. Statistical Consulting Group. How can I get margins for a multiply imputed survey logit model? Stata FAQ. URL: https://stats.idre.ucla.edu/stata/faq/how-can-i-get-margins-for-a-multiply-imputed-survey-logit-model/.
41. Barsuk JH, Cohen ER, Feinglass J, McGaghie WC, Wayne DB. Use of simulation-based education to reduce catheter-related bloodstream infections. Arch Intern Med 2009;169:1420–3.
42. Weinstein RA, Milstone AM, Passaretti CL, Perl TM. Chlorhexidine: expanding the armamentarium for infection control and prevention. Clin Infect Dis 2008;46:274–81.
43. Andrade RM, Alarcón GS, Fernández M, Apte M, Vilá LM, Reveille JD, et al, for the LUMINA Study Group. Accelerated damage accrual among men with systemic lupus erythematosus: XLIV. Results from a multiethnic US cohort. Arthritis Rheum 2007;56:622–30.
44. Schwartzman-Morris J, Putterman C. Gender differences in the pathogenesis and outcome of lupus and of lupus nephritis. Clin Dev Immunol 2012;2012:604892.
45. Bernatsky S, Linehan T, Hanly JG. The accuracy of administrative data diagnoses of systemic autoimmune rheumatic diseases. J Rheumatol 2011;38:1612–6.
46. Lim SS, Jamal A, Bayakly R, Tong L, Drenkard C. Georgia Lupus Registry: accuracy of hospital discharge data in identifying systemic lupus erythematosus. Arthritis Rheum 2007;54 Suppl:S505.
47. Toson B, Harvey LA, Close JC. The ICD-10 Charlson comorbidity index predicted mortality but not resource utilization following hip fracture. J Clin Epidemiol 2015;68:44–51.

Arthritis Care & Research
Vol. 73, No. 10, October 2021, pp 1451–1455
DOI 10.1002/acr.24367

American College of Rheumatology
Empowering Rheumatology Professionals

BRIEF REPORT

Applying the 2019 European Alliance of Associations for Rheumatology/American College of Rheumatology Lupus Criteria to Patients From the LUMINA Cohort: Results From the Multiethnic, Multicenter US Cohort

Manuel F. Ugarte-Gil,[1] Guillermo J. Pons-Estel,[2] Guillermina B. Harvey,[3] Luis M. Vilá,[4] Russell Griffin,[5] and Graciela S. Alarcón[6]

Objective. To evaluate the performance of the 2019 European Alliance of Associations for Rheumatology (EULAR)/American College of Rheumatology (ACR) criteria for systemic lupus erythematosus (SLE) in terms of earlier SLE classification in comparison to the ACR or the Systemic Lupus International Collaborating Clinics (SLICC) criteria.

Methods. Patients from a multiethnic, multicenter cohort, the Lupus in Minorities: Nature versus Nurture cohort, where SLE was defined using the 1982/1997 ACR criteria were included. Demographic, clinical, and immunologic criteria were compared among the 2019 EULAR/ACR and the 1982/1997 ACR and the 2012 SLICC timing categories.

Results. The 2019 EULAR/ACR criteria allowed an earlier SLE classification in 13.3% of patients (mean 0.66 years) and 15.3% of patients (mean 0.63 years) compared to the 1982/1997 ACR and the 2012 SLICC criteria, respectively. Patients accruing the 2019 EULAR/ACR criteria later than the 1982/1997 ACR criteria had a lower disease activity, were less likely to have positivity to anti–double-stranded DNA and anti-Sm, as well as lupus nephritis classes II or V; they were more likely to have mucocutaneous manifestations, serositis, leukopenia, and antiphospholipid antibodies positivity. These differences were less pronounced when compared to the 2012 SLICC criteria

Conclusion. The 2019 EULAR/ACR criteria classified SLE patients earlier than the 2 other criteria sets in real-life clinical practice scenarios in a relatively small proportion of the patients. However, these criteria could allow earlier classification of a subset of patients with a more severe disease.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a complex disease, and its diagnosis requires clinical expertise. However, for research purposes and clinical trials, criteria are needed to recruit patients with similar/comparable clinical and laboratory characteristics. The 1982 American College of Rheumatology (ACR) criteria (1) modified in 1997 (2), albeit never validated, have been widely used worldwide; however, they have several limitations. The Systemic Lupus International Collaborating Clinics (SLICC) proposed a new set of criteria in 2012, including some of the manifestations and laboratory tests that were not included in the ACR criteria (3).

Recently, the European Alliance of Associations for Rheumatology (EULAR) and the ACR have joined efforts and proposed a new set of criteria. Of importance, these new criteria have an entry criterion, that the patient has to be antinuclear antibody (ANA) positive, and all the criteria have different weights for a patient to be classified as having SLE (4).

The objective of the current study was to examine whether patients from a multiethnic, multicenter US cohort would be classified earlier using the EULAR/ACR criteria than the older ACR or

[1]Manuel F. Ugarte-Gil, MD, MSc: Hospital General Guillermo Almenara Irigoyen, EsSalud, and Universidad Científica del Sur, Lima, Perú; [2]Guillermo J. Pons-Estel, MD, PhD: Centro Regional de Enfermedades Autoinmunes y Reumáticas and Hospital Provincial de Rosario, Rosario, Argentina; [3]Guillermina B. Harvey, PhD: Universidad Nacional de Rosario, Rosario, Argentina; [4]Luis M. Vilá, MD: University of Puerto Rico Medical Sciences Campus, San Juan, Puerto Rico; [5]Russell Griffin, PhD: University of Alabama at Birmingham; [6]Graciela S. Alarcón, MD, MPH: University of Alabama at Birmingham and Universidad Peruana Cayetano Heredia, Lima, Perú.

Drs. Ugarte-Gil and Pons-Estel contributed equally to this work.

Dr. Ugarte-Gil has received research grants from Pfizer and Janssen. Dr. Pons-Estel has received consulting and speaking fees from Pfizer, GlaxoSmithKline, Janssen, and Sanofi (less than $10,000 each). No other disclosures relevant to this article were reported.

Address correspondence to Manuel F. Ugarte-Gil, MD, MSc, Universidad Científica del Sur, Avenida Panamericana Sur Km 19, Villa El Salvador, Peru 15067. Email: mugarte@cientifica.edu.pe.

Submitted for publication December 30, 2019; accepted in revised form June 23, 2020.

SIGNIFICANCE & INNOVATIONS

- In this multiethnic multicenter cohort, the 2019 European Alliance of Associations for Rheumatology/American College of Rheumatology criteria allowed for classifying patients earlier than with using the previous criteria set in a relatively small proportion of patients.
- These criteria could allow earlier classification of a subset of patients with more severe disease.

SLICC criteria, in order to validate this new set of criteria. In addition, characteristics of the patients would be compared based on the timing of the EULAR/ACR versus the ACR and SLICC criteria.

PATIENTS AND METHODS

The Lupus in Minorities: Nature versus Nurture (LUMINA) cohort has been amply described in the literature (5,6). LUMINA patients were recruited using the 1982/1997 ACR classification criteria, and they could have had up to 5 years of disease duration (from the date of criteria diagnosis). We must point out that some of the clinical and laboratory manifestations included on the new 2019 EULAR/ACR criteria have not been recorded in this cohort's database; therefore, they could not be included in these analyses (arthralgia, fever, alopecia, delirium, acute pericarditis, and complement levels). The clinical and laboratory variables were measured in all patients at the time of entry into the cohort and every 6 months for

Table 1. Characteristics at the time of EULAR/ACR–based classification in patients from the LUMINA cohort classified at the same time, earlier, or later than classifications based on the 1982/1997 ACR criteria*

EULAR/ACR criteria and items	EULAR/ACR at the same time (n = 368, 61.9%)	EULAR/ACR earlier (n = 79, 13.3%)	EULAR/ACR later (n = 147, 24.7%)	*P*
Demographic				
Race or ethnicity				<0.0001
Hispanic (Texas)	78 (21.2)	19 (24.1)	17 (11.6)	–
White	86 (23.4)	18 (22.8)	58 (39.5)	–
African American	158 (42.9)	31 (39.2)	38 (25.9)	–
Hispanic (Puerto Rico)	46 (12.5)	11 (13.9)	34 (23.1)	–
Others	NA	NA	NA	–
Sex				0.0762
Female	329 (89.4)	66 (83.5)	137 (93.2)	–
Male	39 (10.6)	13 (16.5)	10 (6.8)	–
Age at enrollment, mean ± SD years	35.4 ± 12.2	36.4 ± 13.2	37.8 ± 13.0	0.1310
Clinical				
SLAM score at enrollment, mean ± SD	9.7 ± 6.1	8.3 ± 5.3	8.1 ± 5.0	0.0048
ACR/EULAR clinical domains				–
Fever	NA	NA	NA	NA
Acute cutaneous or malar rash	148 (40.2)	9 (11.4)	99 (67.3)	<0.0001
Subacute cutaneous lupus or discoid rash	35 (9.5)	1 (1.3)	26 (17.7)	0.0004
Oral ulcers	97 (26.4)	0 (0.0)	76 (51.7)	<0.0001
Nonscarring alopecia	NA	NA	NA	NA
Synovitis	260 (70.7)	47 (59.5)	102 (69.4)	0.1494
Seizures	20 (5.4)	4 (5.1)	4 (2.7)	0.4175
Psychosis	10 (2.7)	0 (0.0)	4 (2.7)	0.3330
Delirium	NA	NA	NA	NA
Acute pericarditis	NA	NA	NA	NA
Pleural or pericardial effusion	122 (33.2)	12 (15.2)	49 (33.3)	0.0054
Thrombocytopenia	47 (12.8)	6 (7.6)	17 (11.6)	0.4305
Autoimmune hemolysis	20 (5.4)	8 (10.1)	12 (8.2)	0.2328
Leukopenia	142 (38.6)	8 (10.1)	61 (41.5)	<0.0001
Proteinuria	85 (23.1)	14 (17.7)	24 (16.3)	0.1802
Renal biopsy class II or V	27 (7.3)	10 (12.7)	5 (3.4)	0.0333
Renal biopsy class III or IV	14 (3.8)	3 (3.8)	8 (5.4)	0.6917
Immunologic				
aCL >40 or LAC positive	54 (14.7)	3 (3.8)	32 (21.8)	0.0014
Low C3 or C4	NA	NA	NA	NA
Low C3 and C4	NA	NA	NA	NA
Anti-Sm	120 (32.6)	19 (24.1)	22 (15.0)	0.0001
Anti–double-stranded DNA	183 (49.7)	43 (54.4)	46 (31.3)	0.0003

* Values are the number (%) unless indicated otherwise. aCL = anticardiolipin; EULAR/ACR = European Alliance of Associations for Rheumatology/American College of Rheumatology; LAC = lupus anticoagulant; LUMINA = Lupus in Minorities: Nature versus Nurture; NA = not available; SLAM = Systemic Lupus Activity Measure.

the first year and yearly thereafter. Investigators were asked to evaluate each manifestation at the entry visit and at every visit until the criteria were present and to establish precisely the dates of disease onset, diagnosis, fulfillment of ACR SLE criteria, and the first appearance of each clinical manifestation.

Demographic, clinical, and immunologic criteria were compared among the 2019 EULAR/ACR and the 1982/1997 ACR or the 2012 SLICC timing categories. Categorical variables were compared using chi-square and continuous variables with analysis of variance. A *P* value less than 0.05 was set as the level of statistical significance. The statistical analyses were performed using SAS software, version 9.4.

RESULTS

For these analyses of 640 LUMINA patients, 594 (ACR 1982/1997 criteria) (Table 1) and 595 (SLICC criteria) (Table 2) were included. In all, 46 patients (7.2%) were excluded because they did not meet the 2019 EULAR/ACR criteria, even though an average of 16 months had elapsed between the time patients met the 1982/1997 ACR and 2012 SLICC criteria and the time they entered the LUMINA cohort. There were no differences in terms of sex between the included and excluded patients; however, among the excluded patients there was a higher proportion of White patients (40.0% versus 27.3%) and a lower proportion of Hispanic patients from Texas (8.9% versus 19.4%) than among the patients included. The excluded patients were also older (mean ± SD age 40.7 ± 12.8 years versus 36.1 ± 12.5 years) but did not differ in disease activity as measured by the Systemic Lupus Activity Measure (SLAM) score (mean ± SD 9.7 ± 6.8 versus 9.4 ± 5.8).

Comparison between the 2019 EULAR/ACR and the 1982/1997 ACR criteria. Comparing the 2019 EULAR/ACR to 1982/1997 ACR criteria, a majority of patients (n = 368, 61.9%) met both criteria at the same time, whereas 79 (13.3%) met the 2019 EULAR/ACR criteria earlier (mean 0.66 years), and 147 (24.7%) met the 2019 EULAR/ACR criteria later (mean 1.48 years) (Table 1). Those who met the EULAR/ACR criteria later were more likely to be White ($P < 0.0001$). In addition, these patients had a lower mean SLAM score at enrollment ($P = 0.0048$) and were less likely to be positive for the anti-Sm ($P = 0.0001$) and anti–double-stranded DNA (anti-dsDNA; $P = 0.0003$) antibodies but more likely to be antiphospholipid antibody positive ($P = 0.0014$) (immunologic factors). Assessing ACR/EULAR clinical domains, those who met the EULAR/ACR criteria later were more likely to have acute cutaneous or malar rash ($P < 0.0001$), subacute cutaneous lupus ($P = 0.0004$), and oral ulcers ($P < 0.0001$). Those who met the EULAR/ACR criteria earlier were less likely to have leukopenia ($P < 0.0001$) but were more likely to have renal biopsy classes II or V ($P = 0.0333$).

Comparison between the 2019 EULAR/ACR and the 2012 SLICC criteria. Similarly, when the 2019 EULAR/ACR criteria were compared to the 2012 SLICC criteria, the majority of the patients (n = 428, 71.9%) met both sets of criteria at the same time, 91 patients (15.3%) met them earlier (mean 0.63 years), and 76 (12.8%) met them later (mean 1.37 years) (Table 2). Those who met the EULAR/ACR criteria later were less likely to be African American ($P < 0.0001$) and were most likely to have acute cutaneous or malar rash ($P = 0.0213$) and subacute cutaneous lupus ($P = 0.0004$). Those who met the EULAR/ACR criteria earlier had the lowest mean SLAM score at enrollment ($P = 0.0039$) and were least likely to have oral ulcers ($P < 0.0001$) and pleural or pericardial effusion ($P < 0.0001$); in addition, when compared to those who met both criteria at the same time, those meeting EULAR/ACR criteria earlier or later had a lower probability of renal biopsy classes II or V ($P = 0.0018$). Among those patients who did not achieve the EULAR/ACR criteria, 21 of 46 were ANA negative. The criteria attained in these patients are shown in Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24367/abstract.

DISCUSSION

This new set of criteria, developed jointly by EULAR and the ACR, represents a truly major effort in this field; however, we need more information to know how these criteria will perform in real-life scenarios across several ethnic groups and geographical regions. In this study, performed in a multiethnic US cohort, the 2019 EULAR/ACR criteria apparently allowed an earlier classification of SLE only in a relatively small proportion of patients, compared with the 1982/1997 ACR and the 2012 SLICC criteria: only 13.3% and 15.3% were classified earlier using the 2019 EULAR/ACR criteria, respectively. On the other hand, 7.2% of patients did not achieve these new criteria. Nevertheless, we need to take into consideration that the database from LUMINA did not include all the items noted in the new criteria, which may have acted against their performance. These criteria need to be evaluated longitudinally by different groups around the world and not only from existing databases.

Among the demographic variables, African American patients from LUMINA tended to be classified earlier and White patients later. Regarding sex, a clear pattern did not emerge in terms of an earlier classification, except that LUMINA male patients tended to be classified earlier compared to the 1982/1997 ACR criteria but not compared to the 2012 SLICC criteria.

When the clinical manifestations present in patients were taken into account, patients who achieved the 2019 EULAR/ACR criteria earlier than the 1982/1997 ACR criteria had a lower frequency of milder disease manifestations (such as mucocutaneous, articular, or leukopenia) and a higher frequency of lupus nephritis classes II or V, anti-Sm and anti-dsDNA antibodies and a

Table 2. Characteristics at the time of EULAR/ACR–based classification in patients from the LUMINA cohort classified at the same time, earlier, or later than classifications based on the SLICC criteria*

EULAR/ACR criteria and items	EULAR/ACR at the same time (n = 428, 71.9%)	EULAR/ACR earlier (n = 91, 15.3%)	EULAR/ACR later (n = 76, 12.8%)	*P*
Demographic				
Race or ethnicity				<0.0001
Hispanic (Texas)	91 (21.3)	13 (14.3)	10 (13.2)	–
White	115 (26.9)	25 (27.5)	23 (30.3)	–
African American	177 (41.4)	32 (35.2)	18 (23.7)	–
Hispanic (Puerto Rico)	45 (10.5)	21 (23.1)	25 (32.9)	–
Others	NA	NA	NA	–
Sex				0.2066
Female	378 (88.3)	83 (91.2)	72 (94.7)	–
Male	37 (10.7)	8 (8.8)	13 (7.5)	–
Age at enrollment, mean ± SD years	35.6 ± 12.5	36.7 ± 11.8	38.1 ± 13.6	0.2661
Clinical				
SLAM score at enrollment, mean ± SD	9.6 ± 6.1	7.6 ± 3.9	8.2 ± 5.2	0.0039
ACR/EULAR clinical domains				–
Fever	NA	NA	NA	NA
Acute cutaneous or malar rash	175 (40.9)	38 (41.8)	44 (57.9)	0.0213
Subacute cutaneous lupus or discoid rash	47 (11.0)	1 (1.1)	15 (19.7)	0.0004
Oral ulcers	134 (31.3)	2 (2.2)	37 (48.7)	<0.0001
Nonscarring alopecia	NA	NA	NA	NA
Synovitis	298 (69.6)	65 (71.4)	47 (61.8)	0.3423
Seizures	25 (5.8)	2 (2.2)	1 (1.3)	0.1078
Psychosis	13 (3.0)	0 (0.0)	1 (1.3)	0.1808
Delirium	NA	NA	NA	NA
Acute pericarditis	NA	NA	NA	NA
Pleural or pericardial effusion	156 (36.4)	9 (9.9)	18 (23.7)	<0.0001
Thrombocytopenia	57 (13.3)	5 (5.5)	8 (10.5)	0.1026
Autoimmune hemolysis	32 (7.5)	3 (3.3)	5 (6.6)	0.3510
Leukopenia	156 (36.4)	22 (24.2)	33 (43.4)	0.0253
Proteinuria	101 (23.6)	7 (7.7)	15 (19.7)	0.0030
Renal biopsy class II or V	40 (9.3)	0 (0.0)	2 (2.6)	0.0018
Renal biopsy class III or IV	22 (5.1)	0 (0.0)	3 (3.9)	0.0846
Immunologic				
aCL >40 or LAC positive	66 (15.4)	3 (3.3)	20 (26.3)	0.0002
Low C3 or C4	NA	NA	NA	NA
Low C3 and C4	NA	NA	NA	NA
Anti-Sm	136 (31.8)	13 (14.3)	12 (15.8)	0.0002
Anti–double-stranded DNA	213 (49.8)	34 (37.4)	25 (32.9)	0.0055

* Values are the number (%) unless indicated otherwise. aCL = anticardiolipin; EULAR/ACR = European Alliance of Associations for Rheumatology/American College of Rheumatology; LAC = lupus anticoagulant; LUMINA = Lupus in Minorities: Nature versus Nurture; NA = not available; SLAM = Systemic Lupus Activity Measure; SLICC = Systemic Lupus International Collaborating Clinics.

higher SLAM score, suggesting that these criteria could be quite useful in subsets of patients with more severe disease. Similarly, those who achieved the 2019 EULAR/ACR criteria earlier than the SLICC criteria had a lower frequency of mucocutaneous involvement and serositis but not a clear pattern in which severe manifestations were seen, probably because the renal biopsy has a special weight in the SLICC criteria similar to the EULAR/ACR criteria.

Consistent with our results, other groups have found an association between these criteria and the severity of the disease. The Grupo Latino Americano de Estudio del Lupus cohort has reported that patients classified earlier with these new criteria have a more severe disease (7). And in a London cohort, a higher score in the EULAR/ACR criteria predicted a higher damage accrual (8). However, the accuracy of these new criteria does not seem to be better than the previous ones, in particular compared with the SLICC criteria (9,10).

Our study has some limitations. First, the patients studied were adults, so whether these new criteria perform differently in pediatric patients cannot be stated. Second, the possibility that not all the manifestations listed in the 2019 EULAR/ACR criteria were recorded prevented us from reaching more firm conclusions. Third, and as already noted, all patients had satisfied the 1982/1997 ACR criteria to enter the cohort; therefore, patients with fewer than 4 criteria were not eligible for LUMINA and thus we do not know whether they could have been classified as

having SLE with the new criteria. Finally, the date at which each criteria manifestation had occurred was based on the information available; possibly these dates do not have the precision that could have been derived from obtaining these data with the specific purpose of assessing all 3 sets of criteria.

In summary, we found that the EULAR/ACR criteria achieved the goal of classifying LUMINA patients earlier only in a small proportion of the patients; however, the criteria seemed to classify earlier a more severe subset of patients. These new criteria need to be evaluated longitudinally in several populations across the world, especially in patients with early disease. This need is particularly important for the Hispanic population, which was underrepresented in the derivation and validation of the EULAR/ACR criteria cohorts.

ACKNOWLEDGMENTS

The authors thank our LUMINA patients and staff who made this cohort possible.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. Ugarte-Gil 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. Ugarte-Gil, Pons-Estel, Griffin, Alarcón.
Acquisition of data. Pons-Estel, Vilá, Alarcón.
Analysis and interpretation of data. Ugarte-Gil, Pons-Estel, Harvey, Vilá, Griffin, Alarcón.

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. Aringer M, Costenbader K, Daikh D, Brinks R, Mosca M, Ramsey-Goldman R, et al. 2019 European League Against Rheumatism/American College of Rheumatology classification criteria for systemic lupus erythematosus. Arthritis Rheumatol 2019;71:1400–12.
5. Alarcon GS, Friedman AW, Straaton KV, Moulds JM, Lisse J, Bastian HM, et al. Systemic lupus erythematosus in three ethnic groups. III: a comparison of characteristics early in the natural history of the LUMINA cohort. LUpus in MInority populations: NAture vs. Nurture. Lupus 1999;8:197–209.
6. Alarcón GS, Roseman J, Bartolucci AA, Friedman AW, Moulds JM, Goel N, et al. Systemic lupus erythematosus in three ethnic groups. II: features predictive of disease activity early in its course. Arthritis Rheum 1998;41:1173–80.
7. Pons-Estel GJ, Ugarte-Gil MF, Harvey G, Wojdyla D, Quintana R, Saurit V, et al. Applying the 2019 EULAR/ACR lupus criteria to patients from an established cohort: a Latin American perspective. RMD Open 2020;6:e001097.
8. Carneiro AC, Ruiz MM, Freitas S, Isenberg D. A comparison of three classification criteria sets for systemic lupus erythematosus: a study looking at links to outcome and mortality. Arthritis Care Res (Hoboken) 2020;72:1611–4.
9. Dahlstrom O, Sjowall C. The diagnostic accuracies of the 2012 SLICC criteria and the proposed EULAR/ACR criteria for systemic lupus erythematosus classification are comparable. Lupus 2019;28:778–82.
10. Rodrigues Fonseca A, Felix Rodrigues MC, Sztajnbok FR, Gerardin Poirot Land M, Knupp Feitosa de Oliveira S. Comparison among ACR1997, SLICC and the new EULAR/ACR classification criteria in childhood-onset systemic lupus erythematosus. Adv Rheumatol 2019;59:20.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 73, No. 10, October 2021, pp 1456–1460
DOI 10.1002/acr.24335

BRIEF REPORT

Defining Minimum Clinically Important Changes for the Patient Activity Scale II

Joshua F. Baker,[1] Patti Katz,[2] and Kaleb Michaud[3]

Objective. To define the minimum clinically important improvement (MCII) and minimum clinically important worsening (MCIW) for the Patient Activity Scale II (PAS-II; range 0–10), a recommended patient-reported outcome measuring rheumatoid arthritis disease activity.

Methods. Data were taken from Forward, The National Databank for Rheumatic Diseases, from four 6-month data collection periods. Both anchor-based and distribution-based methods were used to estimate the MCII and MCIW. Anchor-based analyses used comparisons of pain and general health to the previous 6 months. Distribution-based analyses used 0.5 and 0.35 SDs. We stratified analyses based on the PAS-II score (above/below 3.7), hypothesizing that the MCII and MCIW would depend on the baseline score. To assess construct validity, we evaluated the odds of achieving the MCII in patients receiving new therapies.

Results. In the overall sample, for pain and general health anchor questions, the MCIW was 0.50 and 0.55, respectively. The MCII was defined as 0.39 and 0.45, respectively, for pain and general health. The MCIW for anchor-based methods among participants with low disease activity was 1.10 (1.09/1.11 [pain/general health]), while the MCII for those with moderate-to-high disease activity was 1.09 (1.15/1.02 [pain/general health]). Distribution-based methods for 0.5 and 0.35 SD were 1.08 and 0.76, respectively, for pain and general health. There was fair-to-excellent agreement with clinically important differences in assessments of pain and disability. Patients receiving new treatments had 30% greater odds of achieving the MCII.

Conclusion. The minimum important change in PAS-II score was approximately 0.5. Among participants with a moderate-to-high PAS-II score , the MCII was 1.1, and among participants with low disease activity, the MCIW was 1.1.

INTRODUCTION

The American College of Rheumatology management guidelines have recommended the use of quantitative disease activity measures to guide the management of patients with rheumatoid arthritis (RA) and to facilitate the achievement of well-defined treatment targets (1). Multiple disease activity measures have been accepted, and among these are disease activity measures that consist of patient-reported outcomes, such as the Patient Activity Scale II (PAS-II) (2).

The PAS-II consists of a composite assessment of RA disease activity that combines multiple domain-level scores for physical function, pain, and overall well-being and has been validated for clinical use. The score ranges from 0 to 10, with higher scores representing higher patient-reported disease activity. This practical measure can be used in clinical practice or in research studies that use survey instruments to quantify disease impact. An important construct to evaluate the impact of disease interventions is the minimum clinically important change, which represents the change in the construct that indicates a typically important change to an individual with the condition. Where research studies often focus on statistical significance, the construct of a clinically important change can be helpful when aiming to understand the meaning of the effect of exposures or interventions on the

The contents of this work do not represent the views of the Department of the Veterans Affairs or the US Government.

Dr. Baker's work was supported by a Veterans Affairs Clinical Science Research and Development Merit Award (I01 CX001703).

[1]Joshua F. Baker, MD, MSCE: Philadelphia Veterans Affairs Medical Center and Perelman School of Medicine, University of Pennsylvania, Philadelphia; [2]Patti Katz, PhD: University of California San Francisco; [3]Kaleb Michaud, PhD: University of Nebraska Medical Center, Omaha, Nebraska, and Forward, The National Databank for Rheumatic Diseases, Wichita, Kansas.

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

Address correspondence to Joshua F. Baker, MD, MSCE, Division of Rheumatology, Department of Medicine, 5 White Building, 3400 Spruce Street, Hospital of the University of Pennsylvania, Philadelphia, PA 19104. Email: bakerjo@uphs.upenn.edu.

Submitted for publication March 13, 2020; accepted in revised form May 22, 2020.

SIGNIFICANCE & INNOVATIONS

- The minimum clinically important change for the Patient Activity Scale II was ±0.5.
- The minimum clinically important worsening was 1.1 among those with low disease activity, and the minimum clinically important improvement was 1.1 among those with high disease activity.
- Patients initiating new treatments for rheumatoid arthritis had approximately 30% greater odds of an improvement as large as the clinically important improvement.

patient experience. While the minimum clinically important change in a number of other disease activity measures has been defined, it has not been defined for the PAS-II.

We aimed to determine the minimum clinically important worsening (MCIW) and minimum clinically important improvement (MCII) for the PAS-II among patients with RA in a large patient registry using anchor-based methods and distribution-based methods. Because minimum clinically important differences (MCIDs) may differ according to current health states, we also aimed to determine whether the MCIW and MCII were different among patients who reported low or high PAS-II scores at baseline (3). Finally, we illustrated the agreement of the constructs with other clinical changes and described the impact of treatment on the achievement of the MCII for the PAS-II.

MATERIALS AND METHODS

Study setting. Data were used from Forward, The National Databank for Rheumatic Diseases, a longitudinal observational study that follows patients with comprehensive questionnaires every 6 months. Participants in Forward are recruited primarily from rheumatologists, and diagnoses are provided by the rheumatologists. A minority of participants are enrolled from other sources, in which case diagnoses may be confirmed by participants' physicians or self-reported. More than 93% of the participants in each period had physician-confirmed RA. All participants have the option of completing the semi-annual questionnaire online, as a mailed paper questionnaire, or by telephone interview. Data shown in these analyses span four 6-month data collection periods: A, January 2017 (n = 3,680); B, July 2017 (n = 3,504); C, January 2018 (n = 3,737); and D, July 2018 (n = 3,102). All Forward procedures were approved by the Via Christi Institutional Review Board, and all participants provided consent to participate.

Measures. The PAS-II is a well-validated, self-reported assessment of function, pain, and overall health that has been endorsed by the American College of Rheumatology (1). This measure as well as its components were collected on each questionnaire (2,4). Components of the PAS-II include the Health Assessment Questionnaire II (HAQ-II), a measure of function in everyday activities, combined with results from a visual analog scale (VAS) assessment of pain, and a VAS assessment of overall disease activity. Low disease activity was defined as ≤3.7, as previously described (4).

Statistical analysis. Baseline characteristics of the study population from the first 6-month collection period (January 2017) were described. Both anchor-based and distribution-based methods were used to estimate the MCII and MCIW (5). For anchor-based analyses, the primary anchors used were comparisons of pain and general health to 6 months ago (e.g., "Compared to 6 months ago, would you say your pain is: much better now, somewhat better now, about the same, somewhat worse, or much worse?"). Differences in PAS-II scores were calculated for each pair of consecutive administrations, yielding 3 change periods (period A to period B, B to C, and C to D). The mean change in PAS-II scores of individuals falling into each response category for the anchor items (e.g., "much worse," "somewhat worse") were then calculated for each change period and averaged over the 3 change periods.

Effect sizes (mean change/SD of baseline, Cohen's *d*) were calculated for each group (6,7). Mean changes in PAS-II scores and effect sizes within each response category were averaged over the 4 change periods. Effect sizes of 0.2–0.50 were considered small, 0.50–0.80 moderate, and >0.80 large. Effect sizes <0.20 were considered negligible. The mean change of individuals responding somewhat worse was used as the estimate for the MCID for worsening (MCIW); the mean change of individuals responding somewhat better was used as the estimate for the MCID for improvement (MCII) (8).

For the distribution-based calculations, we used the standardized error of measurement, which reflects the precision of measurement and can be interpreted as the smallest difference likely to reflect a true difference rather than measurement error, and we also used 0.5 and 0.35 SDs (9). Distribution-based estimates were then averaged over the 4 administrations. In addition, we stratified analyses based on a PAS-II score above or below the low disease cut point (above/below 3.7) (10). We hypothesized that the MCII and MCIW would be dependent on the baseline PAS-II score (3).

Among all participants in Forward with available data for the PAS-II and components, we compared changes in results of the HAQ-II and pain VAS (range 1–10) among those who reached the MCII and MCIW for the PAS-II. We also explored the agreement between the MCII and MCIW for PAS-II and the MCID previously defined for pain (2.0) and HAQ-II (0.22) (11,12).

We also evaluated the likelihood of achieving an MCII based on our newly defined criteria among patients in Forward who had a baseline PAS-II score of >3.7 and who initiated treatments including biologic therapy, methotrexate, or prednisone, compared to

Table 1. Minimum important change defined by anchor-based methods, showing data only for the somewhat better and somewhat worse categories*

	Total		PAS-II score ≤3.7		PAS-II score >3.7	
Comparison†	Somewhat worse	Somewhat better	Somewhat worse	Somewhat better	Somewhat worse	Somewhat better
Pain						
A–B	0.53 (0.25)	−0.47 (−0.22)	1.03 (0.96)	−0.10 (−0.09)	0.13 (0.10)	−1.08 (−0.86)
B–C	0.53 (0.25)	−0.47 (−0.22)	1.17 (1.10)	−0.02 (−0.02)	0 (0)	−1.19 (−0.95)
C–D	0.56 (0.26)	−0.43 (−0.20)	1.06 (1.00)	−0.01 (−0.01)	0.16 (0.13)	−1.18 (−0.93)
Mean	0.55 (0.25)	−0.45 (−0.21)	1.09 (1.02)	−0.04 (−0.04)	0.10 (0.08)	−1.15 (−0.92)
General health						
A–B	0.54 (0.25)	−0.40 (−0.19)	1.10 (1.03)	−0.05 (−0.05)	0.10 (0.08)	−0.92 (−0.74)
B–C	0.53 (0.25)	−0.47 (−0.22)	1.17 (1.10)	0.02 (0.02)	0.09 (0.07)	−1.28 (−1.02)
C–D	0.44 (0.21)	−0.31 (−0.14)	1.05 (0.99)	0.02 (0.02)	0.07 (0.06)	−0.85 (−0.67)
Mean	0.50 (0.23)	−0.39 (−0.18)	1.11 (1.04)	0.00 (0.00)	0.09 (0.07)	−1.02 (−0.81)

* Values are the change in Patient Activity Scale II (PAS-II) score (Cohen's *d*: effect size).
† A–D are four 6-month periods of observations in 2017 through 2018.

those who did not receive new therapy. Observations from Forward where the PAS-II score was not >3.7 or where key demographic data were missing were excluded. Multivariable logistic regression models incorporating generalized estimating equations adjusted these analyses for baseline PAS-II score and for age, sex, race, and disease duration and predicted the probability of response in each setting to demonstrate the absolute differences.

RESULTS

The baseline characteristics of the study population during period A (n = 3,860) are shown in Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24335/abstract. The population average age was mean ± SD 64.9 ± 12.0 years, with a disease duration of mean ± SD 20.8 ± 12.7 years. The mean PAS-II score was low on average (mean ± SD 3.2 ± 2.2). The population was 83.1% female and 91.3% White.

For pain- and health-related anchor questions, the MCIW value was defined as approximately 0.50 and 0.55, respectively (Table 1), while the MCII value was defined as 0.39 and 0.45, respectively. These changes represented small effect sizes. Among participants in low disease activity at baseline (PAS-II score ≤3.7), the MCIW values for pain- and health-related anchor questions were 1.09 and 1.11, respectively. Among participants in high disease activity (PAS-II score >3.7), the MCII values for pain- and health-related anchor questions were 1.15 and 1.02, respectively. These changes represented large effect sizes. In contrast, the MCIW values for pain- and health-related anchor questions were 0.10 and 0.09, respectively, for patients with high disease activity, and the MCII values for pain- and health-related anchor questions were –0.04 and 0.00, respectively, for those in low disease activity. These changes represent very small effect sizes. The MCII and MCIW values were slightly smaller for participants age >65 years (see Supplementary Table 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24335/abstract), but were highly similar by sex.

Distribution-based methods resulted in a minimum clinically important change of 1.08 and 0.76, for 0.5 and 0.35 SD, respectively, when the sample was considered as a whole (Table 2). When the sample was examined stratified by PAS-II score, estimates were about half the size. Results were similar in each 6-month data collection period.

Among 70,294 observations from Forward with a baseline PAS-II score of >3.7, those participants who met the MCII (1.1) for the PAS-II in a subsequent time period had much greater improvements in pain (mean ± SD –2.89 ± 2.13 versus 0.34 ± 1.75; $P < 0.00001$) and HAQ-II scores (mean ± SD –0.34 ± 0.67 versus

Table 2. Minimum important change defined by distribution-based methods, by PAS-II level*

	Total		PAS-II score ≤3.7		PAS-II score >3.7	
Six-month period†	0.5 SD	0.35 SD	0.5 SD	0.35 SD	0.5 SD	0.35 SD
A	1.08	0.75	–	–	–	–
B	1.07	0.75	0.54	0.37	0.63	0.44
C	1.07	0.75	0.53	0.37	0.63	0.44
D	1.11	0.78	0.53	0.37	0.64	0.44
Mean	1.08	0.76	0.53	0.37	0.63	0.44

* PAS-II = Patient Activity Scale II.
† A–D are four 6-month periods of observations in 2017 through 2018.

0.042 ± 0.54; $P < 0.00001$) over the same interval. Among 71,732 observations with baseline PAS-II scores of ≤3.7, those who met the MCIW for the PAS-II had much greater worsening of pain (mean ± SD 2.69 ± 2.15 versus –0.12 ± 1.18; $P < 0.00001$) and HAQ-II scores (mean ± SD 0.39 ± 0.67 versus –0.029 ± 0.51; $P < 0.0001$) over the same interval.

Among those with a PAS-II score of >3.7, there was moderate and weak agreement between the MCII for PAS-II and a decrease equal to the MCID for pain ($\kappa = 0.60$, $P < 0.0001$) and HAQ-II ($\kappa = 0.28$, $P < 0.0001$). Among those with a PAS-II score of ≤3.7, there was also moderate and weak agreement between the MCIW for PAS-II and an increase equal to the MCID for pain ($\kappa = 0.63$, $P < 0.0001$) and HAQ-II ($\kappa = 0.28$, $P < 0.0001$).

Among 30,738 observations in 7,167 participants with a PAS-II score of >3.7, the initiation of biologic therapy, methotrexate, and prednisone was associated in each case with a significantly greater probability of achieving the MCII (improvement >1.1 units) compared to participants who did not report initiating new therapy over the same observation period (biologic therapy odds ratio [OR] 1.31 [95% confidence interval (95% CI) 1.22–1.40], $P < 0.001$; methotrexate OR 1.23 [95% CI 1.08–1.40], $P = 0.001$; prednisone OR 1.25 [95% CI 1.12–1.40], $P < 0.001$) (Figure 1).

DISCUSSION

In this study, we used anchor-based and distribution-based methods to define the MCII and MCIW for the PAS-II. On average, patients whose PAS-II score worsened by approximately 0.5 were more likely to report that their pain and general health were worse, while those whose score improved by approximately 0.5 were more likely to report improvement in pain and general health. In the overall sample, distribution-based methods yielded similar values.

The MCII and MCIW were found to vary significantly based on the baseline value. Participants with low disease activity appeared to have a ceiling on improvement and required a larger increase to define worsening, i.e., an increase of 1.1. In contrast, participants considered to be in moderate or high disease activity appeared to have a floor on worsening and required larger decreases to define improvement, i.e., a decrease of 1.1. These observations suggest that more significant worsening must occur among those with low disease activity for it to be considered important to the patient. Similarly, among those with higher disease activity, a more significant improvement must occur for the patient to believe that they have improved. The value of 1.1 was somewhat larger than that observed for distribution-based methods within these subgroups. Similar findings have been observed in recent studies defining the minimum important differences in other measures of pain interference and RA disease activity (13,14).

The MCII and MCIW are important constructs for clinical research studies that aim to determine the smallest change that is relevant to the patient experience. In particular, minimum important changes are important when evaluating changes over time, as in a longitudinal disease registry or in clinical data when evaluating the impact of quality interventions. Thus, the determination of the minimum clinically important change is an important advance for clinical research studies that use patient-reported outcomes measures. The stratification of the analysis by baseline disease activity is an important advance of the current study that emphasizes the need for interpretation of change in the PAS-II score (and likely other patient-reported outcomes) in the context of the baseline value of the score. Similar observations have been made for other disease activity measures (13), but stratified analyses have not consistently been performed for other composite disease activity scores and may be of value. Notably, the MCII and MCIW defined for the PAS-II represent large effect sizes. The large effect sizes observed here are consistent with other disease activity measures and suggests that only relatively large changes in these outcomes may be important to patients.

There was weak-to-moderate agreement between the MCII and MCIW defined for the PAS-II and previously defined MCIDs

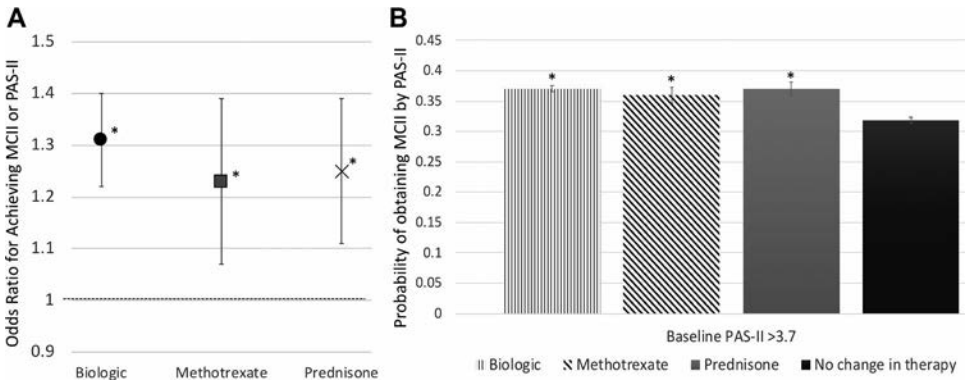


Figure 1. **A**, Odds ratio for achieving a change in the Patient Activity Scale II (PAS-II) score is consistent with the minimum clinically important improvement (MCII) for patients who started biologic therapy, methotrexate, or prednisone, compared to those who did not start new therapy, adjusted for age, sex, race, and baseline PAS-II score. Whiskers indicate 95% confidence intervals. **B**, Predicted probability of achieving a change as large as the MCII based on these models. * = $P < 0.01$ compared to no change in therapy, adjusted for age, sex, race, and baseline PAS-II score.

for pain and HAQ-II, respectively. While the agreement supports the idea that these changes are often associated with clinically important changes in other related constructs, the PAS-II composite score would not be expected to perfectly correlate with its individual components. In other words, while pain is a component of the PAS-II, the PAS-II is different because it combines pain with a measure of function and overall well-being.

The MCII and MCIW defined here are also comparable to MCII defined for the Routine Assessment of Patient Index Data (RAPID3), a similar patient-reported outcome measure. Prior studies have defined the MCII in several studies as between 3.5 and 3.8 on a 30-point scale (i.e., 1.16 and 1.26 on a 10-point scale) (15). The MCII we have defined for the PAS-II among those with moderate-to-high scores (1.1) is on a 10-point scale and is roughly comparable in scale to the MCII previously defined for the RAPID3.

This study also confirmed that patients who receive effective therapies for their disease are more likely to improve by the MCII compared to those who do not receive new therapy. While not surprising, these data help illustrate how the MCII might be used in observational or interventional research to better interpret the importance and clinical impact of interventions to treat RA. This study was not designed to accurately quantify effects of treatment; however, it supports the construct validity of the MCII for use in studies aimed at that purpose.

A limitation of the current study is that it may not be completely generalizable to other patient populations. The results should be validated in other study cohorts with different study populations and correlated with improvements in other clinical disease activity measures that are not readily available in this cohort. Our anchor questions might not accurately capture clinically important changes in this outcome, and future studies may help confirm our results using different anchor questions. The strength of the cohort is the large sample over multiple time points and the use of important anchors to directly asses patient values.

In conclusion, we defined minimum clinically important change for the PAS-II as a change in the score of 0.5 units. Among participants with moderate-to-high PAS-II scores, the MCII was estimated to be 1.1, and among participants with low disease activity, the MCIW was 1.1. These values were somewhat larger than those observed for distribution-based methods. The characterization of clinically meaningful changes in disease activity is important for clinical research studies and clinical settings where this disease assessment is used.

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. Baker 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. Baker, Katz, Michaud.
Acquisition of data. Baker, Katz, Michaud.
Analysis and interpretation of data. Baker, Katz, Michaud.

REFERENCES

1. 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;68:1–26.
2. Wolfe F, Michaud K, Pincus T. A composite disease activity scale for clinical practice, observational studies, and clinical trials: the patient activity scale (PAS/PAS-II). J Rheumatol 2005;32:2410–5.
3. Aletaha D, Funovits J, Ward MM, Smolen JS, Kvien TK. Perception of improvement in patients with rheumatoid arthritis varies with disease activity levels at baseline. Arthritis Rheum 2009;61:313–20.
4. 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.
5. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008;61:102–9.
6. Cella D, Lai J, Jensen S, Christodoulou C, Junghaenel D, Reeve B, et al. PROMIS fatigue item bank had clinical validity across diverse chronic conditions. J Clin Epidemiol 2016;73:128–34.
7. Cook K, Jensen S, Schalet B, Beaumont J, Amtmann D, Czajkowski S, et al. PROMIS measures of pain, fatigue, negative affect, physical function, and social function demonstrated clinical validity across a range of chronic conditions. J Clin Epidemiol 2016;73:89–102.
8. Bellamy N, Hochberg M, Tubach F, Martin-Mola E, Awada H, Bombardier C, et al. Development of multinational definitions of minimal clinical important improvement and patient acceptable symptomatic state in osteoarthritis. Arthritis Care Res (Hoboken) 2015;67:972–80.
9. Norman GR, Wyrwich KW, Patrick DL. The mathematical relationship among different forms of responsiveness coefficients. Qual Life Res 2007;16:815–22.
10. 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.
11. Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: dimensions and practical applications. Health Qual Life Outcomes 2003;1:20.
12. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. Eur J Pain 2004;8:283–91.
13. Curtis JR, Yang S, Chen L, Pope JE, Keystone EC, Haraoui B, et al. Determining the minimally important difference in the Clinical Disease Activity Index for improvement and worsening in early rheumatoid arthritis patients. Arthritis Care Res (Hoboken) 2015;67:1345–53.
14. Katz P, Kannowski CL, Sun L, Michaud K. Estimation of minimally important differences and patient acceptable symptom state scores for the Patient-Reported Outcomes Measurement Information System pain interference short form in rheumatoid arthritis. ACR Open Rheumatol 2020;2:320–9.
15. Ward MM, Castrejon I, Bergman MJ, Alba MI, Guthrie LC, Pincus T. Minimal clinically important improvement of routine assessment of patient index data 3 in rheumatoid arthritis. J Rheumatol 2019;46:27–30.

Arthritis Care & Research
Vol. 73, No. 10, October 2021, pp 1461–1469
DOI 10.1002/acr.24358

American College of Rheumatology
Empowering Rheumatology Professionals

Treatment Sequences After Discontinuing a Tumor Necrosis Factor Inhibitor in Patients With Rheumatoid Arthritis: A Comparison of Cycling Versus Swapping Strategies

Aliza R. Karpes Matusevich,[1] Zhigang Duan,[2] Hui Zhao,[2] Lincy S. Lal,[1] Wenyaw Chan,[1] María E. Suarez-Almazor,[2] Sharon H. Giordano,[2] J. Michael Swint,[3] and Maria A. Lopez-Olivo[2]

Objective. To evaluate the sequences of tumor necrosis factor inhibitors (TNFi) and non-TNFi used by rheumatoid arthritis (RA) patients whose initial TNFi therapy has failed, and to evaluate effectiveness and costs.

Methods. Using the Truven Health MarketScan Research database, we analyzed claims of commercially insured adult patients with RA who switched to their second biologic or targeted disease-modifying antirheumatic drug between January 2008 and December 2015. Our primary outcome was the frequency of treatment sequences. Our secondary outcomes were the time to therapy discontinuation, drug adherence, and drug and other health care costs.

Results. Among 10,442 RA patients identified, 36.5% swapped to a non-TNFi drug, most commonly abatacept (54.2%). The remaining 63.5% cycled to a second TNFi, most commonly adalimumab (41.2%). For subsequent switches of therapy, non-TNFi were more common. Patients who swapped to a non-TNFi were significantly older and had more comorbidities than those who cycled to a TNFi ($P < 0.001$). Survival analysis showed a longer time to discontinuation for non-TNFi than for TNFi (median 605 days compared with 489 days; $P < 0.001$) when used after initial TNFi discontinuation, but no difference in subsequent switches of therapy. Although non-TNFi were less expensive for adherent patients, cycling to a TNFi was associated with lower costs overall.

Conclusion. Even though patients are more likely to cycle to a second TNFi than swap to a non-TNFi, those who swap to a non-TNFi are more likely to persist with the therapy. However, cycling to a TNFi is the less costly strategy.

INTRODUCTION

The discovery of tumor necrosis factor inhibitors (TNFi) and other biologic and targeted synthetic therapies has brought new hope to patients with rheumatoid arthritis (RA). These drugs are an important treatment option after failure of conventional synthetic disease-modifying antirheumatic drugs (DMARDs). However, biologic and targeted therapies are associated with increased adverse events and can cost over $20,000 per year (1).

Over the course of their lifetime, most patients are required to switch medication several times owing to the adverse events of the drug or lack or loss of efficacy in managing symptoms. A systematic review of studies of TNFi discontinuation rates, based on registry and administrative databases, calculated a mean discontinuation rate of 27% (range 23–32%) after 1 year, increasing to 52% (range 46–57%) after 5 years (2).

There are 2 basic approaches following initial TNFi failure: cycling (switching to another TNFi) or swapping (i.e., to a drug with another mechanism of action), but there is no consensus regarding the most cost-effective therapeutic option yet. Time to discontinuation of treatment, calculated from administrative data sets, has become an acceptable proxy for effectiveness in the

Presented in part at the Annual Scientific Meeting of the American College of Rheumatology, Atlanta, GA, November 2019.

Presented by Dr. Karpes Matusevich in partial fulfillment of the requirements for a PhD degree, The University of Texas, School of Public Health, Houston, Texas.

Dr. Karpes Matusevich's work was supported by an Investigator Award from the Rheumatology Research Foundation. Dr. Lopez-Olivo's work was supported by an Investigator Award from the Rheumatology Research Foundation.

[1]Aliza R. Karpes Matusevich, PhD, Lincy S. Lal, PhD, Wenyaw Chan, PhD: School of Public Health, The University of Texas Health Science Center at Houston; [2]Zhigang Duan, MS, Hui Zhao, PhD, María E. Suarez-Almazor, MD, PhD, Sharon H. Giordano, MD, MPH, Maria A. Lopez-Olivo, MD, PhD: The University of Texas MD Anderson Cancer Center, Houston; [3]J. Michael Swint, PhD: School of Public Health and McGovern School of Medicine, The University of Texas Health Science Center at Houston.

Dr. Suarez-Almazor has received consulting fees from Pfizer, AbbVie, Eli Lilly and Company, Agile Therapeutics, and Amag Pharmaceuticals (less than $10,000 each). No other disclosures relevant to this article were reported.

Address correspondence to Aliza R. Karpes Matusevich, PhD, 341 Kyle Drive, Coppell, TX 75019 (email: aliza@karpes.org); or to Maria A. Lopez-Olivo, MD, PhD, Health Services Research, The University of Texas MD Anderson Cancer Center, Unit 1465, 1400 Pressler Street, Houston, TX 77230-1402 (email: amlopezo@mdanderson.org).

Submitted for publication December 17, 2019; accepted in revised form June 9, 2020.

SIGNIFICANCE & INNOVATIONS

- After their disease failed to respond to a first tumor necrosis factor inhibitor (TNFi), just under two-thirds of patients received a different TNFi, most commonly adalimumab. More than one-third of the patients switched to a non-TNFi, most commonly abatacept.
- For subsequent switches of therapy, non-TNFi were more common.
- Persistence was longer for non-TNFi than for TNFi when given as the next agent, after initial TNFi discontinuation. However, cycling was associated with lower drug costs.

absence of randomized clinical trials (3,4). Many studies have calculated survival times and the cost of various treatment strategies on the basis of the date of use, but these studies have been limited in terms of length of follow-up (5–8) and sample size (6,9–11). Only 1 study has investigated all 10 drugs approved by the US Food and Drug Administration as of 2017 (9). Furthermore, existing studies have used a limited lead time, so that differentiating between second and subsequent therapies used after initial TNFi discontinuation is difficult, instead categorizing treatment as the first or nonfirst drug used (5,8,10,11). The objectives of the current study were to describe real world sequences of TNFi and non-TNFi, time to drug discontinuation, and drug and other health care costs for adult patients with RA whose initial TNFi therapy failed.

PATIENTS AND METHODS

Data source. This retrospective cohort study used individual-level, de-identified, fully adjudicated health care claims information from employers and health plans collected from 1998 through 2016 in the Truven Health MarketScan Commercial Claims and Encounters Database. These data represent the health care claims of commercially insured employees and their dependents for active employees, early retirees, Consolidated Omnibus Budget Reconciliation Act continuees, and Medicare-eligible retirees with employer-provided Medicare Supplemental plans. Only plans where both the Medicare-paid amounts and the employer-paid amounts were available on the claims were included in our data set. Unlike Medicaid data, this source consolidates claims on a national, rather than state, level, and unlike Medicare data alone, it includes a wider range of patient ages. The MarketScan claims databases are fully compliant with the Health Insurance Portability and Accountability Act of 1996 (12); hence, a waiver from institutional review board approval was granted.

Study cohort. We used a validated claims-based algorithm (13–16), using at least 2 claims >2 months apart with RA diagnosis codes (International Classification of Diseases, Ninth Revision, Clinical Modification: 714.x; International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Clinical Modification: M05.x, M06.x) to identify adults (age ≥18) with RA who received their first TNFi between January 1, 2008, and December 31, 2015. All patients were required to have at least 1 year of continuous enrollment prior to the first claim for a TNFi and at least 1 year of enrollment after initiation of the second drug.

The drugs received by RA patients were identified by the National Drug Code in pharmacy claims or by Healthcare Common Procedure Coding System (HCPCS) codes in inpatient or outpatient claims (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24358/abstract). Claims with a zero or negative allowed payment were excluded. If a patient's index claim was deleted, we removed the patient from analysis.

From this initial cohort, we included only those who subsequently switched to a new drug of interest between January 1, 2008, and December 31, 2015. This timeframe was chosen to maximize sample size and the number of drugs available in the market (certolizumab and golimumab were approved in 2009, subcutaneous abatacept was approved in July 2011, and tofacitinib was approved in November 2012).

We excluded patients with overlapping treatment periods with biologic and targeted synthetic DMARDS (bDMARDs and tsDMARDs) defined as >1 drug within the effective period for that drug, because both American and European guidelines explicitly discourage this concomitant dual therapy (17,18). Furthermore, we excluded patients with RA who had diagnoses of non-RA indications for biologic drugs (ankylosing spondylitis, chronic lymphocytic leukemia, non-Hodgkin lymphoma, Crohn's disease, juvenile idiopathic arthritis, multiple sclerosis, polyarteritis nodosa, psoriasis, psoriatic arthritis, spondyloarthropathy, systemic lupus erythematosus, ulcerative colitis, or chronic relapsing granulomatosis with polyangiitis), as well as those patients with severe comorbidities involving immune suppression, such as HIV, organ transplantation, and malignancies (see Supplementary Table 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24358/abstract).

We defined the date of the first claim of the bDMARD or tsDMARD following discontinuation of initial TNFi as the index date. For the sake of convenience, we refer to these as subsequent drugs. We classified the cohort by mechanism of action of the next drug used: a different TNFi (cyclers) or non-TNFi (swappers).

Study measures. The MarketScan enrollment file provided data on sex, age, geographic region, and insurance type. We calculated the Deyo-Charlson comorbidity score from claims in the 6 months before the index date (19). We defined treatment endpoints for each subsequent treatment as switched, discontinued, or continued treatment until the end of follow-up. End of follow-up included patients who died, lost coverage, or changed insurance carrier. Patients were considered to have switched treatment if

they received a new prescription of a different bDMARD or tsDMARD. The end date of the previous treatment was defined as the first claim date of the subsequent treatment.

We determined a patient to have discontinued treatment if there was no claim for ≥180 days after the last prescription. In those cases, we defined the end date as the last claim date plus days' supply for prescription claims. For claims with an HCPCS code, the days' supply was imputed as the dosing interval for intravenous administration as stated in the product insert. For drugs administered subcutaneously only, the subcutaneous dosing interval was used. In cases where the dosing interval was variable, the smallest interval was used. Previous studies (20–30) used gaps of 30–90 days to determine drug discontinuation, but using gaps of 30–90 days precludes the possibility of patients stopping treatment owing to remission (31), surgery, or adverse events and restarting after flare, recovery from surgery, or the adverse event has resolved. Many studies reported patients restarting TNFi drugs after 140–207 days (32–34). We chose 180 days on the basis of our preliminary results showing that >25% of patients had gaps longer than 90 days. We deemed patients as continuing if they persisted on the same prescription until the end of their enrollment or of the study period. We determined the frequency of patients using different drug sequences to establish the most commonly used treatment patterns after initial TNFi failure.

Last, we calculated 2 categories of costs, comparing between patients who cycled or swapped after TNFi failure: direct drug-related costs, consisting of drug acquisition costs for the drugs of interest; and other health care costs, consisting of all other claims, including drug administration costs, costs for medications not analyzed here, hospital admissions, emergency department and health provider visits, pathology, and radiology. To control for adherence differences that may be attributed to adverse events and financial burden (35), we also compared costs between adherent and nonadherent patients. Adherent patients were those with a medication possession ratio of >80%. For oral and subcutaneous drugs, the medication possession ratio was calculated as the total number of days' supply within the 6-month period, divided by 183 days. For intravenous drugs, which do not have the days' supply variable, we followed Popp et al and defined adherence as receiving at least 80% of the expected doses, based on the dosing schedules for these drugs (see Supplementary Table 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24358/abstract) (36). All costs were adjusted to value at 2016 using the medical care component of the consumer price index (37).

Statistical analysis. We compared baseline demographic and clinical characteristics among the TNFi cycler and non-TNFi swapper groups using chi-square tests for categorical variables and *t*-tests for continuous variables. Six-month health care costs were calculated for the first and second 180-day postindex period by aggregating payment for individual claims for each of the drugs used after discontinuation of the initial TNFi. We calculated these costs for all patients, as well as for subgroups of adherent and nonadherent patients. Median and mean costs were calculated, and *t*-tests were used to compare costs between patients in different groups.

Total rates of switching, discontinuation, and continuation were estimated separately for the study cohort. Drug survival times of agents used after initial TNFi discontinuation were estimated using the Kaplan-Meier method and compared between the TNFi cycler and non-TNFi swapper groups. Switching and discontinuation were considered treatment failure events and

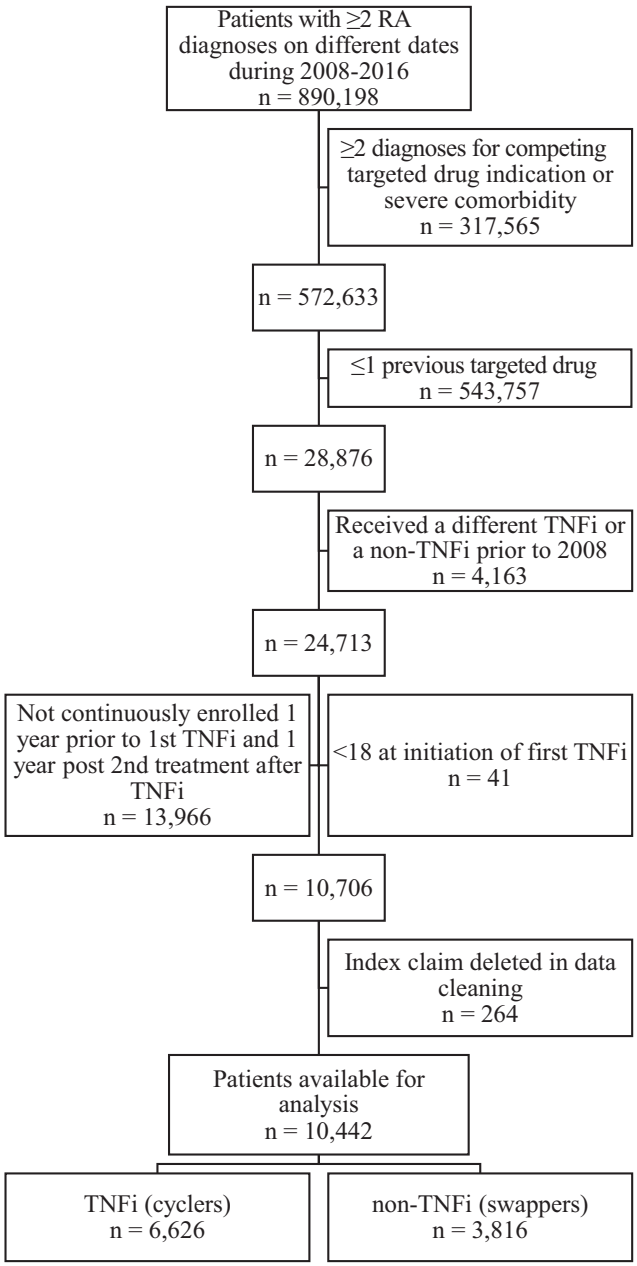


Figure 1. Patient selection flowchart. RA = rheumatoid arthritis; TNFi = tumor necrosis factor inhibitor.

were analyzed combining both types of events and separately. Time to treatment switch was calculated from the first date of the subsequent treatment after TNFi failure to the switch date. The time to treatment switch was censored at the patients' final enrollment date. Time to treatment discontinuation was calculated from the first date of the subsequent treatment after TNFi failure to the date the treatment was stopped. Cox proportional hazards models were used to determine other predicting variables for drug survival. *P* values less than 0.05 were considered statistically significant; all data analysis was conducted using SAS Enterprise Guide, version 7.15.

RESULTS

Baseline characteristics. A total of 10,442 patients with a mean ± SD follow-up time of almost 3 years (1,059 ± 583.1 days) met the study criteria (Figure 1). Of these, 6,626 patients (63.5%) cycled to a new TNFi and 3,816 (36.5%) swapped to a drug with a different mechanism of action. Patients who swapped to non-TNFi drugs were significantly older (53.6 years compared with 51.1 years; $P < 0.001$) and had higher Deyo-Charlson scores (8.4% with 2 or more comorbidities compared with 4.6%; $P < 0.001$). Their mean total follow-up time was also shorter than that of patients who cycled (1,023.4 days compared with 1,079.9 days; $P < 0.001$). There were significant differences between the cycling and swapping groups in terms of year of index claim, region, and health insurance plan type, but not sex (Table 1).

Sequences. Etanercept (n = 4,551 patients) and adalimumab (n = 3,305 patients) accounted for 43.6% and 31.7% of the TNFi used for the first time (before failure), respectively. Overall, although TNFi were most often prescribed as a second treatment for RA in patients whose initial TNFi failed, non-TNFi were most commonly used as subsequent treatment (third, fourth, etc.) for both cyclers and swappers (Figure 2). The most common TNFi drugs used after initial TNFi discontinuation were adalimumab (41.2% of cyclers) and etanercept (24.3% of cyclers) (see Supplementary Table 4, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24358/abstract). Slightly more than half of cyclers (52.5%) subsequently switched to a third drug, and the most common of these were abatacept (30.1% of swappers) and etanercept (14.2%) (see Supplementary Table 5, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24358/abstract).

More than half of patients who swapped after the initial TNFi discontinuation (54.3%) switched to abatacept. Less than half (45.9%) went on to a subsequent switch, of which 18.5% received tocilizumab and 11.8–14.3% received etanercept, tofacitinib, or adalimumab. Overall, approximately 25% of both cyclers and swappers who discontinued treatment did not switch to a new biologic or targeted DMARD (see Supplementary Table 5, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24358/abstract).

Survival analysis. Table 2 shows the overall proportion of patients who switched, discontinued, or continued on subsequent treatment after initial TNFi failure. The treatment switching rate was higher in the cycler compared to the swapper group (52.5% versus 45.9%), while the treatment continuation rate was lower in the cycler compared to the swapper group (33.4% versus 39.6%; $P < 0.001$). Similar results were observed when considering only participants who switched (excluding participants

Table 1. Demographic characteristics of patients in the cohort (n = 10,442)*

Variable	Cyclers	Swappers	*P*
Patients, no.	6,626	3,816	–
Age, mean ± SD years	51.10 ± 11.6	53.64 ± 11.9	<0.001†
Female sex	79.4	80.7	NS
Deyo-Charlson score			<0.001†
0	80.4	73.2	–
1	15.0	18.3	–
≥2	4.6	8.4	–
Region			0.004†
North Central	23.2	24.7	–
Northeast	15.5	15.7	–
South	40.5	40.4	–
West	19.4	17.2	–
Unknown	1.4	2.0	–
Health insurance plan type			<0.001†
Comprehensive	8.1	12.2	–
Exclusive provider organization	1.1	1.2	–
Health maintenance organization	13.3	9.4	–
Point of service	8.1	7.4	–
Preferred provider organization	57.3	58.3	–
Point of service, capitated	0.36	0.31	–
Consumer-directed health plan	5.6	5.6	–
High-deductible health plan	2.6	2.2	–
Unknown	3.5	3.6	–
Year of first TNFi			<0.001†
2008	26.9	31.0	–
2009	14.1	12.1	–
2010	14.2	12.3	–
2011	11.7	11.5	–
2012	10.2	10.1	–
2013	11.2	12.3	–
2014	8.7	8.2	–
2015	3.0	2.5	–
Adherent patients			–
First 6 months	53.6	52.8	NS
Second 6 months	33.2	34.7	NS
Follow-up time, mean ± SD days	1,079.9 ± 590.21	1,023.4 ± 568.84	<0.0001†

* Values are the percentage unless indicated otherwise. NS = not significant, TNFi = tumor necrosis factor inhibitor.
† Statistically significant.

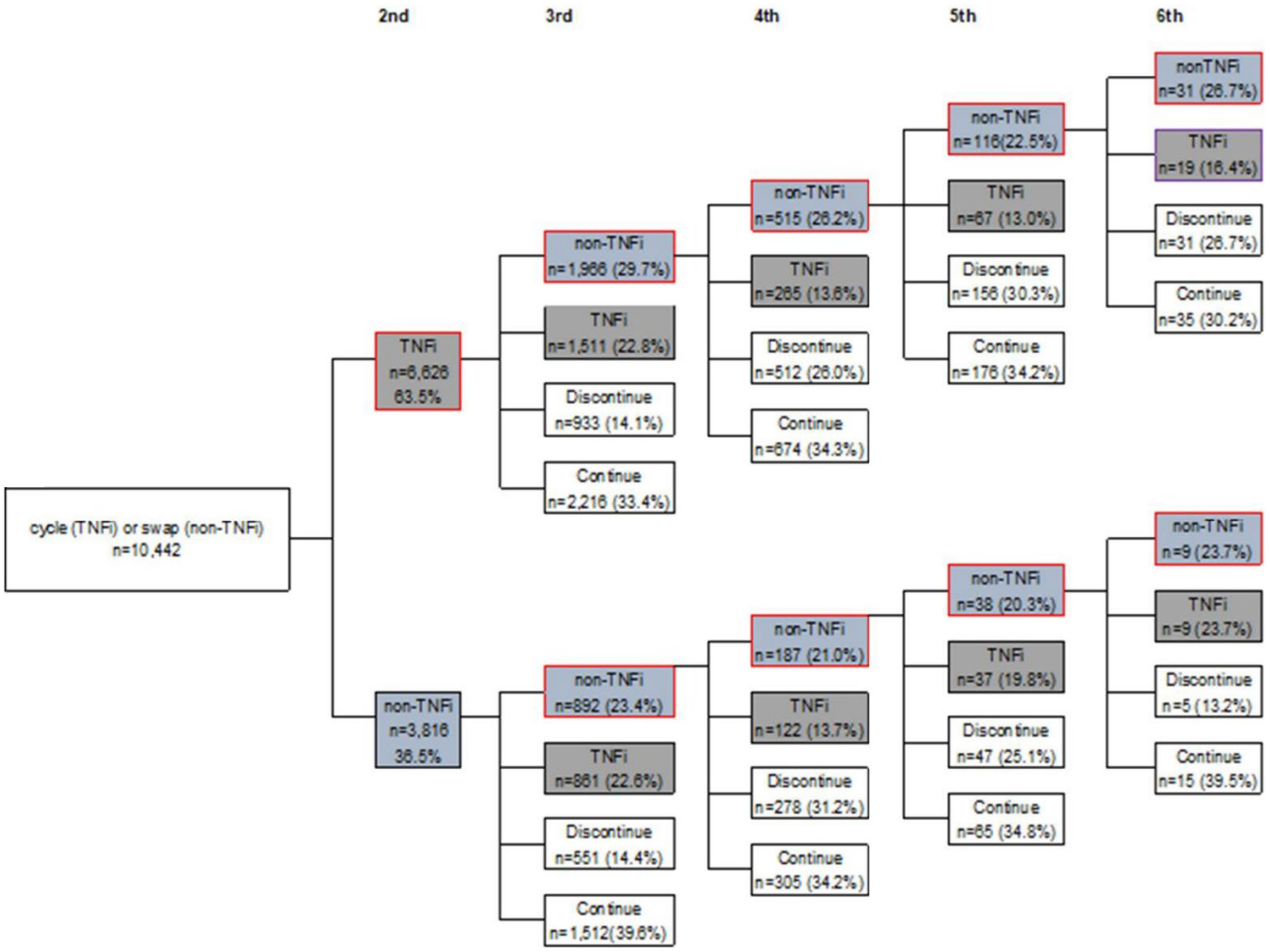


Figure 2. Most common sequences by drug class, with only the most common sequences included. For example, the figure does not show the fourth subsequent treatment for those who had a tumor necrosis factor inhibitor (TNFi) for the first, second, and third time after initial TNFi discontinuation. Totals per treatment sequence can be seen in Supplementary Tables 1–8, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24358/abstract.

who discontinued): the swappers had lower switching rates than cyclers (79.5% versus 85.1%; $P < 0.001$). However, for participants who discontinued, no differences were observed between cyclers and swappers (89.1% versus 88.8%; $P = 0.732$).

Figure 3 shows the second treatment (next agent used after TNFi failure) drug survival rate over time. Cycling patients had a higher rate of switching and discontinuation ($P < 0.001$). The estimated median drug survival time was 489 days (95% confidence interval [95% CI] 463–508) for cyclers and 605 days (95% CI 565–665) for swappers (see Supplementary Table 6, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24358/abstract).

In the multivariate Cox regression analysis (Table 3), the patients in the swapper group had a lower risk of switching their treatment (hazard ratio 0.87 [95% CI 0.82–0.91]; $P < 0.0001$). A recent year of TNFi initiation and higher Charlson-Deyo score were associated with a higher risk of switching treatment. Regional differences were also noted. There was no significant difference in time to switch for subsequent drug used. Also, no differences were observed when comparing the risk of discontinuing treatment among cyclers and swappers (see Supplementary Table 7, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24358/abstract).

Cost per treated patient. Mean costs across most categories were significantly lower for patients who cycled to a second TNFi (see Supplementary Table 8, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24358/abstract). However, among patients with a medication possession ratio of >80%, mean drug costs were lower for non-TNFi swappers in the second 6-month period ($14,455 versus $15,655; $P < 0.001$).

Table 2. Switch, discontinuation, and continuation rates of second drug used after initial TNFi failure*

	Total no.	Switch	Continuation	Discontinuation	*P*
Drug group					
Cyclers	6,628	3,477 (52.5)	2,216 (33.4)	933 (14.1)	<0.001†
Swappers	3,816	1,753 (45.9)	1,512 (39.6)	551 (14.4)	–
Individual drug					
Adalimumab	2,732	1,463 (53.6)	854 (31.3)	415 (15.2)	<0.001†
Certolizumab	738	409 (55.4)	224 (30.4)	105 (14.2)	–
Etanercept	1,612	807 (50.1)	585 (36.3)	220 (13.6)	–
Golimumab	855	451 (52.7)	299 (35.0)	105 (12.3)	–
Infliximab	689	347 (50.4)	254 (36.9)	88 (12.8)	–
Abatacept	2,073	1,023 (49.3)	763 (36.8)	287 (13.8)	–
Anakinra	16	9 (56.3)	2 (12.5)	5 (31.3)	–
Rituximab	539	203 (37.7)	230 (42.7)	106 (19.7)	–
Tocilizumab	640	283 (44.2)	280 (43.8)	77 (12.0)	–
Tofacitinib	548	235 (42.9)	237 (43.2)	76 (13.9)	–

* Values are the number (%) unless indicated otherwise. TNFi = tumor necrosis factor inhibitor.
† Statistically significant.

DISCUSSION

This claims-based analysis assessed treatment sequences, time to discontinuation, and costs for 10,442 patients for up to 8 years. We found that most patients cycled to a second TNFi after their initial TNFi discontinuation, but time to switch was longer for those who swapped to non-TNFi. Non-TNFi were used subsequently more frequently. We also found that patients who began their first TNFi in later calendar years, when there was a greater variety of choices, had shorter times to switch. Patients in the western and southern part of the US were also more likely to switch treatment earlier. Costs tended to be lower for TNFi.

Our results corroborate those of previous authors, who found that non-TNFi drugs were associated with increased treatment persistence despite being prescribed less often (5,7,9–11,38,39). Although we reported lower drug costs for adherent swappers, as in other studies (7,8,10,11), we found that other categories of costs favored TNFi cycling. This finding is possibly due to higher unit costs as well as the higher number of drugs administered intravenously only among the non-TNFi group. Intravenous administration requires a clinic visit and thus incurs more charges. Despite other studies reporting improved adherence among swappers (5,10), we found similar adherence between cyclers and swappers.

This is the first study including a large sample size (previous studies using Truven data analyzed 1,577 to 6,945 patients receiving therapy with a biologic drug after TNFi failure), with extended follow-up times (previous studies using Truven data followed patients for 3 to 5 years), and including all 10 targeted DMARDs available on the market at the end of the period studied. In addition, this is the first study to look at third and subsequent agents used after TNFi discontinuation. The most important strength is the clear identification of treatment used after TNFi discontinuation rather than nonfirst-treatment or continuing treatment. Previous studies that made this differentiation were limited by other factors, such as reliance on self-reporting (40), small sample size (n < 350) (39,41), follow-up of <3 years (32,42), or few drugs (42,43).

As with any data source, MarketScan claims data have limitations. Some have to do with the nature of claims data and others with the nature of the MarketScan sample population. The

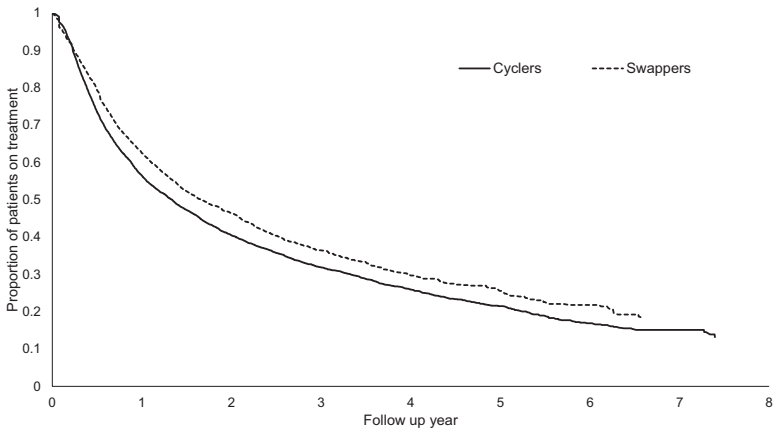


Figure 3. Kaplan Meier survival curve for cycling versus swapping.

Table 3. Time-to-treatment switch for second biologic or targeted synthetic agent used after initial TNFi (next agent after first failure) estimated by multivariable Cox regression*

	HR (95% CI)	*P*
Drug group		
TNFi cycler	Ref.	–
Non-TNFi swapper	0.87 (0.82–0.91)†	<0.0001†
Year of TNFi initiation		
2008	Ref.	–
2009	1.09 (1.01–1.18)†	0.035†
2010	1.12 (1.03–1.21)†	0.006†
2011	1.32 (1.21–1.43)†	<0.0001†
2012	1.24 (1.13–1.36)†	<0.0001†
2013	1.30 (1.19–1.42)†	<0.0001†
2014	1.52 (1.38–1.67)†	<0.0001†
2015	1.89 (1.63–2.20)†	<0.0001†
Age, years		
<40	Ref.	–
41–50	1.01 (0.94–1.09)	0.783
51–60	0.95 (0.88–1.02)	0.179
61–65	0.92 (0.83–1.02)	0.131
≥66	0.96 (0.86 –1.07)	0.484
Sex		
Women	Ref.	–
Men	0.94 (0.89–1.00)	0.054
Insurance plan		
Preferred provider organization	Ref.	–
Consumer-directed health plan	0.86 (0.74–1.01)	0.068
Comprehensive	0.68 (0.51–0.90)†	0.007†
Exclusive provider organization	0.90 (0.77–1.04)	0.145
High-deductible health plan	0.88 (0.75–1.03)	0.099
Health maintenance organization	0.90 (0.79–1.03)	0.114
Point of service	1.17 (0.77–1.77)	0.458
Point of service, capitated	0.78 (0.66–0.92)†	0.004†
Unknown	0.98 (0.80–1.19)	0.808
Region		
Northwest	Ref.	–
Northeast	1.02 (0.94–1.10)	0.663
South	1.09 (1.02–1.16)†	0.011†
West	1.11 (1.03–1.19)†	0.010†
Unknown	0.94 (0.76–1.16)	0.550
Deyo comorbidity score		
0	Ref.	–
1	1.04 (0.97–1.11)	0.299
≥2	1.16 (1.05–1.29)†	0.005†

* 95% CI = 95% confidence interval; HR = hazard ratio; Ref. = reference; TNFi = tumor necrosis factor inhibitor.
† Statistically significant at $P < 0.05$.

usefulness of all administrative data sets is constrained in that their purpose is to support reimbursement and not to serve as a research tool; therefore, there is no information regarding baseline disease activity, disease severity, or response to treatment. The lack of clinical and demographic information precludes propensity score matching, which could theoretically compensate for channeling bias, whereby specific groups of patients may be more likely to receive (or not receive) certain drugs than others. Channeling bias could cause results to be incorrectly attributed to the drug instead of unmeasured characteristics of the patients. In addition, because analysis is based on claims, we cannot know whether patients are taking the medications as prescribed.

Although multivariable modeling does control for some patient characteristics, the nonrandomized allocation of the study groups and baseline heterogeneity introduces bias and confounding. Our study was limited by the number of covariates analyzed compared with similar studies, such as concurrent and pre-index claim use of conventional synthetic DMARDs, pre-index claim costs, and a greater number of treatment effectiveness criteria.

Accuracy is also a concern, in that the existing diagnosis and procedure codes may be subject to upcoding or miscoding or may simply be missing if they are not reimbursable. In that vein, Fisher et al (44) reported errors in recording days' supply; this type of error in recording or coding supports using a more conservative (i.e., shortest possible) cutoff to determine treatment failure.

The MarketScan database also specifically underrepresents medium and small firms in favor of large employers, and the sample is not random, possibly leading to biases and impaired generalizability. The sample may also undercount the newer drugs because claims for newly licensed medications use a nonspecific HCPCS code (e.g., J3490 and J3590) until a unique HCPCS code specific to each drug is assigned, a process that can take up to 2 years. However, because physicians tend to prescribe more familiar drugs first, we believe that undercounting newer drugs is unlikely to impact results significantly.

Last, although time to drug discontinuation is commonly used as a surrogate marker for efficacy, other factors also influence retention rates, such as cost (in terms of absolute cost and patient copayments), insurance coverage, access to alternative treatments, and patient/provider preferences (45). Studies have shown that the threshold of disease activity lowers before patients switch treatments over time (45). In addition, the time to switching treatments tends to decrease over time (2,29,40); specifically, the rate of discontinuations due to inefficacy increases with no concomitant change in discontinuation rate due to adverse events (46). These findings support the contention that the availability of more choices leads to increased treatment switching rates.

Areas for future research include expanding the covariates used in the analysis while preserving the long follow-up time, as well as analyzing clinical databases that will allow for better matching of patients using more pertinent characteristics such as seromarker status. Another avenue for study is to determine how, if at all, the reasons for switching affect time to discontinuation of subsequent treatment. Additional data analysis is required first to corroborate our finding regarding discontinuation of biologic and targeted treatments altogether, second, to examine these issues for patients who are first exposed to non-TNFi,

and, third, to determine what alternative treatments patients are prescribed. We would also like to further examine the reasons for differences in costs between the treatment options. Finally, determining how insurance company coverage policies such as prior authorization requirements and copayments influence switching and discontinuation would be helpful. For example, Medicare enrollees receiving state assistance tend to use injectable biologics more than infusion agents due to lower out of pocket payments (1,47).

In conclusion, this claims-based analysis of commercially insured patients adds to the knowledge base by demonstrating how patients with RA change treatment over an extended period. Our analysis showed that TNFi are more frequently prescribed after initial TNFi discontinuation, and non-TNFi after that. We also found that costs tend to be lower for cyclers, with the exception of drug costs for adherent patients, which were lower for swappers. Furthermore, we showed that patients who swapped to a drug with a different mechanism of action have longer times to discontinuation than those who cycled to a second TNFi, supporting the use of non-TNFi biologic DMARDs in patients whose first TNFi failed. However, patient-specific clinical factors, not available in administrative databases, are needed for more unequivocal evidence.

ACKNOWLEDGMENTS

The authors thank Xerxes Pundole, MD, PhD, and Erica Goodoff in the Department of Scientific Publications at The University of Texas MD Anderson Cancer Center for their assistance with revising the article. They also thank Siddarth Karanth, MD, PhD, and Adriana Ordoñez, PhD, in the Department of Epidemiology and Disease Control, for their input into statistical methods.

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. Karpes Matusevich 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. Karpes Matusevich, Lal, Chan, Suarez-Almazor, Swint, Lopez-Olivo.

Acquisition of data. Karpes Matusevich, Duan, Zhao, Giordano, Lopez-Olivo.

Analysis and interpretation of data. Karpes Matusevich, Duan, Zhao, Lopez-Olivo.

REFERENCES

1. Yazdany J, Dudley RA, Chen R, Lin GA, Tseng CW. Coverage for high-cost specialty drugs for rheumatoid arthritis in Medicare Part D. Arthritis Rheumatol 2015;67:1474–80.
2. Souto A, Maneiro JR, Gomez-Reino JJ. Rate of discontinuation and drug survival of biologic therapies in rheumatoid arthritis: a systematic review and meta-analysis of drug registries and health care databases. Rheumatology (Oxford) 2016;55:523–34.
3. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol 2005;58:323–37.
4. Hudson M, Tascilar K, Suissa S. Comparative effectiveness research with administrative health data in rheumatoid arthritis. Nat Rev Rheumatol 2016;12:358–66.
5. Bonafede MM, Curtis JR, McMorrow D, Mahajan P, Chen CI. Treatment effectiveness and treatment patterns among rheumatoid arthritis patients after switching from a tumor necrosis factor inhibitor to another medication. Clinicoecon Outcomes Res 2016;8:707–15.
6. Harnett J, Gerber R, Gruben D, Koenig AS, Chen C. Evaluation of real-world experience with tofacitinib compared with adalimumab, etanercept, and abatacept in RA patients with 1 previous biologic DMARD: data from a U.S. administrative claims database. J Manag Care Spec Pharm 2016;22:1457–71.
7. Harnett J, Wiederkehr D, Gerber R, Gruben D, Koenig A, Bourret J. Real-world evaluation of TNF-inhibitor utilization in rheumatoid arthritis. J Med Econ 2016;19:91–102.
8. Zhou ZY, Griffith J, Du EX, Chin D, Betts KA, Ganguli A. Economic burden of switching to a non-tumor necrosis factor inhibitor versus a tumor necrosis factor inhibitor biologic therapy among patients with rheumatoid arthritis. Adv Ther 2016;33:807–23.
9. Wei W, Knapp K, Wang L, Chen CI, Craig GL, Ferguson K, et al. Treatment persistence and clinical outcomes of tumor necrosis factor inhibitor cycling or switching to a new mechanism of action therapy: real-world observational study of rheumatoid arthritis patients in the United States with prior tumor necrosis factor inhibitor therapy. Adv Ther 2017;34:1936–52.
10. Chastek B, Becker LK, Chen CI, Mahajan P, Curtis JR. Outcomes of tumor necrosis factor inhibitor cycling versus switching to a disease-modifying anti-rheumatic drug with a new mechanism of action among patients with rheumatoid arthritis. J Med Econ 2017;20:464–73.
11. Chastek B, Chen CI, Proudfoot C, Shinde S, Kuznik A, Wei W. Treatment persistence and healthcare costs among patients with rheumatoid arthritis changing biologics in the USA. Adv Ther 2017;34:2422–35.
12. Butler Quint J. Health research data for the real world: the MarketScan Databases. Ann Arbor (MI): Truven Health Analytics; 2015.
13. Curtis J, Schabert VF, Yeaw J, Korn J, Quach C, Harrison DJ, et al. Using a validated algorithm to evaluate the effectiveness of biologics for rheumatoid arthritis in a commercial claims database. Value Health 2013;16:A224.
14. Widdifield J, Bernatsky S, Paterson JM, Tu K, Ng R, Thorne JC, et al. Accuracy of Canadian health administrative databases in identifying patients with rheumatoid arthritis: a validation study using the medical records of rheumatologists. Arthritis Care Res (Hoboken) 2013;65:1582–91.
15. Zhang J, Xie F, Chen L, Greenberg JD, Curtis JR. Evaluation of a methodological approach to determine timing of rheumatoid arthritis disease onset using administrative claims data. Arthritis Rheumatol 2014;66 Suppl 10.
16. Ng B, Aslam F, Petersen NJ, Yu HJ, Suarez-Almazor ME. Identification of rheumatoid arthritis patients using an administrative database: a Veterans Affairs study. Arthritis Care Res (Hoboken) 2012;64:1490–6.
17. Singh JA, Saag KG, Bridges SL Jr, Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. Arthritis Rheumatol 2016;68:1–26.
18. Smolen JS, Landewe R, Breedveld FC, Buch M, Burmester G, Dougados M, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2013 update. Ann Rheum Dis 2014;73:492–509.
19. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol 1992;45:613–19.
20. Tkacz J, Ellis L, Bolge SC, Meyer R, Brady BL, Ruetsch C. Utilization and adherence patterns of subcutaneously administered anti-tumor necrosis factor treatment among rheumatoid arthritis patients. Clin Ther 2014;36:737–47.

21. Johnston S, McMorrow D, Farr AM, Juneau P, Ogale S. Comparison of healthcare costs between rheumatoid arthritis patients treated with infused biologics after switching from another biologic. Drugs Real World Outcomes 2015;2:99–109.
22. Kievit W, Fransen J, Adang EM, den Broeder AA, Bernelot Moens HJ, Visser H, et al. Long-term effectiveness and safety of TNF-blocking agents in daily clinical practice: results from the Dutch Rheumatoid Arthritis Monitoring register. Rheumatology (Oxford) 2011;50:196–203.
23. Bonafede M, Fox KM, Watson C, Princic N, Gandra SR. Treatment patterns in the first year after initiating tumor necrosis factor blockers in real-world settings. Adv Ther 2012;29:664–78.
24. Bonafede M, Johnson BH, Princic N, Shah N, Harrison DJ. Cost per patient-year in response using a claims-based algorithm for the 2 years following biologic initiation in patients with rheumatoid arthritis. J Med Econ 2015;18:376–89.
25. Bonafede M, Joseph GJ, Shah N, Princic N, Harrison DJ. Cost of tumor necrosis factor blockers per patient with rheumatoid arthritis in a multistate medicaid population. Clinicoecon Outcomes Res 2014;6:381–8.
26. Bonafede MM, Gandra SR, Watson C, Princic N, Fox KM. Cost per treated patient for etanercept, adalimumab, and infliximab across adult indications: a claims analysis. Adv Ther 2012;29:234–48.
27. Borah BJ, Huang X, Zarotsky V, Globe D. Trends in RA patients' adherence to subcutaneous anti-TNF therapies and costs. Curr Med Res Opin 2009;25:1365–77.
28. Ogale S, Hitraya E, Henk HJ. Patterns of biologic agent utilization among patients with rheumatoid arthritis: a retrospective cohort study. BMC Musculoskelet Disord 2011;12:204.
29. Yazici Y, Krasnokutsky S, Barnes JP, Hines PL, Wang J, Rosenblatt L. Changing patterns of tumor necrosis factor inhibitor use in 9074 patients with rheumatoid arthritis. J Rheumatol 2009;36:907–13.
30. Schabert VF, Watson C, Joseph GJ, Iversen P, Burudpakdee C, Harrison DJ. Costs of tumor necrosis factor blockers per treated patient using real-world drug data in a managed care population. J Manag Care Pharm 2013;19:621–30.
31. Gomez-Puerta JA, Hernandez MV, Sanchez-Alonso F, Yoshida K, Sanmarti R, Solomon DH, et al. Predictors of discontinuation of biologic DMARD therapy due to remission in patients with rheumatoid arthritis in a national registry. Arthritis Rheumatol 2014;66 Suppl 10.
32. Yeaw J, Watson C, Fox KM, Schabert VF, Goodman S, Gandra SR. Treatment patterns following discontinuation of adalimumab, etanercept, and infliximab in a US managed care sample. Adv Ther 2014;31:410–25.
33. Bonafede M, Watson C, Fox KM, Princic N, Gandra SR. Tumor necrosis factor blocker treatment patterns after discontinuation within the first year of therapy initiation in rheumatoid arthritis patients in a real-world managed care setting. Arthritis Rheum 2011;63 Suppl 10.
34. Schmeichel-Mueller C, Buysman E, Bolge S, Ingham M, McKenzie RS. Definitions of anti-TNF discontinuation may impact understanding of real-world utilization patterns. Value Health 2011;14:A71.
35. Steiner JF. Measuring adherence with medications: time is of the essence. Pharmacoepidemiol Drug Saf 2016;25:333–5.
36. Popp RA, Rascati K, Davis M, Patel U. Refining a claims-based algorithm to estimate biologic medication effectiveness and cost per effectively treated patient with rheumatoid arthritis. Pharmacotherapy 2018;38:172–80.
37. United States Bureau of Labor. U.S. Bureau of Labor's Medical Consumer Price Index. Washington DC: United States Bureau of Labor; 2017. URL: https://beta.bls.gov/dataViewer/view.
38. Wilke T, Mueller S, Lee SC, Majer I, Heisen M. Drug survival of second biological DMARD therapy in patients with rheumatoid arthritis: a retrospective non-interventional cohort analysis. BMC Musculoskelet Disord 2017;18:332.
39. Choquette D, Bessette L, Alemao E, Haraoui B, Postema R, Raynauld JP, et al. Persistence rates of abatacept and TNF inhibitors used as first or second biologic DMARDs in the treatment of rheumatoid arthritis: 9 years of experience from the Rhumadata(R) clinical database and registry. Arthritis Res Ther 2019;21:138.
40. Ramiro S, Landewe R, van der Heijde D, Harrison D, Collier D, Michaud K. Discontinuation rates of biologics in patients with rheumatoid arthritis: are TNF inhibitors different from non-TNF inhibitors? RMD Open 2015;1:e000155.
41. Favalli EG, Biggioggero M, Marchesoni A, Meroni PL. Survival on treatment with second-line biologic therapy: a cohort study comparing cycling and swap strategies. Rheumatology (Oxford) 2014;53:1664–8.
42. Markenson JA, Gibofsky A, Palmer WR, Keystone EC, Schiff MH, Feng J, et al. Persistence with anti-tumor necrosis factor therapies in patients with rheumatoid arthritis: observations from the RADIUS registry. J Rheumatol 2011;38:1273–81.
43. Meissner B, Trivedi D, You M, Rosenblatt L. Switching of biologic disease modifying anti-rheumatic drugs in patients with rheumatoid arthritis in a real world setting. J Med Econ 2014;17:259–65.
44. Fisher A, Bassett K, Wright JM, Brookhart MA, Freeman H, Dormuth CR. Comparative persistence of the TNF antagonists in rheumatoid arthritis: a population-based cohort study. PLoS One 2014;9:e105193.
45. Zhang J, Shan Y, Reed G, Kremer J, Greenberg JD, Baumgartner S, et al. Thresholds in disease activity for switching biologics in rheumatoid arthritis patients: experience from a large US cohort. Arthritis Care Res (Hoboken) 2011;63:1672–9.
46. Gomez-Reino JJ, Rodriguez-Lozano C, Campos-Fernandez C, Montoro M, Descalzo MA, Carmona L, et al. Change in the discontinuation pattern of tumour necrosis factor antagonists in rheumatoid arthritis over 10 years: data from the Spanish registry BIOBADASER 2.0. Ann Rheum Dis 2012;71:382–5.
47. Zhang J, Xie F, Delzell E, Chen L, Kilgore ML, Yun H, et al. Trends in the use of biologic agents among rheumatoid arthritis patients enrolled in the US Medicare program. Arthritis Care Res (Hoboken) 2013;65:1743–51.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 73, No. 10, October 2021, pp 1470–1478
DOI 10.1002/acr.24362

Randomized Controlled Trial of Patient Education Tools for Patients With Rheumatoid Arthritis

Maria A. Lopez-Olivo,[1] Heather Lin,[1] Tara Rizvi,[2] Andrea Barbo Barthel,[3] Aparna Ingleshwar,[4] Jude K. A. des Bordes,[5] Maria Jibaja-Weiss,[6] Robert J. Volk,[1] and Maria E. Suarez-Almazor[1]

Objective. The present study was undertaken to evaluate the efficacy of 2 educational tools for patients with rheumatoid arthritis (RA) by comparing a newly developed video tool, including storylines and testimonials, combined with a written booklet to the same written booklet alone.

Methods. We conducted a randomized controlled trial. Our primary outcome was disease knowledge. Secondary outcomes were decisional conflict, self-efficacy, effective health care management, and satisfaction. Outcomes were measured before and after reviewing the materials, and 3 and 6 months later. Linear mixed-effects models were performed to evaluate changes over time.

Results. In total, 221 participants received an educational video and booklet (n = 111) or a booklet alone (n = 110). The mean age was 50.8 years, mean disease duration was 4.8 years, 85% were female, and 24% had limited health literacy levels. Within groups, most outcomes improved between baseline and follow-up, but there were no statistically significant differences across groups. Patients receiving the video and booklet were more likely than those receiving the booklet alone to rate the presentation as excellent for providing information about the impact of RA, medication options, evidence about medications, benefits of medication, and self-care options. Factors significantly associated with greater improvements in knowledge and decisional conflict from baseline to 6 months included limited health literacy, lower educational level, and shorter disease duration.

Conclusion. Regardless of the delivery method, outcomes were improved up to 6 months after educational materials were delivered. Our findings support the implementation of self-administered educational materials in clinical settings, as they can result in sustained improvements in disease knowledge and decisional conflict.

INTRODUCTION

Patient education is an integral part of clinical practice in rheumatology. It enables patients to adapt and cope with the effects of rheumatic diseases and treatments. However, patient education is not always a routine part of practice. Few controlled trials of educational materials for patients with rheumatoid arthritis (RA) have been reported, and the findings have varied (1–8).

Behavioral interventions, group education, and provider-led patient education may be difficult to implement in clinical settings, and for simpler, self-administered tools, there is uncertainty about which educational delivery formats (ranging from written material to individualized web-based information) are most effective in improving health outcomes in chronic disease (9). Presenting information to patients may not lead to improvements in knowledge if the materials are not engaging or suited for the individual patient. Furthermore, the vast majority of printed educational materials are not suited for populations with limited health literacy because most are written at a higher level than that recommended by governmental guidelines (10). Greater rates of limited health literacy are observed in minority groups and elderly patients, requiring more careful attention to the content, format, and mode of administration of educational materials (11,12).

ClinicalTrials.gov identifier: NCT01698762.

Supported by the Agency for Health Research and Quality (grant 1R18HS019354) and in part by The University of Texas, MD Anderson Cancer Center Duncan Family Institute for Cancer Prevention and Risk Assessment (grant to Dr. Volk) and the NIH (National Cancer Institute grant P30-CA-016672).

[1]Maria A. Lopez-Olivo, MD, PhD, Heather Lin, PhD, Robert J. Volk, PhD, Maria E. Suarez-Almazor, MD, PhD: The University of Texas, MD Anderson Cancer Center, Houston; [2]Tara Rizvi, MD: Houston Rheumatology & Arthritis Specialists, Memorial Hermann, Houston, Texas; [3]Andrea Barbo Barthel, MS: Yale New Haven Hospital, Center for Outcomes Research and Evaluation, New Haven, Connecticut; [4]Aparna Ingleshwar, BDS, MPH: University of Iowa College of Dentistry, Iowa City; [5]Jude K. A. des Bordes, MBChB, DrPH: The University of Texas McGovern Medical School, Houston; [6]Maria Jibaja-Weiss, EdD: Baylor College of Medicine, Houston, Texas.

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

Address correspondence to Maria E. Suarez-Almazor, MD, PhD, The University of Texas, MD Anderson Cancer Center, 1515 Holcombe Boulevard, Unit 1467, Houston, TX 77030. Email: MSalmazor@mdanderson.org.

Submitted for publication March 23, 2020; accepted in revised form June 16, 2020.

SIGNIFICANCE & INNOVATIONS

- We compared the efficacy in improving knowledge of a newly developed multimedia patient education tool in English and Spanish about therapeutic options for patients with rheumatoid arthritis, incorporating video modeling combined with written materials versus written materials alone.
- Short, self-administered, patient education materials are effective in improving educational outcomes in patients with rheumatoid arthritis regardless of the mode of delivery.
- Our findings support the implementation of self-administered educational materials in clinical settings, as they require few resources to administer and result in sustained improvements in disease knowledge and decisional conflict.

Entertainment education, and in particular, video modeling, has been evaluated in various health conditions (13–16). Video modeling demonstrates health behaviors that are considered desirable through visual presentations including soap operas or serial dramas using actors with narratives that permit audiences to identify and relate with the characters and the changes they undergo. This technique is reported to increase positive self-care behaviors, improve short-term knowledge, decrease anxiety, and increase cooperation (17). Additionally, video modeling can be useful in educating populations with limited health literacy or with impairments affecting their ability to read printed materials (18,19).

For this study, we compared the efficacy of a multimedia patient education tool regarding therapeutic options for patients with RA, incorporating video modeling combined with written materials versus written materials alone in improving knowledge. We also evaluated the effects in decisional conflict, self-efficacy, behaviors to participate in the patient's own health care, and satisfaction with the materials. We hypothesized that the combination of video and written materials would lead to better outcomes.

PATIENTS AND METHODS

Study design and oversight. To report the results, we used the extension of the Consolidated Standards of Reporting Trials statement that addresses randomized trials of nonpharmacologic treatments (20). The current study was a randomized controlled trial with a 6-month follow-up. Participants were recruited from 5 outpatient clinics in 3 Houston area medical facilities (Kelsey-Seybold, Harris Health System, and MD Anderson Cancer Center) and through local newspaper advertisements from March 2013 through January 2014. The study was approved by the institutional review board or relevant research committee at each participating center. Preliminary results of this study were presented at the annual meeting of the American College of Rheumatology (21,22).

Participants. Participants were patients ages ≥18 years with a diagnosis of RA made by a rheumatologist and a disease duration of <10 years. The disease duration criterion was chosen because patients with longstanding disease are more likely to have acquired specific knowledge about their disease and have a number of other issues, such as surgical needs and comorbidities, not specifically discussed in the tools.

Interventions. One group received a newly developed video tool and similar written information in a paper booklet (video plus booklet), whereas the other group only received the booklet. The booklet was a consumer guide developed by the Health Care Program from the Agency for Healthcare Research and Quality (23). All materials (video and written materials) were available in English and Spanish. Details about the development, content, process to ensure the use of lay language, and pilot testing of the video are published elsewhere (24,25). Briefly, the video tool was structured as a series of dramatized episodes (26) within a common storyline depicting a main character with RA. Each episode was linked to a learning module providing patients with factual information about their condition and treatment options, similar to the information in the booklet. The total length of time of the video was 20 minutes. The video is available to the public (English: https://www.youtube.com/watch?v=oK6pCoYT_rk&t=570s; Spanish: https://mediaplayer.mdanderson.org/video-full/D2F7AF2B-B725-4E17-B282-AC777F6F3504).

Randomization. After completion of a baseline questionnaire, participants were randomly assigned in unequal allocation blocks to 1 of 2 study groups using an automated web-based institutional system. Participants were randomly assigned in a 1:1 ratio, stratified by site and by language preference of the participant. The random sequence was concealed from the 4 research staff who enrolled participants and the statistician (HL or ABB) conducting the analysis up until the implementation of the intervention due to the open-label nature of the study.

Implementation. After randomization and baseline assessment, patients were given time on-site to review the materials before the clinical encounter. They completed a questionnaire either immediately after on-site, or within 1 week of participation, mailing the questionnaire back. Participants were allowed to take the educational materials with them.

Follow-up assessments were conducted 3 and 6 months after inclusion. Patients were mailed self-report questionnaires, along with a stamped envelope for questionnaire return. An attempt was made to remind patients who had not returned their questionnaires at least once by phone to complete them. To ensure questionnaire completion, we offered patients who did not complete their questionnaires within 2 weeks the opportunity to either complete the questionnaires over the phone or to be met at the clinic or their home by study personnel to collect the

questionnaires. A modest monetary reimbursement was offered to all participants to compensate them for their time after the completion of each questionnaire.

Outcomes. The primary efficacy end point was knowledge about RA and therapeutic options immediately after reviewing the educational materials and at the 3- and 6-month follow-up visit compared with baseline. We adapted an RA knowledge questionnaire to reflect the key learner content covered in the video tool. The knowledge questionnaire has good psychometric properties (internal consistency r = 0.72–0.94, and test–retest r = 0.81) and has been previously validated (27–30). It includes 10 questions, and the score is the sum of correct items (final scores ranging 0–10). We used the effect-size approach (a distribution-based method) to calculate the minimum clinically important difference (MCID). A standardized mean difference of 0.50 (a moderate effect size) was considered the MCID.

Secondary efficacy end points were also assessed at baseline, 3 months, and 6 months and included the Decisional Conflict Scale, the Arthritis Self-Efficacy Scale, and the Effective Consumer Scale. The Decisional Conflict Scale measures the degree to which someone is conflicted or unclear about a particular health care choice. In our study, we used the Decisional Conflict Scale to learn if after being exposed to the educational materials patients felt more informed and clearer about what was most important to them when making decisions about their treatment. We used the low literacy version of the instrument and modified it to assess only the subscales relevant to our study, "informed" and "values clarity," as our intervention did not involve making a health care decision. Scores range from 0 (feels extremely informed/clear about personal values) to 100 (feels extremely uninformed/unclear about personal values) (31). The Decisional Conflict Scale was also measured immediately after review. The Arthritis Self-Efficacy Scale measures the belief in one's own capability to perform tasks or cope with adversity (32,33). It is an 8-item scale, with scores ranging 0–100, higher values indicating greater self-efficacy. The Effective Consumer Scale is a 17-item instrument that measures individuals' perception of their skills and behaviors in effectively managing their health care (34). The scores range 0–100, where higher scores indicate better health care management (35–37).

Patients rated the acceptability and their satisfaction with the educational tools immediately after reviewing the materials. Acceptability of the materials was measured using the Ottawa Acceptability Scale, including clarity, balance, length, and ease of use (38). Patients' satisfaction with the content, ease of use, and format of the materials was assessed using an instrument consisting of 9 items that asked patients to rate the video tool on the basis of its content, ease of use, transportability, and format (39,40).

Several measures were collected at baseline to serve as analytic covariates, including demographic information (age, sex, race/ethnicity, education, employment status, and marital status), and disease duration in years. Health literacy was assessed using a single item, "How confident are you filling out medical forms

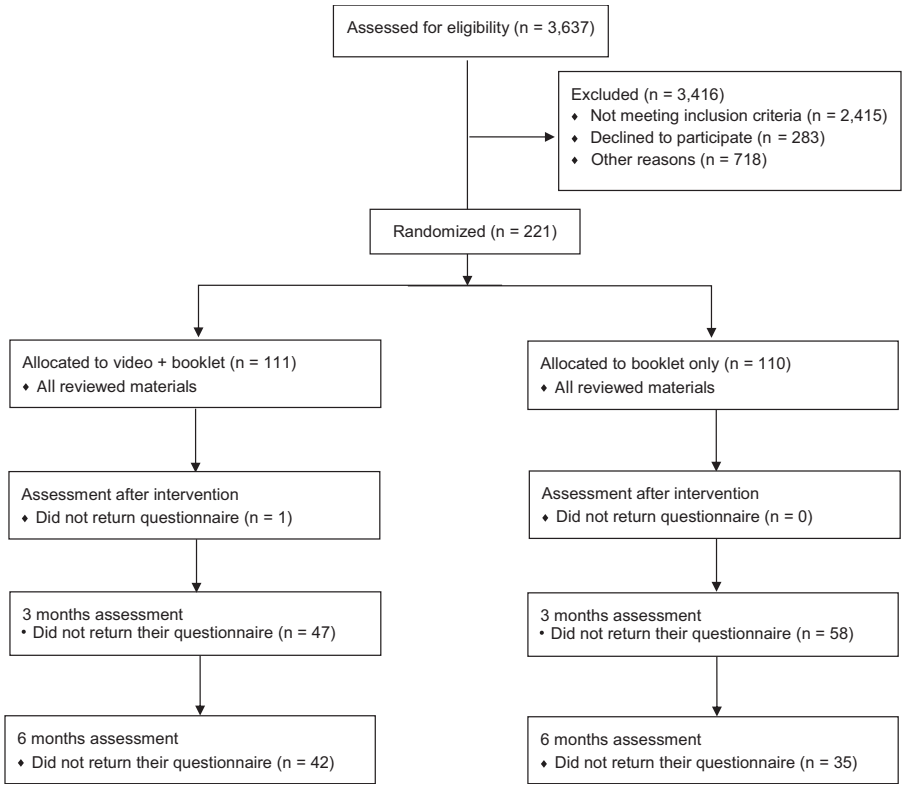


Figure 1. Consolidated Standards of Reporting Trials 2010 flow diagram.

by yourself?", which was developed by Chew et al to measure health literacy (41). The responses range from "extremely" to "not at all." For limited health literacy level, we used a cutoff point of at least "somewhat" in English-speaking participants and at least "a little bit" in Spanish-speaking participants, as has been proposed previously (41,42). Last, we administered the Health Assessment Questionnaire, a measure of physical function widely used in studies of patients with RA (43).

Sample size. The target sample size was estimated a priori to be 220 (110 per group). This sample size allowed for 81% power to detect a difference of 0.46 points on a scale of 0–10 (Cohen's $d = 0.33$, i.e., a small effect size) in a design with 3 repeated measurements that had a compound symmetry covariance structure when the SD is 1.41, a correlation between observations on the same subject of 0.6, and using 2-sided tests ($\alpha = 0.05$). The parameters employed were derived from a Cochrane systematic review of decision aids using knowledge as the outcome measure (44).

Statistical analyses. All analyses were performed on an intent-to-treat (ITT) basis, i.e., all patients who were randomized to receive educational materials were accounted for in the analysis according to the intervention that they were scheduled to receive. Missing data on outcome measures at a given time point were imputed by the mean of the observed data at the corresponding time point.

Given the longitudinal nature of the outcome measures, linear mixed-effect models were used to study the changes of the outcome measures over time to take the intrapatient correlation into account and to compare the changes in the outcome scores (follow-up period minus pre-randomization) between the groups (45).

Linear regression models were used to assess the relationship between the intervention and changes in outcome scores and the effect of covariates of interest. Analyses were conducted for differences before and immediately after reviewing the educational materials and before and after 6 months. The interactions between the group allocation and covariates (age, sex, race/ethnicity, education level, language in which the questionnaire was answered, health literacy, disease duration) were examined first. All 7 independent variables and interaction terms with $P < 0.10$ from above were included in the initial step of model selection. Subgroup analyses were performed in the presence of interaction between independent variable and group allocation. For all analyses, P values less than 0.05 (2-sided) were considered significant. SAS software was used to perform the analyses.

RESULTS

Of 504 patients who were approached, 283 refused to participate. In total, 221 were randomized to receive the video tool combined with the booklet (n = 111) or the booklet only (n = 110). The flow of participants through each stage is shown in Figure 1. Specifically, 23 were recruited through advertisements, 160 from county clinics, and 38 from other hospitals. A total of 116 patients (52%) (64 from the video plus booklet group, and 52 from the booklet only group) returned their questionnaires at 3 months, and 144 patients (65%) (69 from the video plus booklet group, and 75 from the booklet only group) returned their questionnaires at 6 months.

Baseline demographic and clinical characteristics. Table 1 shows baseline characteristics of the patients. No statistically significant differences were found between groups. The mean ± SD age was 50.8 ± 13.3 years, mean ± SD disease

Table 1. Baseline patient characteristics*

Characteristic	Total cohort (n = 221)	Video + booklet (n = 111)	Booklet only (n = 110)
Mean ± SD age, years	50.8 ± 13.3	49.8 ± 13.0	51.8 ± 13.5
Female sex	187 (85)	95 (86)	92 (84)
Race/ethnicity			
White	44 (20)	25 (23)	19 (17)
Black or African American	44 (20)	19 (17)	25 (23)
Hispanic	124 (56)	60 (54)	64 (58)
Other	8 (4)	6 (5)	2 (2)
Marital status, married/living together	119 (54)	65 (59)	54 (49)
Educational attainment			
Less than high school diploma or equivalent	81 (37)	40 (36)	41 (37)
High school diploma or equivalent or associate degree	102 (46)	51 (46)	51 (46)
Bachelor's degree or higher	36 (16)	19 (17)	17 (15)
Language of questionnaire, English	119 (54)	60 (54)	59 (54)
Employment status, employed	78 (35)	35 (32)	43 (39)
Mean ± SD disease duration, years	4.8 ± 2.7	4.5 ± 2.7	5.1 ± 2.7
Health literacy, limited health literacy	54 (24)	24 (22)	30 (27)
Mean ± SD no. of medications	2.7 ± 1.4	2.7 ± 1.3	2.7 ± 1.5
Mean ± SD Health Assessment Questionnaire score†	0.6 ± 0.6	0.6 ± 0.6	0.7 ± 0.6

* Values are the number (%) unless indicated otherwise. Percentages may not add up to 100% owing to rounding.
† Higher score indicates more difficulty with physical function.

duration was 4.8 ± 2.7 years, 85% were female, 24% had limited health literacy levels, and 54% answered the questionnaire in English. The time taken to review the video and the booklet in the group receiving both tools ranged 25–45 minutes compared to 10–30 minutes taken to read the booklet alone in the other group.

Disease knowledge. Outcomes across time points are shown in Table 2. Mean ± SD knowledge scores significantly increased immediately after participants reviewed the educational materials (video plus booklet 5.5 ± 2.1 to 7.6 ± 1.5; $P < 0.0001$) and booklet only 5.5 ± 2.1 to 7.2 ± 2.0; $P < 0.0001$), and at 3 and 6 months compared with baseline in both groups (Table 2). Both groups achieved the MCID (the standardized mean differences were 0.86 and 0.81 for the video plus booklet and booklet alone groups, respectively). However, no significant differences were observed in improvement in knowledge scores between the groups immediately after they reviewed the educational material (video plus booklet 7.6 ± 1.5 versus booklet only 7.2 ± 2.0; $P = 0.07$) and at 3 and 6 months ($P = 0.73$ and 0.74, respectively).

Decisional conflict. Decisional conflict scores decreased immediately after participants reviewed the educational materials, at 3 months, and at 6 months compared with baseline in both groups. There were no differences in score changes in total decisional conflict between the 2 groups immediately after they reviewed the educational material, at 3 months, and at 6 months. Similar results were observed with the "informed" and "values clarity" decisional conflict subscales.

Self-efficacy. Compared with the baseline assessment, both groups had a higher mean self-efficacy score at 6 months. However, there were no statistically significant differences in the changes of the scores across groups at 3 or 6 months.

Table 2. Outcome measures across time*

Outcome and group	Baseline	After intervention	3 months†	6 months†	P†
Knowledge questionnaire‡					
Video + booklet	5.5 ± 2.1	7.6 ± 1.5	7.2 ± 1.3	7.3 ± 1.3	<0.0001
Booklet only	5.5 ± 2.1	7.2 ± 2.0	7.3 ± 1.0	7.2 ± 1.2	<0.0001
P§		0.07	0.73	0.74	
Total decisional conflict score (higher score, higher decisional conflict)					
Video + booklet	42.9 ± 34.6	13.6 ± 21.9	24.8 ± 22.8	23.0 ± 19.7	<0.0001
Booklet only	45.5 ± 34.3	22.7 ± 27.2	25.4 ± 19.5	25.2 ± 20.9	<0.0001
P§		0.13	0.70	0.94	
Informed decisional conflict subscore (higher score, more uninformed)					
Video + booklet	45.2 ± 36.0	14.3 ± 22.1	25.9 ± 23.7	23.2 ± 20.6	<0.0001
Booklet only	49.6 ± 36.6	26.4 ± 30.5	27.8 ± 20.7	25.4 ± 21.3	<0.0001
P§		0.09	0.63	0.67	
Values clarity decisional conflict subscore (higher score, more unclear about personal values)					
Video + booklet	39.4 ± 37.9	12.6 ± 25.0	23.0 ± 24.5	22.5 ± 21.7	<0.0001
Booklet only	39.3 ± 39.3	17.3 ± 28.2	21.8 ± 21.3	24.8 ± 23.2	≤0.0002
P§		0.34	0.85	0.67	
Self-efficacy score (higher score, higher confidence)					
Video + booklet	56.3 ± 27.0		55.5 ± 21.5	62.9 ± 16.7	0.76
Booklet only	55.5 ± 24.1		56.1 ± 15.9	61.3 ± 17.2	0.78
P§			0.68	0.79	
Effective Consumer Scale score (0–100; higher score, better disease management)					
Video + booklet	74.8 ± 17.4		77.4 ± 12.3	79.1 ± 10.9	0.83
Booklet only	76.4 ± 17.1		76.6 ± 9.7	77.9 ± 12.9	0.32
P§			0.25	0.23	

* Values are the unadjusted mean ± SDs unless indicated otherwise. Analysis is based on intent-to-treat population.
† Within groups, the *P* value by *t*-test comparing baseline with follow-up score in PROC MIXED procedure in SAS was significant for all outcomes ($P < 0.01$) except self-efficacy (at 3 months, $P = 0.76$ and 0.78 for the 2 groups, respectively) and Effective Consumer Scale score (at 3 months, $P = 0.067$ and 0.83 for the 2 groups, respectively, and at 6 months, $P = 0.32$ for the booklet group).
‡ Seven of the 10 questions required the patient to check 2 answers. In this case, each correct choice was scored 0.5. If a patient checked 3, with 2 of them the correct answers, a score of 0.75 was assigned for that particular question. Three questions had only 1 correct answer for 1 point each. If a patient checked 2, with 1 of them the correct answer, a score of 0.75 was assigned for that particular question.
§ *P* values are by 2-sample *t*-test comparing the groups in terms of change in outcome between a follow-up period and baseline using the ESTIMATE statement in PROC MIXED procedure in SAS, unless stated otherwise.

Effective Consumer Scale. We observed better scores in consumer effectiveness in both groups at 6 months. However, the observed differences in the changes of the scores across groups were not statistically significant.

Acceptability of the materials. Patients in the video plus booklet group were more likely than those in the booklet only group to rate the presentation as excellent for providing information about the following items: impact of RA (56% versus 37%), medication options (62% versus 43%), evidence about medications (49% versus 32%), benefits of medication (54% versus 37%), and self-care options (48% versus 26%) ($P < 0.05$ for all). Also, more patients receiving the video plus booklet found the length of the material presented to be "just right" compared to those receiving the booklet alone (92% versus 80%; $P = 0.03$).

Satisfaction with educational tool. Most patients in both groups gave favorable responses to all evaluation questions. No significant differences in response options were observed between the 2 groups.

Determinants of improvement. *Knowledge.* Table 3 shows the predictors of knowledge improvement from baseline to immediately after intervention or at 6 months. Being male, having a shorter disease duration, and being Hispanic (compared with being White) were predictive of greater knowledge improvement immediately after participants received the educational materials. At 6 months, significant predictors of greater knowledge improvement were limited health literacy at baseline, lower educational level, and having a shorter disease duration. Baseline knowledge scores were similar across these subpopulations except for the health literacy groups and disease duration (see Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24362/abstract). Participants with limited health literacy and shorter disease duration had lower baseline knowledge scores than their counterparts but caught up to the others either immediately after reviewing the educational materials or at 6 months.

Decisional conflict. Significant predictors of less decisional conflict immediately after participants reviewed the educational materials included younger age and lower education. At 6 months, significant predictors of less decisional conflict were limited health literacy, lower educational level (less than high school diploma compared with bachelor's degree), and shorter disease duration. Factors associated with the decisional conflict "informed" and "values clarity" subscales are shown in Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24362/abstract. Baseline decisional conflict scores were similar across these subpopulations except for the health literacy groups (see Supplementary Figure 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24362/abstract). Participants with limited health literacy had greater decisional conflict at baseline than participants with adequate health literacy.

Effective Consumer Scale. The predictors of better effective consumer scale scores at 6 months were lower disease duration and limited health literacy ($P < 0.001$ for both covariates) (see Supplementary Table 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24362/abstract). Participants with limited health literacy and shorter disease duration had lower baseline effective consumer scale scores than their counterparts, but at 6 months, scores were similar across these subpopulations (see Supplementary Figure 1, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24362/abstract).

DISCUSSION

In this study, regardless of the delivery method, patients with RA showed improved outcomes that persisted up to 6 months after

Table 3. Factors associated with knowledge improvement (mean difference from baseline in total number of correct answers)*

Determinant	Knowledge improvement after intervention		Knowledge improvement at 6 months	
	β ± SE	*P*	β ± SE	*P*
Video tool + booklet (ref. = booklet alone)	0.44 ± 0.23	0.06	0.08 ± 0.24	0.75†
Female (ref. = male)	−0.77 ± 0.32	0.02†	–	–
Disease duration, years	−0.09 ± 0.04	0.04†	−0.12 ± 0.05	0.01†
Ethnicity (ref. = White)				
Black or African American	0.26 ± 0.36	0.47	–	–
Hispanic	0.78 ± 0.30	0.01†	–	–
Other	−0.56 ± 0.64	0.39		
Education level (ref. = less than high school)				
Bachelor's degree or higher	–	–	−1.9 ± 0.38	<0.001†
High school diploma or equivalent	–	–	−0.95 ± 0.29	<0.001†
Adequate health literacy (ref. = limited)	–	–	−0.93 ± 0.31	<0.001†

* Immediately after or within a week. Ref. = reference.
† Significant.

reviewing educational materials. The video tool combined with the reading material was equally as effective as the reading material alone in improving knowledge scores or decisional conflict; however, patient ratings were significantly higher for the multimedia tool. This finding is important, as in the study, all participants were asked and were given time to review the materials. Yet, in a clinical setting and when patients are on their own, they may be more likely to review materials that are more appealing. Our study, nevertheless, was not designed to address patients' adherence to reviewing the materials, but rather to examine the impact of the tools after review.

We found that several patient characteristics were associated with knowledge improvement. While there was some variation across time points, Hispanic participants, patients with limited health literacy or lower educational status, and those with shorter disease duration were more likely to have greater improvement in knowledge than their counterparts. Independent factors associated with less decisional conflict at 6 months included limited health literacy, lower educational level, and shorter disease duration. Some of these associations could be attributed to lower scores at baseline in patients with specific characteristics compared to their counterparts (i.e., participants with limited health literacy and shorter disease duration). These findings suggest that future patient educational tools should take into account baseline patient characteristics for best performance, and that patients at disadvantaged educational levels can catch up to others who are more informed.

Although substantial information on behavioral and nurse-led educational interventions in RA exists (46), there is little knowledge of the effect of self-administered, short educational materials, which are the most common materials provided in clinical settings, as they require few resources to administer. Specific studies on self-administered educational tools (books, pamphlets, workbooks, and computerized lessons) for patients with RA are scarce, most comparing having an educational tool to not receiving any type of health information and primarily showing that educational tools are effective (5,47). However, when comparing different self-administered educational tools, setting and delivery methods may not be as important. Similar to our results, a trial comparing multimedia educational material versus printed materials in patients with RA reported no differences between groups 1 month postintervention in self-reported adherence, illness perception, or functional status (4). Nonetheless, use of audiovisual delivery educational tools may be beneficial in situations where specific behaviors need to be learned, which also improves resource utilization. A trial that compared a web video on methotrexate self-injection combined with nurse guidance versus nurse guidance alone found that teaching time was reduced for the group of participants receiving the video intervention (3).

Although our video had features of decision support, we considered it primarily an educational tool because it does not incorporate all elements of the International Patient Decision Aid Standards. For example, we did not provide information about the outcome probabilities associated with the different treatment options or include a step-by-step tool to make a decision. Adding enhanced decision-making tools may also increase benefits. In a study comparing the same consumer guide that we used with a written decision aid, participants were assigned to either receive the booklet, an adapted low literacy, multilingual medication guide, or a low literacy, multilingual medication guide plus a multilingual decision aid provided during the medical encounter. Educational materials were seen by the participants prior to their routine clinical visit.

Immediately after the visit, patients receiving the combination of the adapted guide plus decision aid had higher knowledge scores. However, at 6 months, no significant differences were observed between groups in long-term clinical outcomes (disease activity or functional status) except for worse self-reported adherence in the group receiving the adapted guide alone. In a subgroup analysis of participants for whom a medication change was reported, those receiving the decision aid had significant improvement in knowledge and reduced decisional conflict (1). Another study, which compared a pharmaceutical booklet on etanercept with a long and a short decision aid, found that the patients who were compared and received any of the decision aid versions had greater knowledge scores immediately after reading the educational materials than those receiving the simple booklet. No other differences in outcomes were observed (2). Our study did not include a decision aid, as it was primarily designed to evaluate general educational materials on RA rather than to aid patients facing specific health decisions.

Our patient educational tool focused on the optimal use of effective disease-modifying antirheumatic drugs (traditional and biologic) and on the acquisition of disease knowledge and management, with design considerations for poor readers. The video tool was developed following a systematic and rigorous process that used an innovative approach to incorporate narratives and stories that contextualized the information and engaged the user in a program that was both didactic and entertaining (24,25).

Although our findings illustrate how so-called "edutainment" can work in RA, the findings may be limited by the following considerations. First, we did not include patients with longstanding disease, but only patients with a disease duration of <10 years. Patients with longstanding RA are likely to have different educational needs and priorities than those with disease of shorter duration (48). Second, the attrition rate at 3 and 6 months was high and could have increased the probability of a Type II error. However, the nonresponse rates are consistent with those reported in other educational studies and in tailored behavioral programs (49,50). Third, the lack of blinding (i.e., the fact that patients were aware of the assignment) may have affected the results. Educational studies are difficult to blind owing to practical issues, especially with 2 very distinct delivery methods. It is possible that the interaction between research staff and participants could have influenced the

responses, although this interaction only occurred at enrollment, as the 3- and 6-month assessments were primarily conducted by mailed self-response questionnaires. Fourth, as with most educational interventions, results from participants may differ from those of patients who declined to participate, thus decreasing the generalizability of the findings.

Our study provides evidence that self-administered, patient educational materials are effective in improving educational outcomes in patients with RA regardless of the mode of delivery. The newly developed video incorporating entertainment and education was liked better by participants than the printed booklet, suggesting that patients may be more likely to engage with multimedia tools. These are systematically developed materials following a rigorous method (25) that are easy to implement in a clinic. Our findings support the implementation of self-administered educational materials in clinical settings, as they can result in sustained improvements in disease knowledge and decisional conflict.

ACKNOWLEDGMENTS

We thank Vincent Richards (The University of Texas, MD Anderson Cancer Center) for contributions during the acquisition of the data and Kenneth Saag (The University of Alabama at Birmingham School of Public Health) and Amye Leong (Healthy Motivation and the Global Alliance for Musculoskeletal Health of the Bone and Joint Decade) for input provided during the development of the multimedia patient educational tool.

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. Suarez-Almazor 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. Jibaja-Weiss, Volk, Suarez-Almazor.
Acquisition of data. Lopez-Olivo, Rizvi, Ingleshwar, des Bordes.
Analysis and interpretation of data. Lopez-Olivo, Lin, Barbo, Ingleshwar, des Bordes, Suarez-Almazor.

REFERENCES

1. Barton JL, Trupin L, Schillinger D, Evans-Young G, Imboden J, Montori VM, et al. Use of low-literacy decision aid to enhance knowledge and reduce decisional conflict among a diverse population of adults with rheumatoid arthritis: results of a pilot study. Arthritis Care Res (Hoboken) 2016;68:889–98.
2. Martin RW, Enck RD, Tellinghuisen DJ, Eggebeen AT, Birmingham JD, Head AJ. Comparison of the effects of a pharmaceutical industry decision guide and decision aids on patient choice to intensify therapy in rheumatoid arthritis. Med Decis Making 2017;37:577–88.
3. Katz SJ, Leung S. Teaching methotrexate self-injection with a web-based video maintains patient care while reducing healthcare resources: a pilot study. Rheumatol Int 2015;35:93–6.
4. Unk JA, Brasington R. Efficacy study of multimedia rheumatoid arthritis patient education program. J Am Assoc Nurse Pract 2014;26:370–7.
5. Mohammad A, Kilcoyne A, Bond U, Regan M, Phelan M. Methotrexate information booklet study 2008. Clin Exp Rheumatol 2009;27:649–50.
6. Walker D, Adebajo A, Heslop P, Hill J, Firth J, Bishop P, et al. Patient education in rheumatoid arthritis: the effectiveness of the ARC booklet and the mind map. Rheumatology (Oxford) 2007;46:1593–6.
7. Barlow JH, Wright CC. Knowledge in patients with rheumatoid arthritis: a longer term follow-up of a randomized controlled study of patient education leaflets. Br J Rheumatol 1998;37:373–6.
8. Wetstone SL, Sheehan TJ, Votaw RG, Peterson MG, Rothfield N. Evaluation of a computer based education lesson for patients with rheumatoid arthritis. J Rheumatol 1985;12:907–12.
9. Cooper H, Booth K, Fear S, Gill G. Chronic disease patient education: lessons from meta-analyses. Patient Educ Couns 2001;44:107–17.
10. National Institutes of Health. Clear communication: an NIH health literacy initiative. 2011. URL: https://www.nihgov/institutes-nih/nih-office-director/office-communications-public-liaison/clear-communication.
11. U.S. Department of Health and Human Services. America's health literacy: why we need accessible health information. Office of Disease Prevention and Health Promotion Health Communication Activities. 2008. URL: https://health.gov/communication/literacy/issuebrief/.
12. National Network of Libraries of Medicine. Health literacy. 2016. URL: https://nnlm.gov/guides/intro-health-literacy.
13. Carpenter DJ, Gatchel RJ, Hasegawa T. Effectiveness of a videotaped behavioral intervention for dental anxiety: the role of gender and the need for information. Behav Med 1994;20:123–32.
14. Cull A, Miller H, Porterfield T, Mackay J, Anderson ED, Steel CM, et al. The use of videotaped information in cancer genetic counselling: a randomized evaluation study. Br J Cancer 1998;77:830–7.
15. Dunn RA, Shenouda PE, Martin DR, Schultz AJ. Videotape increases parent knowledge about poliovirus vaccines and choices of polio vaccination schedules. Pediatrics 1998;102.
16. Gagliano ME. A literature-review on the efficacy of video in patient education. J Med Educ 1988;63:785–92.
17. Krouse HJ. Video modelling to educate patients. J Adv Nurs 2001;33:748–57.
18. Kim S, Love F, Quistberg DA, Shea JA. Association of health literacy with self-management behavior in patients with diabetes. Diabetes Care 2004;27:2980–2.
19. Volandes AE, Paasche-Orlow MK, Barry MJ, Gillick MR, Minaker KL, Chang Y, et al. Video decision support tool for advance care planning in dementia: randomised controlled trial. BMJ 2009;338:b2159.
20. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P, Group C. Methods and processes of the CONSORT Group: example of an extension for trials assessing nonpharmacologic treatments. Ann Intern Med 2008;148:W60–6.
21. Lopez-Olivo MA, Ingleshwar A, Volk R, Barbo A, Jibaja-Weiss M, Lin H, et al. Multimedia patient education tool for patients with rheumatoid arthritis [abstract]. Arthritis Rheumatol 2014;66 Suppl:182.
22. Lopez-Olivo MA, Barbo A, Rizvi T, Volk R, Lin H, Suarez-Almazor ME. Six-month effects of a multimedia patient education tool in patients with rheumatoid arthritis. a randomized controlled trial [abstract]. Arthritis Rheumatol 2015;67 Suppl 10. URL: https://acrabstracts.org/abstract/six-month-effects-of-a-multimedia-patient-education-tool-in-patients-with-rheumatoid-arthritis-a-randomized-controlled-trial/.
23. Robinson S, Schechtel M, Dahlstrom C, Chou R, Goei M, Bianco T, et al, for the Eisenberg Center at Oregon Health & Science University. Rheumatoid arthritis medicines: a guide for adults. In: Donahue KE, Gartlehner G, Jonas DE, Lux LJ, Thieda P, Jonas B, et al, editors. Comparative effectiveness review summary guides for consumers. Rockville (MD): Agency for Healthcare Research and Quality; 2005.
24. Lopez-Olivo MA, Ingleshwar A, Volk R, Barbo A, Jibaja-Weiss M, Suarez-Almazor ME. Development of multimedia patient education tools (MM-PtET) for osteoarthritis (OA), osteoporosis (OP) and

rheumatoid arthritis patients (RA) [abstract]. Arthritis Rheumatol 2014;66 Suppl:S882.

25. Lopez-Olivo MA, Ingleshwar A, Volk RJ, Jibaja-Weiss M, Barbo A, Saag K, et al. Development and pilot testing of multimedia patient education tools for patients with knee osteoarthritis, osteoporosis, and rheumatoid arthritis. Arthritis Care Res (Hoboken) 2018;70:213–20.
26. Singhal A, Cody MJ, Rogers EM, Sabido M. Entertainment-education and social change. Mahwah (NJ): Lawrence Erlbaum Associates; 2004.
27. Hennell SL, Brownsell C, Dawson JK. Development, validation and use of a patient knowledge questionnaire (PKQ) for patients with early rheumatoid arthritis. Rheumatology (Oxford) 2004;43:467–71.
28. Hill J, Bird HA, Hopkins R, Lawton C, Wright V. The development and use of Patient Knowledge Questionnaire in rheumatoid arthritis. Br J Rheumatol 1991;30:45–9.
29. Jennings F, Toffolo S, de Assis MR, Natour J. Brazil Patient Knowledge Questionnaire (PKQ) and evaluation of disease-specific knowledge in patients with rheumatoid arthritis. Clin Exp Rheumatol 2006;24:521–8.
30. Minnock P, Fitzgerald O, Bresnihan B. Quality of life, social support, and knowledge of disease in women with rheumatoid arthritis. Arthritis Rheum 2003;49:221–7.
31. Ottawa Hospital Research Institute. User manual: decisional conflict scale. 2006. URL: https://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_Decisional_Conflict.pdf.
32. Mueller A, Hartmann M, Mueller K, Eich W. Validation of the arthritis self-efficacy short-form scale in German fibromyalgia patients. Eur J Pain 2003;7:163–71.
33. 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.
34. Kristjansson E, Tugwell PS, Wilson AJ, Brooks PM, Driedger SM, Gallois C, et al. Development of the effective musculoskeletal consumer scale. J Rheumatol 2007;34:1392–400.
35. Bremander A, Wikstrom I, Larsson I, Bengtsson M, Hagel S, Strombeck B. Cultural adaptation, validity, reliability and responsiveness of the Swedish version of the effective musculoskeletal consumer scale (EC-17). Musculoskeletal Care 2012;10:43–50.
36. Hamnes B, Garratt A, Kjeken I, Kristjansson E, Hagen KB. Translation, data quality, reliability, validity and responsiveness of the Norwegian version of the Effective Musculoskeletal Consumer Scale (EC-17). BMC Musculoskelet Disord 2010;11:21.
37. Santesso N, Rader T, Wells GA, O'Connor AM, Brooks PM, Driedger M, et al. Responsiveness of the Effective Consumer Scale (EC-17). J Rheumatol 2009;36:2087–91.
38. O'Connor AM, Cranney A, for the Ottawa Hospital Research Institute. User manual: acceptability. 2000. URL: https://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_Acceptability.pdf.
39. Barry MJ, Fowler FJ Jr, Mulley AG Jr, Henderson JV Jr, Wennberg JE. Patient reactions to a program designed to facilitate patient participation in treatment decisions for benign prostatic hyperplasia. Med Care 1995;33:771–82.
40. Volk RJ, Cass AR, Spann SJ. A randomized controlled trial of shared decision making for prostate cancer screening. Arch Fam Med 1999;8:333–40.
41. 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.
42. Sarkar U, Schillinger D, Lopez A, Sudore R. Validation of self-reported health literacy questions among diverse English and Spanish-speaking populations. J Gen Intern Med 2011;26:265–71.
43. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. J Rheumatol 1982;9:789–93.
44. O'Connor AM, Bennett CL, Stacey D, Barry M, Col NF, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2009:CD001431.
45. Jiang J. Linear and generalized linear mixed models and their applications. Springer; 2007.
46. Hawley DJ. Psycho-educational interventions in the treatment of arthritis. Baillieres Clin Rheumatol 1995;9:803–23.
47. Kars Fertelli T. Effects of education about rheumatoid arthritis and sexuality on the sexual problems of women with rheumatoid arthritis. Clin Nurs Res 2019:1054773819858493.
48. Ndosi M, Adebajo A. Patient education in rheumatoid arthritis: is the needs-based approach the way forward? Clin Rheumatol 2015;34:1827–9.
49. Giraudet-Le Quintrec JS, Mayoux-Benhamou A, Ravaud P, Champion K, Dernis E, Zerkak D, et al. Effect of a collective educational program for patients with rheumatoid arthritis: a prospective 12-month randomized controlled trial. J Rheumatol 2007;34:1684–91.
50. Hammond A, Bryan J, Hardy A. Effects of a modular behavioural arthritis education programme: a pragmatic parallel-group randomized controlled trial. Rheumatology (Oxford) 2008;47:1712–8.

Arthritis Care & Research
Vol. 73, No. 10, October 2021, pp 1479–1489
DOI 10.1002/acr.24344

American College of Rheumatology
Empowering Rheumatology Professionals

Preferences for Self-Management and Support Services in Patients With Inflammatory Joint Disease: A Danish Nationwide Cross-Sectional Study

Nanna Maria Hammer,[1] Caroline A. Flurey,[2] Kim Vilbæk Jensen,[1] Lena Andersen,[1] and Bente Appel Esbensen[3]

Objective. To explore preferences for self-management and support services in patients with inflammatory joint disease (IJD) and to investigate whether these preferences differ by age, sex, diagnosis, and disease duration.

Methods. We used a nationwide cross-sectional online survey for patients with rheumatoid arthritis, psoriatic arthritis, and axial spondyloarthritis. Descriptive statistics were applied to explore preferences and to test for differences according to the different subgroups of patients.

Results. The questionnaire was completed by 664 patients. Younger patients indicated greater interest in 1-to-1 discussions with psychologists or another patient, educational sessions, events, and online services, and older patients indicated greater interest in talks by researchers. More women than men indicated interest in health professionals' 1-to-1 discussions, occupational therapists' question-and-answer (Q and A) sessions, physical activity, and informational websites. Patients with axial spondyloarthritis tended to indicate the most interest in the different services, and patients with rheumatoid arthritis the least interest, reaching statistical significance regarding discussion groups about IJD experiences, 1-to-1 discussions with psychologists or another patient, Q and A with another patient, stress/anger management, and online patient communication. More patients with short rather than long disease duration indicated interest in 1-to-1 discussions with rheumatologists or nurses, organized talks with experienced patients, and online services for patient communication and stories.

Conclusion. Patients with IJD report various needs regarding self-management and support services, including 1-to-1 services traditionally delivered as part of usual care, but also talks, physical activity, and educational and online services. Although preferences differed across age, sex, diagnosis, and disease duration, all subgroups indicated great need for support, with only small differences in their top preferences.

INTRODUCTION

Despite advancements in the medical treatment of patients with inflammatory joint disease (IJD) in recent years, diseases like rheumatoid arthritis (RA), psoriatic arthritis (PsA), and axial spondyloarthritis (SpA) continue to have a major impact on patients' physical and psychosocial functioning and wellbeing. IJD has been associated with pain, fatigue, and physical disability (1) and might lead to considerable psychological distress. This distress includes increased levels of anxiety and depression (2), with the prevalence of depressive disorder in people with RA estimated at 13–20% (3). Furthermore, the occurrence of psychological distress that does not fulfil the diagnostic criteria of anxiety and depression is estimated to be as high as 65% in people with RA (4).

Depression or psychological distress can increase the burden on the health care system, with repeated consultations, reduced treatment adherence, and poor treatment outcomes (5,6). However, there are several psychological factors, which have an impact on adaptation to IJD, that are amenable to intervention (7). Meeting patients' support needs can improve their quality of life and result in economic benefits, in terms of a reduction in the economic burden on the health care system (8).

Supported by the Danish Rheumatism Organization (R141-A4016).

[1]Nanna Maria Hammer, OT, MSc, Kim Vilbæk Jensen, Lena Andersen: Copenhagen Center for Arthritis Research and Rigshospitalet, Glostrup, Denmark; [2]Caroline A. Flurey, MSc, PhD, CPsychol, FHEA: University of the West of England, Bristol, UK; [3]Bente Appel Esbensen, RN, MSciN, PhD: Copenhagen Center for Arthritis Research and Rigshospitalet, Glostrup, and University of Copenhagen, Copenhagen, Denmark.

Dr. Esbensen has received speaking fees from Pfizer and Eli Lilly and Company (less than $10,000 each). No other disclosures relevant to this article were reported.

Address correspondence to Bente Appel Esbensen, RN, MSciN, PhD, Rigshospitalet, Valdemar Hansens Vej 17, DK-2600 Glostrup, Denmark. Email: bente.appel.esbensen@regionh.dk.

Submitted for publication August 22, 2019; accepted in revised form May 22, 2020.

SIGNIFICANCE & INNOVATIONS

- Patients with inflammatory joint disease (IJD) report various preferences for self-management and support services, including both 1-to-1 services traditionally delivered as part of usual care, and organized talks, physical activity, and educational and online informational services.
- There are some differences in preferences for self-management and support services across subgroups of patients with IJD (i.e., age, sex, diagnosis, and disease duration), but the reported interest for the services is high among all the subgroups, with only small differences in top preferences.
- Most patients with IJD report a wish for flexibility in the delivery of self-management and support services, i.e., no fixed commitment and accessibility according to their current needs.

Self-management has increasingly been recognized as an important contributor to the adaptation process of living with a chronic disease, including IJD. It includes dealing with the pharmacologic management, as well as role and emotional management, and taking on responsibility for one's own situation. Self-management, which also covers patient education, is more than simple adherence to treatment, but also incorporates psychological and social management of living with chronic illness (9). Self-management interventions and patient education that aim to increase patient involvement in disease management and improve psychological health, coping, and health-promoting behaviors, are recommended as an essential part of the nonpharmacologic management of IJD (10,11). Such interventions have demonstrated beneficial effects on knowledge, adherence, coping, physical functioning, and psychological status in patients with IJD (7,10,12–15). However, these effects are not convincingly demonstrated in long-term follow-up (12,14,15), and only a smaller proportion of patients are found to make use of existing arthritis self-management programs (16,17). Previous research demonstrates that patients with IJD report varying needs of support in relation to managing the emotional, practical, and social consequences of their illness (18–21), with approximately two-thirds of patients indicating interest in self-management and support services and patient education (18,20). However, unmet needs regarding information, education, emotional support, and self-management strategies are reported (22–24), and such unmet health care needs are associated with poorer health-related quality of life (23).

To provide appropriate support for patients with IJDs, understanding their needs and preferences for the type of support, content, timing, and mode of delivery is important (10,11). Patients with IJD are not a homogenous group; therefore, exploring whether different subgroups of patients experience different needs and preferences is also important (10). By understanding whether differences in preferences for self-management and support services are driven by characteristics such as age, sex, diagnosis, or disease duration, enabling the development of patient-centered services tailored to the patients' needs and wishes, may be possible. Therefore, the aim of this study was to explore preferences for self-management and support services in patients with IJD and to determine whether these preferences differ by age, sex, diagnosis, and disease duration.

MATERIALS AND METHODS

Study design, setting, and participants. The study was conducted as a nationwide cross-sectional, descriptive, observational study. Data were collected through a questionnaire set up in the online survey program SurveyXact. Participants were recruited consecutively across Denmark from March through May 2016 (a total of 2 months). Invitations to participate in the study were accessible through informational material with a link and QR code to access the online questionnaire. The questionnaire could also be requested in paper format. The informational material was distributed through the Danish Rheumatism Organization, local arthritis networks/groups, diagnosis networks, and the rheumatology departments of hospitals across the country. Eligible participants were age ≥18 years and diagnosed with RA, PsA, or axial SpA. All eligible participants who completed the online questionnaire were included in the study.

Outcomes and instruments. Sociodemographic information included sex, age, marital status, educational level, and occupation. Disease and medication information collected included diagnosis, disease duration, current treatment with disease-modifying antirheumatic drugs, and the presence of comorbidities. Disease activity was measured using the Patient-based Disease Activity Score 2 (PDAS2) (25) in participants with RA, and the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) (26) in participants with axial SpA. Participants with PsA answered both the PDAS2 and BASDAI, given that PsA can manifest as both a peripheral and axial disease.

Physical functional status was assessed using the Health Assessment Questionnaire disability index (27), including 2 questions on more complex activities (i.e., walking 3 km and participation in recreation and sports) from the Multidimensional Health Assessment Questionnaire (MDHAQ) (28). Psychological functional status was assessed using the 3 psychological items (concerning sleep, anxiety, and depression) from the MDHAQ (28). Symptoms of pain and patient global assessment (i.e., evaluation of the overall impact of arthritis on everyday life) were assessed using visual analog scales (range 0–100), and fatigue severity, effect, and coping were assessed using the Bristol Rheumatoid Arthritis Fatigue numerical rating scales (29).

Data on preferences for self-management and support services were collected using a questionnaire developed in

the UK (30), where participants indicated whether or not they would be interested in participating in 30 different forms of self-management and support services, within 8 overall categories: discussion groups with other people who have IJD, 1-to-1 discussions about coping with IJD, question and answer (Q and A) sessions to ask questions about any aspects of IJD, organized talks or lectures, physical activity sessions, educational group sessions, events to increase public understanding of IJD, and online services (30). The questionnaire also contained questions on preferences regarding practical issues in relation to participation in self-management and support services (30), and preferred timing of support (18).

Parts of the questionnaire that were not already available in a Danish version (including PDAS2 and the questions about self-management and support preferences) were translated from English to Danish following guidelines for the translation and cross-cultural adaptation of patient-reported outcome measures (31). In addition, data on illness acceptance and coping strategies were collected and will be reported in a separate article.

Statistical analysis. Data were presented as means ± SDs or as medians with interquartile ranges and frequencies, where appropriate. Between-group differences according to diagnosis were tested using independent *t*-test/Kruskal-Wallis test for continuous data and chi-square/Fisher's exact test for categorical data, with the significance value set at *P* less than 0.05 (2-tailed). Preferences for self-management and support services were presented as the number and the proportion of participants indicating interest in participation in each individual service. Likewise, preferences regarding practical issues in relation to participation in the services were descriptively presented.

To test for differences in preferences for self-management and support services according to age (divided according to the population mean, i.e., age ≤50/>50 years), sex (male/female), diagnosis (RA/PsA/axial SpA), and disease duration (≤5/>5 years), chi-square/Fisher's exact test were applied. The significance value was set at *P* less than 0.05 (2-tailed). All analyses were performed using SPSS software, version 22.

Patient research partners and ethics. As recommended by the European Alliance of Associations for Rheumatology (32), 2 patients (KVJ [male] and LA [female]), both diagnosed with RA, were included as equal research partners in all relevant phases of the study. The study has been reported to the Danish Ethics Committee of the Capital Region (#16016528). The questionnaire was completed with full anonymity, without collecting any personally identifiable information. On the questionnaire's first page, detailed information about the study was provided, and the participants consented to participate by continuing past that page. The study was conducted in compliance with the Helsinki Declaration.

RESULTS

Characteristics. The characteristics of the participants are shown in Table 1. A total of 664 patients with a median disease duration of 10 years completed the questionnaire, of which 354 (53%) had RA, 180 (27%) had PsA, and 130 (20%) had axial SpA. In total, 565 of the participants (85%) were female, and the mean age was 50 years.

Preferences for self-management and support services. The overall most popular self-management and support services were online services for information about symptoms, treatment, and self-management of IJD (91%), 1-to-1 discussions with a rheumatologist (89%), nurse or physical therapist (both 83%), organized talks by researchers (83%), educational sessions to help with managing symptoms of IJD (80%), organized talks by lifestyle experts (77%), Q and A session with a rheumatologist (77%), and physical activity to improve fitness (77%) (Figure 1).

Preferences for self-management and support services according to age. Significantly more participants in the younger age group (≤50 years), compared with the older group (>50 years) were interested in participating in 1-to-1 discussions with a psychologist or counselor (72% versus 50% [$P < 0.001$]) or an experienced IJD patient (64% versus 55% [$P = 0.02$]), and in educational sessions about stress/anger (63% versus 46% [$P < 0.001$]) and symptom management (84% versus 77% [$P = 0.03$]), while significantly more of the older participants indicated interest in organized talks by research experts (87% versus 80% [$P = 0.03$]), when compared to the younger participants. Also, more of the younger participants were interested in online services and in taking part in events to increase public understanding of IJD (Table 2).

Preferences for self-management and support services according to sex. Overall, a greater proportion of women indicated interest in the various self-management and support services when compared to men, but only a few of these differences reached statistical significance (Table 2). Thus, significantly more women than men indicated interest in 1-to-1 discussions with various health professionals (physical therapist: 84% versus 74% [$P = 0.01$]; occupational therapist: 74% versus 57% [$P < 0.01$]; psychologist or counselor: 64% versus 47% [$P < 0.01$]), and in Q and A sessions with an occupational therapist (62% versus 48% [$P = 0.01$]). Furthermore, more women indicated interest in physical activity sessions to develop skills (74% versus 57% [$P < 0.01$]) and as an organized game (70% versus 57% [$P = 0.01$]), and more women were interested in online services for information about IJD (92% versus 85% [$P = 0.04$]) when compared to the men (Table 2).

Preferences for self-management and support services according to diagnosis. Overall, the participants with axial SpA tended to indicate more interest in the different services, while the participants with RA tended to indicate

Table 1. Characteristics of the study population*

Variable	All (n = 664)	RA (n = 354)	PsA (n = 180)	Axial SpA (n = 130)	*P*
Sociodemographic characteristic					
Female sex, no. (%)	565 (85.1)	324 (91.5)	149 (82.8)	92 (70.8)	<0.01
Age, mean ± SD years	50.0 ± 12.9	51.7 ± 13.0	50.1 ± 12.4	45.1 ± 12.2	<0.01
Marital status (n = 662), no. (%)†					
Married/civil partnership	365 (55.1)	190 (53.8)	110 (61.1)	65 (50.4)	NS
Partner, living together	100 (15.1)	45 (12.7)	26 (14.4)	29 (22.5)	–
Partner, not living together	29 (4.4)	17 (4.8)	5 (2.8)	7 (5.4)	–
Single	111 (16.8)	67 (19.0)	22 (12.2)	22 (17.1)	–
Divorced	40 (6.0)	23 (6.5)	13 (7.2)	4 (3.1)	–
Widowed	17 (2.6)	11 (3.1)	4 (2.2)	2 (1.6)	–
Education (n = 660), no. (%)†					
Lower secondary	58 (8.8)	29 (8.2)	20 (11.1)	9 (7.0)	NS
Higher secondary	289 (43.8)	154 (43.8)	84 (46.7)	51 (39.8)	–
Tertiary	313 (47.4)	169 (48.0)	76 (42.2)	68 (53.1)	–
Occupation (n = 656), no. (%)†					
Full-time work	176 (26.8)	102 (29.1)	45 (25.1)	29 (22.8)	0.01
Part-time work	155 (23.6)	85 (24.3)	38 (21.2)	32 (25.2)	–
Unemployed, due to arthritis	95 (14.5)	40 (11.4)	34 (19.0)	21 (16.5)	–
Unemployed, for other/unknown reasons	68 (10.4)	27 (7.7)	21 (11.7)	20 (15.7)	–
Retired	132 (20.1)	79 (22.6)	37 (20.7)	16 (12.6)	–
Student	30 (4.6)	17 (4.9)	4 (2.2)	9 (7.1)	–
Disease and medications					
Disease duration (n = 663), years†	10.0 (4.0–16.0)	10.0 (4.0–17.0)	9.0 (5.0–15.0)	10.0 (5.0–20.0)	NS
Biologic DMARD (yes), no. (%)	211 (31.8)	105 (29.7)	46 (25.6)	60 (46.2)	<0.01
Comorbidities (any), no. (%)	315 (47.4)	160 (45.2)	98 (54.4)	57 (43.8)	NS
Disease activity§					
PDAS2 (2.7–7.9) (RA + PsA, n = 535)	4.2 (3.5–5.0)	4.0 (3.3–4.9)	4.6 (3.8–5.2)	NA	<0.01
BASDAI (0–10) (axial SpA + PsA, n = 310)	5.8 (4.0–7.1)	NA	5.8 (3.9–7.2)	5.8 (4.2–6.9)	NS
Functional status§					
MDHAQ physical (0–10)	2.3 (1.0–4.0)	2.3 (0.7–4.0)	2.7 (1.3–4.3)	2.0 (1.3–3.7)	0.01
MDHAQ psychological (0–9.9)	3.3 (1.1–4.4)	2.2 (1.1–3.6)	3.3 (2.2–4.4)	3.3 (1.9–5.5)	<0.01
Symptoms: VAS scales and fatigue‡					
VAS pain (0–100)	50.0 (23.0–70.0)	38.5 (18.8–65.0)	58.0 (30.3–75.0)	60.0 (34.3–74.0)	<0.01
VAS global (0–100)	55.5 (28.0–75.0)	47.0 (20.0–69.3)	63.0 (36.0–77.0)	66.0 (34.5–78.0)	<0.01
BRAF-NRS severity (0–10)	7.0 (5.0–8.0)	7.0 (5.0–8.0)	7.0 (5.0–8.0)	7.0 (5.0–8.0)	NS
BRAF-NRS effect (0–10)	7.0 (4.0–8.0)	6.0 (4.0–8.0)	7.0 (5.0–8.0)	7.0 (5.0–8.0)	<0.01
BRAF-NRS ability to cope (0–10)	6.0 (4.0–8.0)	6.0 (4.0–8.0)	6.0 (4.0–8.0)	6.0 (4.0–8.0)	NS

* Values are the median (interquartile range) unless indicated otherwise. *P* values are from tests of differences between the 3 diagnoses (one-way analysis of variance/Kruskal-Wallis test for continuous data and chi-square/Fisher's exact test for categorical data). BASDAI = Bath Ankylosing Spondylitis Disease Activity Index; BRAF-NRS = Bristol Rheumatoid Arthritis Fatigue numerical rating scales; DMARD = disease-modifying antirheumatic drug; MDHAQ = Multidimensional Health Assessment Questionnaire; NA = not applicable; NS = nonsignificant difference; PDAS2 = Patient-based Disease Activity Score 2; PsA = psoriatic arthritis; RA = rheumatoid arthritis; SpA = spondyloarthritis; VAS = visual analog scale.
† The number is less than 664 due to missing data.
‡ A higher score indicates higher levels of disease activity (PDAS2 and BASDAI), higher levels of functional disability (MDHAQ physical) and psychological distress (MDHAQ psychological), higher levels of pain (VAS pain) and perceived overall impact of disease (VAS global), higher levels of fatigue severity (BRAF-NRS severity) and fatigue's effect on life (BRAF-NRS effect), and fewer problems in coping with fatigue (BRAF-NRS coping).

the least interest (Table 3). Statistically significant differences between the diagnoses were found with regard to discussion groups about experiences of IJD (RA: 54%, PsA: 62%, axial SpA: 68% [$P = 0.01$]), 1-to-1 discussions with a psychologist or counselor (RA: 58%, PsA: 62%, axial SpA: 70% [$P < 0.05$]) and experienced IJD patients (RA: 54%, PsA: 64%, axial SpA: 68% [$P < 0.01$]), Q and A sessions with experienced IJD patients (RA: 51%, PsA: 60%, axial SpA: 68% [$P < 0.01$]), educational sessions about managing stress and anger (RA: 49%, PsA: 61%, axial SpA: 64% [$P < 0.01$]), and online services to communicate with other patients about worries/frustrations (RA: 47%, PsA: 54%, axial SpA: 61% [$P = 0.03$]) and practical issues (RA: 60%, PsA: 67%, axial SpA: 72% [$P = 0.04$]) (Table 3).

Preferences for self-management and support services according to disease duration. Statistically significantly more participants with short disease duration (≤5 years) indicated interest in 1-to-1 discussions with a rheumatologist (92% versus 87% [$P = 0.04$]) or nurse (88% versus 80% [$P = 0.01$]), organized talks with experienced IJD patients (69% versus 59% [$P = 0.02$]), and online services to read other patients' stories (80%

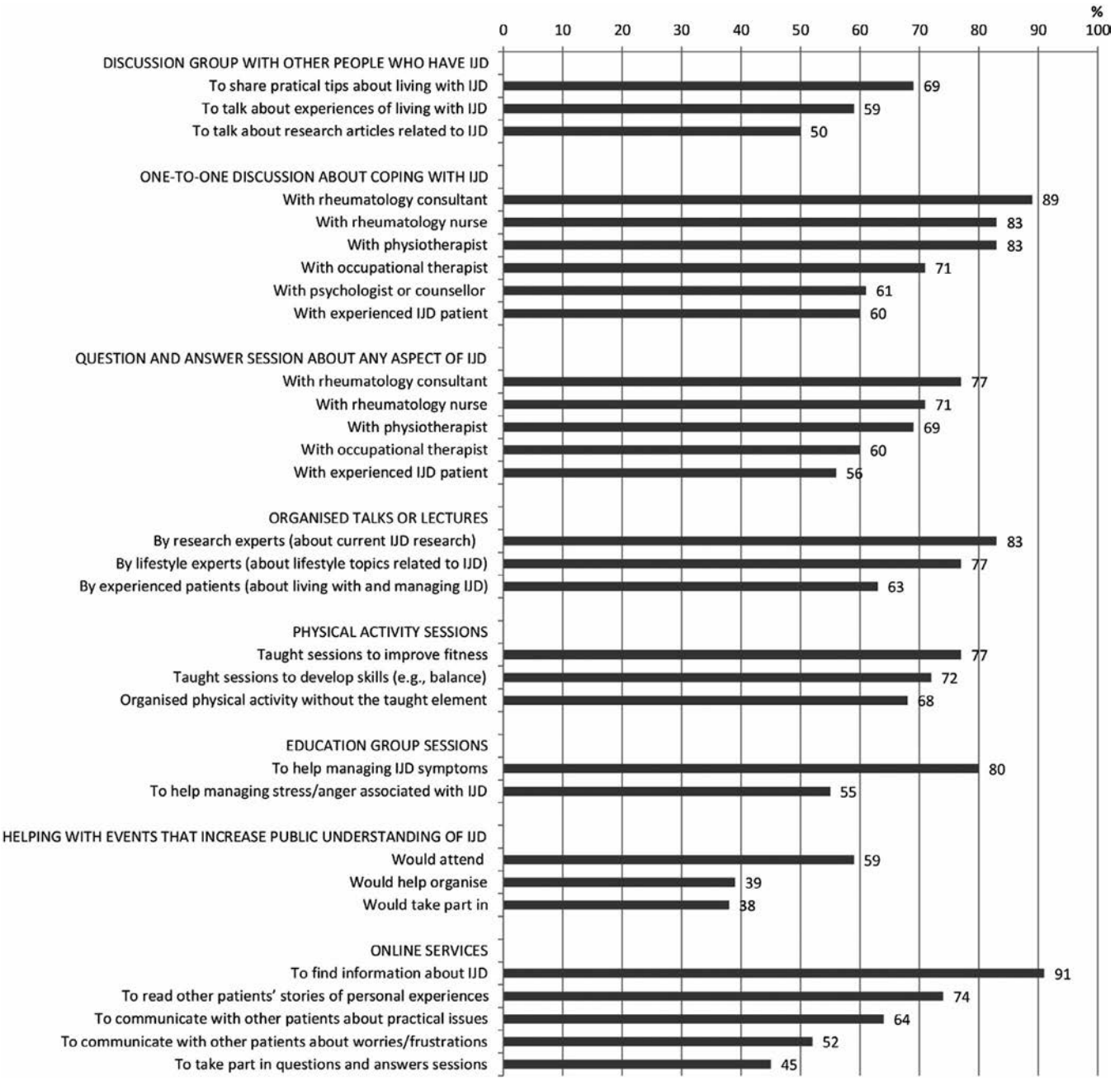


Figure 1. Preferences for self-management and support services. IJD = inflammatory joint disease.

versus 71% [$P < 0.01$]) and to communicate with other patients about worries or frustrations (58% versus 49% [$P = 0.04$]), when compared to the participants with a disease duration of >5 years (Table 3).

Preferences regarding practical issues. Most participants preferred the self-management and support services to be in the afternoon or evening and the vast majority (n = 536, 81%) preferred not to give a fixed commitment to the services, but for there to be an advertised timetable, allowing them to attend according to their available time and needs. The most frequently reported answers regarding timing of support were that the support preferably should be available "whenever needed" (n = 465, 70%), and "within the first 6 months of diagnosis" (n = 322, 48%). The most frequently reported answer regarding group sex mix was "no preferences" (n = 322, 48%). In total, 168 (25%) preferred the groups to be for people with IJD only, while the rest would either like their own friends or family to attend (n = 284, 43%), or reported that others could bring relatives (n = 212, 32%). The most frequently reported motivators, factors making it more likely to attend a self-management or support service, were being invited by the rheumatologist (74%) or nurse (71%) (Table 4).

Table 2. Preferences for self-management and support services according to age and sex*

Self-management and support services	All (n = 664)	Age ≤50 years (n = 334)	Age >50 years (n = 330)	Age *P*	Sex Men (n = 99)	Sex Women (n = 565)	Sex *P*
Discussion group with other people who have IJD							
To talk about experiences of living with IJD	392 (59)	208 (62)	184 (56)	NS	51 (52)	341 (60)	NS
To share practical tips about living with IJD	459 (69)	237 (71)	222 (67)	NS	63 (64)	396 (70)	NS
To talk about research articles related to IJD	335 (50)	163 (49)	172 (52)	NS	45 (45)	290 (51)	NS
One-to-one discussion about coping with IJD							
With rheumatology consultant	588 (89)[2]	296 (89)[2]	292 (88)[1]	NS	85 (86)[1]	503 (89)[2]	NS
With rheumatology nurse	552 (83)[4]	282 (84)[3]	270 (82)[4]	NS	84 (85)[2]	468 (83)[5]	NS
With physical therapist	549 (83)[5]	279 (84)[4]	270 (82)[4]	NS	73 (74)	476 (84)[3]	0.01
With occupational therapist	474 (71)	244 (73)	230 (70)	NS	56 (57)	418 (74)	<0.01
With psychologist or counselor	406 (61)	240 (72)	166 (50)	<0.001	47 (47)	359 (64)	<0.01
With experienced IJD patient	396 (60)	214 (64)	182 (55)	0.02	54 (55)	342 (61)	NS
Question and answer session about any aspect of IJD							
With rheumatology consultant	513 (77)	256 (77)	257 (78)	NS	75 (76)[4]	438 (78)	NS
With rheumatology nurse	470 (71)	240 (72)	230 (70)	NS	74 (75)	396 (70)	NS
With physical therapist	457 (69)	230 (69)	227 (69)	NS	65 (66)	392 (69)	NS
With occupational therapist	400 (60)	204 (61)	196 (59)	NS	48 (48)	352 (62)	0.01
With experienced IJD patient	375 (56)	201 (60)	174 (53)	NS	53 (54)	322 (57)	NS
Organized talks or lectures							
By lifestyle experts (about lifestyle topics related to IJD)	510 (77)	254 (76)	256 (78)	NS	75 (76)[4]	435 (77)	NS
By experienced patients (about living with and managing IJD)	415 (63)	214 (64)	201 (61)	NS	59 (60)	356 (63)	NS
By research experts (about current IJD research)	554 (83)[3]	268 (80)	286 (87)[3]	0.03	83 (84)[3]	471 (83)[4]	NS
Physical activity sessions							
Taught sessions to develop skills (e.g., balance)	475 (72)	230 (69)	245 (74)	NS	56 (57)	419 (74)	<0.01
Taught sessions to improve fitness	514 (77)	256 (77)	258 (78)	NS	71 (72)	443 (78)	NS
Organized physical activity without the taught element	450 (68)	223 (67)	225 (68)	NS	56 (57)	394 (70)	0.01
Education group sessions							
To help manage stress/anger associated with IJD	365 (55)	212 (63)	153 (46)	<0.001	50 (51)	315 (56)	NS
To help manage IJD symptoms	532 (80)	279 (84)[4]	253 (77)	0.03	74 (75)	458 (81)	NS
Helping with events that increase public understanding of IJD							
Would attend	391 (59)	204 (61)	187 (57)	NS	52 (53)	339 (60)	NS
Would take part in	251 (38)	143 (43)	108 (33)	<0.01	31 (31)	220 (39)	NS
Would help organize	259 (39)	133 (40)	126 (38)	NS	41 (41)	218 (39)	NS
Online services							
To find information about IJD†	603 (91)[1]	312 (93)[1]	291 (88)[2]	0.02	84 (85)[2]	519 (92)[1]	0.04
To read other patients' stories of personal experiences	490 (74)	257 (77)	233 (71)	NS	71 (72)	419 (74)	NS
To take part in question and answer sessions	302 (45)	170 (51)	132 (40)	<0.01	40 (40)	262 (46)	NS
To communicate with other patients about worries/frustrations	345 (52)	199 (60)	146 (44)	<0.001	43 (43)	302 (53)	NS
To communicate with other patients about practical issues	425 (64)	223 (67)	200 (61)	NS	55 (56)	370 (65)	NS

* Values are the number (%), with the ranking (1–5) of most popular self-management and support services within each group shown in brackets. *P* values are from tests of differences between groups (chi-square/Fisher's exact test). IJD = inflammatory joint disease; NS = nonsignificant difference.
† Includes information about symptoms, treatment, and self-management of IJD.

DISCUSSION

We found 1-to-1 discussions with health professionals to be especially popular among participants, which is unsurprising, because this practice reflects usual care. Our result supports previous research (33,34), including results from a UK study using the same support preferences questionnaire (30). However, the questionnaire did not ask participants whether they would like further 1-to-1 sessions in addition to usual care, so whether they

Table 3. Preferences for self-management and support services according to diagnosis and disease duration*

Self-management and support services	Diagnosis				Disease duration		
	RA (n = 354)	PsA (n = 180)	Axial SpA (n = 130)	*P*	≤5 years (n = 229)	>5 years (n = 434)	*P*
Discussion group with other people who have IJD							
To talk about experiences of living with IJD	192 (54)	111 (62)	89 (68)	0.01	145 (63)	247 (57)	NS
To share practical tips about living with IJD	235 (66)	126 (70)	98 (75)	NS	162 (71)	297 (68)	NS
To talk about research articles related to IJD	175 (49)	86 (48)	74 (57)	NS	114 (50)	221 (51)	NS
One-to-one discussion about coping with IJD							
With rheumatology consultant	311 (88)[2]	160 (89)[2]	117 (90)[1]	NS	211 (92)[2]	376 (87)[2]	0.04
With rheumatology nurse	286 (81)[5]	159 (88)[3]	107 (82)[3]	NS	202 (88)[3]	349 (80)[5]	0.01
With physical therapist	291 (82)[4]	152 (84)[5]	106 (82)[4]	NS	186 (81)	362 (83)[4]	NS
With occupational therapist	255 (72)	132 (73)	87 (67)	NS	164 (72)	309 (71)	NS
With psychologist or counselor	204 (58)	111 (62)	91 (70)	<0.05	145 (63)	261 (60)	NS
With experienced IJD patient	192 (54)	116 (64)	88 (68)	<0.01	148 (65)	247 (57)	NS
Question and answer session about any aspect of IJD							
With rheumatology consultant	264 (75)	143 (79)	106 (82)[4]	NS	184 (80)	328 (76)	NS
With rheumatology nurse	241 (68)	136 (76)	93 (72)	NS	172 (75)	297 (68)	NS
With physical therapist	236 (67)	130 (72)	91 (70)	NS	156 (68)	300 (69)	NS
With occupational therapist	210 (59)	117 (65)	73 (56)	NS	136 (59)	263 (61)	NS
With experienced IJD patient	179 (51)	108 (60)	88 (68)	<0.01	138 (60)	236 (54)	NS
Organized talks or lectures							
By lifestyle experts (about lifestyle topics related to IJD)	269 (76)	136 (76)	105 (81)	NS	178 (78)	332 (76)	NS
By experienced patients (about living with and managing IJD)	209 (59)	118 (66)	88 (68)	NS	157 (69)	257 (59)	0.02
By research experts (about current IJD research)	296 (84)[3]	149 (83)	109 (84)[2]	NS	188 (82)[4]	365 (84)[3]	NS
Physical activity sessions							
Taught sessions to develop skills (e.g., balance)	255 (72)	124 (69)	96 (74)	NS	159 (69)	315 (73)	NS
Taught sessions to improve fitness	268 (76)	144 (80)	102 (78)	NS	176 (77)	337 (78)	NS
Organized physical activity without the taught element	232 (66)	126 (70)	92 (71)	NS	160 (70)	290 (67)	NS
Education group sessions							
To help manage stress/anger associated with IJD	172 (49)	110 (61)	83 (64)	<0.01	132 (58)	233 (54)	NS
To help manage IJD symptoms	272 (77)	154 (86)[4]	106 (82)[4]	NS	187 (82)[5]	344 (79)	NS
Helping with events that increase public understanding of IJD							
Would attend	205 (58)	104 (58)	82 (63)	NS	126 (55)	265 (61)	NS
Would take part in	133 (38)	61 (34)	57 (44)	NS	87 (38)	164 (38)	NS
Would help organize	140 (40)	60 (33)	59 (45)	NS	98 (43)	161 (37)	NS
Online services							
To find information about IJD†	319 (90)[1]	167 (93)[1]	117 (90)[1]	NS	214 (93)[1]	389 (90)[1]	NS
To read other patients' stories of personal experiences	251 (71)	139 (77)	100 (77)	NS	183 (80)	306 (71)	<0.01
To take part in question and answer sessions	160 (45)	82 (46)	60 (46)	NS	109 (48)	193 (44)	NS
To communicate with other patients about worries/frustrations	168 (47)	98 (54)	79 (61)	0.03	132 (58)	213 (49)	0.04
To communicate with other patients about practical issues	212 (60)	120 (67)	93 (72)	0.04	154 (67)	271 (62)	NS

* Values are the number (%), with the ranking (1–5) of most popular self-management and support services within each group shown in brackets. *P* values are from tests of differences between groups (chi-square/Fisher's exact test). IJD = inflammatory joint disease; NS = nonsignificant difference; PsA = psoriatic arthritis; RA = rheumatoid arthritis; SpA = spondyloarthritis.
† Includes information about symptoms, treatment, and self-management of IJD.

would prefer more time with their rheumatology health professionals is unclear.

Participants also indicated a preference for services not delivered in standard hospital care, such as organized talks, educational sessions on symptom management, and online informational services. This finding supports a recent integrative review of the experiences and needs of patients with RA, which found that patients report receiving inadequate information from health professionals and would prefer more information on treatment, symptoms, and self-management strategies (24). Also, patients report receiving insufficient information from their health professionals on nonmedical issues, such as pain, flares, or psychosocial matters (34). Approaches to managing chronic illness are shifting from the traditional provider-patient relationship to a paradigm in which

Table 4. Preferences regarding practical issues in relation to self-management and support services (n = 664)*

Variable	Value
Time of day†	
Early morning (pre to 9:00 AM)	25 (4)
Morning (9:00 AM to midday)	184 (28)
Lunchtime (midday to 2:00 PM)	158 (24)
Afternoon (2:00 to 5:00 PM)	245 (37)
Evening (after 5:00 PM)	290 (44)
Frequency of group†	
Single 1-time group	66 (10)
Fixed time period	187 (28)
No fixed commitment, advertised time table	536 (81)
Group sex mix†	
Own sex only	47 (7)
Mixed: equal number of men/women	131 (20)
Mixed: more of own sex	21 (3)
Mixed: opposite sex could outnumber own	259 (39)
No preference at all	322 (48)
Other people	
A service for people with IJD only	168 (25)
Would like to invite a friend/family member	284 (43)
Would not bring friend/family, but others could	212 (32)
Motivators†	
An appointment letter	303 (46)
Invitation from rheumatologist	493 (74)
Invitation from rheumatology nurse	472 (71)
Reimbursement of travel costs	286 (43)
Money or vouchers for attendance	94 (14)
Location away from the hospital	272 (41)
Timing of support†	
Within 6 months of being diagnosed	322 (48)
First 6–12 months after diagnosis	161 (24)
Between 12 and 18 months after diagnosis	77 (12)
During a flare	197 (30)
Whenever I feel I need it	465 (70)
Would not use at any time	38 (6)
Other	18 (3)

* Values are the number (%). IJD = inflammatory joint disease.
† Multiple answers possible.

individuals with chronic conditions play a key role in guiding their care, in partnership with health care providers (35). Therefore, patients might experience unmet needs beyond what is already offered in usual care, whether that is different approaches with a more equal power balance with their health care providers, or different services not delivered by health professionals, or as part of usual care. All of this information should be considered in the future planning and development of self-management and support services for patients with IJDs. However, based on our findings, questions regarding the optimal content and setting for such services remain unanswered, and thus further research that investigates these issues is warranted.

Our findings suggest that younger, as opposed to older, patients wanted more support on psychological and self-management issues. Qualitative research in patients with RA has underlined how younger patients express worries about their future, including family life, ability to cope, and identity as an independent person (36). Furthermore, younger age is significantly associated with feelings of guilt and shame in people with RA (37). In addition, the younger participants in our study were significantly more interested in online services, which supports previous research in chronic diseases that suggests younger people are more likely to use the Internet for health-related information seeking (38,39), which might be due to technological development, whereby younger generations might be more accustomed to online services. While the younger participants indicated greater interest in most of the self-management and support services, the older patients showed a greater interest in the more factual, organized talks with research experts. However, despite the abovementioned differences in preferences, the top 5 most popular services did not differ much between the 2 age groups.

In the current study, female participants indicated greater interest in the vast majority of self-management and support services, supporting findings using the same support preferences questionnaire with RA patients in the UK (30). However, although not statistically significant, men were more interested than women in 1-to-1 or Q and A sessions with a rheumatology nurse, organized talks with research experts, or helping to organize events to raise awareness of IJD. This result reflects previous research in RA and other long-term conditions, which suggests that support services for men should have a practical focus (40,41) and provide opportunities to gain new information (42,43).

Evidence from the obesity literature suggests that male-only groups differ qualitatively from mixed-sex groups, with different levels of engagement and styles of language (44,45). Furthermore, a study comparing sex interactions in online forums for breast cancer (aimed at women) and prostate cancer (aimed at men) found that quantitatively, women dominated both forums. Qualitatively, while the men made attempts to adapt their communication to the norms of the opposite sex, the women did not (46). Thus, although only 7% of the participants in our study reported preferring the support groups to include their own sex only, providing men with IJDs with an all-male intervention may be important, to enable them to engage according to masculine norms.

It has been suggested that diagnosis is not associated with support and health care needs when comparing RA with PsA (20) and ankylosing spondylitis (23). However, in our study, we found several significant differences in support needs between the diagnosis groups, with the axial SpA patients overall tending to indicate the greatest interest in support services and RA patients the least. This finding could possibly reflect the fact that patients with RA to a greater extent already feel that they have access to necessary support, because they are a larger patient group (1), and until now more research has been conducted on their needs than on the needs of patients with other IJDs (10,13). Participants with axial SpA more frequently reported a preference for sharing experiences, talking about emotions with another patient and 1-to-1 sessions with a psychologist or counselor. One possible explanation for this preference could be that many patients with axial SpA have had a "difficult journey" due to delays in reaching a diagnosis

(47). Even though such delays have decreased significantly over recent years (48), participants with axial SpA in our study might still have experienced a delay in diagnosis, potentially causing worry and frustration (47), which might lead to a greater need for emotional support.

In our study, newly diagnosed participants indicated a greater interest in most services, with several differences reaching statistical significance, compared to patients with >5 years since diagnosis. Especially the 1-to-1 discussions with rheumatology health professionals and services based on talks and sharing of information and experiences with other patients were preferred among newly diagnosed participants. Additionally, almost half our participants indicated that support should be provided in the first 6 months of diagnosis. These findings support a recent meta-analysis in RA suggesting that psychological support interventions may be more effective for patients with a shorter illness duration (7). Possibly patients are more likely to see themselves as "expert patients" (49,50) over time, and they thus might experience a decreasing need for information and advice with longer disease duration. However, although participants with shorter disease duration tended to show greater interest in most of the self-management and support services when compared to those with longer disease duration, the majority of the latter also indicated interest in most of the services. This finding indicates that patients continue to have support needs even after 5 years of diagnosis.

Overall, there tended to be a preference for flexibility in the timing and delivery of support and self-management services among the patients. In total, 70% of the patients indicated that such services should be available whenever they needed them. Furthermore, 81% of the patients indicated a preference for flexibility in the delivery of the services, preferring services to be offered that they could access as required, rather than committing to a fixed program. Such findings call for considering a greater extent of customization regarding content, timing, and delivery of services that should preferably be tailored to the individual's fluctuating needs and wishes for flexibility.

Notably, our findings revealed that especially the younger participants with shorter disease duration and participants with axial SpA seemed interested in getting tips and sharing experiences with other patients with IJD. A need may exist to look beyond the health professional and patient organizations to deliver self-management and support services and to further investigate the potential need to also offer expert patients/mentors and peer-to-peer services. However, even given that our findings revealed several statistically significant differences in preferences between the subgroups of participants (i.e., according to age, sex, diagnosis, and disease duration), these findings should be interpreted with caution due to the multiple comparisons performed and the resulting risk of type 1 errors. Furthermore, the most popular services differed only a little between the groups, and the findings indicated a need for support across all the subgroups. Therefore, our findings suggest that the different subgroups share preferences for the same self-management and support services, but whether such services should preferably be divided according to age, sex, diagnosis, or disease duration lies beyond the scope of this study and merits further exploration.

Strengths of the current study include the large sample of a nationwide and broad range of patients with IJD, covering 3 of the major IJD diagnosis groups. Furthermore, the inclusion of patient research partners throughout all relevant phases of the study has greatly contributed to ensuring the relevance of the study and its results to patients. However, there are several limitations that need to be considered in the interpretation and generalizability of the study's findings. First, the use of self-reported outcome measures comprises a risk for response biases, including potential social desirability bias and misclassification of diagnoses. The questionnaire we used to collect data on preferences for self-management and support services was originally developed in male patients with RA. Our findings that a great proportion of women reported interest in the various services might indicate that the included self-management and support services also cover the needs of the female patients. However, women's needs and preferences may be insufficiently covered by the questionnaire. Importantly, our recruitment, which was primarily through the Danish Rheumatism Organization and online, is likely to have led to a highly selected sample consisting of mainly the more resourceful patients and those with an overall greater interest in self-management and support services.

Notably, most of our participants were well educated and in a relationship, and only a small proportion of men (15%) took part in the study. Therefore, in our sample, we might have missed the patients most in need of support. All of the abovementioned has an impact on the representativeness of our sample and thereby also generalizability of the findings to other patients with IJD.

In conclusion, patients with IJD report various needs regarding self-management and support services. Although these preferences were found to differ across age, sex, diagnosis, and disease duration, all subgroups indicated a great need for support, and the top preferences did not differ much across the groups. This is the first step in exploring the needs and preferences for self-management and support services among Danish patients with IJD. However, in terms of targeted planning and development of both existing and future self-management and support services for patients with IJD, further research is warranted to explore the optimal timing, content, and setting for these services, as well as the potential benefits of services targeting specific subgroups of patients.

ACKNOWLEDGMENTS

The authors thank all of the participating patients. The authors also thank Dr. Emma Dures, University of the West of England, Bristol, UK for allowing us to use her question regarding preferred timing of support. Furthermore, we also thank the Danish Rheumatism Organization and the rheumatology departments across the country for the invaluable help in recruiting patients.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. Esbensen 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. Hammer, Flurey, Jensen, Andersen, Esbensen.

Acquisition of data. Hammer, Esbensen.

Analysis and interpretation of data. Hammer, Esbensen.

REFERENCES

1. Bergman MJ. Social and economic impact of inflammatory arthritis. Postgrad Med 2006; Spec No: 5–11.
2. Treharne GJ, Lyons AC, Booth DA, Kitas GD. Psychological well-being across 1 year with rheumatoid arthritis: coping resources as buffers of perceived stress. Br J Health Psychol 2007;12: 323–45.
3. Sheehy C, Murphy E, Barry M. Depression in rheumatoid arthritis: underscoring the problem. Rheumatology (Oxford) 2006;45:1325–7.
4. Vriezekolk J, Eijsbouts A, Evers A, Stenger A, Van Den Hoogen F, van Lankveld W. Poor psychological health status among patients with inflammatory rheumatic diseases and osteoarthritis in multidisciplinary rehabilitation: need for a routine psychological assessment. Disabil Rehabil 2010;32:836–44.
5. Covic T, Tyson G, Spencer D, Howe G. Depression in rheumatoid arthritis patients: demographic, clinical, and psychological predictors. J Psychosom Res 2006;60:469–76.
6. Hider SL, Tanveer W, Brownfield A, Mattey DL, Packham JC. Depression in RA patients treated with anti-TNF is common and under-recognized in the rheumatology clinic. Rheumatology (Oxford) 2009;48:1152–4.
7. Astin JA, Beckner W, Soeken K, Hochberg MC, Berman B. Psychological interventions for rheumatoid arthritis: a meta-analysis of randomized controlled trials. Arthritis Rheum 2002;47:291–302.
8. Sharpe L, Allard S, Sensky T. Five-year followup of a cognitive–behavioral intervention for patients with recently-diagnosed rheumatoid arthritis: effects on health care utilization. Arthritis Rheum 2008;59:311–6.
9. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48:177–87.
10. Zangi HA, Ndosi M, Adams J, Andersen L, Bode C, Boström C, et al. EULAR recommendations for patient education for people with inflammatory arthritis. Ann Rheum Dis 2015;74:954–62.
11. Fautrel B, Pham T, Gossec L, Combe B, Flipo RM, Goupille P, et al. Role and modalities of information and education in the management of patients with rheumatoid arthritis: development of recommendations for clinical practice based on published evidence and expert opinion. Joint Bone Spine 2005;72:163–70.
12. Riemsma RP, Kirwan JR, Taal E, Rasker JJ. Patient education for adults with rheumatoid arthritis. Cochrane Database Syst Rev 2003;2:CD003688.
13. Iversen MD, Hammond A, Betteridge N. Self-management of rheumatic diseases: state of the art and future perspectives. Ann Rheum Dis 2010;69:955–63.
14. Niedermann K, Fransen J, Knols R, Uebelhart D. Gap between short- and long-term effects of patient education in rheumatoid arthritis patients: a systematic review. Arthritis Rheum 2004;51:388–98.
15. Albano MG, Giraudet-Le Quintrec JS, Crozet C, d'Ivernois JF. Characteristics and development of therapeutic patient education in rheumatoid arthritis: analysis of the 2003–2008 literature. Joint Bone Spine 2010;77:405–10.
16. Centers for Disease Control and Prevention. Monitoring progress in arthritis management: United States and 25 states, 2003. MMWR Morb Mortal Wkly Rep 2005;54:484–8.
17. Koehn CL, Esdaile JM. Patient education and self-management of musculoskeletal diseases. Best Pract Res Clin Rheumatol 2008;22:395–405.
18. Dures E, Almeida C, Caesley J, Peterson A, Ambler N, Morris M, et al. Patient preferences for psychological support in inflammatory arthritis: a multicentre survey. Ann Rheum Dis 2016;75: 142–7.
19. Radford S, Carr M, Hehir M, Davis B, Robertson L, Cockshott Z, et al. "It's quite hard to grasp the enormity of it": perceived needs of people upon diagnosis of rheumatoid arthritis. Musculoskeletal Care 2008;6:155–67.
20. Drăgoi RG, Ndosi M, Sadlonova M, Hill J, Duer M, Graninger W, et al. Patient education, disease activity and physical function: can we be more targeted? A cross sectional study among people with rheumatoid arthritis, psoriatic arthritis and hand osteoarthritis. Arthritis Res Ther 2013;15:R156.
21. Zuidema RM, Repping-Wuts H, Evers AW, Van Gaal BG, Van Achterberg T. What do we know about rheumatoid arthritis patients' support needs for self-management? A scoping review. Int J Nurs Stud 2015;52:1617–24.
22. Leung YY, Tam LS, Lee KW, Leung MH, Kun EW, Li EK. Involvement, satisfaction and unmet health care needs in patients with psoriatic arthritis. Rheumatology (Oxford) 2009;48:53–6.
23. Kjeken I, Dagfinrud H, Mowinckel P, Uhlig T, Kvien TK, Finset A. Rheumatology care: involvement in medical decisions, received information, satisfaction with care, and unmet health care needs in patients with rheumatoid arthritis and ankylosing spondylitis. Arthritis Rheum 2006;55:394–401.
24. Poh LW, He HG, Lee CS, Cheung PP, Chan WC. An integrative review of experiences of patients with rheumatoid arthritis. Int Nurs Rev 2015;62:231–47.
25. Leung AM, Farewell D, Lau CS, Choy EH. Defining criteria for rheumatoid arthritis patient-derived disease activity score that correspond to Disease Activity Score 28 and Clinical Disease Activity Index based disease states and response criteria. Rheumatology (Oxford) 2016;55:1954–8.
26. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol 1994;21:2286–91.
27. Bruce B, Fries JF. The Health Assessment Questionnaire (HAQ). Clin Exp Rheumatol 2005;23:S14–8.
28. Pincus T, Yazici Y, Bergman M. Development of a multi-dimensional health assessment questionnaire (MDHAQ) for the infrastructure of standard clinical care. Clin Exp Rheumatol 2005;23:S19–28.
29. Nicklin J, Cramp F, Kirwan J, Greenwood R, Urban M, Hewlett S. Measuring fatigue in rheumatoid arthritis: a cross-sectional study to evaluate the Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional questionnaire, visual analog scales, and numerical rating scales. Arthritis Care Res (Hoboken) 2010;62:1559–68.
30. Flurey CA, Hewlett S, Rodham K, White A, Noddings R, Kirwan JR. Coping strategies, psychological impact, and support preferences of men with rheumatoid arthritis: a multicenter survey. Arthritis Care Res (Hoboken) 2018;70:851–60.
31. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health 2005;8:94–104.

32. De Wit MP, Berlo SE, Aanerud GJ, Aletaha D, Bijlsma JW, Croucher L, et al. European League Against Rheumatism recommendations for the inclusion of patient representatives in scientific projects. Ann Rheum Dis 2011;70:722–6.
33. Cunha-Miranda L, Costa L, Ribeiro JS. NEAR study: Needs and Expectations in Rheumatoid ARthritis. Do we know our patients needs? Acta Reumatol Port 2010;35:314–23.
34. Barlow JH, Cullen LA, Rowe IF. Educational preferences, psychological well-being and self-efficacy among people with rheumatoid arthritis. Patient Educ Couns 2002;46:11–9.
35. Grady PA, Gough LL. Self-management: a comprehensive approach to management of chronic conditions. Am J Public Health 2014;104:e25–31.
36. Lempp H, Scott D, Kingsley G. The personal impact of rheumatoid arthritis on patients' identity: a qualitative study. Chronic Illn 2006;2:109–20.
37. Ten Klooster PM, Christenhusz LC, Taal E, Eggelmeijer F, van Woerkom JM, Rasker JJ. Feelings of guilt and shame in patients with rheumatoid arthritis. Clin Rheumatol 2014;33:903–10.
38. Ayers SL, Kronenfeld JJ. Chronic illness and health-seeking information on the Internet. Health (London) 2007;11:327–47.
39. Oh YS, Cho Y. Examining the relationships between resources and online health information seeking among patients with chronic diseases and healthy people. Soc Work Health Care 2015;54:83–100.
40. Flurey CA, Hewlett S, Rodham K, White A, Noddings R, Kirwan JR. "You obviously just have to put on a brave face": a qualitative study of the experiences and coping styles of men with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2017;69:330–7.
41. Galdas P, Darwin Z, Kidd L, Blickem C, McPherson K, Hunt K, et al. The accessibility and acceptability of self-management support interventions for men with long term conditions: a systematic review and meta-synthesis of qualitative studies. BMC Public Health 2014;14:1230.
42. Seymour-Smith S. "Blokes don't like that sort of thing": men's negotiation of a "troubled" self-help group identity. J Health Psychol 2008;13:785–97.
43. Bell K, Lee J, Foran S, Kwong S, Christopherson J. Is there an "ideal cancer" support group? Key findings from a qualitative study of three groups. J Psychosoc Oncol 2010;28:432–49.
44. Young MD, Morgan PJ, Plotnikoff RC, Callister R, Collins CE. Effectiveness of male-only weight loss and weight loss maintenance interventions: a systematic review with meta-analysis. Obes Rev 2012;13:393–408.
45. Robertson C, Archibald D, Avenell A, Douglas F, Hoddinott P, van Teijlingen E, et al. Systematic reviews of and integrated report on the quantitative, qualitative and economic evidence base for the management of obesity in men. Health Technol Assess 2014;18:v–vi.
46. Seale C. Gender accommodation in online cancer support groups. Health (London) 2006;10:345–60.
47. Madsen M, Jensen KV, Esbensen BA. Men's experiences of living with ankylosing spondylitis: a qualitative study. Musculoskeletal Care 2015;13:31–41.
48. Sørensen J, Hetland ML. Diagnostic delay in patients with rheumatoid arthritis, psoriatic arthritis and ankylosing spondylitis: results from the Danish nationwide DANBIO registry. Ann Rheum Dis 2015;74:e12.
49. Donaldson L. Expert patients usher in a new era of opportunity for the NHS. BMJ 2003;326:1279–80.
50. Boulet LP. The expert patient and chronic respiratory diseases. Can Respir J 2016;2016:9454506.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 73, No. 10, October 2021, pp 1490–1501
DOI 10.1002/acr.24520

Development of a New International Antiphospholipid Syndrome Classification Criteria Phase I/II Report: Generation and Reduction of Candidate Criteria

Medha Barbhaiya,[1] Stephane Zuily,[2] Yasaman Ahmadzadeh,[3] Mary-Carmen Amigo,[4] Tadej Avcin,[5] Maria Laura Bertolaccini,[6] D. Ware Branch,[7] Guilherme de Jesus,[8] Katrien M. J. Devreese,[9] Camille Frances,[10] David Garcia,[11] Francis Guillemin,[12] Steven R. Levine,[13] Roger A. Levy,[14] Michael D. Lockshin,[1] Thomas L. Ortel,[15] Surya V. Seshan,[16] Maria Tektonidou,[17] Denis Wahl,[2] Rohan Willis,[18] Ray Naden,[†] Karen Costenbader,[19] and Doruk Erkan,[1] on behalf of the New APS Classification Criteria Collaborators

Objective. An international multidisciplinary initiative, jointly supported by the American College of Rheumatology and European Alliance of Associations for Rheumatology, is underway to develop new rigorous classification criteria to identify patients with high likelihood of antiphospholipid syndrome (APS) for research purposes. The present study was undertaken to apply an evidence- and consensus-based approach to identify candidate criteria and develop a hierarchical organization of criteria within domains.

Methods. During phase I, the APS classification criteria steering committee used systematic literature reviews and surveys of international APS physician scientists to generate a comprehensive list of items related to APS. In phase II, we reviewed the literature, administered surveys, formed domain subcommittees, and used Delphi exercises and nominal group technique to reduce potential APS candidate criteria. Candidate criteria were hierarchically organized into clinical and laboratory domains.

Results. Phase I generated 152 candidate criteria, expanded to 261 items with the addition of subgroups and candidate criteria with potential negative weights. Using iterative item reduction techniques in phase II, we initially reduced these items to 64 potential candidate criteria organized into 10 clinical and laboratory domains. Subsequent item reduction methods resulted in 27 candidate criteria, hierarchically organized into 6 additive domains (laboratory, macrovascular, microvascular, obstetric, cardiac, and hematologic) for APS classification.

Conclusion. Using data- and consensus-driven methodology, we identified 27 APS candidate criteria in 6 clinical or laboratory domains. In the next phase, the proposed candidate criteria will be used for real-world case collection and further refined, organized, and weighted to determine an aggregate score and threshold for APS classification.

Supported jointly by the American College of Rheumatology and the European Alliance of Associations for Rheumatology. Dr. Barbhaiya's work was supported by the Rheumatology Research Foundation (Investigator award). Dr. Bertolaccini's work was supported by King's College London British Heart Foundation (Centre for Excellence award RE/18/2/34213). Dr. Costenbader's work was supported by the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases grant K24-AR-066109).

[1]Medha Barbhaiya, MD, MPH, Michael D. Lockshin, MD, Doruk Erkan, MD, MPH: Barbara Volcker Center for Women and Rheumatic Diseases, Hospital for Special Surgery, Weill Cornell Medicine, New York, New York; [2]Stephane Zuily, MD, MPH, PhD, Denis Wahl, MD, PhD: Vascular Medicine Division and Regional Competence Center for Rare Auto-Immune Diseases, Université de Lorraine, Inserm, DCAC, and CHRU-Nancy, Nancy, France; [3]Yasaman Ahmadzadeh, MD: Hospital for Special Surgery, New York, New York; [4]Mary-Carmen Amigo, MD, FACP: ABC Medical Center, Mexico City, Mexico; [5]Tadej Avcin, MD, PhD: Children's Hospital, University Medical Center, University of Ljubljana, Ljubljana, Slovenia; [6]Maria Laura Bertolaccini, PhD: King's College London British Heart Foundation Centre of Excellence, London, UK; [7]D. Ware Branch, MD: University of Utah Health, Salt Lake City; [8]Guilherme de Jesus, MD, MSc: Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil; [9]Katrien M. J. Devreese, MD, PhD: Ghent University Hospital, Ghent University, Ghent, Belgium; [10]Camille Frances, MD: Tenon Hospital, Paris, France; [11]David Garcia, MD: University of Washington, Seattle; [12]Francis Guillemin, MD, PhD: CIC Clinical Epidemiology, APEMAC and CHRU, Inserm, Université de Lorraine, Nancy, France; [13]Steven R. Levine, MD: Downstate Stroke Center, State University of New York Downstate Health Sciences University, Kings County Hospital Center, and Maimonides Medical Center/Jaffe Stroke Center, Brooklyn, New York; [14]Roger A. Levy, MD, PhD: Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil, and GlaxoSmithKline, Upper Providence, Pennsylvania; [15]Thomas L. Ortel, MD, PhD: Duke University Medical Center, Durham, North Carolina; [16]Surya V. Seshan, MBBS: Weill Cornell Medicine, New York, New York; [17]Maria Tektonidou, MD, PhD: University of Athens, Athens, Greece; [18]Rohan Willis, MD: University of Texas Medical Branch, Galveston; [19]Karen Costenbader, MD, MPH: Brigham and Women's Hospital, Boston, Massachusetts.

[†]Dr. Naden is deceased.

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

Address correspondence to Medha Barbhaiya, MD, MPH, Barbara Volcker Center for Women and Rheumatic Diseases, Hospital for Special Surgery, 535 East 70th Street, New York, NY 10021. Email: barbhaiyam@hss.edu.

Submitted for publication July 7, 2020; accepted in revised form November 24, 2020.

SIGNIFICANCE & INNOVATIONS

- During the item generation and item reduction phases of an international multidisciplinary initiative that is underway to develop new rigorous classification criteria to identify patients with high likelihood of antiphospholipid syndrome (APS), we generated 261 items, which were reduced to 27 potential APS candidate criteria.
- Using data- and consensus-driven methodology, we identified 27 APS candidate criteria organized into 6 domains (laboratory, macrovascular, microvascular, obstetric, cardiac, and hematologic), which will be further refined, organized, and weighted to determine an aggregate score and threshold for APS classification.
- This work demonstrates a potential role for clinical components not previously included in the APS Sapporo classification criteria.

INTRODUCTION

Antiphospholipid syndrome (APS) is a systemic autoimmune disease characterized by thrombosis and/or pregnancy morbidity in patients with persistent antiphospholipid antibodies (aPLs). Classification of APS for clinical trials and studies is currently based on clinical and laboratory criteria described in the Sapporo classification criteria published in 1999 (1), validated in 2000 (2), and revised in 2006 (3), now known as the revised Sapporo APS classification criteria (or Sydney criteria).

The original Sapporo APS classification criteria included clinical (vascular thrombosis or pregnancy morbidity) and laboratory items (persistent lupus anticoagulant [LAC] test and/or IgG/M anticardiolipin antibody [aCL] positivity with at least 2 tests performed at least 6 weeks apart) (1). In the revised criteria (3), modifications were made including IgG/M anti–β_2-glycoprotein I (anti-β_2GPI) antibodies as a new laboratory test, wider time interval between serologic testing (at least 12 weeks instead of 6 weeks), clarification of a time interval between serology and clinical manifestations (maximum of 5 years), and specification of laboratory assay titer threshold and non-criteria aPL-manifestation definitions. Over the past decade, substantial evidence has accumulated describing additional clinical and laboratory manifestations associated with aPLs (4–7). Additionally, new methodologically rigorous and data-driven approaches to address biases and develop a robust set of classification criteria have been published (8–12).

In 2016, the 15th International Congress on Antiphospholipid Antibodies task force on APS classification conducted a needs assessment survey of international physicians with expertise in APS. Respondents indicated the need for the following components in a new APS classification criteria system: 1) capturing the full spectrum of clinical and laboratory manifestations of disease; 2) distinguishing APS from other comorbidities; 3) weighting certain clinical factors more than others; and 4) including a strong evidence basis for definitions of aPL positivity or pregnancy morbidity.

An international effort was thus initiated with the overall goal of developing a new APS classification system, which is jointly supported by the American College of Rheumatology (ACR) and the European Alliance of Associations for Rheumatology (EULAR). The objective was to develop new APS classification criteria that will demonstrate excellent face, discriminant, and construct validity. We aimed to incorporate the use of bias reduction strategies to arrive at a system with the highest sensitivity, specificity, and positive predictive value for the likelihood of APS against the gold standard of expert consensus while retaining face validity for classification. Here, we summarize the overall methodology and results from the first 2 phases of our 4-phase APS classification criteria development initiative.

MATERIALS AND METHODS

The APS classification criteria development involves a 4-phase methodology, as previously used in other classification systems of rheumatic disease. Details of the overall methodology, as well as the APS steering committee development and experts and stakeholders involved in phase I/II, are provided in Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24520/abstract.

Phase I item generation. During phase I, the APS classification criteria steering committee used surveys and literature review to generate all potential candidate criteria associated with APS. First, we asked 54 collaborators in an emailed survey to list all features that, in their experience, occur as part of the aPL/APS laboratory and clinical spectrum. The purpose of this question was to help identify potential candidate criteria with positive

Table 1. Phase I/II methodology of the new antiphospholipid syndrome (APS) classification criteria development

Phase I: Item generation
Part A
Item generation survey with open-ended questions (54 collaborators)
Part B
Item expansion to incorporate negatively weighted responses and APS subgroups (20 steering committee members)
Literature screening for thrombosis risk factors and additional criteria not identified by survey responses
Phase II: Item reduction
Part A
Item reduction survey A with Likert scale (61 collaborators) (low-specificity items [Likert score <1] eliminated)
Systematic literature reviews and meta-analyses
Part B
Item reduction survey B (19 steering committee members) (low-specificity items [Likert score <2] eliminated)
Systematic literature reviews and meta-analyses

weight. Respondents were encouraged to consider their real-life experience with aPL-positive patients rather than focus only on the current APS classification criteria. Responses were systematically clustered by organ system to avoid duplication and improve interpretability. Second, we asked participants to provide potential criteria that, if present, would lead a physician to question the diagnosis of APS, i.e., candidate criteria with negative weight, whose presence points away from APS but does not rule it out. These potential negatively weighted items were also organized into separate categories. Third, we asked whether physicians consider patients with APS in different subpopulations (Tables 1 and 2). Three of the core planning group members (MB, SZ, DE) additionally compiled and screened an article reference library (n = 26) including mostly recent APS-related systematic reviews and meta-analyses (see Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24520/abstract) to include additional items related to aPL/APS.

We analyzed the comprehensive list of potential aPL-associated manifestations as follows: first, we clustered the generated candidate criteria into meaningful categories by organ system; second, each potential criterion included in the list was checked for reliability and precision in its measurement; and third, similar or highly correlated criteria were eliminated to avoid redundant data.

We then used the following strategies to further expand the list of potential criteria: 1) for items that were submitted with subcategories during the phase I survey (e.g., age below or above 55 years; stroke with or without hypertension; aCL titer < or >40 units), we expanded the list to consider all potential items within subgroups individually; and 2) the steering committee systematically reviewed the results of the additional phase I survey designed to identify candidate criteria with potential negative weight and aPL/APS subgroups.

Phase II item reduction. Item reduction was guided by the following principles: 1) the criteria remaining after phase II should demonstrate good face, construct, and discriminant validity; 2) items with low sensitivity or specificity, poor reliability, redundancy, or insufficient feasibility should be removed; and 3) items should be organized into separate domains. This phase was also an iterative process guided by literature reviews, surveys, Delphi exercises, steering committee communications (email, teleconference, in-person meetings), and the development of domain subcommittees.

Literature reviews. To aid in their decision-making, the core planning group provided phase II survey respondents and steering committee members with the article reference library (n = 26) (see Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24520/abstract). Additionally, 5 different teams (under the guidance of SZ and DE) worked on meta-analyses of selected other manifestations of aPL, i.e., livedo, thrombocytopenia, hemolytic anemia, aPL nephropathy, and cardiac valve disease. These meta-analyses, together with others, helped guide the steering committee and collaborators during the development efforts of the new classification criteria. To better clarify arterial and venous thrombosis risk factors, the core planning group evaluated major recent international thrombosis and cardiovascular disease (CVD) guidelines and conferred with North American and European cardiologists. In addition, our pathology domain subcommittee (renal pathologist [SVS] and rheumatologists [MB, MT, DE]) and vascular medicine specialist (SZ), in collaboration with a nephrologist, have been leading an effort to better characterize the definition of aPL-associated nephropathy, a poorly defined but specific and clinically relevant pathologic manifestation related to APS.

Item reduction surveys A/B and Delphi exercises. We administered 2 surveys for item reduction in phase II (A and B), summarized in Table 2. In survey A, 61 collaborators and selected steering committee members were emailed a survey to assess each item generated in phase I based on the specificity of each feature in differentiating APS from other similar conditions using a Likert scale (range –5 to +5; where –5 = extremely strongly against APS, i.e., more likely related to another disease entity; 0 = not for or against APS; and +5 = extremely specific for APS). Mean ± SD survey scores for each item were calculated. Items were ranked from highest to lowest mean score. Low specificity items (Likert score <1) were eliminated when also agreed upon by the steering committee. In survey B, the steering committee members (n = 20) were asked to score previously identified items from survey A using a Likert scale (range –5 to +5) based on how specifically each item differentiated APS from other similar conditions. Items were then ranked again by their mean ± SD survey scores, and again, low specificity items (Likert score <2) were eliminated after discussion and agreement by the steering committee. The threshold for elimination of low specificity items by Likert score was intentionally slightly higher for survey B (score <2) compared to survey A (score <1) to improve specificity.

Domain identification, organization, and assessment of face and content validity. After completion of each survey (A and B), core planning group and steering committee members reviewed survey responses, survey scores, and literature reviews; nominal group technique (via in-person, email, and teleconference meetings) was used to organize and group highest scoring items into separate domains. We created domains under the guidance of classification criteria methodology experts (RN and KC) using basic principles for domain identification, consistent with previous classification criteria (13) and suggesting the need for the following: 1) no more than 8–10 domains; 2) separate domains allowing for an additive classification criteria system; 3) hierarchy of candidate criteria within each domain based on item specificity; 4) the highest specificity item within each domain to be scored; and 5) scores from each domain to be summed (i.e.,

Table 2. Phase I/II item generation and item reduction surveys during the new antiphospholipid syndrome (APS) classification criteria development*

	Survey question and response option	Goal	Response rate†	Respondents	Results summarized	Final no. of candidate criteria
Phase I						
Question 1	Question: "Describe all features (historical, clinical, laboratory, radiological, and pathological) that in your experience can occur as part of aPL/APS spectrum." Response option: open-ended.	Identify potential candidate criteria with positive weight	41/54 (76)	Of 41, 18 were rheumatologists, 5 hematologists, 5 clinical immunologists, 5 nephrologists, cardiologists, or neurologists, 4 internists, 2 pediatric rheumatologists, and 2 obstetricians.	Distribution of potential aPL/APS spectrum candidate criteria by system: neurologic (n = 25); laboratory (aPL) (n = 23); obstetric (n = 16); dermatologic (n = 15); renal (n = 12); vascular (n = 10); cardiac (n = 9); laboratory (non-aPL) (n = 9); other (n = 7); hematology (n = 5); pulmonary (n = 5); gastrointestinal (n = 4); musculoskeletal (n = 4); endocrinologic (n = 3); ophthalmology (n = 2); auditory (n = 2); family history (n = 1)	152 items (based on question 1) expanded to 261 to include negatively weighted criteria and subgroups. Of note, some items were reported as both positive and negative candidate criteria; we included those items in the expanded list.
Question 2	Question: "Describe all features (historical, clinical, laboratory, radiological, and pathological) or concomitant diseases that, if present, would make you question the diagnosis of APS even if aPL tests are positive." Response option: open-ended.	Identify potential candidate criteria with negative weight	41/54 (76)	Of 41, 18 were rheumatologists, 5 hematologists, 5 clinical immunologists, 5 nephrologists, cardiologists, or neurologists, 4 internists, 2 pediatric rheumatologists, and 2 obstetricians.	Distribution of potential negative weight criteria (categories): laboratory (aPL vs. non-aPL); other thrombosis risk factors; infections; traditional CVD risk factors; autoimmune diseases; malignancy; neurologic; histologic; elderly; thrombotic microangiopathies; radiologic; medications; lupus manifestations; strong family history; bleeding; other	152 items (based on question 1) expanded to 261 to include negatively weighted criteria and subgroups. Of note, some items were reported as both positive and negative candidate criteria; we included those items in the expanded list.
Question 3	Question: "When you consider the diagnosis of APS, do you think of APS patients in different sub-populations? If so, please describe how you categorize these patients." Response option: open-ended.	Categorize candidate criteria	41/54 (76)	Of 41, 18 were rheumatologists, 5 hematologists, 5 clinical immunologists, 5 nephrologists, cardiologists, or neurologists, 4 internists, 2 pediatric rheumatologists, and 2 obstetricians.	Subpopulations of APS patients categorized by age, clinical manifestations, aPL profile, risk level	152 items (based on question 1) expanded to 261 to include negatively weighted criteria and subgroups. Of note, some items were reported as both positive and negative candidate criteria; we included those items in the expanded list.

(Continued)

Table 2. (*Cont'd*)

	Survey question and response option	Goal	Response rate†	Respondents	Results summarized	Final no. of candidate criteria
Phase II						
Survey A	Question: "Consider 2 patients who are exactly the same except that one has the clinical feature presented and the other does not. Please rate each feature in terms of how strong this feature is in differentiating APS from other similar conditions, i.e., specific for APS" using a Likert scale (–5 to +5). Response option: "Please rate each item based on a scale of –5 to +5, with +5 being extremely specific for APS."	Item reduction by elimination of low specificity items and organizing higher specificity items into separate domains	43/61 (71)	Of 43, 22 were rheumatologists, 4 hematologists, 4 nephrologists, cardiologists, neurologists, or vascular specialists, 4 internists, 3 clinical immunologists, 3 pediatric rheumatologists, 2 pediatric hematologists, and 1 obstetrician.	See Table 3	132 items (reduced to 64 items and 10 domains when overlapping items were eliminated)
Survey B	"Consider 2 patients who are exactly the same except that one has the clinical feature presented and the other does not. Please rate each feature in terms of how strong this feature is in differentiating APS from other similar conditions, i.e., specific for APS" using a Likert scale (–5 to +5). Response option: "Please rate each item based on a scale of –5 to +5, with +5 being extremely specific for APS."	Further item reduction to goal of ~30 candidate criteria	19/19 (100)	Of 19, 8 were rheumatologists, 2 hematologists, 2 cardiologists or vascular medicine specialists, 2 immunologists, 2 obstetricians, 1 pediatric, rheumatologist, 1 neurologist, and 1 classification criteria methodologist.	See Table 4	27 items and 6 domains

* aPL = antiphospholipid antibody; CVD = cardiovascular disease.
† No./total no. (%).

additive domains) to produce a final APS classification criteria score.

In-person meeting preceding the 2018 ACR annual congress. Proposed domains and candidate criteria within domains were presented and further refined by the steering committee before the ACR annual congress (October 20, 2018, Chicago, IL), which also included patient participation. Changes to the number and content of domains and criteria within domains were made in response to group consensus. Literature was also presented to discuss sensitivity, prevalence, and specificity of items. Challenges with drawing conclusions related to APS from the literature on the general population were also discussed, as were the relevance of definitions used in past classification criteria. In order to define more precisely, to cluster, and to organize items, the steering committee also recommended forming domain subcommittees to review the literature and come to consensus on the definition, reliability, and precision of each potential criterion in its measurement. Additionally, the core planning group proposed classification entry criteria (i.e., minimum criteria required for consideration of APS classification) to the steering committee, and the concept of risk stratifying venous and arterial thrombosis in aPL-positive patients in the context of venous thromboembolism and CVD risk factors was discussed to increase APS specificity based on phase I and II survey results.

Other important considerations relevant to phase I/II. Following this meeting of the steering committee, we administered a separate survey asking the steering committee to assess and rate proposed entry criteria and risk factors for thrombosis. Additional questions were included regarding the definitions of aPL "persistence," aPL "positivity," whether LAC testing in the setting of anticoagulation should be included in the classification criteria scoring, which aPL laboratory tests to include in the entry criteria, and whether a separate classification system should exist for primary APS versus APS-associated with other systemic autoimmune diseases.

RESULTS

Phase I item generation. Based on our phase I item generation survey and literature review, we generated a comprehensive list of potential aPL-associated manifestations, in particular those occurring at the time of APS diagnosis (Table 2). Survey responses were originally clustered based on organ systems (n = 152) (see Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24520/abstract) and then expanded to incorporate negatively weighted responses and aPL/APS subgroups (n = 261) (see Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24520/abstract). We checked each potential criterion for reliability, redundancy, and precision in its measurement. For example, thrombotic stroke was defined as a focal clinical neurologic event confirmed by neuroimaging studies such as computed tomography or magnetic resonance imaging (14). Additionally, similar criteria (e.g., deep vein thrombosis [DVT] and thromboembolism) or highly correlated criteria (e.g., DVT and pulmonary embolism) were merged to avoid redundant data. The reported subpopulations of APS patients, grouped based on age, clinical manifestations, aPL profile, and risk level, are shown in Table 2.

The core planning group subsequently restructured the phase I results. In total, 152 candidate criteria, identified during phase I, were expanded to 261 with the addition of subgroups and candidate criteria with potential negative weight (Table 1). The phase I literature screening of the article reference library did not yield additional candidate criteria that were not already included in the comprehensive list generated by the surveys. Potential criteria with subgroups (e.g., age <55 years, stroke without hypertension, or an aCL titer >40 units) were expanded into individual criteria. We also added potential negatively weighted criteria to the list of 152 candidate criteria, including but not limited to infections (e.g., HIV, hepatitis B or C, or syphilis), other thrombosis risk factors (i.e., surgery, immobilization, or infective endocarditis), and items that overlapped with positive criteria (e.g., false-positive venereal disease research laboratory test for syphilis).

Phase II item reduction. *Literature reviews and additional studies.* Thrombosis risk factors. Based on phase I survey results and steering committee consensus to consider arterial and venous events in the presence or absence of thrombosis risk factors, we reviewed major recent international guidelines that emphasized the need to incorporate cardiovascular risk factors (e.g., hypertension and hypercholesterolemia) for arterial thrombosis events (15–20) and venous thromboembolism risk factors (e.g., oral contraceptive use) for venous events (21–24). The committee agreed that these risk factors will be further defined and structured during phase III.

Meta-analyses. Meta-analyses performed in patients with systemic lupus erythematosus (SLE) with and without aPL demonstrated the increased odds ratios (ORs) of various clinical manifestations, including cardiac valvular disease (OR 3.13 [95% confidence interval (95% CI) 2.31–4.24]), livedo (OR 3.36 [95% CI 2.49–4.55]), thrombocytopenia (OR 2.48 [95% CI 2.10–2.98]), hemolytic anemia (OR 3.22 [95% CI 2.40–3.42]), and renal impairment (OR 2.9 [95% CI 1.9–4.3]) in aPL-positive SLE patients compared to aPL-negative SLE patients (6,25–26). These studies provided additional evidence that the presence of aPL may increase the likelihood of the above clinical manifestations.

Additional studies. The pathology domain subcommittee evaluated real-world renal biopsies to better characterize aPL nephropathy. The committee concluded that biopsy reports inconsistently use aPL nephropathy–related terminology, and the use of aPL nephropathy–related terms varies among pathologists

Table 3. Phase II survey A results and subsequent steering committee and domain subcommittee decisions and rationale for item reduction during the new antiphospholipid syndrome (APS) classification criteria development*

Candidate criteria and domains	Committee consensus supported by the literature review for items eliminated (despite score >2) or retained (despite score <2)
Cardiopulmonary	
Retained: CV vegetation, CV thickening, AH†	AH: Considered distinct entity associated with aPL (36); retained for phase 3 microvascular domain.
Eliminated: Pulmonary hypertension (TE),‡ heart failure due to ischemia	Pulmonary hypertension (TE): Commonly due to pulmonary embolism in APS, which is redundant with the item VTE (chronic TE pulmonary hypertension [group 4]) (37,38).
Dermatologic	
Retained: LRa, livedoid vasculopathy	
Eliminated: LRe,‡ atrophie blanche de Milan, pyoderma gangrenosum–like skin ulcers, anetoderma	LRe: Relatively common in general population; not as specific as LRa for APS classification (39,40).
Hematologic	
Retained: Thrombocytopenia (mild), thrombocytopenia (severe)†	Thrombocytopenia (severe): Given that other more likely explanations than aPL exist, the committee proposed evaluating cutoff thresholds (mild–moderate–severe) in phase 3 to improve specificity (41,42).
Eliminated: Microangiopathic hemolytic anemia, positive Coombs test results without hemolytic anemia	
Neurology	
Retained: TIA without ARF (age <55 years), TIA without ARF (age ≥55 years),† TIA with ARF (age <55 years),† TIA with ARF (age ≥55 years),† acute ischemic encephalopathy (age <55 years)	TIA: Retained for phase 3 independent of age and ARF, with plan to further analyze in context of age and ARF during case collection.
Eliminated: Seizure/epilepsy,‡ atypical multiple sclerosis–like disease, migraine responsive to anticoagulation, chorea, cerebral white matter lesions, cognitive dysfunction, longitudinal extensive transverse myelitis, multiinfarct dementia	Seizure/epilepsy: Rarely associated with aPL; mainly secondary to stroke in aPL-positive patients; poor definability (5,43).
Obstetric	
Retained: Fetal loss (single†/recurrent/consecutive), stillbirth, early pregnancy loss (recurrent†), preeclampsia (early/severe/mild/late†), HELLP syndrome†, eclampsia, intrauterine growth restriction†	Obstetric morbidity: Restructured domain for evaluation of individual items during phase 3 case collection (44,45).
Eliminated: Chorea gravidarum, pregnancy-induced hypertension	
Vascular	
Retained: AT without ARF (age <55 years), AT with ARF (age <55 years),† AT without ARF (age ≥55 years),† AT with ARF (age ≥55 years),† VTE without ARF (age <55 years), VTE with ARF (age <55 years),† VTE without ARF (age ≥55 years),† VTE with ARF (age ≥55 years),† SVT without ARF (age <55 years), SVT with ARF (age <55 years),† SVT without ARF (age ≥55 years),† SVT with ARF (age ≥55 years)†	AT, VTE, and SVT: SVT considered to have lower specificity for aPL than AT or VTE (5,46). However, all 3 retained for phase 3 independent of age and risk factors; committee agreed to include age thresholds and additional thrombosis risk factors during case collection.
Renal/abdominal	
Retained: Nephrotic syndrome, adrenal hemorrhage	
Pathology	
Retained: Thrombosis/infarction without vasculitis, PC,† FIH,† focal cortical atrophy	PC and FIH: PC is increasingly associated with aPL (36). FIH is considered the most frequent chronic lesion in primary aPL nephropathy (47). Retained and restructured both items for phase 3 data collection under microvascular domain.
Laboratory part I	
Retained: Lupus anticoagulant test	
Laboratory part II	
Retained: IgG/IgM aCL,† IgG/IgM anti-β_2GPI antibodies†	IgM aCL and anti-β_2GPI: Wide variation in survey scores for IgM aPL ELISA based on titer level and single/persistent positivity. IgM aPL ELISA is noted to have lower specificity than IgG for APS. Final decision was to collect detailed aPL ELISA isotype and titer information during phase 3 (48,49).
Eliminated: Anti-DI antibodies,‡ anti-PS/PT,‡ IgA aCL, IgA anti-β_2GPI, antiprothrombin antibodies	Anti-DI and anti-PS/PT: Limited commercial availability; additional research needed to define feasibility, clinical correlation, and standardization.

* aCL = anticardiolipin antibodies; AH = alveolar hemorrhage; anti-β_2GPI = anti–β_2-glycoprotein I; anti-DI = antidomain I; anti-PS/PT = anti–phosphatidylserine/prothrombin complex; aPL = antiphospholipid antibody; ARF = additional risk factor; AT = arterial thrombosis; CV = cardiac valve; ELISA = enzyme-linked immunosorbent assay; FIH = fibrous interstitial hyperplasia; HELLP = hemolysis, elevated liver enzymes, and low platelets (syndrome); LRa = livedo racemosa; LRe = livedo reticularis; PC = pulmonary capillaritis; SVT = superficial venous thrombosis; TE = thromboembolic; TIA = transient ischemic attack; VTE = venous thromboembolism.

† Any item scoring <2 but retained based on committee discussions. The reason for retention was based on committee agreement, supported by survey Likert score >2, and is explained in the second column.

‡ Any item scoring >2 but eliminated based on committee discussions. The reason for elimination was based on committee agreement, supported by survey Likert score <2 (low specificity), and is explained in the second column.

while reporting the biopsy findings of kidney involvement in aPL-positive patients (data to be published separately).

Item reduction surveys A/B and Delphi exercises. Phase II, survey A, mean Likert scale scores for item specificity ranged from –2.29 to 4.65; items of low specificity (score <1) were eliminated. The higher specificity items (n = 132) were organized into 7 clinical (cardiopulmonary, dermatologic, hematologic, neurologic, obstetric, renal/abdominal, and vascular), 2 laboratory (aPL immunoassays and lupus anticoagulant test), and 1 pathologic domain (Tables 2 and 3). Items overlapping or describing similar concepts were combined to arrive at n = 64. Phase II, survey B, mean item scores ranged from –1.05 to 4.79; items of low specificity (score <2) were eliminated (Tables 2 and 3). Each item was considered by the steering committee, and in certain situations, the committee agreed to retain or eliminate criteria regardless of their survey score for further evaluation in phase III, as described in Table 3. Both survey A and B results demonstrated very different weights for macrovascular domain outcomes with or without additional thrombosis risk factors.

Domain identification, organization, and assessment of face and content validity. After phase II survey A and B completion, based on subsequent literature reviews, nominal group technique, steering committee and domain subcommittee discussions, and 1 in-person face-to-face steering committee meeting (see below), 64 items identified in phase II survey A were reduced to 27 candidate criteria (Table 4). Through literature review and domain subcommittee expert-based discussions, face and content validities of each candidate criterion were assessed and determined. Specificity of a candidate criterion for APS was determined to be contingent on assessment in a non-APS population using literature review, when available, and/or expert opinion. Consensus was achieved among the steering committee for the following: 1) only score a particular criterion if no other more likely cause exists; 2) clinical criteria can occur on 1 occasion and need not occur simultaneously with other clinical criteria; and 3) only the highest weighted criterion in each domain would be counted toward the total APS score. The core planning group and steering committee members adhered to the recommendations outlined by the classification criteria methodology experts and hierarchically organized the highest scoring candidate criteria into 6 separate and additive domains (Table 4).

In-person meeting prior to the 2018 ACR annual congress. During this face-to-face meeting, the steering committee agreed that candidate clinical criteria must be interpreted in the context of a clinically acceptable aPL profile. Thus, all members voted in favor of proposed permissive, rather than restrictive, entry criteria requiring fulfillment of at least 1 laboratory criterion (aPL positivity [LAC test, IgG/M aCL, or IgG/M anti-β_2GPI positivity above the normal laboratory range] at any time) and 1 clinical criterion (identified from the clinical domains). The group also discussed and eliminated items with low specificity (i.e., phase II survey A Likert score <1) or poor definability; items overlapping or describing similar concepts were combined (Table 3). The committee also agreed with the need for further evaluation of the following: 1) the definition of aPL laboratory "persistence" and titer level cutoffs; 2) use of an age cutoff; 3) and phase I and phase II survey A results suggesting that thrombotic events with other risk factors should be weighted differently than those without risk factors. The steering committee agreed that there are not enough data supporting the inclusion of the IgA isotype in the classification criteria until there is better understanding of the pathogenic and prognostic significance. This decision was based on the following: 1) our phase II survey B results demonstrating low scores for aCL or IgA anti-β_2GPI regardless of titer or persistence (Likert score <1); 2) multiple studies demonstrating the lack of predictive value of isolated IgA positivity for APS manifestations, including a study led by members of our laboratory domain subcommittee team (27,28); and 3) the lack of feasibility and standardization (29–31). During and after the pre-ACR 2018 in-person meeting, the proposed candidate criteria within domains were further refined by domain subcommittees.

Table 4. Phase II survey B results with proposed domains and items (for phase III case collection) during the new antiphospholipid syndrome classification criteria development*

Proposed domains (n = 6)	Proposed items (n = 27)
Candidate laboratory criteria	
aPL testing, coagulation-based functional assays	Lupus anticoagulant test
aPL testing, solid-phase assays	IgG anticardiolipin antibody, IgM anticardiolipin antibody; IgG anti–β_2-glycoprotein I, IgM anti–β_2-glycoprotein I
Candidate clinical criteria	
Macrovascular	Superficial vein thrombosis, venous thromboembolism, arterial thrombosis, transient ischemic attack
Microvascular	Livedo racemosa, livedoid vasculopathy, adrenal hemorrhage or plexus thrombosis, acute ischemic encephalopathy, cardiac microvascular disease, pulmonary hemorrhage, acute aPL nephropathy, chronic aPL nephropathy
Obstetric	Pregnancy loss <10 weeks (w) of gestation, fetal death between 10 w to <16 w of gestation, fetal death between 16 w to 34 w of gestation, preeclampsia with severe features <34 w of gestation, placental insufficiency with severe features <34 w of gestation
Cardiac valve disease	Noninfectious valve vegetation, thickening
Hematologic	Platelet count $<20 \times 10^9$ per liter, platelet count $20–130 \times 10^9$ per liter, platelet count $131–150 \times 10^9$ per liter

* aPL = antiphospholipid antibody.

Other important considerations relevant to phase I/II. Based on phase II survey C results and additional steering committee discussions, the following consensus decisions were also made: 1) additional thrombosis risk factors with low specificity were eliminated (72–89% agreed with proposed thrombosis and CVD risk factors and definitions); 2) entry criteria would be used to identify the relevant patient population to whom the classification criteria would be applied; 3) 84% voted in favor of introducing a time restriction between a positive aPL test and the aPL-related clinical event in the entry criteria; 4) 89% agreed that persistence of laboratory testing should be defined as at least 12 weeks apart; 5) 79% agreed that timing of aPL testing with respect to known active infections and malignancy should be included in the assessment; and 6) 100% agreed with including LAC testing based on International Society on Thrombosis and Haemostasis guidelines (32).

DISCUSSION

The first 2 phases of the APS classification criteria development support the concept that clinical manifestations associated with aPL are heterogeneous and complex. We used a data-driven, consensus-based approach to identify and organize 27 candidate criteria for classification of APS into 6 separate domains using previously established methodology and incorporating bias-reduction techniques. Phase I/II involved a dedicated, multidisciplinary team of ~80 physicians, investigators, epidemiologists, classification criteria methodologists, and patients worldwide, allowing for diversity of opinions and experiences. In the next phase, these proposed candidate criteria will be used for real-world case collection and further refined, hierarchically organized, and weighted so that an aggregate score and threshold for APS classification can be determined.

Our methodology builds on recently published rheumatic disease classification criteria, including that of SLE, systemic sclerosis, rheumatoid arthritis, and IgG4-related disease, but also includes novel concepts (33–35). Given our findings that physicians view aPL-positive patients in different subpopulations (e.g., by thrombosis provoking risk factors, age, and aPL profile), consideration of these factors provides a unique risk stratification aspect to the new classification criteria development, which differs from previous rheumatic disease or prior APS classification criteria (3). Additionally, the need to evaluate candidate clinical criteria in the context of a clinically significant aPL profile emphasizes the important role of entry criteria and may alter the structure of the final threshold determination by suggesting the need for separate clinical and laboratory summation scores.

Classification criteria, in contrast to diagnostic criteria, are used to identify relatively homogeneous groups of patients for inclusion in clinical trials and observational studies; thus, the goal of the new APS classification criteria is to better standardize patients for APS research. Given the heterogeneity of aPL-related clinical manifestations and laboratory tests, our methodology ultimately aims to develop a weighted scoring system to increase the validity and performance of APS classification criteria. Through expert- and data-driven methods, we captured the wide spectrum of APS manifestations, which were eventually refined to more specific manifestations. In particular, inclusion of microvascular, hematologic, and cardiac valve disease domains will improve the sensitivity and generalizability of the new classification criteria.

Given the major morbidity and mortality risk related to APS and currently limited treatment options, the development and validation of stringent, high-performing classification criteria is critical to improve understanding of APS epidemiology and clinical outcomes. Additionally, research on the immunologic, genetic, and diagnostic aspects of APS has been evolving rapidly. Fortunately, our new classification criteria methodology will allow us to modify the domains in the future, e.g., to include new commercially available laboratory tests or clinical criteria shown to be highly specific for APS.

Our study has a number of potential biases and limitations. First, a major source of potential bias in classification criteria development is one of circularity of reasoning, which may result when the same experts who contribute to development of classification criteria also provide patient data (in phase III) and participate in validation (phase IV). Additionally, we asked experts to consider the clinical spectrum of aPL/APS during phase I/II and not to use the previously published Sapporo criteria as a decision-making guide. As one bias reduction strategy, we ensured that the majority of collaborators involved in phase I/II would be mutually distinct from those who would be involved in future phases. Second, our literature reviews revealed a relative dearth of studies evaluating the performance characteristics (e.g., sensitivity and specificity) or prevalence of certain candidate criteria. However, by using expert-based surveys, nominal group technique and Delphi exercises, forming domain subcommittees, and conducting our own meta-analyses and additional studies, we aimed to address these gaps in knowledge and to minimize its effect on the validity and reliability of the data.

Following the work described above, phase III will refine the definitions of entry criteria and relative candidate criteria and collect real-world, international potential APS cases. This next phase aims to use Multi-Criteria Decision Analysis methodology and 1000Minds software. We will test our criteria for reliability and applicability, employ a systematic methodology for weighting and threshold determination, test the performance of the scoring system at the margin between included and excluded cases, and use cohorts for testing and validation, which deliberately include the gray zone where classification is difficult.

In conclusion, substantial advances in the methodology for classification criteria development have occurred since formulation of the original APS Sapporo classification criteria. By employing a new methodology used for the development of classification

criteria for other rheumatic diseases, and by incorporating novel disease-specific strategies, we anticipate that the new APS classification criteria will have excellent face validity, criterion validity, and performance.

ACKNOWLEDGMENTS

The APS Classification Criteria Steering Committee dedicates this work to the late Raymond P. Naden, MD. Dr. Naden helped lead the development of numerous rheumatic disease classification criteria, including ours; without him, the work presented in this manuscript would not have been possible. The APS Classification Criteria Steering Committee also thanks our ACR/EULAR liaison, Ronald van Vollenhoven, MD (Amsterdam University Medical Center, Amsterdam, The Netherlands) for help and guidance, Ayten Yazici, MD (Kocaeli University School of Medicine, Kocaeli, Turkey) for phase I analysis, Maxime Tagavi, MD (Brugmann Hospital, Brugmann, Belgium) for collaboration to better define antiphospholipid antibody nephropathy, Joann Vega (Hospital for Special Surgery, New York, NY) for administrative assistance, and our patient representatives.

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. Barbhaiya 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. Barbhaiya, Zuily, Ahmadzadeh, Naden, Costenbader, Erkan.

Acquisition of data. Barbhaiya, Zuily, Amigo, Avcin, Bertolaccini, Branch, de Jesus, Devreese, Frances, Garcia, Guillemin, Levine, Levy, Lockshin, Ortel, Seshan, Tektonidou, Wahl, Willis, Costenbader, Erkan.

Analysis and interpretation of data. Barbhaiya, Zuily, Amigo, Avcin, Bertolaccini, Branch, de Jesus, Devreese, Frances, Garcia, Guillemin, Levine, Levy, Lockshin, Ortel, Seshan, Tektonidou, Wahl, Willis, Naden, Costenbader, Erkan.

ADDITIONAL DISCLOSURES

Author Levy is an employee of GlaxoSmithKline.

REFERENCES

1. Wilson WA, Gharavi AE, Koike T, Lockshin MD, Branch DW, Piette JC, et al. International consensus statement on preliminary classification criteria for definite antiphospholipid syndrome: report of an international workshop. Arthritis Rheum 1999;42:1309–11.
2. Lockshin MD, Sammaritano LR, Schwartzman S. Validation of the Sapporo criteria for antiphospholipid syndrome. Arthritis Rheum 2000;43:440–3.
3. Miyakis S, Lockshin MD, Atsumi T, Branch DW, Brey RL, Cervera R, et al. International consensus statement on an update of the classification criteria for definite antiphospholipid syndrome (APS). J Thromb Haemost 2006;4:295–306.
4. Bertolaccini ML, Amengual O, Andreoli L, Atsumi T, Chighizola CB, Forastiero R, et al. 14th International Congress on Antiphospholipid Antibodies Task Force. Report on antiphospholipid syndrome laboratory diagnostics and trends. Autoimmun Rev 2014;13:917–30.
5. Abreu MM, Danowski A, Wahl DG, Amigo MC, Tektonidou M, Pacheco MS, et al. The relevance of "non-criteria" clinical manifestations of antiphospholipid syndrome: 14th International Congress on Antiphospholipid Antibodies Technical Task Force report on antiphospholipid syndrome clinical features. Autoimmun Rev 2015;14:401–14.
6. Zuily S, Regnault V, Selton-Suty C, Eschwege V, Bruntz JF, Bode-Dotto E, et al. Increased risk for heart valve disease associated with antiphospholipid antibodies in patients with systemic lupus erythematosus: meta-analysis of echocardiographic studies. Circulation 2011;124:215–24.
7. Erkan D, Derksen R, Levy R, Machin S, Ortel T, Pierangeli S, et al. Antiphospholipid Syndrome Clinical Research Task Force report. Lupus 2011;20:219–24.
8. Johnson SR, Goek ON, Singh-Grewal D, Vlad SC, Feldman BM, Felson DT, et al. Classification criteria in rheumatic diseases: a review of methodologic properties. Arthritis Rheum 2007;57:1119–33.
9. Johnson SR, Tomlinson GA, Hawker GA, Granton JT, Feldman BM. Methods to elicit beliefs for Bayesian priors: a systematic review. J Clin Epidemiol 2010;63:355–69.
10. Johnson SR, Naden RP, Fransen J, van den Hoogen F, Pope JE, Baron M, et al. Multicriteria decision analysis methods with 1000Minds for developing systemic sclerosis classification criteria. J Clin Epidemiol 2014;67:706–14.
11. Aggarwal R, Ringold S, Khanna D, Neogi T, Johnson SR, Miller A, et al. Distinctions between diagnostic and classification criteria? Arthritis Care Res (Hoboken) 2015;67:891–7.
12. Fries JF, Hochberg MC, Medsger TA Jr, Hunder GG, Bombardier C, for the American College of Rheumatology Diagnostic and Therapeutic Criteria Committee. Criteria for rheumatic disease: different types and different functions. Arthritis Rheum 1994;37:454–62.
13. 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.
14. Sacco RL, Kasner SE, Broderick JP, Caplan LR, Connors JJ, Culebras A, et al. An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2013;44:2064–89.
15. Authors/Task Force Members, ESC Committee for Practice Guidelines (CPG), ESC National Cardiac Societies. 2019 ESC/EAS guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. Atherosclerosis 2019;290:140–205.
16. Shah N, Kelly AM, Cox N, Wong C, Soon K. Myocardial infarction in the "young": risk factors, presentation, management and prognosis. Heart Lung Circ 2016;25:955–60.
17. Cosentino F, Grant PJ, Aboyans V, Bailey CJ, Ceriello A, Delgado V, et al. 2019 ESC guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD. Eur Heart J 2020;41:255–323.
18. Singh A, Collins BL, Gupta A, Fatima A, Qamar A, Biery D, et al. Cardiovascular risk and statin eligibility of young adults after an MI: partners YOUNG-MI registry. J Am Coll Cardiol 2018;71:292–302.
19. Yang J, Biery DW, Singh A, Divakaran S, DeFilippis EM, Wu WY, et al. Risk factors and outcomes of very young adults who experience myocardial infarction: the Partners YOUNG-MI registry. Am J Med 2020;133:605–12.e1.
20. Collet JP, Zeitouni M, Procopi N, Hulot JS, Silvain J, Kerneis M, et al. Long-term evolution of premature coronary artery disease. J Am Coll Cardiol 2019;74:1868–78.
21. Lim W, Le Gal G, Bates SM, Righini M, Haramati LB, Lang E, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: diagnosis of venous thromboembolism. Blood Adv 2018;2:3226–56.
22. Kearon C, Ageno W, Cannegieter SC, Cosmi B, Geersing GJ, Kyrle PA. Categorization of patients as having provoked or unprovoked venous thromboembolism: guidance from the SSC of ISTH. J Thromb Haemost 2016;14:1480–3.
23. Konstantinides SV, Meyer G, Becattini C, Bueno H, Geersing GJ, Harjola VP, et al. 2019 ESC guidelines for the diagnosis and

management of acute pulmonary embolism developed in collaboration with the European Respiratory Society (ERS). Eur Heart J 2020;41:543–603.
24. Gregson J, Kaptoge S, Bolton T, Pennells L, Willeit P, Burgess S, et al. Cardiovascular risk factors associated with venous thromboembolism. JAMA Cardiol 2019;4:163–73.
25. Chock YP, Moulinet T, Dufrost V, Erkan D, Wahl D, Zuily S. Antiphospholipid antibodies and the risk of thrombocytopenia in patients with systemic lupus erythematosus: a systematic review and meta-analysis. Autoimmun Rev 2019;18:102395.
26. Ünlü O, Zuily S, Erkan D. The clinical significance of antiphospholipid antibodies in systemic lupus erythematosus. Eur J Rheumatol 2016;3:75–84.
27. Vlagea A, Pascual-Salcedo D, Álvarez Doforno R, Lavilla P, Diez J, Padilla Merlano B, et al. IgA anti-β2 glycoprotein I antibodies: experience from a large center. Thromb Res 2018;162:38–43.
28. Chayoua W, Yin DM, Kelchtermans H, Moore GW, Gris JC, Musiał J, et al. Is there an additional value in detecting anticardiolipin and anti–β_2 glycoprotein I IgA antibodies in the antiphospholipid syndrome? Thromb Haemost 2020;120:1557–68.
29. Pérez D, Martínez-Flores JA, Serrano M, Lora D, Paz-Artal E, Morales JM, et al. Evaluation of three fully automated immunoassay systems for detection of IgA anti-beta 2-glycoprotein I antibodies. Int J Lab Hematol 2016;38:560–8.
30. Tebo AE, Willis R, Jaskowski TD, Guerra M, Pierangeli SS, Salmon J, et al. Clinical significance and correlations between anti–β_2 glycoprotein I IgA assays in antiphospholipid syndrome and/or systemic lupus erythematosus. Clin Chim Acta 2016;460:107–13.
31. Martínez-Flores JA, Serrano M, Alfaro J, Mora S, Paz-Artal E, Morales JM, et al. Heterogeneity between diagnostic tests for IgA anti-beta2 glycoprotein I: explaining the controversy in studies of association with vascular pathology. Anal Chem 2013;85:12093–8.
32. Pengo V, Tripodi A, Reber G, Rand JH, Ortel TL, Galli M, et al. Update of the guidelines for lupus anticoagulant detection. J Thromb Haemost 2009;7:1737–40.
33. Van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2013;65:2737–47.
34. 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.
35. Aringer M, Costenbader K, Daikh D, Brinks R, Mosca M, Ramsey-Goldman R, et al. 2019 European League Against Rheumatism/American College of Rheumatology classification criteria for systemic lupus erythematosus. Arthritis Rheumatol 2019;71:1400–12.
36. Stoots SA, Lief L, Erkan D. Clinical insights into diffuse alveolar hemorrhage in antiphospholipid syndrome. Curr Rheumatol Rep 2019;21:56.
37. Galiè N, Humbert M, Vachiery JL, Gibbs S, Lang I, Torbicki A, et al. 2015 ESC/ERS guidelines for the diagnosis and treatment of pulmonary hypertension: the Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS): Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). Eur Heart J 2016;37:67–119.
38. Wilkens H, Konstantinides S, Lang IM, Bunck AC, Gerges M, Gerhardt F, et al. Chronic thromboembolic pulmonary hypertension (CTEPH): updated recommendations from the Cologne Consensus Conference 2018. Int J Cardiol 2018;272s:69–78.
39. Francès C, Niang S, Laffitte E, le Pelletier F, Costedoat N, Piette JC. Dermatologic manifestations of the antiphospholipid syndrome: two hundred consecutive cases. Arthritis Rheum 2005;52:1785–93.
40. Ordi-Ros J, Sáez-Comet L, Pérez-Conesa M, Vidal X, Riera-Mestre A, Castro-Salomó A, et al. Rivaroxaban versus vitamin K antagonist in antiphospholipid syndrome: a randomized noninferiority trial. Ann Intern Med 2019;171:685–94.
41. Cuadrado MJ, Mujic F, Muñoz E, Khamashta MA, Hughes GR. Thrombocytopenia in the antiphospholipid syndrome. Ann Rheum Dis 1997;56:194–6.
42. Finazzi G. The Italian Registry of Antiphospholipid Antibodies. Haematologica 1997;82:101–5.
43. Yelnik CM, Kozora E, Appenzeller S. Non-stroke central neurologic manifestations in antiphospholipid syndrome. Curr Rheumatol Rep 2016;18:11.
44. Chighizola CB, Andreoli L, de Jesus GR, Banzato A, Pons-Estel GJ, Erkan D. The association between antiphospholipid antibodies and pregnancy morbidity, stroke, myocardial infarction, and deep vein thrombosis: a critical review of the literature. Lupus 2015;24:980–4.
45. De Jesús GR, Benson AE, Chighizola CB, Sciascia S, Branch DW. Sixteenth international congress on antiphospholipid antibodies task force: report on obstetric antiphospholipid syndrome. Lupus 2020:961203320954520.
46. Wahl DG, Guillemin F, de Maistre E, Perret-Guillaume C, Lecompte T, Thibaut G. Meta-analysis of the risk of venous thrombosis in individuals with antiphospholipid antibodies without underlying autoimmune disease or previous thrombosis. Lupus 1998;7:15–22.
47. Nochy D, Daugas E, Droz D, Beaufils H, Grünfeld JP, Piette JC, et al. The intrarenal vascular lesions associated with primary antiphospholipid syndrome. J Am Soc Nephrol 1999;10:507–18.
48. Chayoua W, Kelchtermans H, Gris JC, Moore GW, Musiał J, Wahl D, et al. The (non-) sense of detecting anti-cardiolipin and anti–$\beta_2$glycoprotein I IgM antibodies in the antiphospholipid syndrome. J Thromb Haemost 2020;18:169–79.
49. Vanoverschelde L, Kelchtermans H, Musial J, de Laat B, Devreese KM. Influence of anticardiolipin and anti–β_2 glycoprotein I antibody cutoff values on antiphospholipid syndrome classification. Res Pract Thromb Haemost 2019;3:515–27.

APPENDIX A: NEW APS CLASSIFICATION CRITERIA COLLABORATORS

The new APS classification criteria collaborators are as follows: Nancy Agmon-Levin (Sheba Medical Center, Tel Aviv, Israel), Cassyanne Aguilar (Children's Hospital of The King's Daughters, Norfolk, VA), Paula Alba (National University of Cordoba, Cordoba, Argentina), Oral Alpan (O&O Alpan, Tysons Corner, VA), Ales Ambrozic (Ljubljana University Medical Center, Ljubljana, Slovenia), Zahir Amoura (Pitie-Salpêtrière Hospital, Paris, France), Danieli Andrade (University of Sao Paulo, Sao Paulo, Brazil), Luis Andrade (Universidade Federal de São Paulo, São Paulo, Brazil), Simone Appenzeller (University of Campinas, São Paulo, Brazil), Bahar Artim Esen (Istanbul University, Istanbul, Turkey), Tatsuya Atsumi (Hokkaido University, Hokkaido, Japan), Yackov Berkun (Hadassah Medical Center, Jerusalem, Israel), Antonio Cabral (University of Ottawa, Ottawa, Ontario, Canada), Guillaume Canaud (Université de Paris, Paris, France), Ricard Cervera (Hospital Clinic, Barcelona, Spain), Pojen Chen (University of California at San Diego, La Jolla), Cecilia Chighizola (University of Milan, Milan, Italy), Rolando Cimaz (University of Florence, Florence, Italy), Hannah Cohen (University College London, London, UK), Nathalie Costedoat-Chalumeau (Cochin Hospital, Paris, France), Mark Crowther (McMaster University, Hamilton, Ontario, Canada), Maria Jose Cuadrado (University Hospital Navarra Madrid, Madrid, Spain), Philip G. de Groot (University Medical Centre, Utrecht, The Netherlands), Philippe de

Moerloose (University Hospital, Geneva, Switzerland), Maria De Sancho (Weill Cornell Medicine, New York, NY), Ronald Derksen (University Hospital, Utrecht, The Netherlands), Reyhan Diz-Kucukkaya (Istanbul, Turkey), Thomas Dörner (Charité University, Berlin, Germany), Paul Fortin (Centre Hospitalier Universitaire de Québec, Quebec City, Quebec, Canada), Bill Giannakopoulos (University of New South Wales, Sydney, New South Wales, Australia), Jose A. Gómez-Puerta (Hospital Clinic, Barcelona, Spain), Emilio B. Gonzalez (University of Texas Medical Branch, Galveston), Murat Inanc (Istanbul University, Istanbul, Turkey), Gili Kenet (Sheba Medical Center, Tel Aviv, Israel), Munther Khamashta (St. Thomas Hospital, London, UK), Martin Kriegel (Yale School of Medicine, New Haven, CT), Steven Krilis (University of New South Wales, Sydney, New South Wales, Australia), Carl Laskin (LifeQuest Centre for Reproductive Medicine, Toronto, Ontario, Canada), Patti Massicotte (University of Alberta, Edmonton, Alberta, Canada), Gale McCarty (Santa Monica, CA), Pier Luigi Meroni (University of Milan, Milan, Italy), Jamal Mikdashi (University of Maryland, College Park), Barry Myones (Texas Children's Hospital, Houston), Vittorio Pengo (University Hospital, Padova, Italy), Michelle Petri (Johns Hopkins Hospital, Baltimore, MD), Robert Roubey (University of North Carolina, Chapel Hill), Lisa Sammaritano (Hospital for Special Surgery, New York, NY), Giovanni Sanna (Guy's and St Thomas' NHS Foundation Trust, London, UK), Savino Sciascia (University of Turin, Turin, Italy), Flavio Signorelli (Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil), Arzu Soybilgic (University of Illinois, Chicago), Angela Tincani (University of Brescia, Brescia, Italy), Scott Woller (University of Utah, Salt Lake City), and Cécile Yelnik (University of Lille, Lille, France).

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 73, No. 10, October 2021, pp 1502–1510
DOI 10.1002/acr.24411

Trends in Systemic Sclerosis Mortality Over Forty-Eight Years, 1968–2015: A US Population–Based Study

Eric Y. Yen,[1] Devanshu R. Singh,[2] and Ram R. Singh[3]

Objective. To identify secular trends associated with systemic sclerosis (SSc) mortality over a 48-year period.

Methods. Using national mortality data compiled by the Centers for Disease Control and Prevention's Wide-Ranging Online Data for Epidemiologic Research, and population data from the US Census Bureau, we calculated an age-standardized mortality rate (ASMR) for SSc and non-SSc (all other causes), and we also calculated the ratio of the SSc ASMR to the non-SSc ASMR for each year from 1968 to 2015. We then used a joinpoint regression model to evaluate mortality trends overall and by sex and race.

Results. From 1968 to 2015, there were 46,798 deaths with SSc recorded as the "underlying" cause of death and 106,058,839 non-SSc deaths. There were an additional 9,063 deaths with SSc recorded as a "contributing" cause of death from 1999 to 2015. Whereas the non-SSc ASMR decreased throughout the 48-year time period, the SSc ASMR increased from 1968 to 2000, followed by decreases each year from 2001 to 2015. The SSc ASMR also decreased for deaths where SSc was a contributing cause from 1999 to 2015. Women and Black persons had higher SSc ASMRs and SSc ASMR to non-SSc ASMR ratios than men and White persons, respectively. Additionally, SSc ASMRs and SSc ASMR to non-SSc ASMR ratios increased at higher rates in women and White persons than in men and Black persons, respectively, during the initial three decades.

Conclusion. Mortality attributable to SSc increased from 1968 to 2000, followed by a steady decline from 2001 to 2015. However, SSc mortality relative to non-SSc mortality remains high. SSc mortality has disproportionately changed by sex and race over the 48-year period assessed in the present study.

INTRODUCTION

Systemic sclerosis (SSc) is a systemic autoimmune disease that causes premature death due to complications from interstitial lung disease, pulmonary arterial hypertension, gastrointestinal dysmotility, renal crisis, and malnutrition (1,2). Over the past 2 decades, advances in the treatment of SSc-associated complications may have affected patient outcomes (2). For example, with the availability of prostanoids for the treatment of pulmonary arterial hypertension, the 2-year survival of patients with SSc-associated pulmonary arterial hypertension has improved from 47% to 71% (3). However, the influence of these advances on SSc mortality trends in the US general population is unknown.

The disease-specific mortality rate is an important measure of the burden of disease. The actual mortality burden of SSc is unknown. Previous studies of SSc mortality were based primarily on deaths in patient cohorts at referral centers (4–7), which does not capture changes in incidence over time and does not reflect the actual burden and trends of SSc mortality in the general population. A few studies have used population-based designs (8,9) but were limited to specific regions, small samples, or relatively short durations. We therefore undertook a population-based observational study of all death counts in the US over 5 decades to examine temporal trends in SSc mortality overall and by sex and race. In order to evaluate SSc mortality in the context of changes in overall mortality

Supported in part by the NIH (grants R01-AI080778 and R01-AR056465 from the National Institute of Allergy and Infectious Diseases and the National Institute of Arthritis and Musculoskeletal and Skin Diseases, respectively). Dr. Yen's work was supported by NIH grants T32-DK-07789 and T32-HD-007512 from the National Institute of Diabetes and Digestive and Kidney Diseases and the National Institute of Child Health and Human Development, respectively.

[1]Eric Y. Yen, MD, MS: University of California David Geffen School of Medicine, Los Angeles; [2]Devanshu R. Singh, BS, BA: University of California David Geffen School of Medicine, Los Angeles, and Whiting School of Engineering and Zanvyl Krieger School of Arts and Sciences, Johns Hopkins University, Baltimore, Maryland; [3]Ram R. Singh, MD: University of California David Geffen School of Medicine, and Jonsson Comprehensive Cancer Center, Los Angeles.

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

Address correspondence to Ram R. Singh, MD, Autoimmunity and Tolerance Laboratory, Division of Rheumatology, University of California, Los Angeles, 1000 Veteran Avenue, Room 32-59, Los Angeles, CA 90095-1670. Email: rrsingh@mednet.ucla.edu.

Submitted for publication September 11, 2019; accepted in revised form August 4, 2020.

SIGNIFICANCE & INNOVATIONS

- Our joinpoint regression analysis of 55,861 systemic sclerosis deaths in the US shows that mortality attributable to systemic sclerosis has steadily decreased in the last decade and a half after continuously increasing over the previous 33 years.
- Systemic sclerosis mortality has disproportionately changed by sex and race over the 48-year period, with an increase shown in in women and White persons and a decrease in men and Black persons.
- The recent, steady improvement in the systemic sclerosis mortality rate could have resulted from advances in the management of its complications, such as pulmonary hypertension and renal crisis.

in the US population, we compared SSc mortality to non-SSc mortality from all other causes.

PATIENTS AND METHODS

Data sources. The Centers for Disease Control and Prevention (CDC) National Vital Statistics System maintains a mortality database that encompasses >99% of deaths of US residents in all 50 US states and Washington, DC. We used the CDC Wide-Ranging Online Data for Epidemiologic Research (WONDER) website application (10) to gather data on SSc deaths from 1968 (the earliest year for which the CDC published county-level mortality data) through 2015.

The underlying cause of death, defined as "the disease or injury that initiated the events resulting in death" is provided on the death certificate as an International Classification of Diseases (ICD) code (11,12). We identified specific ICD codes for SSc (ICD-8 [734.0], ICD-9 [710.1], and ICD-10 [M34]) and used standard methodology to ascertain race (13). We obtained annual death counts of the entire US population and separately by sex (men and women) and race (White, Black, and "other"). Information on Hispanic ethnicity and Asian or Pacific Islander, or American Indian or Alaska Native racial categories is not available before 1999 in the CDC WONDER.

Since 1999, information on the contributing cause of death, defined as "other significant conditions contributing to death but not resulting in the underlying cause" (12), became available on CDC WONDER. To address the possibility that SSc may have directly contributed to death but was not listed as the underlying cause of death, such as patients who died of SSc-related complications, we reanalyzed our mortality data where SSc was listed as a contributing cause of death. For calculation of mortality rates, we obtained the size of the population (total and each group) from the US Census Bureau for each year.

Annual mortality rates. We quantified age-specific crude mortality rates for SSc and non-SSc for each year from 1968 to 2015 as the number of deaths in each year divided by the number of persons in the US general population in that same year. This was done within age strata (Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24411/abstract) for the total US population as well as separately for men and women and each of the 3 racial groups.

To calculate the overall age-standardized mortality rate (ASMR) for the population for each year from 1968 to 2015, we combined the yearly age-specific crude mortality rates with the age distribution of the US population in 2000, as described in Supplementary Table 1 (14). We performed this analysis separately for the total US population, sex, and race for both SSc and non-SSc deaths. We then computed the ratio of the SSc ASMR to the non-SSc ASMR for each year.

Statistical analysis. We used a joinpoint regression model to assess trends in the annual SSc ASMR, non-SSc ASMR, and SSc to non-SSc ASMR ratio. Joinpoint regression analysis identifies changes in trend data by fitting a set of joinpoints, the calendar years at which the change in the slope (of ASMR) is statistically significant, over the entire period. The model then computes the slope (year-to-year percentage change in annual ASMR) and the 95% confidence interval (95% CI) over each linear trend segment

Table 1. Demographic characteristics of systemic sclerosis (SSc) and non-SSc deaths, 1968–2015

Characteristic	SSc deaths*		Non-SSc deaths*		Average population size†	
	No.	%	No.	%	No.	%
Total deaths	46,798		106,058,839		258,208,302	
Sex						
Male	10,150	21.7	54,915,106	51.8	126,273,136	48.9
Female	36,648	78.3	51,143,733	48.2	131,935,166	51.1
Race						
White	38,060	81.3	91,729,629	86.5	214,348,874	83.0
Black	7,630	16.3	12,555,274	11.8	32,518,150	12.6
Other race‡	1,108	2.4	1,773,936	1.7	11,341,278	4.4

* Absolute number of deaths from all 50 US states and Washington, DC. The annual death count ranged from 466 to 1401 for SSc and 1,898,350 to 2,700,382 for non-SSc causes.
† Average annual population derived from US Census Bureau files.
‡ Information on Asian or Pacific Islander, or American Indian or Alaska Native racial categories and on Hispanic ethnicity is not available before 1999 and is therefore not shown.

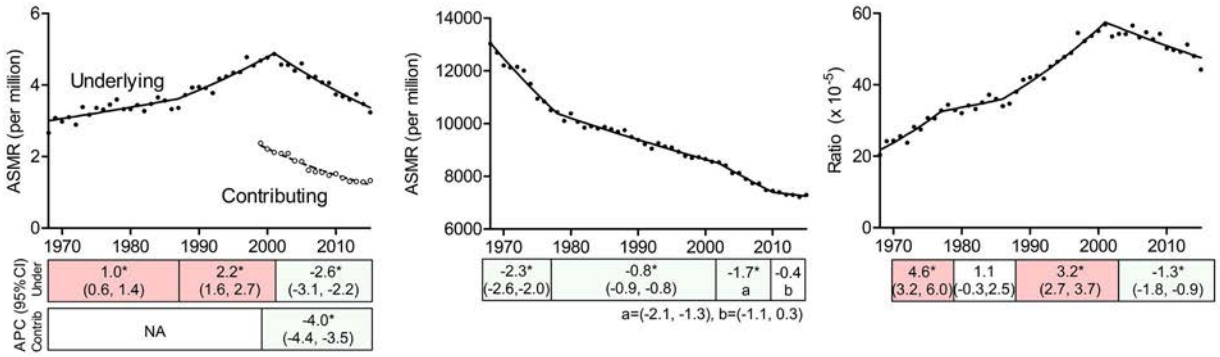


Figure 1. Trends in age-standardized mortality rate (ASMR) for systemic sclerosis (SSc) and non-SSc causes, and ratio of SSc to non-SSc mortality rates, 1968–2015. ASMRs per million persons for SSc causes (left), non-SSc causes (middle), and SSc as a contributing (Contrib) cause of death (left) are shown. Data are displayed per calendar year of death with lines fitted based on joinpoint analysis. The ratio of SSc ASMRs to non-SSc ASMRs ($\times 10^{-5}$) is also shown (right). A positive slope indicates an increased risk for death from SSc versus non-SSc causes, whereas a negative slope indicates a decreased risk of death from SSc. Stacked bars under each panel represent the annual percent change (APC) for each trend in SSc ASMRs, non-SSc ASMRs, and the ratio of SSc ASMRs to non-SSc ASMRs. Each stack is segmented at the year in which the change in slope is statistically significant and is aligned with the trend line. Numbers in each stack denote the APC (95% confidence interval [95% CI]). Red-shaded stacks indicate an increasing trend, unshaded stacks indicate a nonsignificant trend, and light green–shaded stacks indicate a decreasing trend. * = $P < 0.05$ for slope change. NA = not available; Under = underlying cause of death.

between adjacent joinpoints, as previously described (14,15). We calculated the annual percent change (APC) with 95% CIs for each trend, and then determined the average APC with 95% CIs for the entire study period (Supplementary Methods, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24411/abstract).

RESULTS

There were 46,798 deaths with SSc listed as the underlying cause and 106,058,839 non-SSc deaths in the US from 1968 to 2015. Additionally, we identified 9,063 deaths with SSc listed as a contributing cause from 1999 to 2015. The proportions of deaths among women and non-White persons were higher for SSc than for non-SSc (Table 1).

Mortality trends for SSc. The ASMR for SSc was 2.7 (95% CI 2.4, 2.9) per million persons in 1968. The SSc ASMR increased at an APC of 1.0% from 1968 to 1987 and then increased at a higher APC (2.2%) from 1987 to 2001 before decreasing starting in 2001 (APC –2.6% for 2001–2015) (Figure 1). Despite the steady decrease for 15 years, the SSc ASMR was 3.2 (95% CI 3.0, 3.4) per million persons in 2015, which was 18.5% higher in 2015 than in 1968 (Table 2). In contrast, the ASMR for non-SSc decreased continuously between 1968 and 2015 (Figure 1).

To highlight the changes in SSc mortality relative to non-SSc mortality, we calculated the ratio of the SSc ASMR to the non-SSc ASMR (Figure 1). The ratio increased at higher rates between 1968 and 2001, followed by decreases each year starting in 2001, indicating decreases in the proportion of US deaths from SSc during 2001–2015. However, the SSc ASMR to non-SSc ASMR ratio was still 111.6% higher in 2015 than in 1968 (Table 2).

The reduction in SSc ASMR after 2001 could be due to SSc not being recorded as the underlying cause on the death certificates for some patients. To address this possibility, we conducted a joinpoint analysis for deaths where a contributing cause was SSc (Figure 1). Like trends observed for deaths where the underlying cause was SSc, the ASMR for SSc as a contributing cause decreased from 1999 to 2015.

SSc mortality trends by sex and race. The SSc ASMR was higher in women (3.5 [95% CI 3.1, 3.9]) than in men (1.8 [95% CI 1.5, 2.1]) in 1968 (Table 2). From 1968 to the early 2000s, SSc ASMR increased at a higher APC in women (1.1–2.9%) than in men (0.3%), followed by similar decreases observed among both women and men (–2.3% and –3.0%, respectively) (Figure 2). This trend resulted in a cumulative change of +40.0% in women and –22.2% in men (Table 2) at an average APC of +0.5% in women and –0.6% in men over the 48 years (Supplementary Table 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24411/abstract).

Black persons had higher SSc ASMRs than White persons, with SSc ASMRs of 4.9 (95% CI 3.8, 6.1) and 2.4 (95% CI 2.2, 2.6), respectively, in 1968 (Table 2). After that, White persons had higher annual increases in the SSc ASMR than Black persons from 1968 through the early 2000s, when it began to decrease at similar APCs for each race (Figure 2). Over the 48 years, the SSc ASMR increased in White persons (with a cumulative increase of 33.3%), whereas the SSc ASMR decreased in Black persons

Table 2. Cumulative percentage change in SSc ASMR, non-SSc ASMR, and ratio of SSc to non-SSc ASMR, 1986–2015*

	SSc					Non-SSc†			SSc:Non-SSc ASMR ratio		
	1968		2015			1968	2015		1968	2015	
Variable	ASMR per million (95% CI)	No. of deaths	ASMR per million (95% CI)	No. of deaths	% change 1968–2015	ASMR per million	ASMR per million	% change, 1968–2015	Ratio $\times 10^{-5}$	Ratio $\times 10^{-5}$	% change 1968–2015
All	2.7 (2.4, 2.9)	466	3.2 (3.0, 3.4)	1,195	18.5	13,032.9	7,298.1	−44.0	20.7	43.8	111.6
Sex											
Male	1.8 (1.5, 2.1)	145	1.4 (1.2, 1.5)	227	−22.2	16,334.6	8,591.6	−47.4	11.0	16.3	47.9
Female	3.5 (3.1, 3.9)	321	4.9 (4.6, 5.2)	968	40.0	10,422.2	6,215.7	−40.4	33.6	78.8	134.7
Race											
White	2.4 (2.2, 2.6)	378	3.2 (3.0, 3.4)	987	33.3	12,706.2	7,319.7	−42.4	18.9	43.7	131.5
Black	4.9 (3.8, 6.1)	80	3.9 (3.3, 4.5)	166	−20.4	16,118.3	8,463.8	−47.5	30.4	46.1	51.6
Other race	NA‡		1.8 (1.3, 2.5)	42	42	9,29.4	4,253.1	−54.2	NA‡	42.3	NA‡

* 95% CI = 95% confidence interval; ASMR = age-standardized mortality rate; NA = not available.
† Total non–systemic sclerosis (non-SSc) deaths in 1968 and 2015 were 1,928,931 and 2,700,382, respectively.
‡ Data are not shown because of small sample size (<20 SSc deaths) in 1968 in this subpopulation.

(with a cumulative decrease of –20.4%) (Table 2), with an average APC of 0.4% in White persons and –0.4% in Black persons (Supplementary Table 2).

In contrast to the increase-and-decrease trends in SSc ASMRs, non-SSc ASMRs decreased or stayed stable throughout the 48 years in all subpopulations (Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24411/abstract). Consequently, the SSc to non-SSc ASMR ratio initially increased at greater APCs in all demographic subgroups studied (Figure 2B). Starting in the 2000s, the ratio decreased in all subpopulations, although the differences were not statistically significant in Black persons and other races. All subpopulations studied had a relative cumulative increase in the ratio from 1968 to 2015 (Table 2).

DISCUSSION

In addition to large-scale, population-based prospective studies covering almost all SSc deaths in the population, the national death certificates database remains a useful source for an unbiased assessment of the mortality burden from a disease at the population level. This burden of disease data may be useful for healthcare policy planning, resource allocation, identification of high-risk population, and assessment of changes in disease management at the population level. This information cannot be obtained from limited cohort-based studies at referral centers, as the cause-specific mortality may vary due to changes in the incidence of disease, disease severity, or both over time.

Our joinpoint regression analysis of 55,861 SSc deaths in the US shows that mortality attributable to SSc has steadily decreased in the last decade and a half after 33 years of sustained increase from 1968 to 2000. Despite these 15 years of steady improvement, mortality rates for SSc relative to non-SSc were still higher in 2015 than in 1968. The rise-and-decline trend in SSc mortality may reflect changes in disease incidence, SSc recognition, improved evaluation, and/or better management, as illustrated in Figure 3. We explore some possible reasons for changes in SSc mortality rates over five decades below.

First, SSc incidence rates in the US increased from 0.6 new cases per million persons annually in the 1940s (in Tennessee [16]) to 12 cases per million per year (Minnesota) (17), 19 cases per million per year in the 1990s (Michigan) (9), and 28–33 cases per million per year during the 1990s to 2007 (Utah) (18) (Figure 3). Analysis of US administrative health care data sets also suggests a higher prevalence of SSc (0.03–0.05%) from 2001 to 2002 (19) than previously reported in 1991 (0.02%) (9). Taken together, the incidence of SSc in the US increased between the 1960s and 1970s (16,20) and doubled between the 1970s and 1980s (20). SSc incidence then stabilized over the next decade (9). Thus, changes in SSc incidence over time could partially explain the observed rising trends in SSc mortality from the 1960s to the 1990s.

Second, the establishment of SSc classification criteria in 1980 (21) and the introduction of SSc-associated autoantibodies in the 1980s could have contributed to the increase in the number of diagnosed cases, and subsequently, to the attribution of death to SSc. However, the increase in SSc recognition cannot explain the mortality trends in recent years. The development of new criteria for the classification of early SSc in 2001 (22) and identification of the relationship between autoantibodies and prognosis (23,24) appears to have coincided with improvement in SSc mortality (Figure 3). It is possible that the description of skin scores and autoantibodies as predictors of disease course in the 1990s (23,24) might have helped with early intervention, and subsequently, with declining SSc mortality in the 2000s. The development of classification criteria for early SSc in 2001 (22) may have also helped with early diagnosis and treatment of the disease, thus contributing to the improved SSc outcomes observed in recent years.

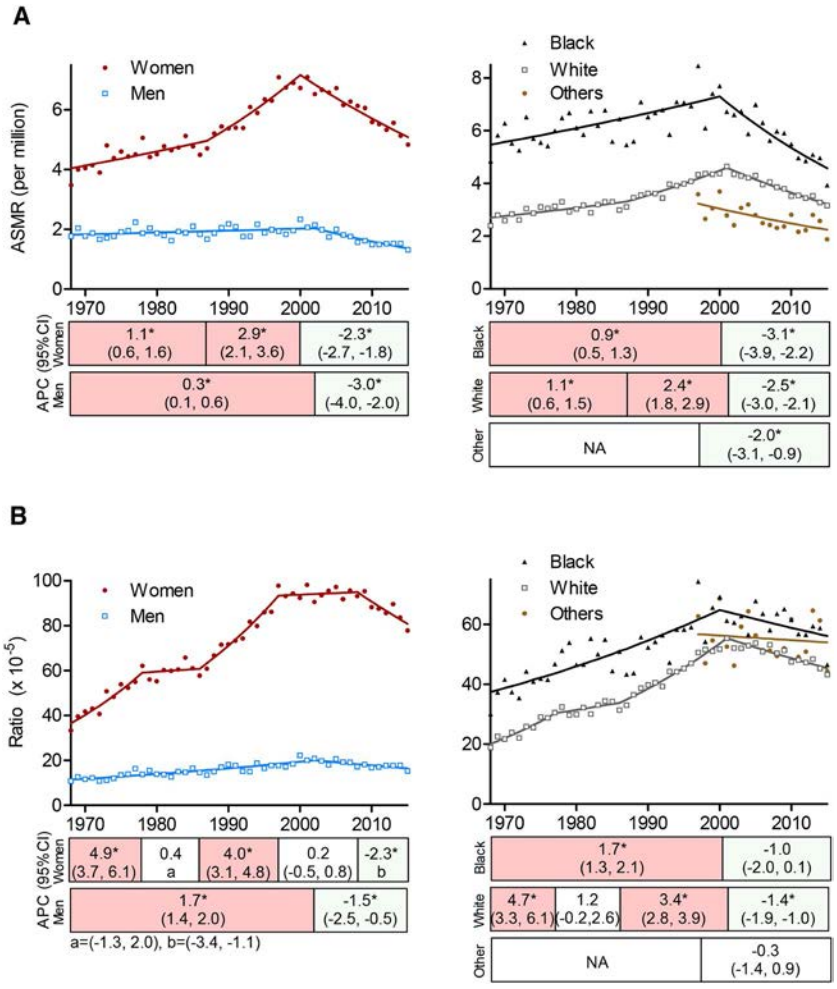


Figure 2. Trends in the systemic sclersosis (SSc) age-standardized mortality rate (ASMR) and in the ratio of SSc mortality rates to non-SSc mortality rates by sex and race, 1968–2015. Results are shown as ASMR for SSc per million persons (**A**) and the ratio of SSc mortality rates to non-SSc mortality rates (**B**). Annual number of SSc deaths ranged 321–1,116 among women, 144–293 among men, 378–1,138 among White persons, and 80–234 among Black persons. Data for other races are shown only for the 1997–2015 period (n = 26–58 deaths) as data before 1997 are unreliable due to small numbers of reported SSc deaths (less than 20) per year in this subpopulation. Stacked bars below each panel represent the annual percent change (APC) for each trend for each subpopulation. See Figure 1 for an explanation of data under each panel. * = $P < 0.05$ for slope change. 95% CI = 95% confidence interval; NA = not available.

Third, medication toxicities in the 1960s to 1980s and new treatments for SSc complications in the 1990s and 2000s might have influenced mortality trends (Figure 3). D-penicillamine was used to treat SSc starting in the 1960s (25) until studies performed in the 1990s highlighted its inefficacy (26) and toxicity (26,27). Additionally, glucocorticoids used in the 1960s to the 1980s were implicated in precipitating SSc renal crisis (28), which was associated with high mortality, until the recognition in the 1990s of the beneficial effects of angiotensin-converting enzyme inhibitors in preventing this complication (29). Furthermore, the benefits of early screening for alveolitis and selected use of cyclophosphamide in SSc-related lung disease were reported in the early 2000s (30,31). Finally, the introduction in the early 2000s of prostanoids, endothelin receptor antagonists (ERAs), and phosphodiesterase 5 inhibitors to treat SSc-associated pulmonary arterial hypertension has led to improved SSc outcomes (2,3,32). Together, multiple factors likely influenced the SSc mortality trends that we observed.

Conflicting results have been published about sex differences in SSc mortality. Standardized mortality ratios for SSc were similar between men and women in some studies (7,33), but were higher in men compared to women in other studies (6). However, the standardized mortality ratios calculated in these studies may not have accounted for a higher incidence of SSc in women, which appears to have increased between 1972 and 1982 (from 13 to 27.6 per million per year). In contrast, the incidence of SSc in men has remained relatively stable over this period (20). These observations may explain our finding that SSc mortality rates, which were higher in women than men in 1968, increased at an even higher rate in women than in men until 2000. Since ASMR depends on both the prevalence and severity of the disease, we cannot exclude the possibility that SSc may be more severe in men than

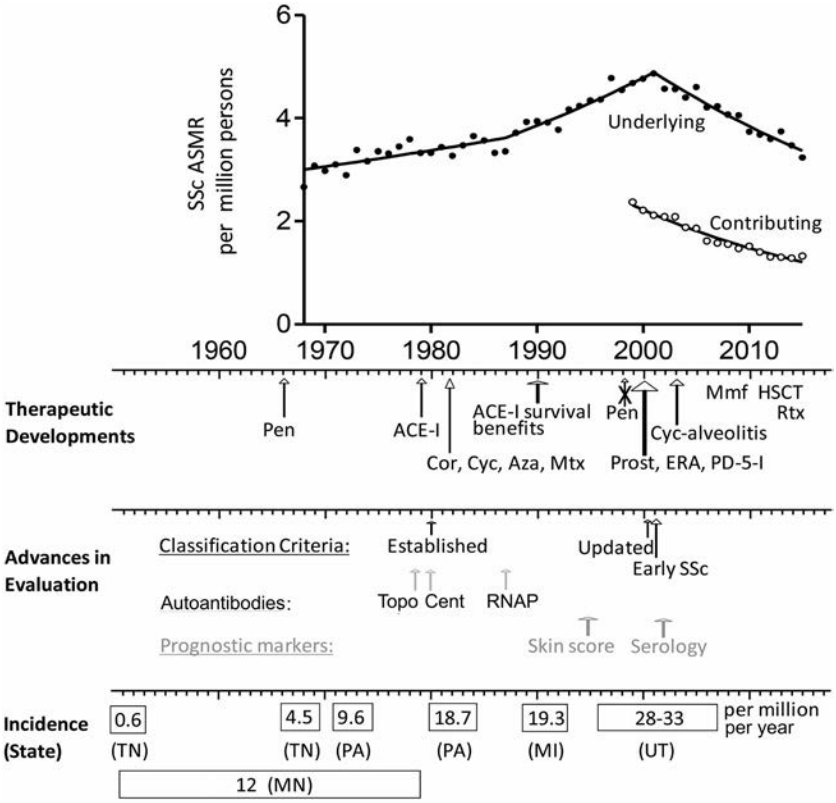


Figure 3. Evaluation and management milestones and changes in incidence of systemic sclerosis (SSc) over time in relation to SSc mortality trends, 1968–2015. Age-adjusted morality rate (ASMR) per million persons for SSc as the underlying or contributing causes of death is shown (graph at top) in relation to the timeline for major therapeutic developments for SSc (top row), the timeline for major advances in SSc evaluation (middle row), and changes in SSc incidence in different US states over time (bottom row). ACE-I = angiotensin-converting enzyme inhibitor; Aza = azathioprine; Cent = anticentromere antibody; Cor = corticosteroids; Cyc = cyclophosphamide; HSCT = hematopoietic stem cell transplant; MI = Michigan; Mmf = mycophenolate mofetil; Mtx = methotrexate; PD-5-I = phosphodiesterase-5 inhibitor; Pen = D-penicillamine; Prost = prostanoids; RNAP = anti-RNA polymerase antibodies; Rtx = rituximab; TN = Tennessee; Topo = anti-topoisomerase (Scl-70) antibody UT = Utah.

women. Accurate SSc prevalence rates across the US are needed to calculate the case-fatality rate that adjusts for differences in SSc prevalence between men and women. It is also plausible that SSc was recognized or recorded on death certificates more often for women than for men from 1968 to 2000. Nevertheless, our data show that the ASMR that represents SSc mortality burden in the general population is higher in women than in men.

As reported previously (8,34), SSc mortality rates were higher in Black persons than in White persons. The potential causes for this disparity in SSc mortality may include higher SSc incidence, more severe disease, and younger age at diagnosis in Black persons (35). However, SSc ASMRs increased at a higher rate in White persons than in Black persons from 1968 until the early 2000s, when it began to decrease at similar rates in both races. Furthermore, SSc mortality in Black persons showed a cumulative decrease between 1968 and 2015 (–20%), whereas it increased in White persons during the same period (+33%). It is unclear, whether from 1968–2001, White persons experienced more severe disease activity, higher prevalence of disease, increased recognition of SSc, or improved recording of SSc on the death certificates as compared to Black persons. Further analyses of mortality by race/ethnicity could not be reliably evaluated using the CDC WONDER database, which does not have information on Hispanic, Asian, or Native American categories before 1999.

Strengths of this study include the use of an unbiased, systematic approach to assess SSc mortality in a large sample size comprising all recorded deaths in the US over 48 years, the use of joinpoint regression analysis as a computational approach to identify trends, and computation of the ratio that compares SSc mortality relative to non-SSc mortality. Calculation of the standardized mortality ratio, reported in previous studies on SSc mortality (4–6), uses an indirect method of adjustment that depends on the age structure of the study population (SSc, in this case) (36). However, the age structure may vary between different study cohorts, across populations of different regions and countries, and over time (37). Thus, the standardized mortality ratios computed for one population may not represent SSc mortality in another population. Therefore, we performed a direct method of adjustment using a standard population to calculate the ASMR for both SSc and non-SSc causes (36).

Our study has limitations. First, the validity of our findings depends on the accuracy of the physicians' coding of causes of

death recorded on death certificates, which is difficult to ascertain. Nevertheless, the strong associations that we detected cannot be explained by a low specificity on the cause of death recorded on death certificates, as random misclassification increases the similarity between the study population and the general population, creating an underestimation of the risk estimates. To the best of our knowledge, there are no reports on misclassification of SSc on the death certificates (i.e., SSc recorded on the death certificate for the decedent that did not have SSc). While misclassification on death certificates has been reported in other autoimmune diseases, it is still rare. For example, for the 731 decedents for whom lupus was recorded on the death certificate, only 2 had lupus erroneously recorded as a cause of death (i.e., decedent did not have lupus) (38). Death certificates are unlike electronic health records and claims databases where a substantial proportion of subjects (about 25% in one study) coded as having SSc did not fulfill criteria for SSc (39,40), likely due to the entry of probable/possible/working diagnoses. In contrast, physicians who encounter patients at the time of death are likely to record the disease that was the most proximate or probable cause of death recorded on death certificates. Thus, misclassification of SSc on the death certificates is less likely to have influenced SSc mortality rates substantially.

Second, SSc might be left off the death certificate in some portion of the 30–50% of deaths among SSc patients that were caused by infection, cancer, and cardiovascular disease (7,33). These proximate causes of death may be perceived to be unrelated to SSc, when in fact, the disease or the medications used for it predispose patients to them. Such underestimation of cause-specific mortality has been reported in other autoimmune diseases. For example, multiple sclerosis was not mentioned on death certificates in 6–27% of patients who died of unrelated causes at an older age (41), and lupus was not recorded on the death certificates of some patients who died of complications such as infections, cardiovascular events, and respiratory diseases (42). Increasing awareness among primary care physicians and internists about the multi-organ complications of autoimmune diseases such as SSc and their varying presentations at the time of death would help assess the actual burden of SSc mortality in the future.

Third, the underreporting of SSc on death certificates might selectively occur in specific subpopulations. For example, SLE was not recorded on the death certificate in older patients, those without health insurance, and those with low education levels in other autoimmune rheumatic diseases (43,44). Nevertheless, the significant differences in SSc mortality by sex and race are less likely to be artifacts from the misclassification of cause of death in any meaningful way because greater underreporting of SSc as the cause of mortality in underprivileged groups would lead to a larger underestimation of SSc mortality in the groups that we found mortality to be higher in, e.g., female patients and Black persons. Our data show that mortality attributable to SSc increased at higher rates in women and White persons relative to men and Black persons, respectively, from the 1970s to the 1990s. This finding raises the possibility of whether factors related to disease evaluation, healthcare delivery, or socioeconomic issues led to a differential recognition and reporting of SSc by sex and race during this period.

Fourth, ICD revisions between ICD-8 and ICD-9 and between ICD-9 and ICD-10 might have influenced the estimation of mortality trends. However, studies that measured the effects of ICD revisions have reported good comparability ratios for disease classification between revisions (45,46).

Finally, changes in the physicians' reporting of SSc as the underlying cause of death versus contributing cause of death over time could have influenced SSc mortality trends. However, analysis of deaths for which SSc was recorded as a contributing cause in 1999–2015 showed a similar pattern, suggesting that errors in the coding of cause of death did not substantially bias the findings at least over the last 15 years.

In conclusion, SSc mortality has begun to improve after the year 2000. Still, the improvement in SSc mortality has not kept up with an improvement in mortality from other causes in the general population. Comprehensive examination using prospective, population-based data could help clarify the mechanisms of the changing disparities in SSc mortality and identify modifiable risk factors that could be altered to improve outcomes.

ACKNOWLEDGMENTS

The authors thank Drs. Roy Altman, John Fitzgerald, Antonio La Cava, Anand Malaviya, and John Rambharose for critically reading 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 published. Drs. Yen and R. Singh had full access to all the study data and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Yen, R. Singh.
Acquisition of data. Yen.
Analysis and/or interpretation of data. Yen, D. Singh, R. Singh.

REFERENCES

1. Kumar A, Malaviya AN, Tiwari SC, Singh RR, Kumar A, Pande JN. Clinical and laboratory profile of systemic sclerosis in northern India. J Assoc Physicians India 1990;38:765–8.
2. Denton CP, Khanna D. Systemic sclerosis. Lancet 2017;390: 1685–99.
3. Williams MH, Das C, Handler CE, Akram MR, Davar J, Denton CP, et al. Systemic sclerosis associated pulmonary hypertension: improved survival in the current era. Heart 2006;92:926–32.
4. Rubio-Rivas M, Royo C, Simeon CP, Corbella X, Fonollosa V. Mortality and survival in systemic sclerosis: systematic review and meta-analysis. Semin Arthritis Rheum 2014;44:208–19.
5. Elhai M, Meune C, Avouac J, Kahan A, Allanore Y. Trends in mortality in patients with systemic sclerosis over 40 years: a systematic review and meta-analysis of cohort studies. Rheumatology (Oxford) 2012;51:1017–26.

6. Hao Y, Hudson M, Baron M, Carreira P, Stevens W, Rabusa C, et al. Early mortality in a multinational systemic sclerosis inception cohort. Arthritis Rheumatol 2017;69:1067–77.
7. Jacobsen S, Halberg P, Ullman S. Mortality and causes of death of 344 Danish patients with systemic sclerosis (scleroderma). Br J Rheumatol 1998;37:750–5.
8. Krishnan E, Furst DE. Systemic sclerosis mortality in the United States: 1979–1998. Eur J Epidemiol 2005;20:855–61.
9. Mayes MD, Lacey JV Jr, Beebe-Dimmer J, Gillespie BW, Cooper B, Laing TJ, et al. Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. Arthritis Rheum 2003;48:2246–55.
10. Centers for Disease Control and Prevention. CDC WONDER online databases. URL: https://wonder.cdc.gov/.
11. Centers for Disease Control and Prevention. Section I - Instructions for classifying the underlying cause of death. ICD-10 Mortality Manual 2a. URL: http://www.cdc.gov/nchs/data/dvs/2a_2015.pdf.
12. Yen EY, Singh RR. Lupus–an unrecognized leading cause of death in young women: population-based study using nationwide death certificates, 2000–2015. Arthritis Rheumatol 2018;70:1251–55.
13. Centers for Disease Control and Prevention. Vital statistics of the United States: mortality. Technical appendix. URL: http://www.cdc.gov/nchs/data/statab/techap99.pdf.
14. Yen EY, Shaheen M, Woo JMP, Mercer N, Li N, McCurdy DK, et al. 46-year trends in systemic lupus erythematosus mortality in the United States, 1968 to 2013: a nationwide population-based study. Ann Intern Med 2017;167:777–85.
15. Clegg LX, Hankey BF, Tiwari R, Feuer EJ, Edwards BK. Estimating average annual per cent change in trend analysis. Statistics in medicine 2009;28:3670–82.
16. Medsger TA Jr, Masi AT. Epidemiology of systemic sclerosis (scleroderma). Ann Intern Med 1971;74:714–21.
17. Michet CJ Jr, McKenna CH, Elveback LR, Kaslow RA, Kurland LT. Epidemiology of systemic lupus erythematosus and other connective tissue diseases in Rochester, Minnesota, 1950 through 1979. Mayo Clin Proc 1985;60:105–13.
18. Frech T, Khanna D, Markewitz B, Mineau G, Pimentel R, Sawitzke A. Heritability of vasculopathy, autoimmune disease, and fibrosis in systemic sclerosis: a population-based study. Arthritis Rheum 2010;62:2109–16.
19. Robinson D Jr, Eisenberg D, Nietert PJ, Doyle M, Bala M, Paramore C, et al. Systemic sclerosis prevalence and comorbidities in the US, 2001–2002. Curr Med Res Opin 2008;24:115766.
20. Steen VD, Oddis CV, Conte CG, Janoski J, Casterline GZ, Medsger TA Jr. Incidence of systemic sclerosis in Allegheny County, Pennsylvania. A twenty-year study of hospital-diagnosed cases, 1963–1982. Arthritis Rheum 1997;40:441–5.
21. Masi AT, Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Arthritis Rheum 1980;23:581–90.
22. LeRoy EC, Medsger TA Jr. Criteria for the classification of early systemic sclerosis. J Rheumatol 2001;28:1573–6.
23. Clements PJ, Hurwitz EL, Wong WK, Seibold JR, Mayes M, White B, et al. Skin thickness score as a predictor and correlate of outcome in systemic sclerosis: high-dose versus low-dose penicillamine trial. Arthritis Rheum 2000;43:2445–54.
24. Ho KT, Reveille JD. The clinical relevance of autoantibodies in scleroderma. Arthritis Res Ther 2003;5:80–93.
25. Harris ED Jr, Sjoerdsma A. Effect of penicillamine on human collagen and its possible application to treatment of scleroderma. Lancet 1966;2:996–9.
26. Clements PJ, Furst DE, Wong WK, Mayes M, White B, Wigley F, et al. High-dose versus low-dose D-penicillamine in early diffuse systemic sclerosis: analysis of a two-year, double-blind, randomized, controlled clinical trial. Arthritis Rheum 1999;42:1194–203.
27. Taneja V, Mehra N, Singh YN, Kumar A, Malaviya A, Singh RR. HLA-D region genes and susceptibility to D-penicillamine-induced myositis. Arthritis Rheum 1990;33:1445–7.
28. Trang G, Steele R, Baron M, Hudson M. Corticosteroids and the risk of scleroderma renal crisis: a systematic review. Rheumatol Int 2012;32:645–53.
29. Steen VD, Costantino JP, Shapiro AP, Medsger TA Jr. Outcome of renal crisis in systemic sclerosis: relation to availability of angiotensin converting enzyme (ACE) inhibitors. Ann Intern Med 1990;113:352–7.
30. Sahhar J, Littlejohn G, Conron M. Fibrosing alveolitis in systemic sclerosis: the need for early screening and treatment. Intern Med J 2004;34:626–38.
31. Tashkin DP, Elashoff R, Clements PJ, Goldin J, Roth MD, Furst DE, et al. Cyclophosphamide versus placebo in scleroderma lung disease. N Engl J Med 2006;354:2655–66.
32. Badesch DB, Tapson VF, McGoon MD, Brundage BH, Rubin LJ, Wigley FM, et al. Continuous intravenous epoprostenol for pulmonary hypertension due to the scleroderma spectrum of disease. A randomized, controlled trial. Ann Intern Med 2000;132:425–34.
33. Bryan C, Howard Y, Brennan P, Black C, Silman A. Survival following the onset of scleroderma: results from a retrospective inception cohort study of the UK patient population. Br J Rheumatol 1996;35:1122–6.
34. Chung L, Krishnan E, Chakravarty EF. Hospitalizations and mortality in systemic sclerosis: results from the Nationwide Inpatient Sample. Rheumatology (Oxford) 2007;46:1808–13.
35. Laing TJ, Gillespie BW, Toth MB, Mayes MD, Gallavan RH Jr, Burns CJ, et al. Racial differences in scleroderma among women in Michigan. Arthritis Rheum 1997;40:734–42.
36. Schoenbach VJ, Rosamond WD. Standardization of rates and ratios. In: Understanding the fundamentals of epidemiology: an evolving text. URL: http://www.epidemiolog.net/evolving/Standardization.pdf.
37. Ahmad OB, Boschi-Pinto C, Lopez AD, Murray CJ, Lozano R, Inoue M. Age Standardization of Rates: A New WHO standard. World Health Organization. URL: http://www.who.int/healthinfo/paper31.pdf.
38. Falasinnu T, Rossides M, Chaichian Y, Simard JF. Do death certificates underestimate the burden of rare diseases? The example of systemic lupus erythematosus mortality, Sweden, 2001–2013. Public Health Rep 2018;133:481–8.
39. Jamian L, Wheless L, Crofford LJ, Barnado A. Rule-based and machine learning algorithms identify patients with systemic sclerosis accurately in the electronic health record. Arthritis Res Ther 2019;21:305.
40. Valenzuela A, Yaqub A, Fiorentino D, Krishnan E, Chung L. Validation of the ICD-9-CM code for systemic sclerosis using updated ACR/EULAR classification criteria. Scand J Rheumatol 2015;44:253–5.
41. Hirst C, Swingler R, Compston DA, Ben-Shlomo Y, Robertson NP. Survival and cause of death in multiple sclerosis: a prospective population-based study. J Neurol Neurosurg Psychiatry 2008;79:1016–21.
42. Thomas G, Mancini J, Jourde-Chiche N, Sarlon G, Amoura Z, Harle JR, et al. Mortality associated with systemic lupus erythematosus in France assessed by multiple-cause-of-death analysis. Arthritis Rheumatol 2014;66:2503–11.

43. Calvo-Alen J, Alarcon GS, Campbell R Jr, Fernandez M, Reveille JD, Cooper GS. Lack of recording of systemic lupus erythematosus in the death certificates of lupus patients. Rheumatology (Oxford) 2005;44:1186–9.
44. Ward MM. Education level and mortality in systemic lupus erythematosus (SLE): evidence of underascertainment of deaths due to SLE in ethnic minorities with low education levels. Arthritis Rheum 2004;51:616–24.
45. Anderson RN, Miniño AM, Hoyert DL, Rosenberg HM. Comparability of cause of death between ICD-9 and ICD-10: preliminary estimates. 2001;49:1–32.
46. Klebba AH, Scott JH. Estimates of selected comparability ratios based on dual coding of 1976 death certificates by the Eight and Ninth Revisions of the International Classification of Diseases. 1980. URL: https://www.cdc.gov/nchs/data/mvsr/supp/mv28_11s.pdf.

Arthritis Care & Research
Vol. 73, No. 10, October 2021, pp 1511–1517
DOI 10.1002/acr.24353

American College of Rheumatology
Empowering Rheumatology Professionals

Lifetime Risk of Primary Shoulder Arthroplasty From 2008 to 2017: A Population-Level Analysis Using National Registry Data

Daisuke Miura,[1] Ljoudmila Busija,[1] Richard S. Page,[2] Richard de Steiger,[3] Michelle Lorimer,[4] and Ilana N. Ackerman[1]

Objective. To estimate the lifetime risk of primary shoulder arthroplasty in Australia and to examine changes over time.

Methods. For this retrospective population-level analysis, de-identified individual-level data on all primary partial shoulder arthroplasty (PSA) and total shoulder arthroplasty (TSA) procedures performed in Australia from 2008 to 2017 (n = 38,868) were obtained from the Australian Orthopaedic Association National Joint Replacement Registry. Population data and life tables were obtained from the Australian Bureau of Statistics. Lifetime risk of primary shoulder arthroplasty was calculated for each year using a standardized formula. Separate calculations were undertaken by sex and for PSA and TSA.

Results. The lifetime risk of shoulder arthroplasty increased significantly over time. For men, this risk more than doubled from 0.78% (95% confidence interval [95% CI] 0.73–0.84) in 2008 to 1.78% (95% CI 1.70–1.86) in 2017. Lifetime risk for women rose from 1.54% (95% CI 1.46–1.62) to 2.88% (95% CI 2.78–2.99) over the study period. This increase was predominantly driven by growth in lifetime risk of TSA. In contrast, lifetime risk of PSA decreased over time, from 0.25% (95% CI 0.22–0.28) in 2008 to 0.11% (95% CI 0.09–0.13) in 2017 for men, and from 0.55% (95% CI 0.51–0.60) to 0.11% (95% CI 0.09–0.13) for women.

Conclusion. By the end of 2017, the lifetime risk of primary shoulder arthroplasty in Australia increased to 1 in 57 for men and 1 in 35 for women. Compared to declining PSA trends, there was substantial growth in TSA use over a decade. These data improve our understanding of the rising national burden of primary shoulder arthroplasty and can assist in planning to meet future surgical demand.

INTRODUCTION

Primary shoulder arthroplasty is an intervention increasingly used for end-stage shoulder disorders that no longer respond to nonoperative treatment. Shoulder arthroplasty is used by orthopedic surgeons to relieve patients from pain and to repair anatomical damage to the glenohumeral joint and surrounding structures (1). The 2 main classes of shoulder arthroplasty include partial shoulder arthroplasty (PSA), which involves replacement of 1 joint surface, or total shoulder arthroplasty (TSA), where both joint surfaces are replaced. Both types of procedures have been shown to be highly cost-effective for improving quality of life due to severe shoulder disease (2–4). One of the most frequent indications for shoulder arthroplasty is glenohumeral osteoarthritis (OA) (5,6), which is a relatively common condition among older people (7,8). Data from published reports and cohort studies indicate that the incidence of shoulder arthroplasty is rising over time in many health care regions and countries (9,10). In particular, the use of reverse TSA (a class of primary TSA that reverses the usual anatomical orientation of the glenohumeral joint) has increased internationally over the past decade (9). Growth in the use of reverse TSA likely relates to expanding clinical indications for this type of

Dr. Ackerman's work was supported by a Victorian Health and Medical Research fellowship awarded by the Victorian Government.

[1]Daisuke Miura, BMedSc (Hons), Ljoudmila Busija, BA (Hons), MSc (AppStat), PhD, Ilana N. Ackerman, BPhysio (Hons), PhD: School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia; [2]Richard S. Page, BMedSci, MBBS, FRACS (Orth), FAOrthA: Barwon Centre for Orthopaedic Research and Education, Barwon Health, St John of God Hospital, and School of Medicine, Deakin University, Geelong, Victoria, and Australian Orthopaedic Association National Joint Replacement Registry, Adelaide, South Australia, Australia; [3]Richard de Steiger: MBBS, FRACS (Orth), PhD: Epworth HealthCare and University of Melbourne, Melbourne, Victoria, and Australian Orthopaedic Association National Joint Replacement Registry, Adelaide, South Australia, Australia; [4]Michelle Lorimer, BSc (Hons): South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia.

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

Address correspondence to Ilana N. Ackerman, BPhysio (Hons), PhD, School of Public Health and Preventive Medicine, 553 St Kilda Road, Melbourne, Victoria 3004, Australia. Email: ilana.ackerman@monash.edu.

Submitted for publication March 2, 2020; accepted in revised form June 9, 2020.

SIGNIFICANCE & INNOVATIONS

- This national-level study, using comprehensive and well-validated registry data, is the first to quantify the lifetime risk of primary shoulder arthroplasty in any country.
- The study methods can be easily reapplied for ongoing population-level surveillance and to facilitate future international comparisons.
- From 2008 to 2017, the lifetime risk of primary total shoulder arthroplasty increased 3-fold for men, while lifetime risk over this period more than doubled for women.
- These data improve our understanding of the rising national burden of primary shoulder arthroplasty and can assist in planning to meet future surgical demand.

surgery, together with improved prostheses and a larger surgical workforce skilled in performing these procedures. The increasing use of primary shoulder arthroplasty is of public health importance given aging populations, the breadth of indications for surgery, and the relatively high cost of each procedure.

Australia has a longstanding national joint replacement registry that collects data on all arthroplasty procedures. These comprehensive data provide an opportunity to understand changes in the use of shoulder arthroplasty at a national level. Calculating the lifetime risk of shoulder arthroplasty quantifies the probability of a person having the procedure within their lifetime. Lifetime risk estimates can serve as a marker for disease burden in a population and can enable health funders and the health workforce to plan for likely future demand. These metrics can also be used by health care professionals to provide patient education and improve engagement with preventative health measures. Previous studies have quantified the lifetime risk of both total knee arthroplasty (TKA) and total hip arthroplasty (THA) in several countries (11–15). To date, this epidemiologic approach to the quantification of disease burden has not been applied to shoulder arthroplasty. This study aimed to estimate the lifetime risk of primary shoulder arthroplasty in Australia and to examine changes in lifetime risk over a 10-year period.

MATERIALS AND METHODS

Study design and ethics approval. This was a retrospective population-level analysis of national registry data. Ethics approval for this study was obtained from the Monash University Human Research Ethics Committee (#17924), Melbourne, Australia. Approval was also obtained from the Data Review Committee of the Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR), Adelaide, Australia. Consent to participate was not required for this study because only de-identified data were used.

Data sources. The AOANJRR collects arthroplasty data from all hospitals in all Australian states and territories that perform arthroplasty surgery. It has a near-complete capture rate (>98.8%) of all hip, knee, and shoulder arthroplasties performed in Australia (16). The AOANJRR first started collecting data on shoulder arthroplasty in 2004, and full national data collection was implemented from 2008; 2017 was the most recent year of data available for this study. AOANJRR data are validated against state and territory health department unit record data using a sequential multilevel matching process. The accuracy of this process indicates a negligible potential for missing data or misclassification of data, even in the early phase of data collection. De-identified individual-level data were obtained from the AOANJRR on all primary PSA and TSA procedures performed in Australia from January 1, 2008 to December 31, 2017 (n = 38,868 procedures), including patient age, patient sex, Australian state/territory, type of surgery (PSA, TSA, unilateral, or bilateral) and prosthesis class, primary shoulder diagnosis (e.g., OA, fracture, rotator cuff arthropathy, rheumatoid or inflammatory arthritis), date of surgery, and side of surgery (left or right). Life table data from 2006–2008 to 2015–2017 (stratified by sex) and data on Australia's population size (by age and sex) for each year from 2008 to 2017 were obtained from the Australian Bureau of Statistics (ABS) (17,18).

Data analysis. We used the same data analysis methods as previous studies that have estimated the lifetime risk of primary hip and knee arthroplasty in Australia and internationally (11–13). A "standardized lifetime risk" approach (19) incorporating age-specific utilization rates was used to calculate the lifetime risk of primary shoulder arthroplasty. Age data were categorized for analysis according to the following prespecified age groups by years: <40, 40–49, 50–59, 60–69, 70–79, and ≥80.

The formula used to calculate lifetime risk is shown in Figure 1. Lifetime risk was calculated for each age group by dividing the total number of people having primary shoulder arthroplasty (PSA plus TSA) procedures within that year (obtained from AOANJRR data) by the age group–specific and sex-specific population for that year (obtained from ABS population data). These rates were then multiplied by the total number of people expected to be alive at the middle of the age group interval (obtained from ABS life table data). A Poisson distribution was assumed when calculating 95% confidence intervals (95% CIs). Separate analyses were undertaken for men and women given known sex-based differences in the prevalence of joint pathologies such as OA (20). Separate analyses were also undertaken for PSA and TSA, given the differing range of clinical indications for these procedures. A subgroup analysis (examining the lifetime risk of reverse TSA arthroplasty and total stemmed arthroplasty, as the 2 most commonly used TSA prostheses) was also undertaken.

Consideration was given to bilateral shoulder arthroplasties (representing 2.4% of all primary shoulder arthroplasties in the

$$\text{Lifetime Risk} = \sum_{j=1}^{6} 10 \times S_j \times R_j$$

j: Indexes the 6 age groups (<40, 40-49, 50-59, 60-69, 70-79 and ≥80 years old)

S_j: Sex-specific population that survives to mid-year of respective age group for each year

R_j: Rate of people who have primary shoulder arthroplasty by sex-specific age group

Figure 1. Lifetime risk formula.

cohort) to avoid the potential overestimation of lifetime risk. Where these were performed, simultaneous bilateral shoulder arthroplasties (performed on the same day) were counted as 1 procedure to avoid potential overestimation of lifetime risk. Where staged (sequential) bilateral shoulder arthroplasty procedures were performed within the same year, only the first procedure was included in the lifetime risk analysis.

Where a clear pattern of linear increase or decrease in lifetime risk was evident, the Prais-Winsten regression method (21), which takes into account correlation between adjacent years, was used to test for linear trend. Lifetime risk estimates were calculated using Excel (Microsoft), version 16.20. Poisson CIs and linear trend analyses were carried out in Stata software, version 15.1.

RESULTS

Characteristics of the primary shoulder arthroplasty cohort. Table 1 summarizes the characteristics of all primary shoulder arthroplasties undertaken in Australia from 2008 to 2017. Of the 38,868 primary shoulder arthroplasties performed over this period, 5,979 (15%) were primary PSA procedures, and the majority (n = 32,889, 85%) were primary TSA procedures. Less than 3% of procedures were undertaken for younger patients (those age <50 years), with people age ≥60 years comprising the greatest proportion of individuals undergoing primary PSA or primary TSA. The sex distribution was similar for primary PSA and primary TSA, with 63.9% and 61.9% of procedures, respectively, performed on women. The most common operative diagnosis for primary PSA was fracture (2,724 of 5,979 cases, 45.6%), while for primary TSA the majority of procedures were performed for glenohumeral OA (21,324 of 32,889 cases, 64.8%). As shown in Table 1, the most frequently used class of primary PSA was hemi-stemmed arthroplasty. For primary TSA, total reverse arthroplasty was the class most commonly used. Most shoulder arthroplasty procedures were unilateral (98%).

Changes in demographics and surgical indications over time. Table 2 compares key demographic variables for primary shoulder arthroplasty in 2008 compared to 2017. The median age at the time of primary PSA decreased from 71 years in 2008 to 63 years in 2017. In contrast, the median age at the time of primary TSA remained consistent at 73 years. There was a notable difference in the proportion of primary PSA procedures performed on female patients over the study period, decreasing from 67% in 2008 to 47% in 2017. However, there was little change in the sex distribution for primary TSA over this time (64% and 61% of procedures were performed for women in 2008 and 2017, respectively).

Ten different primary diagnoses were reported in the AOANJRR data. Only the 5 most common primary diagnoses are reported here because the remaining 5 diagnoses comprised <1% of all primary shoulder arthroplasty cases. Glenohumeral OA was the most common primary diagnosis for both PSA and TSA. Clear shifts in the indications for PSA and TSA were observed over

Table 1. Characteristics of the primary shoulder arthroplasty cohort from 2008 to 2017*

Characteristic	Primary PSA (n = 5,979)	Primary TSA (n = 32,889)
Age, years		
<40	207 (3.5)	89 (0.3)
40–49	339 (5.7)	317 (1.0)
50–59	758 (12.7)	2,115 (6.4)
60–69	1,677 (28.0)	9,225 (27.8)
70–79	1,721 (28.8)	14,253 (43.0)
≥80	1,277 (21.4)	6,890 (20.8)
Sex		
Men	2,157 (36.1)	12,524 (38.1)
Women	3,822 (63.9)	20,365 (61.9)
Diagnosis		
Glenohumeral osteoarthritis	2,376 (39.7)	21,324 (64.8)
Fracture	2,724 (45.6)	3,126 (9.5)
Rotator cuff arthropathy	279 (4.7)	6,840 (20.8)
Avascular necrosis	183 (3.1)	434 (1.3)
Rheumatoid arthritis	106 (1.8)	610 (1.9)
Instability	160 (2.7)	258 (0.8)
Tumor	128 (2.1)	137 (0.4)
Other inflammatory arthritis	22 (0.4)	149 (0.5)
Osteochondritis dissecans	1 (0)	0 (0)
Other	0 (0)	11 (0)
Arthroplasty class		
Hemi-stemmed	4,413 (73.8)	NA
Hemi-resurfacing	1,353 (22.6)	NA
Partial resurfacing	167 (2.8)	NA
Hemi–mid-head	46 (0.8)	NA
Total reverse	NA	19,573 (59.5)
Total stemmed	NA	11,980 (36.4)
Total mid-head	NA	1,128 (3.4)
Total resurfacing	NA	208 (0.6)

* Values are the number (%). NA = not applicable; PSA = partial shoulder arthroplasty; TSA = total shoulder arthroplasty.

Table 2. Key demographic variables for PSA in 2008 and 2017*

Characteristic	Primary PSA		Primary TSA	
	2008	2017	2008	2017
Age in years, median (IQR)	71 (62–80)	63 (51–73)	73 (67–78)	73 (67–78)
Women	577 (67.2)	141 (47.3)	964 (63.5)	3,357 (61.2)
Primary diagnosis				
Osteoarthritis	389 (45.3)	159 (53.4)	1,184 (78.0)	3,113 (56.7)
Fracture	380 (44.2)	78 (26.2)	87 (5.7)	638 (11.6)
Rotator cuff arthropathy	24 (2.8)	10 (3.4)	141 (9.3)	1,487 (27.1)
Rheumatoid arthritis	14 (1.6)	11 (3.7)	48 (3.2)	72 (1.3)
Osteonecrosis	25 (2.9)	20 (6.7)	24 (1.6)	85 (1.5)

* Values are the number (%) unless indicated otherwise. PSA = partial shoulder arthroplasty; TSA = total shoulder arthroplasty.

the study period. The proportion of PSA procedures performed for fracture decreased over time, from 44% in 2008 to 26% in 2017, while the proportion of TSA procedures performed for this reason increased (Table 2). For TSA, the proportion of procedures performed for rotator cuff arthropathy increased 3-fold, from 9% in 2008 to 27% in 2017.

Lifetime risk of primary shoulder arthroplasty. As shown in Table 3, the lifetime risk of shoulder arthroplasty increased significantly over time for both sexes ($P < 0.001$). For men, the lifetime risk more than doubled from 0.78% (95% CI 0.73–0.84) in 2008 to 1.78% (95% CI 1.70–1.86) in 2017. Lifetime risk for women rose from 1.54% (95% CI 1.46–1.62) to 2.88% (95% CI 2.78–2.99) over the study period. To put these estimates in context, the lifetime risk of shoulder arthroplasty for men was ~1 in 130 in 2008, increasing to 1 in 57 in 2017. For women, the lifetime risk increased from 1 in 65 in 2008 to 1 in 35 in 2017.

Lifetime risk of primary PSA. Figure 2 shows the lifetime risk of primary PSA by sex for each year from 2008 to 2017. Both men and women demonstrated a decrease in the lifetime risk of this procedure over the 10-year period. The lifetime risk of primary PSA for men decreased from 0.25% (95% CI 0.22–0.28) in 2008 to 0.11% (95% CI 0.09–0.13) in 2017, representing a relative reduction of 56%. The relative reduction in lifetime risk of primary PSA among women was even greater, with lifetime risk decreasing from 0.55% (95% CI 0.51–0.60) in 2008 to 0.11% (95% CI 0.09–0.13) in 2017 (a relative reduction of 80%). While women initially had a higher lifetime risk than men, both sexes demonstrated an equivalent lifetime risk of PSA by 2017.

Table 3. Lifetime risk of primary shoulder arthroplasty for 2008–2017 by sex*

Year	Men	Women
2008	0.78 (0.73–0.84)	1.54 (1.46–1.62)
2009	0.91 (0.85–0.97)	1.81 (1.73–1.90)
2010	0.97 (0.91–1.03)	1.85 (1.77–1.94)
2011	1.08 (1.02–1.15)	2.05 (1.96–2.14)
2012	1.27 (1.20–1.34)	2.04 (1.95–2.13)
2013	1.21 (1.14–1.28)	2.24 (2.15–2.33)
2014	1.38 (1.31–1.45)	2.32 (2.23–2.42)
2015	1.44 (1.37–1.52)	2.44 (2.34–2.54)
2016	1.64 (1.56–1.72)	2.71 (2.61–2.81)
2017	1.78 (1.70–1.86)	2.88 (2.78–2.99)

* Values are the lifetime risk estimates, presented as percentages (95% confidence interval). Primary shoulder arthroplasty incorporates both partial and total shoulder arthroplasty procedures. $P < 0.001$ for linear trend for women and for men.

Lifetime risk of primary TSA. The lifetime risk estimates for primary TSA for each year from 2008 to 2017 are shown in Figure 3. In contrast to PSA, the lifetime risk of TSA increased over time for both men and women. Notably, from 2008 to 2017 the lifetime risk of primary TSA for men more than tripled, from 0.54% (95% CI 0.49–0.58) in 2008 to 1.68% (95% CI 1.60–1.76). For women, the lifetime risk more than doubled, from 0.99% (95% CI 0.92–1.05) to 2.77% (95% CI 2.67–2.88). Although the lifetime risk of primary TSA was consistently higher for women compared to men, both sexes demonstrated a comparable magnitude of increase over time (a relative increase of 211% for women and 180% for men). A subgroup analysis focusing on reverse TSA and total stemmed arthroplasty (representing the 2 most commonly used prostheses for this cohort, as shown in Table 1) is presented in Supplementary Figures 1 and 2 and Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24353/abstract. While a sharp increase in the lifetime risk of reverse TSA was seen (with annual increases of 8–38% for men and 11–29% for women), there was little overall change in the lifetime risk of total stemmed arthroplasty from 2008 to 2017.

DISCUSSION

This study reports the burden of primary shoulder arthroplasty at a national level in Australia. It draws on established epidemiologic methods that have previously been applied for lower-extremity joint arthroplasty (11,12). To the best of our knowledge, this is the first study to quantify the lifetime risk of primary shoulder arthroplasty at a national level, analyze changes over time, and describe sex-based differences. The data also provide important insights into shifts in the clinical indications for primary

PSA and TSA procedures over a 10-year period. From 2008 to 2017, nearly 40,000 primary shoulder arthroplasty procedures were performed in Australia. A large proportion were performed for diagnoses of glenohumeral OA, proximal humeral fracture, and rotator cuff arthropathy. The incidence of these conditions in many countries is likely to increase in the future given aging populations (22), foreshadowing greater demand for shoulder arthroplasty. The approach used to calculate lifetime risk in our study can be easily reapplied in future years for population-level monitoring. The estimates generated can be used to guide public health policies and to assist in health care resource allocation, supporting health services to plan ahead for the changing demand for primary shoulder arthroplasty.

We examined trends in lifetime risk for primary shoulder arthroplasty overall and separately by procedure type (PSA or TSA). This approach enabled us to provide arthroplasty surgery burden estimates that are relevant to health service providers and policymakers and to identify divergent trends with regard to procedure type. While TSA is the dominant form of shoulder arthroplasty (representing 85% of all primary shoulder arthroplasty procedures in Australia from 2008 to 2017), we contend that investigation of PSA utilization remains important given current clinical interest around its role in younger patients and its use in the treatment of fractures. Our approach is also valuable for future benchmarking because PSA is more widely used in some jurisdictions. We found that the lifetime risk of primary PSA decreased significantly for men and women over the 10-year study period; the lifetime risk for men more than halved and for women this reduced by 80%. The decrease in lifetime risk of primary PSA aligns with US administrative data that showed a reduction in hemiarthroplasties for patients age >55 years from 2002 to 2011 (from 58% to 21% of all shoulder arthroplasties) (23), and with US data from a health insurer cohort that demonstrated a large fall in shoulder hemiarthroplasty cases from 2005 to 2013 (from 40% to 9% of all shoulder arthroplasties) (10). Potential reasons for the declining use of PSA include increased risk of persistent pain (24), higher revision rates at 10 years (16), and lower cost-effectiveness compared to TSA for the treatment of conditions such as glenohumeral OA (2). In contrast to the PSA findings, we found that the lifetime risk of primary TSA for both sexes increased significantly over the same period, tripling for men and more than doubling for women. Other studies using administrative data sets or registry reports have reported increasing counts or rates of primary TSA in the US (6,9,10,25,26), although the timeframes examined and magnitude of growth vary considerably between studies.

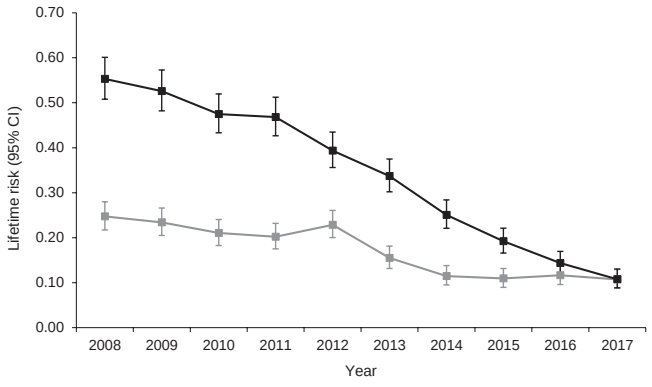


Figure 2. Changes in the lifetime risk of primary partial shoulder arthroplasty from 2008 to 2017. Lifetime risk estimates are presented as percentages with 95% confidence intervals. The black line is for women; the gray line is for men. Linear trend $P < 0.001$ for women and $P = 0.003$ for men.

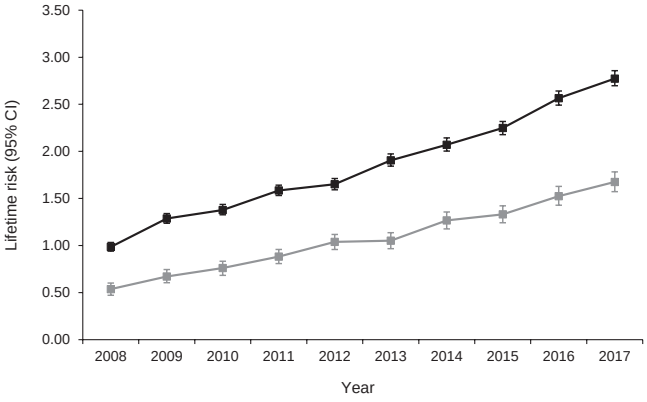


Figure 3. Changes in the lifetime risk of primary total shoulder arthroplasty from 2008 to 2017. Lifetime risk estimates are presented as percentages with 95% confidence intervals. The black line is for women; the gray line is for men. Linear trend $P < 0.001$ for women and for men.

In our study, women consistently displayed a higher lifetime risk of primary TSA at all time points, likely reflecting a higher prevalence of conditions for which this procedure is indicated. The steep increase in lifetime risk of TSA over time probably relates to addressing previously unmet demand for surgery given the increasing availability of orthopedic shoulder surgeons who are trained to perform the procedure and advances in prosthesis design, materials, and outcome monitoring (27). Improvements in perioperative management may mean that these complex surgical procedures can now be more safely performed for older patients. Greater awareness among both patients and health professionals of successful outcomes following TSA may have also contributed to the observed growth. A shift in prosthesis preferences among orthopedic surgeons may have further contributed, including an increasing uptake of reverse TSA for conditions such as glenohumeral OA, fracture, and rotator cuff arthropathy (6,16,28). The increasing uptake of reverse TSA procedures may reflect recent evidence for the superiority of reverse TSA over primary PSA procedures, particularly for older patients. Such benefits include better functional outcomes while demonstrating equivalent prosthesis longevity (29,30). The higher cost of reverse TSA procedures (previously a limiting factor) can be justified by evidence demonstrating greater cost-effectiveness for conditions

traditionally treated with primary PSA (4,31). Our subgroup analysis showed a sharp increase in the lifetime risk of reverse TSA over the study period for both sexes, with little overall change observed for total stemmed procedures. Growth in the use of reverse TSA (associated with expanding clinical indications, improved prostheses, and greater surgical experience in the technique) has likely been a major driver of the rise in lifetime risk of primary shoulder arthroplasty in Australia.

In the absence of comparable lifetime risk data with which to compare our findings, growth in primary shoulder arthroplasty use over the study period can be considered in the context of primary TKA and THA use. TKA and THA procedures are more frequently performed in Australia than shoulder arthroplasty procedures (16), and this difference is reflected in published lifetime risk estimates. In 2013, the lifetime risk of primary TKA and primary THA for men in Australia was 1 in 7 and 1 in 10, respectively. For women, the lifetime risk of primary TKA and primary THA was 1 in 5 and 1 in 8, respectively. As demonstrated in this study, the lifetime risk of primary shoulder arthroplasty was estimated to be 1 in 57 for men and 1 in 35 for women. However, the rate of growth in primary shoulder arthroplasty has clearly exceeded the reported growth in primary lower-extremity arthroplasty. Calculations based on our findings indicate that the average annual change in lifetime risk of primary shoulder arthroplasty was +13% per year for men and +9% per year for women. In contrast, the average annual change in lifetime risk of primary TKA from 2003 to 2013 was +5% per year for men and for women (11). For primary THA, the average annual change from 2003 to 2013 for men and women was +4% and +3%, respectively (12).

This study has several key strengths, including our use of comprehensive, well-validated national data from the AOANJRR that include procedures performed in public and private hospital settings. The large cohort provided a valuable opportunity for calculating lifetime risk with high precision, as reflected in the narrow CIs around these estimates. The methods used to estimate lifetime risk have been previously applied to primary TKA and THA procedures at national and international levels; however, their application to primary shoulder arthroplasty is novel. These methods incorporate life expectancy and all-cause mortality (through the use of life tables) while also taking the age structure of the Australian population into account. This analytic approach represents an advance over traditional methods of quantifying population burden (for example, utilization or incidence rates), which crudely use only the number of surgical procedures and population size.

The limitations of this research should also be acknowledged. The generalizability of our findings to other countries is limited given likely geographic variation in utilization. There is currently a lack of international lifetime risk estimates for primary shoulder arthroplasty, which would be useful for benchmarking purposes. We also acknowledge that our approach only includes procedures that have been performed and does not consider people who may need shoulder arthroplasty (for example, those on surgical waiting lists). This study provides information about the overall use of shoulder arthroplasty for all clinical indications, and it was not intended to examine differential use according to shoulder diagnosis. Finally, this study has estimated population-level risk and is not intended to reflect risks at the individual level, which could vary depending on clinical and demographic characteristics.

In conclusion, this national-level study has identified an increase in the lifetime risk of primary shoulder arthroplasty over a decade, with a clear shift away from use of PSA toward the use of TSA. Changes in the clinical indications for surgery were also observed, with an increasing proportion of TSA procedures now being performed for fracture and rotator cuff disease. These data improve our understanding of the rising burden of primary shoulder arthroplasty in Australia and may aid in informing health funders and health services to plan for the expected continued rise in population demand for primary TSA. The statistical methods used for this study can be easily reapplied in future years for ongoing population-level surveillance and to facilitate a planned international comparison.

ACKNOWLEDGMENT

The authors thank the Australian Orthopaedic Association National Joint Replacement Registry for providing the de-identified data extract used for this study.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. Ackerman 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. Busija, Page, de Steiger, Lorimer, Ackerman.

Acquisition of data. Page, de Steiger, Lorimer.

Analysis and interpretation of data. Miura, Busija, Ackerman.

REFERENCES

1. Page RS, Navarro RA, Salomonsson B. Establishing an international shoulder arthroplasty consortium. J Shoulder Elbow Surg 2014;23:1081–2.
2. Mather RC III, Watters TS, Orlando LA, Bolognesi MP, Moorman CT III. Cost effectiveness analysis of hemiarthroplasty and total shoulder arthroplasty. J Shoulder Elbow Surg 2010;19:325–34.
3. Nwachukwu BU, Schairer WW, McCormick F, Dines DM, Craig EV, Gulotta LV. Arthroplasty for the surgical management of complex proximal humerus fractures in the elderly: a cost-utility analysis. J Shoulder Elbow Surg 2016;25:704–13.
4. Osterhoff G, O'Hara NN, D'Cruz J, Sprague SA, Bansback N, Evaniew N, et al. A cost-effectiveness analysis of reverse total shoulder arthroplasty versus hemiarthroplasty for the management of complex proximal humeral fractures in the elderly. Value Health 2017;20:404–11.
5. Jain NB, Higgins LD, Guller U, Pietrobon R, Katz JN. Trends in the epidemiology of total shoulder arthroplasty in the United States from 1990–2000. Arthritis Rheum 2006;55:591–7.

6. Kim SH, Wise BL, Zhang Y, Szabo Rm. Increasing incidence of shoulder arthroplasty in the United States. J Bone Joint Surg Am 2011;93:2249–54.
7. Oh JH, Chung SW, Oh CH, Kim SH, Park SJ, Kim KW, et al. The prevalence of shoulder osteoarthritis in the elderly Korean population: association with risk factors and function. J Shoulder Elbow Surg 2011;20:756–63.
8. Chillemi C, Franceschini V. Shoulder osteoarthritis. Arthritis 2013;2013:370231.
9. Lübbeke A, Rees JL, Barea C, Combescure C, Carr AJ, Silman AJ. International variation in shoulder arthroplasty. Acta Orthop 2017;88:592–9.
10. Dillon MT, Chan PH, Inacio MC, Singh A, Yian EH, Navarro RA. Yearly trends in elective shoulder arthroplasty, 2005–2013. Arthritis Care Res (Hoboken) 2017;69:1574–81.
11. Ackerman IN, Bohensky MA, de Steiger R, Brand CA, Eskelinen A, Fenstad AM, et al. Substantial rise in the lifetime risk of primary total knee replacement surgery for osteoarthritis from 2003 to 2013: an international, population-level analysis. Osteoarthritis Cartilage 2017;25:455–61.
12. Ackerman IN, Bohensky MA, de Steiger R, Brand CA, Eskelinen A, Fenstad AM, et al. Lifetime risk of primary total hip replacement surgery for osteoarthritis from 2003 to 2013: a multinational analysis using national registry data. Arthritis Care Res (Hoboken) 2017;69:1659–67.
13. Henzell IS, Zhou L, Frampton C, Hooper G, Ackerman I, Young SM. Lifetime risk of primary total knee replacement surgery in New Zealand from 2000 to 2015. N Z Med J 2019;132:48–56.
14. Burn E, Murray DW, Hawker GA, Pinedo-Villanueva R, Prieto-Alhambra D. Lifetime risk of knee and hip replacement following a GP diagnosis of osteoarthritis: a real-world cohort study. Osteoarthritis Cartilage 2019;27:1627–35.
15. Culliford DJ, Maskell J, Kiran A, Judge A, Javaid MK, Cooper C, et al. The lifetime risk of total hip and knee arthroplasty: results from the UK General Practice Research Database. Osteoarthritis Cartilage 2012;20:519–24.
16. Australian Orthopaedic Association National Joint Replacement Registry. Annual report: hip, knee and shoulder arthroplasty. Adelaide: Australian Orthopaedic Association; 2019.
17. Australian Bureau of Statistics. 3302.0.55.001: Life tables, states, territories and Australia, 2015-2017. URL: https://www.abs.gov.au/AUSSTATS/abs@.nsf/allprimarymainfeatures/123B67D527F36A63CA2584A20012C43C?opendocument.
18. Australian Bureau of Statistics. 3101.0: Australian demographic statistics, June 2019. URL: https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/3101.0Main+Features1Jun%202019?OpenDocument.
19. Sasieni PD, Adams J. Standardized lifetime risk. Am J Epidemiol 1999;149:869–75.
20. Chard MD, Hazleman R, Hazleman BL, King RH, Reiss BB. Shoulder disorders in the elderly: a community survey. Arthritis Rheum 1991;34:766–9.
21. Park RE, Mitchell BM. Estimating the autocorrelated error model with trended data. J Econometrics 1980;13:185–201.
22. Australian Bureau of Statistics. Census insights into Australia's ageing population (media release). URL: http://www.abs.gov.au/AUSSTATS/abs@.nsf/mediareleasesbyReleaseDate/39BF03C9400F2E2DCA2581FB0019824A?OpenDocument.
23. Padegimas EM, Maltenfort M, Lazarus MD, Ramsey ML, Williams GR, Namdari S. Future patient demand for shoulder arthroplasty by younger patients: national projections. Clin Orthop Rel Res 2015;473:1860–7.
24. Bjornholdt KT, Brandsborg B, Soballe K, Nikolajsen L. Persistent pain is common 1–2 years after shoulder replacement. Acta Orthop 2015;86:71–7.
25. Day JS, Lau E, Ong KL, Williams GR, Ramsey ML, Kurtz SM. Prevalence and projections of total shoulder and elbow arthroplasty in the United States to 2015. J Shoulder Elbow Surg 2010;19:1115–20.
26. Schwartz BE, Savin DD, Youderian AR, Mossad D, Goldberg BA. National trends and perioperative outcomes in primary and revision total shoulder arthroplasty: trends in total shoulder arthroplasty. Int Orthop 2015;39:271–6.
27. Durchholz H, Salomonsson B, Moroder P, Lambert S, Page R, Audigé L, et al. Core set of radiographic parameters for shoulder arthroplasty monitoring. JB JS Open Access 2019;4:e0025.
28. Day JS, Paxton ES, Lau E, Gordon VA, Abboud JA, Williams GR. Use of reverse total shoulder arthroplasty in the Medicare population. J Shoulder Elbow Surg 2015;24:766–72.
29. Mata-Fink A, Meinke M, Jones C, Kim B, Bell JE. Reverse shoulder arthroplasty for treatment of proximal humeral fractures in older adults: a systematic review. J Shoulder Elbow Surg 2013;22:1737–48.
30. Ferrel JR, Trinh TQ, Fischer RA. Reverse total shoulder arthroplasty versus hemiarthroplasty for proximal humeral fractures: a systematic review. J Orthop Trauma 2015;29:60–8.
31. Kang JR, Sin AT, Cheung EV. Treatment of massive irreparable rotator cuff tears: a cost-effectiveness analysis. Orthopedics 2017;40:e65–76.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 73, No. 10, October 2021, pp 1518–1527
DOI 10.1002/acr.24601

Machine Learning–Based Individualized Survival Prediction Model for Total Knee Replacement in Osteoarthritis: Data From the Osteoarthritis Initiative

Afshin Jamshidi,[1] Jean-Pierre Pelletier,[2] Aurelie Labbe,[3] François Abram,[4] Johanne Martel-Pelletier,[2] and Arnaud Droit[5]

Objective. By using machine learning, our study aimed to build a model to predict risk and time to total knee replacement (TKR) of an osteoarthritic knee.

Methods. Features were from the Osteoarthritis Initiative (OAI) cohort at baseline. Using the lasso method for variable selection in the Cox regression model, we identified the 10 most important characteristics among 1,107 features. The prognostic power of the selected features was assessed by the Kaplan-Meier method and applied to 7 machine learning methods: Cox, DeepSurv, random forests algorithm, linear/kernel support vector machine (SVM), and linear/neural multi-task logistic regression models. As some of the 10 first-found features included similar radiographic measurements, we further looked at using the least number of features without compromising the accuracy of the model. Prediction performance was assessed by the concordance index, Brier score, and time-dependent area under the curve (AUC).

Results. Ten features were identified and included radiographs, bone marrow lesions of the medial condyle on magnetic resonance imaging, hyaluronic acid injection, performance measure, medical history, and knee-related symptoms. The methodologies Cox, DeepSurv, and linear SVM demonstrated the highest accuracy (concordance index scores of 0.85, Brier score of 0.02, and an AUC of 0.87). DeepSurv was chosen to build the prediction model to estimate the time to TKR for a given knee. Moreover, we were able to decrease the features to only 3 and maintain the high accuracy (concordance index of 0.85, Brier score of 0.02, and AUC of 0.86), which included bone marrow lesions, Kellgren/Lawrence grade, and knee-related symptoms, to predict risk and time of a TKR event.

Conclusion. For the first time, we developed a model using the OAI cohort to predict with high accuracy if a given osteoarthritic knee would require TKR, when a TKR would be required, and who would likely progress fast toward this event.

INTRODUCTION

Knee osteoarthritis (OA) is the most common joint disorder and leading cause of disability across the world. Typically, this disease progresses slowly over many years. However, for many subjects with knee OA, the disease progresses rapidly. Recent studies have documented that in a population with no radiographic knee OA, it was estimated that over a 4-year timeframe, the incidence of "accelerated knee OA" ranged from 0.4% to 22.1% (1,2). Davis et al further showed that knees with accelerated OA

This article was prepared using the Osteoarthritis Initiative (OAI), a public-use data set. Its contents do not necessarily reflect the opinions or views of the OAI Study Investigators, the NIH, or the private funding partners of the OAI. The OAI is a public–private partnership between the NIH (contracts N01-AR-2-2258, N01-AR-2-2259, N01-AR-2-2260, N01-AR-2-2261, and N01-AR-2-2262) and private funding partners (Merck Research Laboratories, Novartis Pharmaceuticals, GlaxoSmithKline, and Pfizer) and is conducted by the OAI Study Investigators. Private sector funding for the OAI is managed by the Foundation for the NIH. The authors of this article are not part of the OAI investigative team.

Supported in part by the Osteoarthritis Research Unit of the University of Montreal Hospital Research Centre and the Chair in Osteoarthritis, University of Montreal; CHU de Quebec Research Centre, Laval University; and the Canada First Research Excellence Fund through the TransMedTech Institute.

[1]Afshin Jamshidi, PhD: University of Montreal Hospital Research Centre, Montreal, Quebec, Canada, and Laval University Hospital Research Centre, Montreal, Quebec, Canada; [2]Jean-Pierre Pelletier, MD, Johanne Martel-Pelletier, PhD: University of Montreal Hospital Research Centre, Montreal, Quebec, Canada; [3]Aurelie Labbe, PhD: HEC Montreal, Montreal, Quebec, Canada; [4]François Abram, PhD: ArthroLab Inc., Montreal, Quebec, Canada; [5]Arnaud Droit, PhD: Laval University Hospital Research Centre, Quebec, Canada.

Drs. Martel-Pelletier and Droit contributed equally to this work.

Drs. Pelletier and Martel-Pelletier own stocks or stock options in ArthroLab, Inc. No other disclosures relevant to this article were reported.

Address correspondence to Johanne Martel-Pelletier, PhD, Osteoarthritis Research Unit, University of Montreal Hospital Research Centre, 900 Saint-Denis, Suite R11.412, Montreal, Quebec, Canada H2X 0A9. Email: jm@martelpelletier.ca.

Submitted for publication August 20, 2020; accepted in revised form March 18, 2021.

SIGNIFICANCE & INNOVATIONS

- Individualized prediction time to total knee replacement (TKR) can be done with high accuracy using only 3 features.
- A combination of radiographs, bone marrow lesions, and knee-related symptoms demonstrated the most significant impact in prediction time to a TKR event.
- The developed TKR prediction model, once validated, could guide clinicians to appropriate therapeutic strategies.

have a higher chance of needing a total knee replacement (TKR) compared to those with a more gradual onset or without accelerated knee OA (3).

The initiation and rapid progression of primary OA over time for a given knee is generally unknown. Integrating data for uncovering the complex mechanisms of subjects with knee OA leading to a TKR event will enable objective-driven analytical leads, improve survival prediction (4–6), and help develop better therapeutic strategies for these subjects.

Survival analysis is one of the fundamental tools in the medical domain to identify predictors of time to adverse events and to develop systems for clinical decision support. In knee OA research, survival analysis can be used to predict time of pathologic events such as TKR. However, for this disease, it is not possible to apply traditional survival analysis methods (7) as it involves integrated high-dimensional and nonlinear data structures. For instance, the most commonly used approach for analyzing survival data, the semiparametric Cox proportional hazards regression method (8), demonstrated several drawbacks in analyzing nonlinear data structures, high dimensions, and low sample size data (9). Machine learning/deep learning–based survival prediction models have proven to be a better option in the case of complex data with interactions among the features, as in OA (10).

To date, there have been no comprehensive attempts to identify the most important features and to build a survival-based model for predicting the time to TKR for an individual with OA. Recently, Heisinger et al, with the use of a small cohort (n = 165), applied 14 factors in a 4-year period prior to TKR to predict an individual's need for TKR surgery (11). In addition, an image-based model with knee radiographs to classify individuals with OA who are at high risk of TKR has also been recently developed (12). There are also other works on TKR survival analyses, but all were done for subjects following TKR surgery, not before. For example, the research questions in these studies included the following: How long does a TKR last (13)? What is the subject death rate after TKR surgery (14)? And what are the clinical and radiologic outcomes of TKR (15)?

Our study is the first to look at TKR as the survival outcome. Two specific research questions were addressed: 1) Which features at baseline are most associated with accelerated knee OA leading to TKR? and 2) Can we estimate for a given OA knee, the risk and time to TKR (e.g., remaining useful life [RUL] for a given knee)? To answer these questions, we evaluated the length of time from the date of enrollment until the TKR event (overall survival) by applying feature selection methods to find the most important features in survival analysis of a TKR event and developing a survival prediction model based on the selected features using machine learning–based methods.

PATIENTS AND METHODS

Ethics approval and consent to participate. Ethics approval was obtained by each Osteoarthritis Initiative Cohort (OAI) clinical site (University of Maryland Baltimore Institutional Review Board [IRB], Ohio State University Biomedical Sciences IRB, University of Pittsburgh IRB, and Memorial Hospital of Rhode Island IRB) and the OAI coordinating center (Committee on Human Research at University of California, San Francisco). All patients provided written informed consent for participation in the OAI. The Institutional Ethics Committee Board of the University of Montreal Hospital Research Centre approved the present study.

Data availability statement. All data used in this study are publicly available from the OAI cohort (https://data-archive.nimh.nih.gov/oai/). Additional data may be obtained from the corresponding author upon reasonable request.

Knee selection. Data used in this study were from the OAI cohort. Details of the study cohort are shown in the Supplementary Methods, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24601/abstract.

Selection of predictor features. The predictor features were selected at baseline from the OAI database and from the quantitative determination of knee tissues by magnetic resonance imaging (MRI) (Supplementary Tables 1–3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24601/abstract). Feature selection was performed using the target knee of each subject (see Supplementary Methods for definition of "target knee"). In the present study, 1,431 knees were included in our assessments (Figure 1).

For the feature selection, as the knees with accelerated OA (progressors) have a higher chance of needing a TKR compared to knees without accelerated OA (3), we considered 2 groups: progressors and non-progressors. The OA knee progressor and non-progressor definition from this cohort has been previously described and discussed (16). Data from the knees of 733 progressors and 698 non-progressors were included (Figure 1). The Supplementary Methods further explain how groups were categorized for the present study.

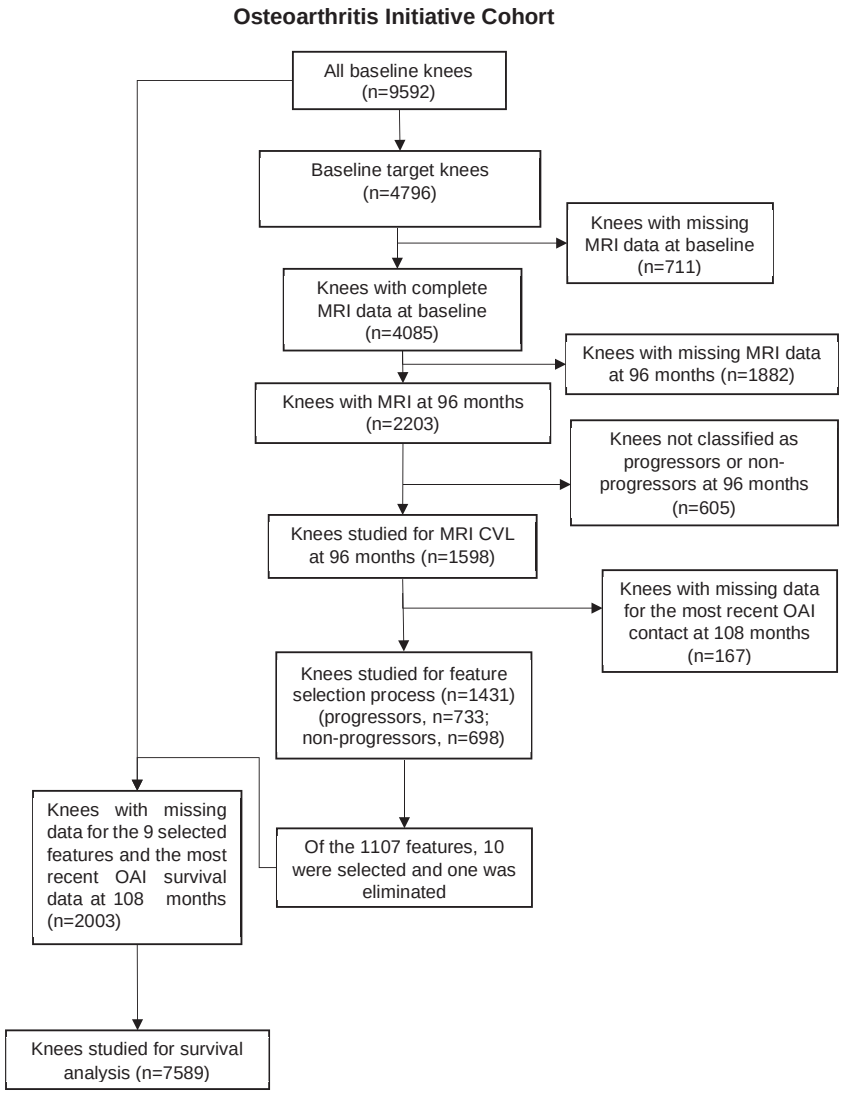


Figure 1. Flow chart of knee selection from the Osteoarthritis Initiative Cohort (OAI) used for assessment in the present study. CVL = cartilage volume loss in the medial tibial plateau; MRI = magnetic resonance imaging.

Next, we used all the features recorded at baseline in the OAI database, which included standard features (clinical, demographic characteristics, anthropometry, among others) as well as other uncommon features (health care access, nutrition, knee MRI data). After data cleaning, 1,107 features remained and were divided into categories and subcategories as described previously (16) and reported in Supplementary Tables 1–3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24601/abstract. Data cleaning methodology is shown in the Supplementary Methods.

Feature selection was performed using the lasso method for variable selection in the Cox regression model. Data from 1,431 knees were divided into training sets (80%) and testing sets (20%) to generate the prediction model and data applied for measuring the accuracy of the developed prediction model, respectively. For extraction of the most important features, the glmnet package was used in R (17). With the use of the Cox model with the lasso method (18), the 10 top baseline features were extracted and served for designing a TKR survival prediction model.

TKR survival. For the survival analysis, we considered all the knees to predict the outcome (time to TKR) (Figure 1). Knees (n = 7,589) with complete data at the final visit (108 months) were included. As a first step, to verify if and which of the most important features described in the literature could individually impact TKR survival, we compared, among the well-known features related to OA, the survival curves of 6 of these features. These characteristics included age, sex, race, body mass index (BMI), Kellgren/Lawrence (K/L) grade (19), and pain measured with the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) (20). Further, Kaplan-Meier analysis with log rank test ($P < 0.050$) was used on the selected features to compare the prognosis power of each feature for risk and time to TKR.

Machine learning methods applied for building survival prediction models. Survival prediction models for TKR events were built with 7,589 knees, which were categorized into a training set (n = 6,071) and a testing set (n = 1,518). The following models in the PySurvival package (21) in Python 3.7 were applied: 1) Cox-PH model (8), 2) DeepSurv/nonlinear model (10), 3) linear multi-task logistic regression (22), 4) neural linear multi-task logistic regression model (23), 5) random survival forest model (24), 6) linear support vector machines (SVM) model (25), and 7) kernel SVM model (26) (for details on models used, refer to the Supplementary Methods, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24601/abstract). Supplementary Figure 1 illustrates the pipeline of the data analysis.

Hyperparameter tuning. For hyperparameter tuning, we applied the GridSearchCV in scikit-learn to determine which values of hyperparameters perform best in each model (Supplementary Table 4). We selected the configuration with the largest validation Concordance index. A description of the concordance index is available in the Supplementary Methods (http://onlinelibrary.wiley.com/doi/10.1002/acr.24601/abstract).

Prediction of the performance evaluation. To compute the prediction performance of the models mentioned above, we applied 3 metrics on the test data set: concordance index, Brier score, and the time-dependent area under the curve (AUC) of the receiver operating characteristic (ROC) (Supplementary Methods).

Overall and knee-specific predictions of TKR survival. To compute the prediction performance of the best model, we compared the time series of the actual and predicted number of knees experiencing a TKR event, for each time (*t*), by calculating the mean and median absolute error and a root mean square error.

To show representative curves of the TKR survival of different conditions, we used 5 knees from the OAI that demonstrated a range of values for each selected feature and applied the selected model to predict their specific survival curves.

RESULTS

Association between clinical/demographic features and TKR survival. Of the 7,589 knees included in the analysis, 413 had a TKR event and 7,176 survived a TKR event (right-censored data, which occurs when the TKR event does not happen by the end of study) at the end of follow-up (3,320 days or 108 months). Supplementary Figure 2A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24601/abstract, illustrates that at 3,000 days, 7,224 knees survived TKR and 365 knees had TKR.

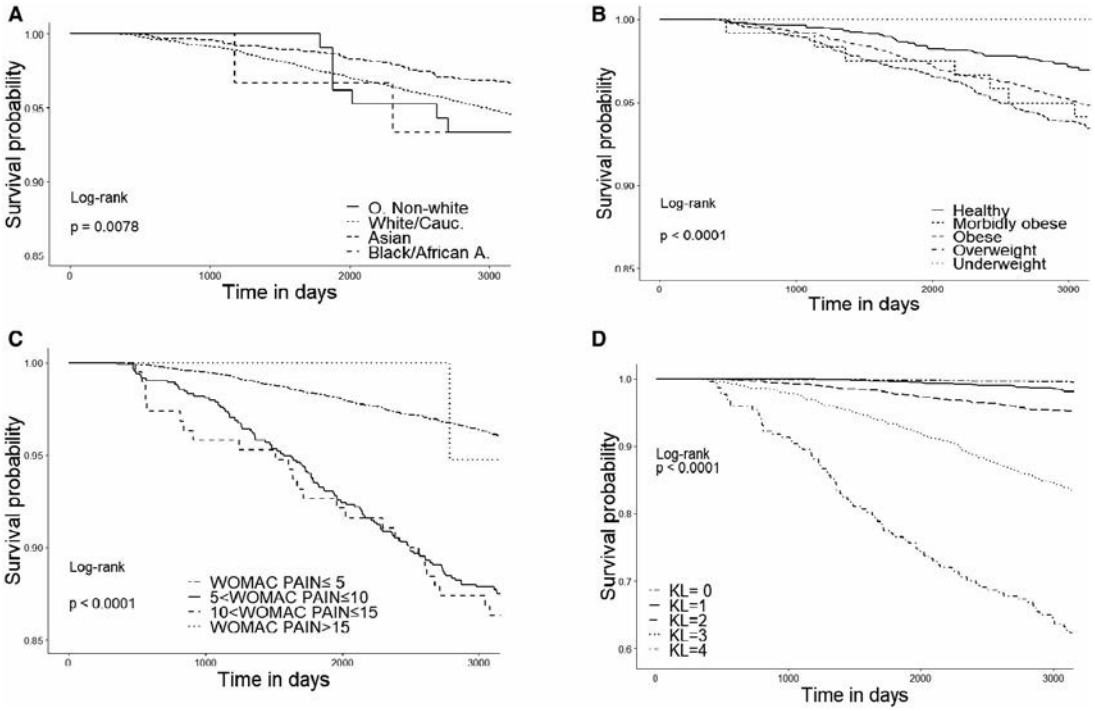


Figure 2. Probability of survival for the most important clinical/demographic features before a total knee replacement event. Probability of survival associated with race (**A**), body mass index (BMI) (**B**), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain score (**C**), and Kellgren/Lawrence (KL) grades (**D**) are shown. In the Osteoarthritis Initiative Cohort, BMI was measured using the following categories: healthy (≤24.9 kg/m^2), morbidly obese (>40 kg/m^2), obese (≤39.9 kg/m^2), overweight (≤29.9 kg/m^2), and underweight (≤18.5 kg/m^2). WOMAC pain was scored as a continued numerical feature (score 0–20 [20]) and categorized into 4 different groups. A. = American; Cauc. = Caucasian, O = non-White, other non-White.

Next, we assessed the association between age, sex, race, BMI, WOMAC pain, and K/L grade features with survival probability before a TKR event and illustrated the most important features (Figure 2) and less important ones (Supplementary Figure 2B).

With regard to race (Figure 2), although the survival curves showed a very small difference between groups in regard to the time to survival, they appeared to be statistically different ($P = 0.0078$). The time to TKR for Asian individuals and other non-White groups is slightly higher than the other 2 groups studied (White/Caucasian and Black/African American). This could be because of the smaller number of patients available in these 2 groups (105 and 60, respectively). In addition, it is believed that the slight difference (about 3%) between the groups with the best and worst survival curves may not be due to less disease, but for other reasons, such as access to surgery and preference of some groups for surgical intervention.

Another major risk factor involved in the OA process is BMI (Figure 2). Data showed that knees of subjects who were morbidly obese, obese, and overweight had a higher chance of experiencing TKR in compared to knees of subjects who were a healthy weight and underweight ($P < 0.0001$). In addition, knees from healthy subjects are slightly more at risk of a TKR event than those from underweight subjects.

For WOMAC pain (20), knees from the groups that had pain scores of 6–10 and 11–15 on the WOMAC have a high chance of a TKR event ($P < 0.0001$) compared to knees from subjects with a lower level of pain (scores 0–5) and knees from subjects who reported a high level of pain (scores 16–20) (Figure 2). At the end of the study, survival of the groups having a score of 6–15 on the WOMAC was ~87%.

K/L grade, a widely used approach for classifying the severity of knee OA with the use of knee radiographs (19), was also assessed. There was a drastic and significant decrease in survival chances of OA knees with a K/L grade of 3 (84% survival at the end of study) ($P < 0.0001$), but more so of knees with a K/L grade of 4, in which the probability of survival decreased to ~60% at the end of the study (Figure 2).

For age, data showed that though there was a statistically significant difference between the 2 groups ($P < 0.0001$), the knees in the group of subjects ages 60 years or older had a lower chance of survival compared to knees from subjects ages 60 years or younger. The probability of survival in both groups was >90% at the end of the study (Supplementary Figure 2B, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24601/abstract). Additionally, there was no significant difference found in survival probability for a TKR event in individuals based on sex (Supplementary Figure 2B).

When comparing the above 6 features together (Figure 2 and Supplementary Figure 2B), it appears that WOMAC pain, and to a higher degree, K/L grade, have a more significant impact on TKR survival, and probability of survival in the other groups was still very high (>90%) at the end of the study.

Building a model for time to TKR event. *Selection of survival-based features.* To build the model for time to TKR event, 1,431 knees were used (Figure 1) to identify the most important TKR survival–based features among 1,107 features overall (Supplementary Table 1). This was performed using the Cox model with the lasso method. The selected top 10 features (Table 1)

Table 1. Top 10 features selected using Cox model with the lasso method for TKR survival prediction*

Priority†	Relative importance	Label	Category
1	1	Severe radiographic damage of the knee at baseline: composite OA grade 4 (quasi–K/L grade [score 0–4])	Radiograph
2‡	0.82	BMLs in the medial condyle (>0.2, with data expressed as BML size in regions)	MRI
3	0.62	Knee radiograph at baseline, with osteophytes and joint space narrowing (both with a score of 2 on a 0–2 scale)	Radiograph
4	0.48	Unable to perform 400-meter walk (excluded for heart rate)	Performance measure
5	0.47	Baseline radiographic knee OA (defined as a K/L grade of ≥2 in the left knee, right knee, or both (45)	Radiograph
6	0.47	Charlson comorbidity index: history of stroke, cerebrovascular accident, blood clot or bleeding in brain, or transient ischemic attack	Medical history
7‡	0.37	K/L grade of 4	Radiograph
8	0.29	Received 1 injection of hyaluronic acid treatment in either knee within the past 6 months	Medication
9‡	0.25	Presence of knee symptoms (sometimes swelling in the last 7 days)	Knee symptoms
10	0.21	Baseline symptomatic knee OA (defined as radiographic OA and frequent knee pain in the same knee [45])	Radiograph

* MRI = magnetic resonance imaging; OA = osteoarthritis; TKR = total knee replacement.
† Priority indicates the importance of the selected features. Relative importance is calculated as the absolute importance of a variable divided by the absolute importance of the most important variable. The bone marrow lesion (BML) value is between 0 (no BMLs) and 1 (the BML extends into the entire bone region) (37,44), with 0.2 corresponding to 20%. Knee symptoms were scored according to the Osteoarthritis Initiative nomenclature (range 0–4, with 0 indicating "never swelling" and 4 indicating "always swelling"), and the Kellgren/Lawrence (K/L) grade was based on a 0–4 scale (with 4 indicating severe radiographic damage of the knee at baseline). For further details about scoring, refer to Figures 2, 3, and Supplementary Figure 3, available on the *Arthritis Care & Research* website at http://onlin elibr ary.wiley.com/doi/10.1002/acr.24601/ abstract.
‡ The final 3 features used for the prediction of TKR survival.

included 5 radiographs, MRI-assessed bone marrow lesions (BMLs) in the medial condyle, a performance measure, medical history, medication use, and knee symptoms (sometimes swelling over the last 7 days).

We then assessed the association of the selected features with survival probability before a TKR event, with the most important features shown in Figure 3 and features of lesser importance shown in Supplementary Figure 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24601/abstract. Of note, a graph for K/L grade was not included in Figure 3 as this information was already shown in Figure 2 and in data described above; however, K/L grade was included in our model that assessed features shown in Figure 3. Data showed that in addition to K/L grade 4 (Figure 2), 4 other features demonstrated a high impact on the TKR event (Figure 3). For 3 of these features (composite OA grade 4, the presence of >0.2 medial condyle BMLs, and osteophytes and joint space narrowing [JSN] both with a score of 2), survival probability at the end of the study was ~65%. For the fourth feature analyzed in this model (having received an injection of hyaluronic acid [HA] in the past 6 months), survival probability was 75%.

For the other features analyzed (Supplementary Figure 3), although a statistical difference was demonstrated (not included the 400-meter walk feature), survival probability at the end of study was >80%.

Development of survival prediction models based on the selected features. Seven machine learning methods were applied. As mentioned above, the 400-meter walk feature was not statistically different (Supplementary Figure 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24601/abstract), and the survival prediction models were developed without this feature included. By using the 9 remaining features (Table 1) and the 7 machine learning methods, data showed a very low Brier score for all of them (0.02), indicating that all models have very good predictive abilities. Cox, DeepSurv, and linear SVM models demonstrated the highest concordance index score (0.85) compared to the random forest (concordance index score of 0.82), kernel SVM (concordance index score of 0.83), and linear/neural multi-task logistic regression models (concordance index score of 0.80).

However, as nonlinear analysis outperforms other models when assessing the huge amount of data in machine learning for finding important patterns and predicting diseases, we eliminated the linear SVM method during the development of

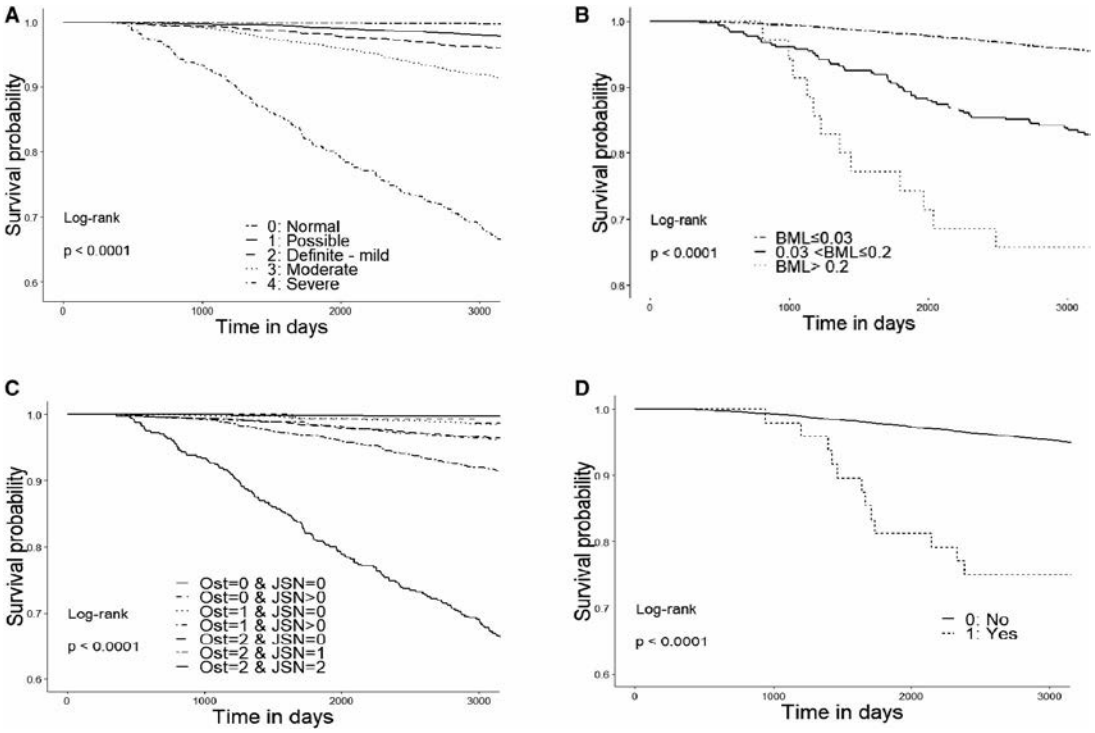


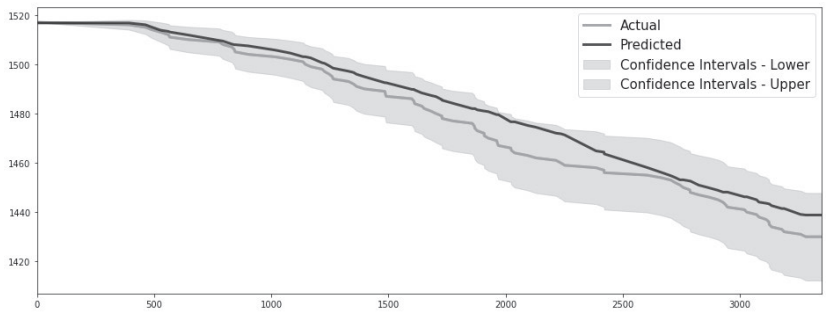
Figure 3. Probability of survival for the most important selected features before a total knee replacement event. Selected features included a composite osteoarthritis (OA) grade (**A**); the presence of bone marrow lesions (BMLs) (data are expressed as BML size in regions/compartments and corresponds to the percentage of BMLs in this region; for example, >0.2 corresponds to 20% [37,44]) (**B**); a composite quasi–Kellgren/Lawrence grade based on a 0–4 scale (with grade 4 indicating severe OA damage of the knee at baseline on radiographs) comprising evidence of joint space narrowing (JSN) and osteophytes (Ost) (with JSN and osteophytes scored on a 0–2 scale, corresponding to severe disease [43]) (**C**); and the history of hyaluronic acid injection (with either knee having received 1 injection of hyaluronic acid within the past 6 months) (**D**).

the survival prediction models. Between the Cox and DeepSurv methods, we chose DeepSurv as it can consider nonlinear interactions between features and was shown to better handle complex data interactions (as an OA data set) among features and outperform other models in general and the Cox model in particular (7,10). Indeed, the assumption of the linear log-risk function in the Cox model may by too simplistic when assessing personalized survival predictions. Therefore, further analysis was performed with the DeepSurv model to better fit OAI survival data with nonlinear log-risk functions. As some of the selected radiographs features, including the composite OA, osteophytes and JSN, baseline radiographic knee OA, and K/L grade 4 (Table 1), are based on a similar measurement, we further analyzed if 1 or some of these features could be removed from the analysis without impairing the prediction model. To this end, we studied, in addition to the other 5 features (features 2, 6, 8, 9, and 10 shown in Table 1), only 1 of the features at a time in the DeepSurv model. Further, to explore if eliminating 1 or more of the above-mentioned 5 features could yield similar statistical indexes, we removed each of the features in a recursive manner from the model. Data showed that we were able to achieve an identical concordance index score (0.85) and Brier score (0.02) using the 3 following features: BMLs in the medial condyle, K/L grade, and knee symptoms (i.e., sometimes swelling in the last 7 days) (Table 1); therefore, for the next steps of analysis we considered only these 3 features.

Overall predictions of TKR survival. We then compared, by using the DeepSurv model, the time series of the actual number of knees experiencing a TKR versus the predicted number of knees experiencing a TKR, as well as the risk of TKR, for each time point. Data showed that the time series of actual and predicted number of knees experiencing a TKR are quite similar in which the predicted values fall in the confidence interval of actual values, with a very low mean absolute error (5.64) and median absolute error (5.10) and root mean square error (6.55) (Figure 4A). This indicates that the average prediction error of the model is very low throughout the entire timeline, in which the average error is about 5 knees for all 3,320 days of the study period.

Knee-specific predictions of TKR survival. To plot the knee-specific survival curves and to estimate individually the TKR event time using the DeepSurv model, we assessed 5 knees from the OAI data set, which were selected according to a range of values for each selected feature. Values for the 3 selected features and TKR survival curves for the 5 knees assessed are shown (Figures 4B and C). Comparison of the survival curve indicates that

A



B

Knees	BML in the medial condyle	KL grade	Knee symptoms sometimes swelling, last seven days
1	0	2	0
2	0.011	2	2
3	0.043	3	1
4	0.214	4	3
5	0.311	4	4

C

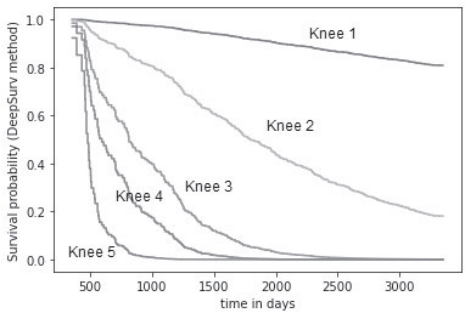


Figure 4. Predictions of total knee replacement survival before a total knee replacement event. **A**, Comparison of actual number of knees requiring total knee replacement versus predicted number of knees. Lower and upper confidence intervals are for the actual curve, and both represent the lower and upper 95% confidence interval. **B**, Baseline values are shown for bone marrow lesions (BMLs) in the medial condyle data, expressed as BML size in regions and corresponding to the percentage of BMLs in this region (37,44); Kellgren/Lawrence (KL) score according to knee osteoarthritis (OA) grade (based on a 0–4 scale, with 4 indicating severe OA damage of the knee at baseline on radiographs); and knee symptoms characterized by swelling in the last 7 days according to the Osteoarthritis Initiative Cohort nomenclature (symptoms were scored on a 0–4 scale, with 0 indicating "never swelling" and 4 indicating "always swelling"). **C**, Knee-specific survival curves.

the model could perfectly predict the TKR event time (Figure 4C). Hence, when the survival curve reaches 0% survival probability, it indicates the approximate TKR event time. Additionally, this could identify the RUL of the knee before TKR. RUL could be calculated as the difference between the enrollment date and the date that the survival curve reaches 0%. Of note, all the curves start from day 357 (the first TKR event recorded in the present study); however, the RUL should be calculated from day 0 until 0% is reached on the survival curve. For example, in the case of knee 5, the RUL is around 1,000 days. Moreover, knees 1 and 2 showed that their time to TKR will be longer than the cohort time limit. This is not unexpected as their baseline values for the selected features are low (Figure 4B).

With these data, one could classify the knees into different groups of TKR event. For example, knees reaching 0% survival before 1,000 days could be considered "high risk" (knees 4 and 5) (Figure 4C), knees that reached 0% survival between 1,000 and 3,000 days could be considered "medium risk" (knee 3), and knees that reached 0% survival after more than 3,000 days could be considered "low risk" (knees 1 and 2).

Time-dependent AUC of the developed model. As TKR is a time-to-event outcome in which the status changes over time, we further looked at the time-dependent AUC. An average AUC of 0.86 is shown in Supplementary Figure 4, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24601/abstract. Of note, the average AUC was 0.87 when using 9 features in the analysis (data are not shown). The highest performance (an AUC of up to 0.99) is achieved at the beginning of events from day 357 until day 450, after which it slightly fluctuates until 1,700 days, then remains stable until almost 2,400 days, and slightly decreases (maximum 2%) until the end of study (3,320 days [108 months]). These data demonstrated that the prediction model is effective in the long-term prediction of TKR until 3,320 days and very effective in predicting TKR until 2,400 days. Therefore, it is possible to predict the TKR event using the 3 selected features mentioned above with high accuracy and a virtually stable AUC score for a long-term period at each time point up to 3,320 days after enrollment.

DISCUSSION

In the present study, we first considered the best-known risk factors in OA for identifying the most important factors that lead to a TKR event. Data revealed that although age, sex, BMI, WOMAC pain, and K/L grade were all significant, a K/L grade of 4 had the highest impact on TKR survival, with low survival probability. The importance of K/L grade was further confirmed by the feature selection using machine learning methodology. This finding was not surprising, as the importance of K/L grade in the prediction of knee OA severity has been known for a long time (19,27). Although these findings provide insightful information, they were not sufficient in the developing a powerful survival prediction model as there could be unknown factors impacting the risk of a TKR event. As such, for the first time, we used 7,589 knees with simultaneous integration of 1,107 features, including standard and uncommon ones. The 10 most influential features were identified, and 9 were used to develop a survival prediction model. Further analysis revealed that 3 of these features were the most influential, which included the presence of BMLs in the medial condyle, K/L grade, and knee symptoms (sometimes swelling in the last 7 days). The Cox model with the lasso method was used for feature selection as it can perfectly deal with multicollinearity issues occurring with OA features, particularly those from MRIs and radiographs (28).

Of the selected features, the most important ones were radiographs and the presence of BMLs in the medial condyle. The BML finding is not unexpected and reaffirms this altered structure prediction of the occurrence of TKR (29–34). Moreover, BMLs in the medial condyle as an indication of the likelihood of a TKR event is well in line with findings showing that this is the area in knee OA where both BMLs and cartilage degeneration are the most frequently affected (35–37). There are medications that have been shown to prevent or reduce the severity of BMLs (38,39), but to the best of our knowledge, there is only one study showing that a bone remodelling therapeutic agent, bisphosphonates, was associated with about a 25% lower risk of TKR (40).

From those selected, 4 features (2 radiographs, BMLs, and HA injection) demonstrated a significant impact on survival analysis by dropping survival probability to 75% or less at the end of study, reaffirming the importance of radiograph and MRI features in survival prediction of TKR. For HA, our findings corroborate with recent studies (41,42) in which HA injection in knees with OA is highly associated with a significant delay in TKR (41). However, caution should be taken in interpreting the role of HA injection prior to TKR, as patients who do not want to undergo TKR may use HA as a substitute. This does not apply only to HA, but to other alternative measures taken by OA patients for whom surgery is not an option for personal or medical reasons.

As conventional survival models such as the Kaplan-Meier curve are not designed to predict an outcome, we considered machine learning–based survival models. By comparing 7 machine learning models, data revealed that DeepSurv was the most appropriate model for the present study. With the DeepSurv method and after further analyses, we were able to reduce the number of features to 3 without compromising the accuracy of the model. By using these selected features, we further estimated for any given knee the time to TKR event/RUL. Of note, this proposed methodology could also be applied to other articulations (e.g., the hip).

The present study has some limitations. First, although we used all possible features (1,107) from one of the most complete databases for OA subjects (the OAI), including standard and uncommon characteristics, other unanticipated features could

putatively also influence the TKR survival time. Second, our model was developed using a cohort in which subjects are at a mild-to-moderate stage of the disease. To ascertain the generalizability of our prediction model, a validation study should be performed with another cohort, preferably with clinical trial OA patients or electronic health research, which will represent more subjects routinely seen by a physician. After validation, the next step of this work will be to develop a predictive tool which could be used to guide clinical decision-making. Finally, it could be said that our model requiring an MRI feature to predict time to TKR is not customary in clinical practice. Although MRI may not be commonly used as a first-line treatment by physicians, these technologies are becoming an increasingly routine part of the investigation of knee OA by subspecialists, such as orthopedic surgeons and rheumatologists. Moreover, the use of MRI for the investigation of knee OA is more accessible as availability improves and the cost of the procedure becomes less expensive.

In the present study, we showed that with the use of the OAI cohort, it is possible to predict with a high degree of certainty when a TKR event would happen for a given OA knee, and who will progress fast toward this event. To the best of our knowledge, this is the first study in which a survival prediction model for a TKR event was built for knee OA by using machine learning methods, applying a survival-based feature selection method, and considering a very large number of features. Another important contribution of this work is the development of a prediction model that estimates the time of the risk of a TKR event for a given knee. Presently, as the time estimate to TKR is an arbitrary feature for clinicians, this developed survival prediction model built with the OAI cohort could, in the future, better guide clinicians to the best therapeutic strategy to improve the survival of knees affected by OA.

ACKNOWLEDGMENTS

The authors would like to thank the participants in the Osteoarthritis Initiative Cohort (OAI) and the Coordinating Center for their work in generating the clinical and radiologic data of the OAI cohort and for making these data publicly available. The authors are grateful to ArthroLab, Inc. for having provided the magnetic resonance imaging data and Santa Fiori for preparing 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 published. Drs. Martel-Pelletier and Jamshidi had full access to all the study data and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Jamshidi, Pelletier, Labbe, Martel-Pelletier, Droit.

Acquisition of data. Jamshidi, Abram, Martel-Pelletier, Droit.

Analysis and/or interpretation of data. Jamshidi, Pelletier, Labbe, Abram, Martel-Pelletier, Droit.

ADDITIONAL DISCLOSURES

Author Abram is an employee of ArthroLab Inc.

REFERENCES

1. Driban JB, Eaton CB, Lo GH, Ward RJ, Lu B, McAlindon TE. Association of knee injuries with accelerated knee osteoarthritis progression: data from the Osteoarthritis Initiative. Arthritis Care Res (Hoboken) 2014;66:16739.
2. Driban JB, Stout AC, Lo GH, Eaton CB, Price LL, Lu B, et al. Best performing definition of accelerated knee osteoarthritis: data from the Osteoarthritis Initiative. Ther Adv Musculoskelet Dis 2016;8:165–71.
3. Davis JE, Liu SH, Lapane K, Harkey MS, Price LL, Lu B, et al. Adults with incident accelerated knee osteoarthritis are more likely to receive a knee replacement: data from the Osteoarthritis Initiative. Clin Rheumatol 2018;37:1115–8.
4. Lu J, Cowperthwaite MC, Burnett MG, Shpak M. Molecular predictors of long-term survival in glioblastoma multiforme patients. PLoS One 2016;11:e0154313.
5. Yousefi S, Amrollahi F, Amgad M, Dong C, Lewis JE, Song C, et al. Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models. Sci Rep 2017;7:11707.
6. Zhu B, Song N, Shen R, Arora A, Machiela MJ, Song L, et al. Integrating clinical and multiple omics data for prognostic assessment across human cancers. Sci Rep 2017;7:16954.
7. Kim DW, Lee S, Kwon S, Nam W, Cha IH, Kim HJ. Deep learning-based survival prediction of oral cancer patients. Sci Rep 2019;9:6994.
8. Cox DR. Regression models and life tables (with discussion). J R Stat Soc B 1972;34:187–220.
9. Hao J, Kim Y, Mallavarapu T, Oh JH, Kang M. Interpretable deep neural network for cancer survival analysis by integrating genomic and clinical data. BMC Med Genomics 2019;12 Suppl 10:189.
10. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res Methodol 2018;18:24.
11. Heisinger S, Hitzl W, Hobusch GM, Windhager R, Cotofana S. Predicting total knee replacement from symptomology and radiographic structural change using artificial neural networks-data from the Osteoarthritis Initiative (OAI). J Clin Med 2020;9.
12. Leung K, Zhang B, Tan J, Shen Y, Geras KJ, Babb JS, et al. Prediction of total knee replacement and diagnosis of osteoarthritis by using deep learning on knee radiographs: data from the Osteoarthritis Initiative. Radiology 2020;296:584–93.
13. Evans JT, Walker RW, Evans JP, Blom AW, Sayers A, Whitehouse MR. How long does a knee replacement last? A systematic review and meta-analysis of case series and national registry reports with more than 15 years of follow-up. Lancet 2019;393:655–63.
14. Hunt LP, Ben-Shlomo Y, Whitehouse MR, Porter ML, Blom AW. The main cause of death following primary total hip and knee replacement for osteoarthritis: a cohort study of 26,766 deaths following 332,734 hip replacements and 29,802 deaths following 384,291 knee replacements. J Bone Joint Surg Am 2017;99:565–75.
15. Puliero B, Favreau H, Eichler D, Adam P, Bonnomet F, Ehlinger M. Total knee arthroplasty in patients with varus deformities greater than ten degrees: survival analysis at a mean ten year follow-up. Int Orthop 2019;43:333–41.
16. Jamshidi A, Leclercq M, Labbe A, Pelletier JP, Abram F, Droit A, et al. Identification of the most important features of knee osteoarthritis progressors using machine learning methods. Ther Adv Musculoskel Dis 2020;12:1–12.
17. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33:1–22.
18. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc B 1996;58:267–88.

19. Kohn MD, Sassoon AA, Fernando ND. Classifications in brief: Kellgren-Lawrence classification of osteoarthritis. Clin Orthop Relat Res 2016;474:1886–93.
20. Bellamy N, Buchanan WW. A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. Clin Rheumatol 1986;5:231–41.
21. Fotso S. PySurvival: Open source package for survival analysis modeling. 2019.
22. Yu CN, Greiner R, Lin HC, Baracos V. Learning patient-specific cancer survival distributions as a sequence of dependent regressors. In: Advances in Neural Information Processing Systems 24 (NIPS 2011). 2011. p. 1845–53.
23. Fotso S. Deep neural networks for survival analysis based on a multi-task framework. arXiv preprints arXiv 2018;1801.05512.
24. Hemant Ishwaran UBK, Eugene H. Blackstone, and Michael S. Lauer. Random survival forests. Ann Appl Stat 2008;2:841–60.
25. Pölsterl S, Navab N, Katouzian A. Fast training of support vector machines for survival analysis. In: Appice A, Rodrigues PP, Costa VS, Gama J, Jorge A, Soares C, editors. Machine Learning and Knowledge Discovery in Databases. Springer International Publishing; 2015. p. 243–59.
26. Pölsterl S, Navab N, Katouzian A. An Efficient Training Algorithm for Kernel Survival Support Vector Machines. The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. ArXiv 2016;1611.07054.
27. Guermazi A, Hayashi D, Roemer F, Felson DT, Wang K, Lynch J, et al. Severe radiographic knee osteoarthritis: does Kellgren and Lawrence grade 4 represent end stage disease? The MOST study. Osteoarthritis Cartilage 2015;23:1499–505.
28. Jamshidi A, Pelletier JP, Martel-Pelletier J. Machine-learning-based patient-specific prediction models for knee osteoarthritis. Nat Rev Rheumatol 2019;15:49–60.
29. Pelletier JP, Cooper C, Peterfy C, Reginster JY, Brandi ML, Bruyere O, et al. What is the predictive value of MRI for the occurrence of knee replacement surgery in knee osteoarthritis? Ann Rheum Dis 2013;72:1594–604.
30. Raynauld JP, Martel-Pelletier J, Dorais M, Haraoui B, Choquette D, Abram F, et al. Total knee replacement as a knee osteoarthritis outcome: predictors derived from a 4-year long-term observation following a randomized clinical trial using chondroitin sulfate. Cartilage 2013;4:219–26.
31. Raynauld JP, Martel-Pelletier J, Haraoui B, Choquette D, Dorais M, Wildi LM, et al. Risk factors predictive of joint replacement in a 2-year multicentre clinical trial in knee osteoarthritis using MRI: results from over 6 years of observation. Ann Rheum Dis 2011;70:1382–8.
32. Dore D, Quinn S, Ding C, Winzenberg T, Zhai G, Cicuttini F, et al. Natural history and clinical significance of MRI-detected bone marrow lesions at the knee: a prospective study in community dwelling older adults. Arthritis Res Ther 2010;12:R223.
33. Tanamas SK, Wluka AE, Pelletier JP, Pelletier JM, Abram F, Berry PA, et al. Bone marrow lesions in people with knee osteoarthritis predict progression of disease and joint replacement: a longitudinal study. Rheumatology (Oxford) 2010;49:2413–9.
34. Scher C, Craig J, Nelson F. Bone marrow edema in the knee in osteoarthrosis and association with total knee arthroplasty within a three-year follow-up. Skeletal Radiol 2008;37:609–17.
35. Singh V, Oliashirazi A, Tan T, Fayyad A, Shahi A. Clinical and pathophysiologic significance of MRI identified bone marrow lesions associated with knee osteoarthritis. Arch Bone Jt Surg 2019;7:211–9.
36. Mattap SM, Aitken D, Wills K, Laslett L, Ding C, Pelletier JP, et al. How do MRI-detected subchondral bone marrow lesions (BMLs) on two different MRI sequences correlate with clinically important outcomes? Calcif Tissue Int 2018;103:131–43.
37. Teichtahl AJ, Cicuttini FM, Abram F, Wang Y, Pelletier JP, Dodin P, et al. Meniscal extrusion and bone marrow lesions are associated with incident and progressive knee osteoarthritis. Osteoarthritis Cartilage 2017;25:1076–83.
38. Laslett LL, Dore DA, Quinn SJ, Boon P, Ryan E, Winzenberg TM, et al. Zoledronic acid reduces knee pain and bone marrow lesions over 1 year: a randomised controlled trial. Ann Rheum Dis 2012;71:1322–8.
39. Martel-Pelletier J, Roubille C, Abram F, Hochberg MC, Dorais M, Delorme P, et al. First-line analysis of the effects of treatment on progression of structural changes in knee osteoarthritis over 24 months: data from the osteoarthritis initiative progression cohort. Ann Rheum Dis 2015;74:547–56.
40. Neogi T, Li S, Peloquin C, Misra D, Zhang Y. Effect of bisphosphonates on knee replacement surgery. Ann Rheum Dis 2018;77:92–7.
41. Altman R, Lim S, Steen RG, Dasa V. Hyaluronic acid injections are associated with delay of total knee replacement surgery in patients with knee osteoarthritis: evidence from a large U.S. health claims database. PLoS One 2015;10:e0145776.
42. Delbarre A, Amor B, Bardoulat I, Tetafort A, Pelletier-Fleury N. Do intra-articular hyaluronic acid injections delay total knee replacement in patients with osteoarthritis: a Cox model analysis. PLoS One 2017;12:e0187227.
43. Guermazi A, Hunter DJ, Li L, Benichou O, Eckstein F, Kwoh CK, et al. Different thresholds for detecting osteophytes and joint space narrowing exist between the site investigators and the centralized reader in a multicenter knee osteoarthritis study: data from the Osteoarthritis Initiative. Skeletal Radiol 2012;41:179–86.
44. Dodin P, Abram F, Pelletier JP, Martel-Pelletier J. A fully automated system for quantification of knee bone marrow lesions using MRI and the osteoarthritis initiative cohort. J Biomed Graph Comput 2013;3:51–65.
45. Tormalehto S, Mononen ME, Aarnio E, Arokoski JPA, Korhonen RK, Martikainen J. Health-related quality of life in relation to symptomatic and radiographic definitions of knee osteoarthritis: data from Osteoarthritis Initiative (OAI) 4-year follow-up study. Health Qual Life Outcomes 2018;16:154.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 73, No. 10, October 2021, pp 1528–1536
DOI 10.1002/acr.24329

Participation in Informal Caregiving Among People With Arthritis: Findings From a Canadian Longitudinal Study on Aging

Jessica M. Wilfong,[1] Anthony V. Perruccio,[2] Dov B. Millstone,[3] and Elizabeth M. Badley[2]

Objective. Despite the joint pain and significant dysfunction that characterizes arthritis, many people with arthritis continue to carry out everyday duties and responsibilities. The objective of the present study was to describe participation in informal caregiving (unpaid assistance to someone with a health issue or limitation) among people with arthritis.

Methods. Analysis of baseline data from the Canadian Longitudinal Study on Aging (CLSA), a nationally representative sample of people ages 45–85 years (n = 21,241), was performed. A questionnaire covering sociodemographic, health, and caregiving variables was completed by each study participant. Caregiving variables examined characteristics of the person who received the most care from the questionnaire respondent, as well as the types of caregiving (e.g., hands-on versus hands-off tasks) and amount of care provided (e.g., hours per week).

Results. There was no difference in the proportion of people with and without arthritis who provided informal care (46%). Individuals with arthritis reported worse health, but this did not affect the likelihood of providing care, nor the types or amount of care provided. Caregivers with and without arthritis were most likely to provide fewer than 7 hours per week of care, and the most common type of care was characterized as hands-off, particularly transportation assistance. Men were just as likely to provide care as women but were less likely to provide high intensity care or perform hands-on tasks.

Conclusion. Despite reporting worse health on average, people with arthritis were just as likely as people without arthritis to provide informal care. The need to provide informal care among people with arthritis may impact their ability to engage in self-management activities for their arthritis.

INTRODUCTION

Arthritis is the most common musculoskeletal disorder and is characterized by joint pain that can cause significant dysfunction. Altogether, musculoskeletal disorders, including arthritis, are one of the leading causes of global years lived with disability (1). The literature on arthritis tends to focus on the impacts of the disease, including disability and dysfunction, as well as risk factors and treatments. Therefore, the narrative surrounding arthritis is often focused on what people with the disease cannot do and how their life is negatively affected. This view neglects an understanding of the everyday duties and responsibilities still carried out by people with arthritis. For example, although arthritis increases the risk of not being in the labor force (2), studies show that the majority of people with arthritis continue to be fully engaged in employment (3). One aspect of participation that has not, to our knowledge, been explored is the role of people with arthritis providing informal care.

Informal caregivers are family and friends who provide unpaid assistance to someone with a health issue or limitation and can include assistance with everything from transportation to medical care. The literature on informal caregiving shows that a substantial proportion of the population participates in this responsibility. According to the 2008/2009 Canadian Community Health Survey

Supported by a service contract from The Arthritis Society of Canada to the Arthritis Community Research and Evaluation Unit. This research was made possible using data collected by the Canadian Longitudinal Study on Aging (CLSA). Funding for the CLSA is provided by the Government of Canada through the Canadian Institutes of Health Research (CIHR) under grant reference LSA 94473 and the Canada Foundation for Innovation. This research has been conducted using the CLSA data set (Baseline Tracking Dataset, version 3.3) under application number 170302. The CLSA is led by Drs. Parminder Raina, Christina Wolfson, and Susan Kirkland.

[1]Jessica M. Wilfong, MPH: Krembil Research Institute, University Health Network, and Arthritis Community Research and Evaluation Unit, Toronto, Ontario, Canada; [2]Anthony V. Perruccio, PhD, Elizabeth M. Badley, DPhil: Krembil Research Institute, University Health Network, and Arthritis Community Research and Evaluation Unit, and University of Toronto, Toronto, Ontario, Canada; [3]Dov B. Millstone, MPH: Arthritis Community Research and Evaluation Unit, Toronto, Ontario, Canada.

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

Address correspondence to Elizabeth M. Badley, DPhil, Krembil Research Institute, Toronto Western Hospital, MP 10-310, 399 Bathurst Street, Toronto, Ontario M5T 2S8, Canada. Email: E.Badley@utoronto.ca.

Submitted for publication April 29, 2020; accepted in revised form May 19, 2020.

SIGNIFICANCE & INNOVATIONS

- The tendency to only focus on the activities that individuals living with arthritis cannot do means we often neglect considering what duties and responsibilities individuals living with arthritis do. We found that people with arthritis participated in daily responsibilities, such as informal caregiving, to the same extent as people without arthritis.
- While individuals with arthritis had on average poorer health status than those without arthritis, this did not affect whether or not those affected by arthritis provided informal care, nor did it affect the types or amount of care they provided compared to individuals without arthritis.
- A similar proportion of men and women with arthritis assumed roles as caregivers, though there were differences in the types and amount of care they provided.

– Healthy Aging (CCHS), an estimated 3.8 million Canadians ages 45 years or older (35%) provide care to a senior individual with a short-term or long-term health condition (4). The prevalence of informal caregiving is estimated to increase as the aging population continues to grow. Studies show that informal caregiving reduces demands on the health care system and allows senior individuals to remain in their homes longer (5,6). However, many studies have also demonstrated that informal caregiving places a burden on the caregivers and that caregivers are at an increased risk of depression, stress, and anxiety as well as worse physical health and self-perceived well-being (7–10). Specifically, it has been suggested that caregivers may experience poorer physical health due to the effects of physical exertion involved in caregiving tasks associated with muscle strain, skeletal injury, aggravation of chronic illness such as arthritis, or other sources of physical discomfort and pain (11–13).

Women are consistently shown to be the predominant providers of informal care compared to men (4,14). Additionally, many studies have found a difference in the type and amount of care provided by men and women (15–17). However, these differences are not observed consistently across studies (14). More research is needed using large population-based samples as most previous caregiving studies examining gender differences are based on convenience samples that have self-selected participation and include very small samples of men (18). This area of study is becoming more important as demographic characteristics and social norms change and men increasingly assume the role of informal caregivers (14,19).

While there is an ever-growing body of literature on informal caregiving, there are no studies we are aware of that examine the role of caregiving among individuals with arthritis. In the present study, we used data from a large population-based health survey to describe informal caregivers with arthritis and compared them to caregivers without arthritis across sociodemographic and health characteristics. Additionally, we described the types and amount of care provided by those with and without arthritis and performed comparisons based on sex. Finally, we examined what sociodemographic and health factors among individuals with arthritis were associated with being a caregiver.

MATERIALS AND METHODS

Study design and setting. Data were obtained from the baseline tracking sample of the Canadian Longitudinal Study on Aging (CLSA) and were collected between September 2011 and May 2014. The CLSA is a longitudinal study that follows ~50,000 community-dwelling Canadians ages 45–85 years over a period of 20 years. The design and recruitment of the study has been described fully elsewhere (20). Briefly, the study consists of two samples. The tracking sample, used in the present study, consists of a nationally representative sample of 21,241 individuals from the 10 Canadian provinces. Participants completed the tracking questionnaire via computer-assisted telephone surveys. Participants were excluded if they could not communicate in English or French, had a cognitive impairment at the time of contact, were a resident of 1 of the 3 Canadian territories, were a full-time member of the Canadian Armed Forces, were a resident in a long-term care institution, or were living on reserves or other Aboriginal settlements. Additionally, no proxy responses were allowed in the present study.

Care activities. Participants were asked to indicate whether they had provided informal care or if they had received informal or formal care in the past 12 months. Informal caregiving was described as "types of assistance you may have provided to other people because of a health condition or limitation." Participants were instructed to only include "assistance you provided to family members, friends, and other people living both inside and outside your household." Informal care receiving was described as "different types of assistance that you may have received because of a health condition or limitation that affects your daily activities." For this section, participants were instructed to only include "assistance from family, friends, or neighbors." Formal care receiving was described as "home care services you may have received because of a health condition or limitation that affects your daily activities." Participants were instructed to only include "services provided by professionals or paid workers" for the formal care section.

Four care activities emerged from the questions: 1) caregivers who reported providing informal care, 2) care recipients who reported receiving informal and/or formal care, 3) both caregivers and care recipients who reported that they both provided and received care, and 4) neither caregivers nor care recipients who reported that they did not provide or receive care.

Characteristics of care provided. Caregivers were then asked a series of questions about to whom they provided care. First, they were asked how many people they had provided informal care to in the past 12 months. Then a series of questions were asked about "the person to whom, in the past 12 months, you have dedicated the most time and resources to assisting." Relationship to the care recipient was categorized as "spouse/partner," "parent/in-law," "other relative," or "friend, neighbor, or other." Residence of the care recipient was categorized as the same household, another household, and "other," with "other" indicating the recipient either resided in a health care institution or had been provided care in the past year and was now deceased. Finally, participants were asked to indicate the intensity (hours per week) and duration (months in the year) of care they provided to the care recipient.

Types of care provided. Participants were asked about what types of assistance they provided to someone because of a health condition or limitation in the past 12 months, including both hands-on and hands-off care. Hands-on care was defined as providing personal care (e.g., eating, dressing, bathing, or toileting) and medical assistance (e.g., taking medicine or nursing care). Hands-off care was defined as managing care (e.g., making appointments), help with household activities (e.g., housework, home maintenance, and outdoor work), transportation assistance (e.g., trips to the doctor or for shopping), or meal preparation (e.g., meal preparation or delivery).

Health variables. Participants were asked about any long-term health conditions diagnosed by a health professional, and participants were considered to have arthritis if they answered yes to having rheumatoid arthritis; osteoarthritis of the knee, hip, or hand; or any other type of arthritis. The other conditions asked about were asthma, chronic obstructive pulmonary disease, high blood pressure, diabetes mellitus, heart disease, stroke/transient ischemic attack, neurologic disorders, migraines, ulcers, bowel disorders, urinary incontinence, cancer, osteoporosis, thyroid disorders, kidney disease, mood disorders, anxiety disorders, and "other" conditions. We categorized the number of additional conditions that a participant reported (excluding arthritis) into 0, 1, 2, and 3 or more categories of co-occurring conditions for descriptive analyses. Body mass index (BMI) was calculated using self-reported height and weight, excluding pregnant women. BMI categories included: under/normal weight (<24.9 kg/m^2), overweight (25–29.9 kg/m^2), or obese (≥30 kg/m^2). Participants were also asked to self-rate their health. Responses were dichotomized as follows: 1) excellent or very good or good and 2) fair or poor.

Participants completed the Center for Epidemiological Studies Short Depression Scale (CES-D 10) and were given a total score (range 0–30). Any score of 10 or higher was considered to signify a "depressed" state (21). Participants also completed the Satisfaction with Life Scale (SWLS), and a total score was calculated (range 5–35) (22). Participants were classified as "dissatisfied" for scores between 5 and 19, and "neutral or satisfied" for scores between 20 and 35 (23). Additionally, participants who identified as not being free of pain or discomfort were asked how many activities their pain or discomfort prevented ("none," "a few," "some," or "most"). Participants were identified as having pain that prevents activities if they responded that at least a few of their activities were limited. Participants were also asked about difficulty performing everyday activities such as lifting their arms above their shoulders, stooping, crouching or kneeling, pushing or pulling large objects, lifting 10 pounds, handling small objects, standing or sitting for long periods of time, standing up after sitting, walking for various amounts of time, making their bed, washing their backs, cutting food with a knife, and, finally, able to withstand some force or impact through their arm. For descriptive analyses, those who reported having any difficulties with at least 2 of the movements were identified as having functional difficulties.

Sociodemographic variables. Age was categorized into 10-year intervals (45–54, 55–64, 65–74, and 75–85 years). Marital status was dichotomized as either single or married/common law married. Household income was separated into 3 categories (<$50,000; $50,000–99,999; and >$100,000). Highest level of education achieved was categorized as high school or less and postsecondary education. Employment was categorized as not working or retired, working 1–19 hours per week, working 20–29 hours per week, and working full-time (>30 hours per week).

Statistical analysis. The distribution of people with and without arthritis who participate in care activities was depicted with a stacked bar graph. Comparisons between groups were made by examining the overlap of 95% confidence intervals (95% CIs). Sociodemographic and health characteristics of informal

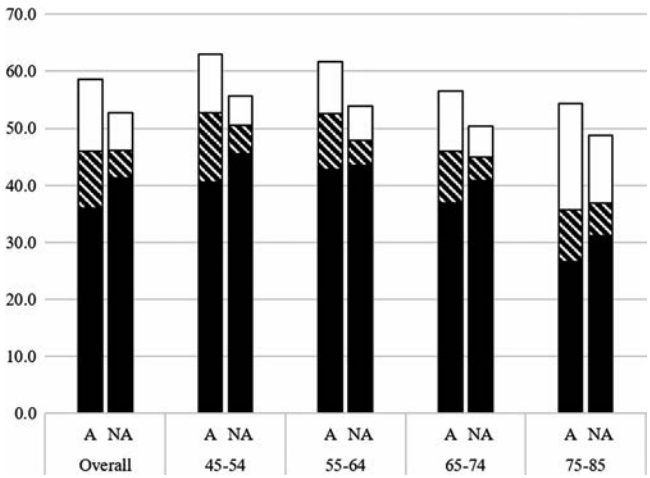


Figure 1. Proportion of caregivers (solid), care recipients (white), and both caregivers and care recipients (stripe) with arthritis (A) and those without arthritis (NA) who provide and receive care overall and by age in years.

caregivers were described for individuals with arthritis and those without, as well as the characteristics and types and amount of care provided by these individuals. Significant differences between caregivers with arthritis and those without were identified by chi-square test. Multivariable Poisson logistic regression analyses were performed to compare the characteristics of caregivers with arthritis and those without arthritis, adjusting for differences in demographic characteristics between the two groups and to test which sociodemographic and health factors were associated with caregiving among individuals with arthritis. Prevalence ratios (PRs) produced by Poisson logistic regression analysis are a more accurate estimate of risk when the outcome is not rare (≥10%) (24). The characteristics and types of care provided were further stratified by sex, and differences were assessed by chi-square tests. Analytic weights provided by the CLSA were used to derive estimates representative of the Canadian population in the 10 provinces.

RESULTS

Proportion of sample providing/receiving care. Over one-third (36%) of the CLSA sample reported having arthritis. Proportions of individuals with arthritis and those without who reported providing and receiving care overall and by age group are shown in Figure 1. Supporting numerical information, including proportions and 95% CIs used to make comparisons between groups, is presented in Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24329/abstract. More people were caregivers than care recipients overall and within each age group. The proportion of care recipients increased with age and was higher for people with arthritis. Including the minority of the study population who were both caregivers and care recipients, almost half were caregivers, with the same proportion (46%) among people with arthritis and those without. Excluding those who both provided and received care, the proportion of people with arthritis who provide informal care (36%) was nonetheless similar to people without arthritis (41%).

Characteristics of caregivers. Characteristics of the caregivers with arthritis and those without are presented in Table 1. Overall, the proportion of women who reported providing informal care was only slightly higher than the proportion of men who reported providing informal care. The mean age of the study population was 63 years, and the majority of individuals were married

Table 1. Demographic characteristics of caregivers with arthritis compared to caregivers without arthritis

	Caregivers with arthritis, %	Caregivers without arthritis, %	*P*
Female sex	62.0	52.0	<0.001
Age, years			
45–54	25.9	46.1	<0.001
55–64	36.2	30.8	–
65–74	24.0	15.9	–
≥75 years	13.9	7.2	–
Married/common law	71.5	77.0	<0.001
Yearly household income			
<$50,000	34.7	24.0	<0.001
$50,000–99,999	37.7	37.7	–
≥$100,000	27.6	38.2	–
Postsecondary education	71.0	75.6	<0.001
Employment status			
Not currently working	55.6	35.9	<0.001
1–19 hours per week	9.2	9.7	–
20–29 hours per week	5.3	7.1	–
≥30 hours per week	29.9	47.3	–
Body mass index			
Normal/underweight	31.6	40.9	<0.001
Overweight	37.8	38.8	–
Obese	30.6	20.3	–
Co-occurring medical conditions			
0	10.9	22.3	<0.001
1	18.8	29.1	–
2	22.5	21.9	–
3+	47.8	26.7	–
Fair/poor self-rated health	17.0	7.5	<0.001
Dissatisfied state on life satisfaction scale	14.1	9.7	<0.001
Depressed state on depression scale	23.0	15.5	<0.001
Presence of pain that limits abilities	37.1	14.6	<0.001
Presence of ≥2 functional difficulties	50.5	18.4	<0.001

Table 2. Health-related characteristics of caregivers with arthritis compared to caregivers without arthritis*

	PR (95% CI) for arthritis	*P*
Body mass index (ref. normal weight/ underweight)		
Overweight	1.11 (1.02–1.22)	0.018
Obese	1.20 (1.09–1.33)	0.000
Self-rated health (ref. excellent/very good/good)		
Fair/poor	0.91 (0.81–1.03)	0.148
Life satisfaction scale (ref. satisfied/neutral)		
Dissatisfied	0.97 (0.86–1.09)	0.602
Depression scale (ref. not depressed)		
Depressed	1.00 (0.91–1.10)	0.981
Number of co-occurring conditions	1.06 (1.03–1.08)	<0.001
Pain that limits activities (ref. No)		
Yes	1.48 (1.36–1.62)	<0.001
Number of functional difficulties	1.09 (1.07–1.11)	<0.001

* Values were adjusted for sex, age, marital status, education, and employment status. Data were calculated by Poisson logistic regression analysis. 95% CI = 95% confidence interval; PR = prevalence ratio.

and had a postsecondary education. Characteristics of informal caregivers with arthritis and without arthritis were generally similar, but there were minor differences. Caregivers with arthritis were more likely to be women and tended to be older, have lower education attainment, and be more likely to not currently work full-time compared to those without arthritis. Caregivers with arthritis were also much more likely to report worse health outcomes. A higher proportion of caregivers with arthritis, compared to those without, reported fair or poor health, life dissatisfaction, and depression.

Table 3. Characteristics of the care provided by individuals with arthritis compared to individuals without arthritis

	Caregivers with arthritis, %	Caregivers without arthritis, %	*P*
Sex of recipient of most care			
Male	31.0	31.5	0.586
Female	69.0	68.5	–
Relationship to recipient of most care			
Spouse/partner	15.3	14.2	<0.001
Parent/in-law	34.6	44.5	–
Other relative	21.2	17.4	–
Friend, neighbor, or other	28.9	23.9	–
Dwelling of recipient of most care			
Outside of home	64.6	64.4	0.989
In home	19.8	19.9	–
Other (e.g., health care institution)	15.6	15.7	–
Number of people cared for			
1 person	54.7	56.3	0.121
More than 1 person	45.3	43.7	–
Hours of care provided per week			
<7 hours	63.7	65.1	0.140
>7 hours	36.4	34.9	–
Duration of care in the past 12 months			
1–4 months	55.1	58.1	0.016
5–8 months	11.0	10.4	–
9–12 months	33.9	31.5	–
Types of care provided			
Hands-off tasks only	58.4	59.4	0.579
Transportation assistance	72.7	73.8	–
Help with household activities	52.5	54.4	–
Meal preparation	45.9	43.8	–
Managing care	31.1	30.7	–
Hands-on tasks only	3.9	3.7	–
Personal care	30.0	29.3	–
Medical assistance	27.1	25.8	–
Both hands-off and hands-on tasks	37.7	36.9	–

Additionally, individuals with arthritis were far more likely to report experiencing pain that limits their activities compared to individuals without arthritis (37% versus 15%, respectively) and were far more likely to report having two or more functional difficulties compared to caregivers without arthritis (51% versus 18%, respectively). Differences in health between individuals with arthritis and those without arthritis persisted after adjustment for differences in demographic characteristics (Table 2). While there were no significant differences in general health (self-rated health, life satisfaction, and depression), individuals with arthritis were significantly more likely to report poorer physical health compared to caregivers without arthritis, with more co-occurring conditions reported (PR 1.06 [95% CI 1.03–1.08]), pain that limits activity (PR 1.48 [95% CI 1.36–1.62]), and number of functional difficulties (PR 1.09 [95% CI 1.07–1.11]). Caregivers with arthritis were also significantly more likely to be overweight or obese.

Characteristics of care provided by individuals with arthritis. Characteristics of care provided among individuals with arthritis and those without arthritis are shown in Table 3. Differences between the two groups were minimal, with individuals with arthritis being slightly less likely to provide care to a parent and slightly more likely to provide a longer duration of care.

Results of regression analysis. The results of the Poisson logistic regression analytical model with the outcome of being a caregiver (yes/no) among individuals with arthritis are presented in Table 4. Being a woman (PR 1.14 [95% CI 1.06–1.23]), earning a moderate income (PR 1.13 [95% CI 1.03–1.24]), and having a higher educational attainment (PR 1.12 [95% CI 1.03–1.22]) were associated with an increased likelihood of being a caregiver, whereas being older (PR 0.84 [95% CI 0.74–0.95] and PR 0.68 [95% CI 0.59–0.78] for the 65–74 years age group and >75 years age group, respectively) and currently working full-time for ≥30 hours per week (PR 0.88 [95% CI 0.79–0.98]) were associated with a decreased likelihood of being a caregiver. Receiving care, number of co-occurring conditions, having pain, and number of functional limitations were not significantly associated with being a caregiver. Results were similar among individuals without arthritis (data not shown).

Characteristics of care provided by sex. Finally, we further investigated the relationship between sex and the characteristics of care provided among individuals with arthritis (Table 5). Women with arthritis who were caregivers were more likely to provide care to a woman, to an extended family member, or to someone outside of their household than men. Additionally, among those with arthritis, women who were caregivers were more likely than men who were caregivers to provide a greater number of hours per week of care and were more likely to report providing care involving hands-on tasks, including personal care and medical assistance. Results were similar among individuals without arthritis (data not shown).

Table 4. Characteristics of individuals with arthritis who are caregivers compared to individuals with arthritis who are not caregivers*

	PR (95% CI) for caregiving	*P*
Care recipient (ref. no)		
Yes	1.06 (0.97–1.17)	0.200
Sex (ref. male)		
Female	1.14 (1.06–1.23)	0.001
Age, years (ref. 45–54 years)		
55–64	0.99 (0.90–1.10)	0.917
65–74	0.84 (0.74–0.95)	0.004
≥75	0.68 (0.59–0.78)	<0.001
Marital status (ref. single)		
Married/common law	1.01 (0.92–1.11)	0.801
Yearly household income (ref. <$50,000)		
$50,000–99,999	1.13 (1.03–1.24)	0.012
≥$100,000	1.08 (0.96–1.21)	0.184
Education (ref. high school or less)		
Postsecondary	1.12 (1.03–1.22)	0.007
Employment status (ref. retired/not working)		
1–19 hours per week	0.90 (0.79–1.03)	0.122
20–29 hours per week	0.92 (0.78–1.09)	0.346
≥30 hours per week	0.88 (0.79–0.98)	0.015
Number of co-occurring conditions	1.00 (0.98–1.03)	0.649
Pain that limits activities (ref. no)		
Yes	0.96 (0.89–1.05)	0.382
Number of functional difficulties, by linear model analysis	1.00 (0.96–1.03)	0.813
Number of functional difficulties, by quadratic model analysis	1.00 (0.99–1.00)	0.207

* Data were calculated by Poisson logistic regression analysis. 95% CI = 95% confidence interval; PR = prevalence ratio.

DISCUSSION

In the present study of a representative population-based sample of Canadians ages 45 years or older, we found minimal differences in the proportion of people with arthritis and those without arthritis who provided informal care. Our findings suggest that the experience of pain and functional difficulties does not affect whether someone with arthritis provides informal care. We also found that there were distinct differences based on sex in the type and amount of care provided by men and women with arthritis who provide informal care. To our knowledge, no study to date has examined specifically the role of caregiving among individuals with arthritis.

In light of the most previous literature that focuses on the limitations of people with arthritis in carrying out activities, it might be assumed that people with arthritis would be less likely to participate in caregiving than those without arthritis. Additionally, previous

Table 5. Characteristics of the care provided by individuals with arthritis according to sex*

	Male caregivers with arthritis	Female caregivers with arthritis	*P*
Sex of recipient of most care			
Male	34.6	28.8	<0.001
Female	65.5	71.2	–
Relationship to recipient of most care			
Spouse/partner	19.1	12.9	<0.001
Parent/in-law	36.0	33.8	–
Other relative	16.2	24.3	–
Friend, neighbor, or other	28.8	29.0	–
Dwelling of recipient of most care			
Outside of home	60.9	66.8	<0.001
In home	23.1	17.8	–
Other (e.g., health care institution)	31.8	15.4	–
Number of people cared for			
1 person	52.2	56.2	0.021
More than 1 person	47.8	43.8	–
Hours of care provided per week			
<7 hours	68.1	61.0	<0.001
>7 hours	31.9	39.1	–
Duration of care in the past 12 months			
1–4 months	57.3	53.8	0.083
5–8 months	9.9	11.7	–
9–12 months	32.8	34.5	–
Types of care provided			
Hands-off tasks only	66.7	53.3	<0.001
Transportation assistance	77.3	69.9	–
Help with household activities	59.1	48.4	–
Meal preparation	35.7	52.2	–
Managing care	26.8	33.7	–
Hands-on tasks only	2.7	4.6	–
Personal care	21.4	35.2	–
Medical assistance	23.0	29.6	–
Both hands-off and hands-on tasks	30.6	42.1	–

work in the literature on caregiving has suggested that poor physical health may decrease the likelihood of someone participating in caregiving activities (12,13), may increase difficulties in managing caregiving responsibilities (25), and may also increase the likelihood of institutionalization of the care recipient (26). Contrary to these observations, we found that despite experiencing much worse physical health, caregivers with arthritis provided similar amounts and types of care as caregivers without arthritis. Additionally, we found that predictors of caregiving did not differ between those with arthritis and those without, and most notably, that requiring care, having co-occurring conditions, and experiencing pain and functional difficulties had no influence over whether someone with arthritis was a caregiver or not. This finding could reflect previous studies that have shown that a large proportion of caregivers feel a lack of choice when taking on the role of a caregiver (27). This lack of choice is theorized to possibly stem from kinship relationships and a sense of obligation and also be due to the availability and accessibility of services and of other family members.

Often due to social and cultural expectations, women tend to make up the majority of informal caregivers (14). These findings are consistent across studies ranging from small convenience samples (18) to studies that examine population data (4). In the present study of population-based data, we found that the proportion of men who reported being informal caregivers was only slightly lower than for women. This finding supports other recent studies which show that with changing social norms, men are increasingly assuming roles as caregivers (14,19).

There has been some debate in the literature with regard to whether there are differences based on sex among caregivers, including the intensity of care and types of care provided (16,18). Our findings support a body of literature that asserts that women are more likely to provide personal and hands-on care, such as help with eating, dressing, and bathing. It has been suggested that women are more likely to take on a primary caregiving role and are less likely to solicit help from secondary informal or formal caregivers. Therefore, they are taking on more hours of care and taking on more complex duties (15,28,29). Our finding that women were more likely to provide more hours of care and more hands-on care may suggest that women are caring for individuals with greater illness and needs. That men were more likely to provide fewer hours of care per week and more hands-off tasks may suggest that they are providing services such as driving individuals to appointments or performing yard work. More investigation

is needed to understand how individual circumstances surrounding the provision of care influences these relationships.

The major strengths of this study are that it utilizes data from a nationally representative survey and focuses on individuals who reported having arthritis and who also reported providing informal care to someone. To our knowledge, this is a group that has not been previously studied. Another strength of the study is that it focuses on the complete range of caregiving, rather than just caregivers to people with one certain type of condition (e.g., people with dementia or patients receiving palliative care). Details on the health of care recipients could not be examined further as the survey we used did not collect any information on the condition of the care recipient.

A limitation of the study is that not all questions in the survey were asked using the same time frame. While questions on caregiving were asked in the context of the past 12 months, health-related information was asked in reference to the time of the survey, thus restricting the opportunity to look at the health status of the caregiver during the time frame of caregiving. Future longitudinal studies are needed to test whether caregiving tasks aggravate arthritis, which has been previously suggested as possibly being a risk of these activities (11–13). Only general questions about caregiving were available in our data set. Further studies of caregiving provided to others by individuals with arthritis should consider the use of validated measurements of disease-specific health status and informal care-related burden to confirm the findings of the current analyses and provide a further base of evidence. As with most population-based surveys, arthritis diagnosis was self-reported, which potentially introduces misclassification. However, for broad population-based work, self-reported arthritis is deemed valid (30,31). Finally, the study was based on cross-sectional data, which means we cannot attribute directionality to the reported associations.

Nearly half of people with arthritis provided informal care to someone, the same proportion as people without arthritis, despite reporting substantial pain, activity limitation, and more co-occurring medical conditions. With the aging population of Canada (and other high-income countries), the number of both people with arthritis and people who require informal care is on the rise. Although informal caregiving provides clear benefits to the health care system, it has also been shown to be physically and emotionally demanding for the caregivers. In the future, it is important for health care teams to understand that people with arthritis may have caregiving obligations, which could impact their ability to engage in arthritis self-management activities and to care for themselves.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, approved the final draft to be published, and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. Dr. Badley had full access to all 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. Perruccio, Badley.
Acquisition of data. Perruccio, Badley.
Analysis and interpretation of data. Wilfong, Perruccio, Millstone, Badley.

REFERENCES

1. Vos T, Allen C, Arora M, Barber RM, Bhutta ZA, Brown A, et al. 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.
2. Sharif B, Garner R, Sanmartin C, Flanagan WM, Hennessy D, Marshall DA. Risk of work loss due to illness or disability in patients with osteoarthritis: a population-based cohort study. Rheumatology (Oxford) 2016;55:861–8.
3. Gignac MA, Cao X, Lacaille D, Anis AH, Badley EM. Arthritis-related work transitions: a prospective analysis of reported productivity losses, work changes, and leaving the labor force. Arthritis Rheum 2008;59:1805–13.
4. Turner A, Findlay L. Informal caregiving for seniors. Health Rep 2012;23:33–6.
5. Canadian Institute for Health Information. Supporting informal caregivers—the heart of home care. Ottawa (ON): Canadian Institute for Health Information; 2010.
6. Jull J. Seniors caring for seniors: examining the literature on injuries and contributing factors affecting the health and well-being of older adult caregivers. Provided to the Public Health Agency of Canada Prepared on behalf of the Canadian Association of Occupational Therapists. 2010.
7. Bastawrous M. Caregiver burden—a critical discussion. Int J Nurs Stud 2013;50:431–41.
8. Capistrant BD. Caregiving for older adults and the caregivers' health: an epidemiologic review. Curr Epidemiol Rep 2016;3:72–80.
9. Ho JS, Bordon J, Wang V, Ceglowski J, Kim DH, Chattillion EA, et al. Reduced activity restriction buffers the relations between chronic stress and sympathetic nervous system activation. J Gerontol B Psychol Sci Soc Sci 2013;69:408–16.
10. Pinquart M, Sörensen S. Differences between caregivers and non-caregivers in psychological health and physical health: a meta-analysis. Psychol Aging 2003;18:250.
11. Pruchno RA, Potashnik SL. Caregiving spouses. Physical and mental health in perspective. J Am Geriatr Soc 1989;37:697–705.
12. Pinquart M, Sorensen S. Correlates of physical health of informal caregivers: a meta-analysis. J Gerontol B Psychol Sci Soc Sci 2007;62:P126–37.
13. Bauer JM, Sousa-Poza A. Impacts of informal caregiving on caregiver employment, health, and family. J Popul Ageing 2015;8:113–45.
14. Sharma N, Chakrabarti S, Grover S. Gender differences in caregiving among family - caregivers of people with mental illnesses. World J Psychiatry 2016;6:7–17.
15. Navaie-Waliser M, Spriggs A, Feldman PH. Informal caregiving: differential experiences by gender. Med Care 2002;40:1249–59.
16. Pinquart M, Sorensen S. Gender differences in caregiver stressors, social resources, and health: an updated meta-analysis. J Gerontol B Psychol Sci Soc Sci 2006;61:P33–45.
17. Yee JL, Schulz R. Gender differences in psychiatric morbidity among family caregivers: A review and analysis. Gerontologist 2000; 40:147–64.
18. Houde SC. Methodological issues in male caregiver research: an integrative review of the literature. J Adv Nurs 2002;40:626–40.
19. Baker KL, Robertson N. Coping with caring for someone with dementia: reviewing the literature about men. Aging Ment Health 2008;12:413–22.

20. Raina PS, Wolfson C, Kirkland SA, Griffith LE, Oremus M, Patterson C, et al. The Canadian longitudinal study on aging (CLSA). Can J Aging 2009;28:221–9.
21. Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). Am J Prev Med 1994;10:77–84.
22. Diener E, Emmons RA, Larsen RJ, Griffin S. The Satisfaction With Life Scale. J Pers Assess 1985;49:71–5.
23. Pavot W, Diener E. The Satisfaction with Life Scale (SWL). Measurement instrument database for the social science. 2013. URL: https://www.midss.org/content/satisfaction-life-scale-swl.
24. McNutt LA, Wu C, Xue X, Hafner JP. Estimating the relative risk in cohort studies and clinical trials of common outcomes. Am J Epidemiol 2003;157:940–3.
25. Navaie-Waliser M, Feldman PH, Gould DA, Levine C, Kuerbis AN, Donelan K. When the caregiver needs care: the plight of vulnerable caregivers. Am J Public Health 2002;92:409–13.
26. McCann JJ, Hebert LE, Bienias JL, Morris MC, Evans DA. Predictors of beginning and ending caregiving during a 3-year period in a biracial community population of older adults. Am J Public Health 2004;94:1800–6.
27. Schulz R, Beach SR, Cook TB, Martire LM, Tomlinson JM, Monin JK. Predictors and consequences of perceived lack of choice in becoming an informal caregiver. Aging Ment Health 2012;16:712–21.
28. Arber S, Ginn J. Gender differences in informal caring. Health Soc Care Community 1995;3:19–31.
29. Del-Pino-Casado R, Frias-Osuna A, Palomino-Moral PA, Ramon Martinez-Riera J. Gender differences regarding informal caregivers of older people. J Nurs Scholarsh 2012;44:349–57.
30. Bombard JM, Powell KE, Martin LM, Helmick CG, Wilson WH. Validity and reliability of self-reported arthritis: Georgia senior centers, 2000–2001. Am J Prev Med 2005;28:251–8.
31. Sacks JJ, Harrold LR, Helmick CG, Gurwitz JH, Emani S, Yood RA. Validation of a surveillance case definition for arthritis. J Rheumatol 2005;32:340–7.

Arthritis Care & Research
Vol. 73, No. 10, October 2021, pp 1537–1543
DOI 10.1002/acr.24357

American College of Rheumatology
Empowering Rheumatology Professionals

BRIEF REPORT

Cost-Effectiveness of Colchicine Prophylaxis for Gout Flares When Commencing Allopurinol

Philip C. Robinson,[1] Nicola Dalbeth,[2] and Peter Donovan[1]

Objective. Colchicine prophylaxis to prevent gout flares when commencing urate-lowering therapy is recommended by international rheumatology society guidelines. Whether this is a cost-effective intervention is currently unknown. Our objective was to perform a cost-effectiveness analysis using both a US cost input model and an Australian cost input model.

Methods. This cost-effectiveness analysis was completed from the point of view of the third-party payer. We used a 2-arm decision tree with 1 arm commencing allopurinol with no colchicine prophylaxis and the other with colchicine prophylaxis. Model inputs were drawn from published literature where available. We completed a univariate and probabilistic sensitivity analysis to confirm the robust nature of the modeling. The time frame for the model was 6 months.

Results. The colchicine prophylaxis arm resulted in a cost of $1,276 and 0.49 quality-adjusted life-years (QALYs), while in the placebo arm the cost was $516 and 0.47 QALYs, with an incremental cost-effectiveness ratio of $34,004 per QALY gained. In Australia, where cost of colchicine was much lower, the colchicine arm dominated the placebo ($208 [Australian] in the colchicine arm versus $415 [Australian] in the placebo). Univariate and probability sensitivity analysis demonstrated that results were robust to changes in input parameters. In the probabilistic sensitivity analysis, the probability of colchicine prophylaxis being the most cost-effective option was 93% in the US and 100% in the Australian setting.

Conclusion. Colchicine prophylaxis to prevent gout flares while commencing allopurinol in gout is very cost-effective.

INTRODUCTION

Gout is the most common form of inflammatory arthritis, affecting ~4% of adults in the US (1). Colchicine is a commonly used drug in both treatment of flares and also the prophylaxis of flares when commencing urate-lowering therapy (ULT) (2). The successful treatment of gout requires serum urate to be lowered below an appropriate target of either 5 mg/dl (0.30 mmoles/liter) or 6 mg/dl (36 mmoles/liter) (3). This usually requires the use of ULT. The introduction of ULT is often associated with an increase in gout flares. Strategies to reduce gout flares during initiation of ULT usually involve the use of antiinflammatory agents such as colchicine, nonsteroidal antiinflammatory drugs, or glucocorticoids. Randomized controlled trials have demonstrated the benefit of this approach in reducing gout flares (4,5), and the use of antiinflammatory prophylaxis when introducing ULT is recommended by international rheumatology society guidelines (6–8).

Colchicine is relatively inexpensive in many jurisdictions, including Australia. In the US, colchicine was given orphan drug status in 2009 for gout and familial Mediterranean fever, and the price increased from cents per tablet to dollars per tablet (9). This change has resulted in a reduction in the prescription of colchicine for newly diagnosed patients with gout and familial Mediterranean fever (10).

At present, it is unknown whether gout flare prophylaxis with colchicine is a cost-effective intervention. The aim of this study

[1]Philip C. Robinson, MBChB, PhD, FRACP, Peter Donovan, MBBS, MSc, FRACP: Royal Brisbane and Women's Hospital and University of Queensland, Brisbane, Queensland, Australia; [2]Nicola Dalbeth, MBChB, MD, FRACP: University of Auckland, Auckland, New Zealand.

Dr. Robinson has received consulting fees and/or speaking fees from AbbVie, Eli Lilly and Company, Janssen, Novartis, Pfizer, Roche, and UCB (less than $10,000 each) and research support from those companies. Dr. Dalbeth has received consulting fees and/or speaking fees from Pfizer, Horizon, Janssen, AbbVie, Horizon, Arthrosi, AstraZeneca, Dyve Biosciences, and Hengrui (less than $10,000 each) and research support from Amgen and AstraZeneca. No other disclosures relevant to this article were reported.

Address correspondence to Philip C. Robinson, MBChB, PhD, FRACP, Department of Rheumatology, Royal Brisbane Hospital, Bowen Bridge Road, Herston, Queensland 4006, Australia. Email: philip.robinson@uq.edu.au.

Submitted for publication December 31, 2019; accepted in revised form June 9, 2020.

SIGNIFICANCE & INNOVATIONS

- Colchicine prophylaxis is very cost-effective for preventing gout flares in both the US and Australian health care systems.
- Despite the increased cost associated with prescribing colchicine in the US health care environment due to its prior orphan drug status, it is still cost-effective.

was to examine the cost-effectiveness of gout flare prophylaxis with colchicine when introducing the ULT agent allopurinol with models using inputs from both the US and Australian health care environment. The US was chosen because drug pricing is generally higher there than in other developed nations, and Australia was chosen because it is an example of a largely single pharmaceutical purchaser resulting in generally lower pricing.

PATIENTS AND METHODS

The reporting of our methods and results of this economic analysis conforms with the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement (11). To conduct this study, we performed a cost–utility analysis using TreeAge Pro 2015 R2 (TreeAge Software). We used a decision tree that assessed a treatment duration of 6 months and modeled 2 different scenarios (Figure 1). We modeled from the perspective of the third-party payer, so costs borne directly by the patient (e.g., out-of-pocket costs or costs associated with driving to the doctor) were not included. The common components were a patient with gout commencing on allopurinol 100 mg and increasing in 100-mg increments, with 50-mg increment increases for those with creatinine clearance of 20–50 ml/minute. The 2 arms then modeled either colchicine 0.6 mg twice a day or placebo. The central model parameters were based on the published trial by Borstad et al for preventing gout flares with allopurinol initiation, which is the only published trial of this approach (4). This trial was a prospective, randomized, double-blind placebo-controlled study that treated 43 patients, 21 in the colchicine group and 22 in the placebo group. The mean age was 63 years; there were 37 men and 6 women. All participants who received the study drug were included in the analysis.

Patients in each arm were modeled to incur gout flares at the same frequency as seen in each arm of the trial by Borstad et al (4). Model assumptions for the costs are shown in Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24357/abstract. Patients in the active colchicine prophylaxis arm had the financial cost of the colchicine and the financial cost of treating flares, including cost for both medication and medical services along with the cost of quality-adjusted life-years (QALYs) incurred by any gout flares and diarrhea. Patients in the placebo arm had no financial colchicine cost but the financial cost of medication for treatment of gout flare and cost of medical services to treat those gout flares, along with the cost of QALYs from gout flares and diarrhea. As per our previous published research, we modeled the cost and location of gout flare treatment in 3 different ways (12). We modeled home/self-treatment, primary care physician (PCP) treatment, and hospital treatment. The proportion of the flares that were treated in each setting was based on model inputs taken from previous published research, which determined that 20% of patients with gout in New Zealand were not treated by PCPs or in hospital, and that 1.3% of patients with gout had their gout flares treated in hospital, leaving 78.7% treated by PCP (2,12–17).

We modeled the costs of treatment of gout flares using both Australian inputs and US inputs in 2019 dollars. For home treatment of gout flares, we modeled the QALY reduction. For PCP flare treatment, in the base case analysis, we modeled the cost of the medical consultation, and laboratory tests were also assumed to have been ordered at 50% of these PCP appointments (1 each of full blood examination, C-reactive protein level, and electrolytes and renal function tests). We also modeled the QALY reduction. For hospital treatment, we modeled the cost of a gout hospital admission in the US and Australia and the QALY reduction. For hospital admission cost calculation, we used diagnosis-related group (DRG) 553 and 554 for "bone diseases & arthropathies with/without major complication or comorbidity." For the base case, we modeled a 50:50 proportional split between these 2 admission costs. For the US model, we used reported Medicare costs for admissions, physician visits costs, and laboratory costs (18). For the Australian model, we used Australian refined DRG cost weights from the Independent Hospital Pricing Authority (v8.0x) for admissions (19) and the Pharmaceutical Benefits Scheme for medication costs (20). We used the Australian Medicare Benefits Scheme for costs associated with physician visits and laboratory tests (21).

In the trial conducted by Borstad et al (4), there was a chance of diarrhea both in the colchicine arm (38%) and placebo arm (4.5%). When patients experienced diarrhea, we modeled them decreasing their medication dose to only 0.6 mg per day. Colchicine-related diarrhea was common in the study by Borstad et al (4) but was never a reason for patients to withdraw from the study, and all participants who experienced it responded to a dose decrease. Colchicine-induced diarrhea was the only modeled adverse event, as this is the only common adverse effect of colchicine observed. This is supported by a recent meta-analysis of randomized controlled trials of colchicine (22). We did not model a change in flare rate based on dose reduction of the colchicine because that information was not included in the results of the study by Bortsad et al.

Data on quality of life were obtained from the study by Perez-Ruiz et al, which details a disutility of 0.0097 and a 7-day duration for a single gout flare based on EuroQol 5-domain questionnaire data (23). This duration of gout flare correlated well with the study by Borstad et al (4), which found gout flare duration of 5.6–6.0 days (4). QALY reductions attributable to diarrhea are not available specifically for colchicine, so we used data on QALY reductions

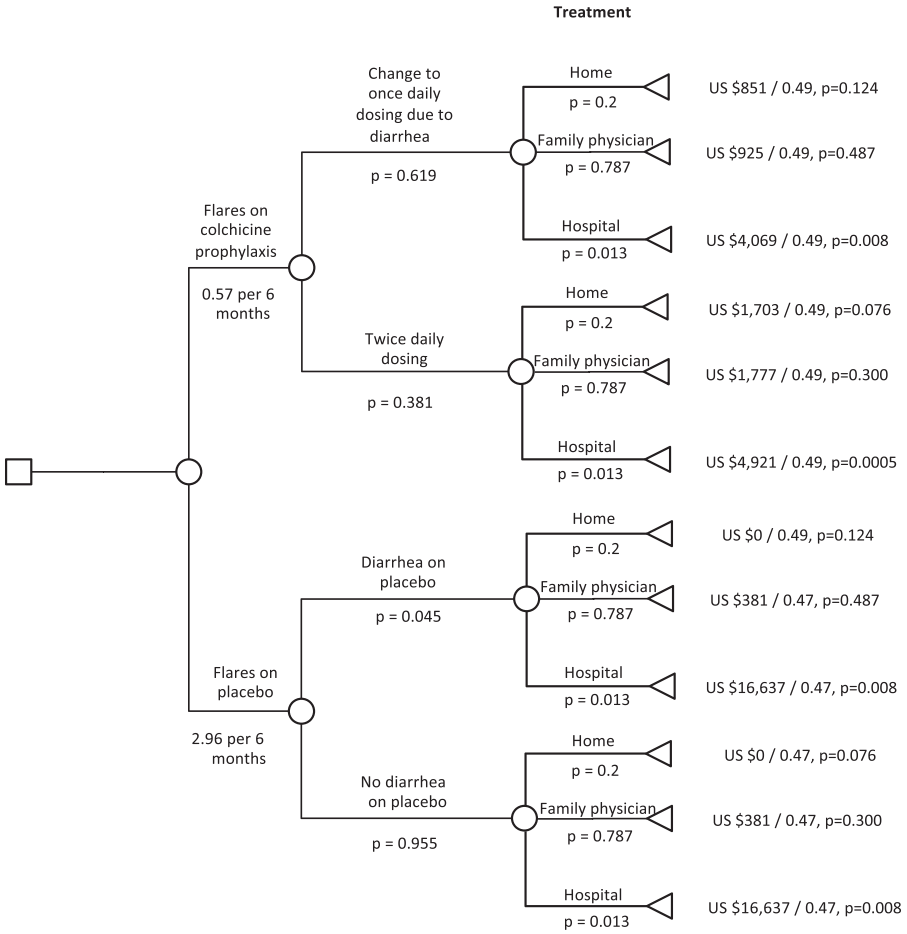


Figure 1. Model diagram of the study. Probabilities at the chance node, plus overall probability, cost, and quality-adjusted life years gained for each arm of the decision tree are displayed. Patients taking colchicine were assumed to have experienced 0.57 flares over 6 months, on average, while patients taking placebo were assumed to have experienced 2.96 flares over 6 months, on average, according to the study by Borstad et al (4). Flare costs were multiplied by the expected flare frequency for each arm. The costs in the colchicine arm for home treatment represent the estimated costs for low dose ($4.73 × 180 = $851) and high dose ($4.73 × 180 × 2 = $1,703) colchicine therapy alone. Square represents decision node; circles represent chance nodes; triangles represent terminal nodes.

with rotavirus infection in hospitalized children, a condition that we speculate is more severe than colchicine diarrhea (24). We therefore estimated a QALY reduction of one-half this value in the base case and included a wide range in the sensitivity analysis up to the value reported in these hospitalized children.

There is a wide variation in the cost of colchicine through the US health care system, so we included a wide range in our sensitivity analysis based on multiple sources. The cost of colchicine was modeled in the base case on the median reported Federal Supply Scheme (FSS) cost from the Veterans Affairs (VA) health system, being $3.91 per tablet, multiplied by the recommended factor of 1.21, giving a base case price of $4.73 per tablet. Guidance provided by the VA on using their data for health economic research recommends multiplying the FSS price by a factor of 1.21 to get a general US health care system price (25). This is based on work from the US Congressional Budget Office, which reports that the Medicaid cost is ~121% of the FSS cost (26). The lowest reported VA health system cost is $0.75 per tablet, and this price was used as the lowest end of the univariate sensitivity analysis (27). The highest FSS cost reported by the VA health system was $6.68 per tablet (27). We then used Lexicomp data, accessed through UpToDate (28), to determine a range of prices in the US health care system as a second way to provide confirmation that the price we were using was appropriate (29). These prices ranged from $6.54–$8.38 per tablet or capsule. We used $8.38 as the upper bound of our univariate sensitivity analysis. Australian colchicine costs are standardized through the single payer system and do not vary ($0.52 [Australian] per tablet).

We performed a univariate sensitivity analysis in which the model outcome was reexamined after changing 1 input and assessing the effect on the outcome of the model. Where available, sensitivity analysis ranges (for both univariate and probabilistic sensitivity analysis [PSA]) were based on published literature, and where the data were not available, we made assumptions, ensuring that all parameters were varied across a reasonable and broad range.

We performed a PSA, which allows multiple parameters to be varied simultaneously. By convention, probabilities were fitted to a β distribution, and the mean to a normal distribution, with costs and utility reductions to a γ distribution (30). Monte Carlo simulation was then employed using random number generation as the seed. This was performed 1,000 times using different values from each distribution. The individual outcomes from these 1,000 trials were then examined. We used a willingness-to-pay (WTP) threshold of $100,000 per QALY gained in the US model in base case analysis but also assessed thresholds of $50,000 and $200,000, as suggested by Neumann et al (31). In the Australian model, we used a WTP threshold of $50,000 [Australian] per QALY gained, as per our previous publication (12). The time horizon for the model was 6 months, given that this was the duration of the trial by Borstad et al (4). While the dose size of colchicine in the US is 0.6 mg, in Australia it is 0.5 mg; therefore, the models differed in this respect. The 2 different doses are generally used interchangeably in clinical practice.

RESULTS

The twice a day colchicine prophylaxis arm resulted in a cost of $1,276 and 0.49 QALYs, while the placebo arm had a cost of $516 and QALY of 0.47, with an incremental cost-effectiveness ratio (ICER) of $34,004 per QALY gained. Cost savings in Australia were more substantial, with a cost of $208 (Australian) in the colchicine arm versus $415 (Australian) in the placebo arm due to lower colchicine cost.

The univariate sensitivity analysis demonstrated that the results were robust to changes in input parameters (Table 1). The model was most sensitive to changes in the cost of colchicine, changes in quality of life associated with a gout flare, and rates of gout flares on placebo (with ICER ranges of –$10,308 to $84,923, $22,405–$70,501, and $21,365–$58,588, respectively).

In PSA in the US model, colchicine treatment was cost-effective compared to placebo in 70% of iterations at a WTP threshold of $50,000 per QALY gained, in 93% of iterations at a threshold of $100,000, and in 98% at a threshold of $200,000 per QALY gained (see Figure 2A for the cost-effectiveness acceptability curve of the US model). The incremental cost-effectiveness scatterplot of the US model is shown in Figure 2B, with a WTP threshold of $100,000 per QALY gained shown. In the Australian model, the colchicine treatment arm was cost-effective in 99.9% of iterations at a WTP threshold of $50,000 [Australian].

DISCUSSION

This study demonstrates that colchicine prophylaxis when commencing allopurinol is cost-effective in both the Australian and US health care settings. The result of this study may also be generalizable to other forms of ULT when patients have similar rates of gout flare and pharmaceutical purchase costs. The analysis also demonstrates that in the US health care system, where colchicine is more expensive than other Western health systems, this approach is still likely to be cost-effective, even if colchicine costs up to $8.38 per tablet. Therefore, the implication of this study is that payers should be willing to fund treatment with colchicine when initiating allopurinol, as it is cost-effective.

The strengths of this study include almost all model parameters coming from previously published gout research or recognized sources for costs associated with the purchase of pharmaceuticals or the cost of health care provision. The sensitivity testing, both univariate and probabilistic, showed that the model is robust to significant changes in the input parameters, including large changes in the cost of colchicine in the US. This research has some limitations. We were unable to find specific disutilities for colchicine-associated diarrhea. We feel that the approach we took was a good estimate of the likely disutility that would result from colchicine-induced diarrhea, but specific studies in colchicine-induced diarrhea would be required to confirm this. Varying the QALY cost of colchicine-induced diarrhea up to 200% of the base case estimate made no difference to the overall outcome of the model, so we are confident that this estimate would not impact substantially on the robustness of the result. Additionally, many ranges of the sensitivity analysis were based on assumptions. However, even our broadly chosen ranges had little effect on the overall results of the models, with colchicine treatment remaining cost-effective despite large changes in the input parameters. There were also only 43 participants in the trial by Borstad et al (4), 21 in the colchicine group, and 22 in the placebo group. This is a small number of participants, but the result of the trial was not equivocal, with 77% of the placebo group experiencing flares, and 33% of the colchicine group experiencing flares. Such a large difference in outcome supports the finding of our cost–utility study because treatment of gout flares incurs substantial cost and disutility. The financial cost and QALY cost of colchicine-induced diarrhea in comparison with this, based on our reported model, is small. While this was a third-party payer cost-effectiveness study, we also acknowledge that there are additional out-of-pocket costs incurred by patients in taking colchicine. Reduced adherence to colchicine would likely lead to an increased probability of flare; however, due to limited information about adherence to colchicine in this context, we have not included this in the model.

Determining the generalizability of the research depends on whether the approach to up-titrating allopurinol is similar to the approach currently employed in usual clinical practice. Borstad et al used a similar approach to what is currently advocated by multiple guidelines (4). Allopurinol was started at 100 mg daily and increased in 100-mg increments. In those with creatinine clearance of 20–50 ml, allopurinol was escalated in 50-mg increments. Serum urate levels were measured at baseline and every 2–3 weeks. This would suggest that up-titration was also performed every 2–3 weeks, but this was not specified in the trial by Borstad et al (4). The American College of Rheumatology 2012

Table 1. Univariate sensitivity analysis*

Parameter	Base case analysis, $	Base case analysis, mean no. of flares	Base case analysis, probability	Base case analysis, QALYs	Sensitivity analysis, $	Sensitivity analysis, % range	Sensitivity analysis, probability range	Results (US), $	Results (Australia)
Costs in US dollars									
Cost of colchicine	473†				75–838†			–10,308 to 84,923‡	Colchicine arm dominates
Hospital admission with major complications	6,988					75–125		32,792–35,216	Colchicine arm dominates
Hospital admission without major complications	4,274					75–125		33,263–34,475	Colchicine arm dominates
Family physician visit	109				81–135			31,820–37,821	Colchicine arm dominates
Pathology									
Blood count	8.63					50–150		33,823–34,185	Colchicine arm dominates
Electrolytes and uric acid	12.81					50–150		33,735–34,273	Colchicine arm dominates
Renal function	9.65					50–150		33,801–34,207	Colchicine arm dominates
Liver function tests	9.08					50–150		33,813–34,195	Colchicine arm dominates
Event probabilities and clinical outcomes									
Mean flares receiving colchicine (in 6 months)		0.57				75–125		30,964–37,446	Colchicine arm dominates
Mean flares receiving placebo (in 6 months)		2.96				75–125		21,365–58,588	Colchicine arm dominates
Need for dose reduction of colchicine to daily			0.62				0.41–0.83	21,387–41,445	Colchicine arm dominates
Probability of high vs. low hospital cost			0.5				0–1	32,121–35,887	Colchicine arm dominates
Diarrhea when receiving colchicine			0.38				0.17–0.60	26,387–41,444	Colchicine arm dominates
Diarrhea when receiving placebo			0.045				0–0.135	33,685–34,166	Colchicine arm dominates
Home treatment of gout flare			0.2				0–0.2	31,248–34,004	Colchicine arm dominates
Hospital treatment of gout flare			0.013			50–150		30,187–37,821	Colchicine arm dominates
Treatment of gout flare by primary care physician			0.787			§			–
Laboratory assessment by primary care physician			0.5				0–1	32,317–34,745	Colchicine arm dominates
Quality of life									
Reduction in quality of life for gouty flare				0.0097		50–150		22,405–70,501	Colchicine arm dominates
Reduction in quality of life with diarrheal episode				0.0038		50–150		33,413–34,616	Colchicine arm dominates

* All dollar amounts are in US dollars. QALYs = quality-adjusted life-years.
† For 100 tablets.
‡ The colchicine arm dominates below $167 for 100 tablets.
§ Dependent on proportions treated at home/hospital.

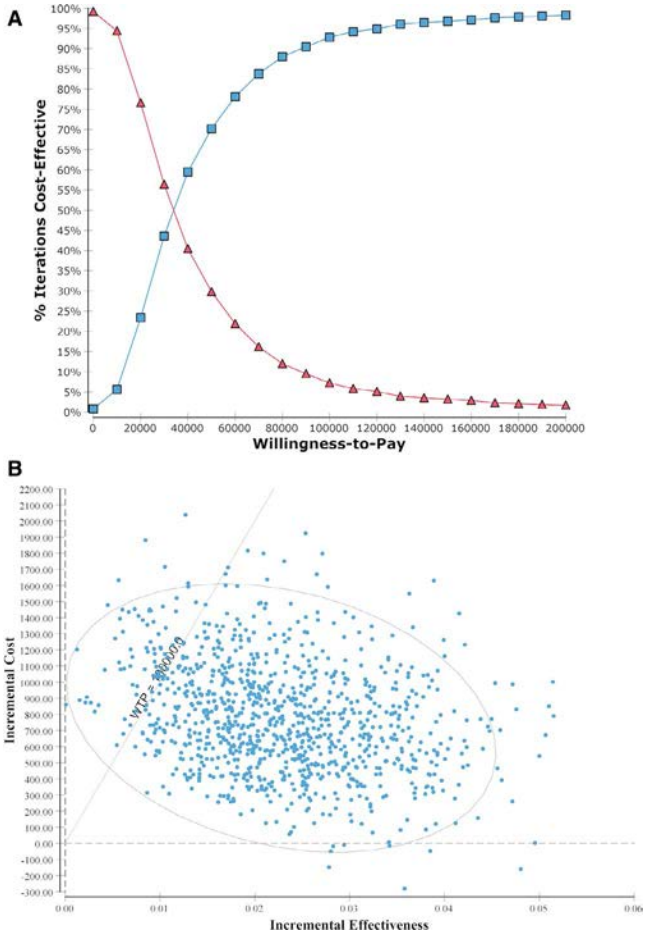


Figure 2. **A**, Cost-efficacy acceptability curve for colchicine (blue) and placebo (red). **B**, Incremental cost-effectiveness scatterplot (each blue dot represents the results of an individual iteration of the simulation; the oval represents the 95% confidence interval). WTP = willingness to pay.

guidelines for the management of gout recommend up-titration of allopurinol every 2–5 weeks (3). Therefore, this trial applied an up-titration regimen in line with what is currently recommended. The rates of flares and frequency of adverse events are consistent with those in other published research (22,32). In conclusion, the present model supports the positive clinical outcome of the study by Borstad et al (4) and suggests that not only is colchicine clinically effective in reducing gout flare while commencing allopurinol but it is also cost-effective.

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. Robinson 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. Robinson, Dalbeth, Donovan.
Acquisition of data. Robinson, Dalbeth, Donovan.
Analysis and interpretation of data. Robinson, Dalbeth, Donovan.

REFERENCES

1. Chen-Xu M, Yokose C, Rai SK, Pillinger MH, Choi HK. Contemporary prevalence of gout and hyperuricemia in the United States and decadal trends: the National Health and Nutrition Examination Survey, 2007–2016. Arthritis Rheumatol 2019;71:991–9.
2. Robinson PC, Taylor WJ, Dalbeth N. An Observational study of gout prevalence and quality of care in a national Australian general practice population. J Rheumatol 2015;42:1702–7.
3. Khanna D, Fitzgerald JD, Khanna PP, Bae S, Singh MK, Neogi T, et al. 2012 American College of Rheumatology guidelines for management of gout. Part 1. Systematic nonpharmacologic and pharmacologic therapeutic approaches to hyperuricemia. Arthritis Care Res (Hoboken) 2012;64:1431–46.
4. Borstad GC, Bryant LR, Abel MP, Scroggie DA, Harris MD, Alloway JA. Colchicine for prophylaxis of acute flares when initiating allopurinol for chronic gouty arthritis. J Rheumatol 2004;31:2429–32.
5. Wortmann RL, Macdonald PA, Hunt B, Jackson RL. Effect of prophylaxis on gout flares after the initiation of urate-lowering therapy: analysis of data from three phase III trials. Clin Ther 2010;32:2386–97.
6. Khanna D, Khanna PP, Fitzgerald JD, Singh MK, Bae S, Neogi T, et al. 2012 American College of Rheumatology guidelines for management of gout. Part 2. Therapy and antiinflammatory prophylaxis of acute gouty arthritis. Arthritis Care Res (Hoboken) 2012;64:1447–61.
7. Hui M, Carr A, Cameron S, Davenport G, Doherty M, Forrester H, et al. The British Society for Rheumatology guideline for the management of gout. Rheumatology (Oxford) 2017;56:e1–20.
8. Richette P, Doherty M, Pascual E, Barskova V, Becce F, Castaneda-Sanabria J, et al. 2016 updated EULAR evidence-based recommendations for the management of gout. Ann Rheum Dis 2017;76:29–42.
9. Murphy SM, Puwanant A, Griggs RC, Consortium for Clinical Investigations of Neurological Channelopathies (CINCH) and Inherited Neuropathies Consortium (INC) Consortia of the Rare Disease Clinical Research Network. Unintended effects of orphan product designation for rare neurological diseases. Ann Neurol 2012;72:481–90.
10. Kesselheim AS, Franklin JM, Kim SC, Seeger JD, Solomon DH. Reductions in use of colchicine after FDA enforcement of market exclusivity in a commercially insured population. J Gen Intern Med 2015;30:1633–8.
11. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. BMJ 2013;346:f1049.
12. Robinson PC, Dalbeth N, Donovan P. The cost-effectiveness of biannual serum urate (SU) monitoring after reaching target in gout: a health economic analysis comparing SU monitoring. J Rheumatol 2018;45:697–704.
13. Robinson PC, Merriman TR, Herbison P, Highton J. Hospital admissions associated with gout and their comorbidities in New Zealand and England 1999–2009. Rheumatology (Oxford) 2013;52:118–26.
14. Winnard D, Wright C, Taylor WJ, Jackson G, Te Karu L, Gow PJ, et al. National prevalence of gout derived from administrative health data in Aotearoa New Zealand. Rheumatology (Oxford) 2012;51:901–9.
15. Jackson G, Wright C, Thornley S, Taylor WJ, Te Karu L, Gow PJ, et al. Potential unmet need for gout diagnosis and treatment: capture-recapture analysis of a national administrative dataset. Rheumatology (Oxford) 2012;51:1820–4.
16. Halpern R, Fuldeore MJ, Mody RR, Patel PA, Mikuls TR. The effect of serum urate on gout flares and their associated costs: an administrative claims analysis. J Clin Rheumatol 2009;15:3–7.
17. Kuo CF, Grainge MJ, Zhang W, Doherty M. Global epidemiology of gout: prevalence, incidence and risk factors. Nat Rev Rheumatol 2015;11:649–62.

18. US Centers for Medicare and Medicaid Services. Medicare 2019. URL: https://www.medicare.gov/index.
19. Independent Hospital Pricing Authority. National hospital cost data collection, AR-DRG cost weight tables V8.0x, round 21 (financial year 2016–17). June 2019. URL: https://www.ihpa.gov.au/publications/national-hospital-cost-data-collection-ar-drg-cost-weight-tables-v80x-round-21.
20. Australian Government Department of Health. The pharmaceutical benefits scheme. URL: http://www.pbs.gov.au/pbs/home.
21. Australian Government Department of Health. Medicare benefits schedule. URL: http://www.mbsonline.gov.au/internet/mbsonline/publishing.nsf/Content/Home.
22. Stewart S, Yang KC, Atkins K, Dalbeth N, Robinson PC. Adverse events during oral colchicine use: a systematic review and meta-analysis of randomised controlled trials. Arthritis Res Ther 2020;22:28.
23. Perez-Ruiz F, Diaz-Torne C, Carcedo D. Cost-effectiveness analysis of febuxostat in patients with gout in Spain. J Med Econ 2016;19:604–10.
24. Rochanathimoke O, Riewpaiboon A, Postma MJ, Thinyounyong W, Thavorncharoensap M. Health related quality of life impact from rotavirus diarrhea on children and their family caregivers in Thailand. Expert Rev Pharmacoecon Outcomes Res 2018;18:215–22.
25. US Department of Veterans Affairs. Determining the cost of pharmaceuticals for a cost-effectiveness analysis. URL: https://www.herc.research.va.gov/include/page.asp?id=pharmaceutical-costs.
26. Congress of the United States Congressional Budget Office. Prices for brand-name drugs under selected federal programs. June 2005. URL: https://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/64xx/doc6481/06-16-prescriptdrug.pdf.
27. US Department of Veterans Affairs. National Acquisition Center (CCST). URL: https://www.vendorportal.ecms.va.gov/nac/Pharma/List.
28. UpToDate. Colchicine. URL: https://www.uptodate.com/contents/colchicine-drug-information.
29. Lexicomp. Lexicomp drug database. URL: https://online.lexi.com/lco/action/login.
30. Briggs AE. Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2006.
31. Neumann PJ, Cohen JT, Weinstein MC. Updating cost-effectiveness: the curious resilience of the $50,000-per-QALY threshold. N Engl J Med 2014;371:796–7.
32. Teoh N, Gamble GD, Horne A, Taylor WJ, Palmano K, Dalbeth N. The challenges of gout flare reporting: mapping flares during a randomized controlled trial. BMC Rheumatol 2019;3:27.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 73, No. 10, October 2021, pp 1544–1551
DOI 10.1002/acr.24348

BRIEF REPORT

Epidemiology, Time Trends, and Outcomes of Serious Infections in Patients With Vasculitis: A Nineteen-Year National Study

Jasvinder A. Singh[1] and John D. Cleveland[2]

Objective. To examine the epidemiology, time trends, and outcomes and types of serious infections in people with vasculitis in the US.

Methods. We identified people with vasculitis who were hospitalized with a primary diagnosis of pneumonia, sepsis/bacteremia, urinary tract infection (UTI), skin and soft tissue infections, or opportunistic infections in the 1998–2016 US National Inpatient Sample. We used adjusted logistic regression to examine the predictors of a hospital stay >3 days, total hospital charges greater than the median, discharge to a non-home setting, and in-hospital mortality.

Results. We noted 111,345 serious infections in patients with vasculitis (14% of all vasculitis hospitalizations). Among the patients, the mean age was 67.3 years, the Deyo-Charlson comorbidity index score was ≥2 in 54%, 37% were male, and 67% were White. The serious infection hospitalization rate per 100,000 US National Inpatient Sample claims in 1998–2000 versus 2015–2016 (and rates of increase) in patients with vasculitis was as follows: overall, 12.14 versus 25.15 (2.1-fold); opportunistic infections, 0.78 versus 0.83 (1.1-fold); skin and soft tissue infections, 1.38 versus 2.52 (1.8-fold); UTI, 0.35 versus 1.48 (4.2-fold); pneumonia, 7.10 versus 6.23 (0.9-fold); and sepsis, 2.53 versus 14.10 (5.6-fold). Pneumonia was the most common serious infection in 1998–2000 (58%) versus sepsis in 2015–2016 (56%). Sepsis, older age, Deyo-Charlson comorbidity index score of ≥2, urban hospital, or medium/large hospital (by number of beds) were associated with higher health care utilization and in-hospital mortality rates; Northeast region and Medicare and Medicaid payer type were associated with higher rates of health care utilization.

Conclusion. Serious infection hospitalization rates are increasing in patients with vasculitis except among those with pneumonia. Sepsis was the most common serious infection in 2015–2016. Several patient and hospital factors are associated with health care utilization and mortality in serious infection hospitalization in vasculitis.

INTRODUCTION

Vasculitis is a systemic autoimmune disease that can rapidly progress to a life-threatening condition. It is associated with high morbidity and mortality (1). Treatment includes immunosuppressive medications, high-dose glucocorticoids, and biologics (2). Not surprisingly, infectious complications are a major morbidity during the treatment of vasculitis and are a leading cause of death. A 2-center UK-based study of 100 people with antineutrophil cytoplasmic antibody–associated vasculitis (AAV), a common subtype of vasculitis, found that 28% of these patients were hospitalized for infection in 2004–2011 (3). The infection hospitalization rate was 2.23 events per person-year in the presence of severe lymphopenia compared to 0.19 events per person-year during periods with no lymphopenia (3). In a 1998–2010 Swedish study of 186 patients with AAV, the risk of severe infection hospitalization was 4.5-fold higher versus age- and sex-matched individuals without vasculitis (4). The rate of hospitalized infection among patients was 116.2 per 1,000 person-years (4). These studies of hospitalized infections in vasculitis provide important

Supported by the University of Alabama at Birmingham.

[1]Jasvinder A. Singh, MBBS, MPH: Birmingham VA Medical Center, Birmingham, Alabama and University of Alabama at Birmingham; [2]John D. Cleveland, MS: University of Alabama at Birmingham.

Dr. Singh has received consulting fees and/or honoraria from Crealta/Horizon, Medisys, Fidia, UBM, Trio Health, Medscape, WebMD, Clinical Care Options, ClearView Health Care Partners, Putnam Associates, Spherix, and Practice Point Communications (less than $10,000 each) and owns stock or stock options in Amarin Pharmaceuticals and Viking Therapeutics. No other disclosures relevant to this article were reported.

Address correspondence to Jasvinder A. Singh, MBBS, MPH, University of Alabama at Birmingham, Faculty Office Tower 805B, 510 20th Street S, Birmingham, AL 35294-0022. Email: Jasvinder.md@gmail.com.

Submitted for publication March 20, 2020; accepted in revised form June 2, 2020.

SIGNIFICANCE & INNOVATIONS

- The overall rate of serious infection hospitalizations has increased in people with vasculitis over the 19-year period from 1998 to 2016.
- Serious infection hospitalizations constituted 12.3% of hospitalizations in 1998–2000 versus 21% of hospitalizations in 2015–2016 in people hospitalized with vasculitis as a secondary diagnosis.
- In 2015–2016, sepsis was the most common serious infection in hospitalized vasculitis patients, compared to 1998–2000, where pneumonia was the most common serious infection.
- Identification of several patient and hospital factors associated with health care utilization and mortality related to serious infection hospitalization outcomes in people with vasculitis can help recognize high-risk groups and provide the data for the development of effective interventions that can improve these outcomes.

insights. Major limitations of these studies were that they consisted of small sample sizes, did not assess health care utilization, used data that are almost 1 decade old, and did not include all types of vasculitides.

Our objective was to assess the epidemiology, time trends, and outcomes of serious infections and their types in patients with vasculitis. In order to study the common types of serious infections in people with vasculitis, we used the US National Inpatient Sample (NIS) to examine the epidemiology and time trends and to assess health care utilization outcomes and in-hospital mortality and evaluate the factors associated with higher health care utilization outcomes and in-hospital mortality rates.

PATIENTS AND METHODS

The NIS is a deidentified, national, all-payer, inpatient health care database that represents a 20% stratified sample of discharges in the US and is a component of the Healthcare Cost and Utilization Project (5). Our analysis used the 1998–2016 NIS data. These data are available from the Agency for Healthcare Research and Quality Healthcare Cost and Utilization Project (HCUPDistributor@ahrq.gov) and can be obtained after completing an on-line Data Use Agreement training session and signing a Data Use Agreement. Individuals are not allowed to distribute these data. The Institutional Review Board at the University of Alabama at Birmingham (X120207004) approved this study.

Serious infections. We identified the 5 types of serious infections by the presence of the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) code in the primary diagnosis position for hospitalization, i.e., the main reason. These types included opportunistic infections (OIs; 010.xx–018.xx, 031.xx, 078.5, 075.xx, 053.xx, 112.4, 112.5, 112.81, 112.83, 130.xx, 136.3, 117.5, 027.0, 039.xx, 117.3, 114.xx, 115.xx, or 116.0); skin and soft tissue infections (SSTI; 040.0, 569.61, 681.xx, 682.xx, 785.4, 728.86, or 035.xx); urinary tract infection (UTI; 590.xx); pneumonia (003.22, 481.0, 513.0, 480.xx, 482.xx, 483.xx, 485.xx, or 486.xx); and sepsis/bacteremia (sepsis; 038.xx or 790.7), as in previous studies (6,7). These infection diagnostic codes had positive predictive values of 70–100% (8–10) and negative predictive value of 84–100% (10) in people with rheumatoid arthritis. We also used the ICD-10-CM codes for infections for the 2015–2016 data because the coding system changed from the ICD-9-CM to ICD-10-CM in 2015 in the US (see Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24348/abstract). Composite serious infection was defined as the presence of any of the 5 serious infections, i.e., was representative of the overall rate.

We identified people with vasculitis in the secondary diagnoses position (any position other than primary) based on the presence of the ICD-9-CM code 446.xx or ICD-10-CM codes M30 or M31. This code-based approach showed a sensitivity of 94% and a specificity of 95% (11).

Statistical analysis. We compared the demographic characteristics of serious infection in people with vasculitis versus those without vasculitis; these characteristics were also compared to all hospitalizations of people with vasculitis, both without any statistical comparisons to avoid multiple testing (a priori decision). Frequencies and rates of the 5 serious infections were assessed per 100,000 NIS claims. We specified all analyses a priori, including time trends in each infection and multivariable-adjusted regression analyses of 3 health care utilization and in-hospital mortality outcomes. Serious infection rates were analyzed for time trends using the Cochran-Armitage test, weighted by the number of hospitalizations in each year.

Unadjusted health care utilization outcomes and in-hospital mortality were assessed for each type of serious infection in people with vasculitis. Multivariable-adjusted logistic regression models were performed for a length of hospital stay of >3 days (median), the total charges above the median (based on the median for the year), discharge status (home versus non-home health care facility [i.e., inpatient, nursing, or rehabilitation]), and in-hospital mortality. Covariates/confounders were adjusted in the multivariable-adjusted regression analyses and included the following: demographic characteristics (age, sex, race/ethnicity, income quartiles per the NIS); comorbidity using the validated Deyo-Charlson comorbidity index based on the presence of ICD-9-CM codes at index admission (including 17 common medical comorbidities such as myocardial infarction, congestive heart failure, cerebrovascular disease, dementia, renal disease, liver disease, chronic pulmonary disease, diabetes mellitus, etc.) (12); insurance type (13,14); and hospital characteristics (region, location/teaching status, number of beds). We categorized age into age groups as <50 years, 50 to <65 years, 65 to <80 years,

Table 1. Characteristics of patients with each serious infection and also with vasculitis*

	OI (n = 5,518)	SSTI (n = 13,908)	UTI (n = 3,146)	Pneumonia (n = 46,940)	Sepsis (n = 41,833)	Composite infection (n = 111,345)
Age, mean ± SE; median	62.3 ± 0.59; 66.2	65.4 ± 0.39; 69.2	63.6 ± 0.91; 70.2	68.6 ± 0.21; 72.8	67.4 ± 0.21; 71.1	67.3 ± 0.14; 71.3
Age category, years						
<50	1,206 (21.9)	2,868 (20.7)	755 (24.0)	6,728 (14.4)	6,575 (15.7)	18,131 (16.3)
50–64	1,254 (22.8)	2,935 (21.1)	424 (13.5)	8,081 (17.3)	8,365 (20.0)	21,059 (19.0)
65–79	2,160 (39.2)	3,984 (28.7)	1,035 (33.0)	16,678 (35.7)	14,579 (34.9)	38,435 (34.6)
≥80	887 (16.1)	4,099 (29.5)	925 (29.5)	15,180 (32.5)	12,245 (29.3)	33,336 (30.0)
Sex						
Male	2,315 (42.1)	5,057 (36.4)	535 (17.0)	17,173 (36.8)	16,415 (39.3)	41,495 (37.4)
Female	3,187 (57.9)	8,826 (63.6)	2,603 (83.0)	29,490 (63.2)	25,348 (60.7)	69,455 (62.6)
Race						
White	3,284 (59.6)	9,687 (69.8)	2,086 (66.5)	30,950 (66.3)	28,236 (67.6)	74,242 (66.9)
Black	370 (6.7)	895 (6.4)	219 (7.0)	2,749 (5.9)	3,323 (8.0)	7,556 (6.8)
Hispanic	533 (9.7)	860 (6.2)	251 (8.0)	3,031 (6.5)	3,059 (7.3)	7,734 (7.0)
Other/missing	1,321 (24.0)	2,444 (17.6)	582 (18.6)	9,941 (21.3)	7,145 (17.1)	21,433 (19.3)
Deyo-Charlson comorbidity index score						
0	1,552 (28.2)	3,549 (25.6)	824 (26.3)	9,742 (20.9)	7,371 (17.6)	23,037 (20.8)
1	1,407 (25.5)	3,927 (28.3)	848 (27.0)	12,773 (27.4)	8,625 (20.7)	27,580 (24.9)
≥2	2,549 (46.3)	6,410 (46.2)	1,466 (46.7)	24,156 (51.8)	25,768 (61.7)	60,349 (54.4)
Income category, percentile						
0–25th	1,032 (19.4)	2,701 (19.9)	592 (19.2)	8,664 (18.9)	8,990 (21.9)	21,980 (20.2)
25–50th	1,231 (23.1)	3,634 (26.8)	891 (28.9)	12,570 (27.5)	10,484 (25.6)	28,810 (26.5)
50–5th	1,553 (29.1)	3,453 (25.5)	722 (23.5)	11,891 (26.0)	10,251 (25.0)	27,872 (25.6)
75–100th	1,514 (28.4)	3,768 (27.8)	874 (28.4)	12,641 (27.6)	11,247 (27.4)	30,044 (27.6)
Insurance						
Private	1,635 (29.7)	2,708 (19.5)	632 (20.1)	8,405 (18.0)	7,624 (18.3)	21,004 (19.0)
Medicare	3,159 (57.5)	9,058 (65.4)	2,076 (66.2)	34,027 (73.0)	29,258 (70.1)	77,578 (70.0)
Medicaid	409 (7.4)	1,280 (9.2)	276 (8.8)	2,691 (5.8)	3,298 (7.9)	7,954 (7.2)
Other	132 (2.4)	390 (2.8)	77 (2.5)	698 (1.5)	680 (1.6)	1,978 (1.8)
Self	163 (3.0)	417 (3.0)	77 (2.4)	779 (1.7)	852 (2.0)	2,288 (2.1)
Hospital region						
Northeast	1,070 (19.4)	3,103 (22.3)	641 (20.4)	9,517 (20.3)	7,322 (17.5)	21,652 (19.4)
Midwest	1,223 (22.2)	3,061 (22.0)	737 (23.4)	11,713 (25.0)	9,776 (23.4)	26,510 (23.8)
South	1,880 (34.1)	4,904 (35.3)	1,090 (34.6)	16,506 (35.2)	14,520 (34.7)	38,899 (34.9)
West	1,345 (24.4)	2,841 (20.4)	678 (21.6)	9,204 (19.6)	10,216 (24.4)	24,284 (21.8)
Hospital location/ teaching						
Rural	432 (7.9)	1,446 (10.7)	372 (12.2)	7,225 (16.0)	3,918 (9.5)	13,394 (12.4)
Urban non-teaching	1,679 (30.8)	5,251 (39.0)	1,154 (37.8)	18,406 (40.8)	15,068 (36.7)	41,559 (38.4)
Urban teaching	3,335 (61.2)	6,765 (50.3)	1,526 (50.0)	19,434 (43.1)	22,109 (53.8)	53,169 (49.2)
Hospital size, by no. of beds						
Small	443 (8.0)	1,945 (14.0)	549 (17.5)	7,068 (15.1)	5,505 (13.2)	15,509 (14.0)
Medium	1,230 (22.3)	3,769 (27.1)	892 (28.4)	12,443 (26.6)	11,525 (27.7)	29,860 (26.9)
Large	3,836 (69.6)	8,178 (58.9)	1,700 (54.1)	27,311 (58.3)	24,613 (59.1)	65,640 (59.1)
Discharge status						
Rehabilitation, inpatient, or nursing facility	1,294 (26.6)	3,036 (22.3)	602 (19.5)	11,196 (25.6)	14,242 (40.8)	30,370 (30.3)
Home	3,571 (73.4)	10,561 (77.7)	2,488 (80.5)	32,470 (74.4)	20,704 (59.2)	69,794 (69.7)
Length of stay, days						
≤3	913 (16.6)	4,877 (35.1)	1,598 (50.9)	14,002 (30.0)	9,555 (22.9)	30,944 (27.9)
>3	4,594 (83.4)	9,009 (64.9)	1,540 (49.1)	32,669 (70.0)	32,209 (77.1)	80,022 (72.1)
Died during hospitalization	628 (11.4)	154 (1.1)	15 (0.5)	2,774 (5.9)	6,562 (15.7)	10,133 (9.1)

(Continued)

Table 1. *(Cont'd)*

	OI (n = 5,518)	SSTI (n = 13,908)	UTI (n = 3,146)	Pneumonia (n = 46,940)	Sepsis (n = 41,833)	Composite infection (n = 111,345)
Length of stay, mean ± SE days; median days	12.7 ± 0.42; 7.9	6.2 ± 0.15; 4.0	4.5 ± 0.16; 3.0	7.1 ± 0.08; 4.6	10.1 ± 0.13; 6.1	8.3 ± 0.07; 5.0
Total hospital charges (USD)						
≤median	1,059 (19.2)	6,041 (43.5)	1,648 (52.5)	15,685 (33.6)	9,724 (23.3)	34,158 (30.8)
>median	4,448 (80.8)	7,845 (56.5)	1,490 (47.5)	30,986 (66.4)	32,039 (76.7)	76,808 (69.2)
Total charge (USD), mean ± SE; median	88,667 ± 5,270; 36,908	28,775 ± 956; 16,681	26,023 ± 1,322; 15,071	38,710 ± 808; 19,040	89,064 ± 1,722; 40,642	58,211 ± 849; 24,699

* Values are the number (%) unless indicated otherwise. OI = opportunistic infection; SSTI = skin and soft tissue infection; USD = US dollar; UTI = urinary tract infection.

and ≥80 years and race/ethnicity as White, Black, Hispanic, and other/missing. Household income was based on the patient's zip code and was categorized from the lowest (poorest) to the highest quartile (wealthiest; 4 quartiles) based on thresholds for each quartile that varied by year as provided by the NIS (15) (e.g., the upper threshold for quartile 1 was $28,999 in 1998 and $39,999 in 2014). We categorized insurance statuses as Medicaid (provides coverage for the low-income and disabled Americans), Medicare (provides health care coverage for Americans ages 65 years or older), private insurance, and self/other (13), as in previous studies (14). Hospital region (Northeast, Midwest, South, West), location/teaching status (rural, urban non-teaching, and urban teaching), and size of hospital (by number of beds; small, medium, large) were other standard NIS variables included in the models.

The University of Alabama at Birmingham's Institutional Review Board (IRB) approved this study, and all investigations were conducted in conformity with ethical principles of research (UAB X120207004). The IRB waived the need for an informed consent for this database study.

RESULTS

Cohort characteristics and comparison of people without vasculitis hospitalized with a serious infection. We noted 111,345 serious infections in people with vasculitis (14% of all hospitalizations with vasculitis as secondary diagnosis; 12.3% of all hospitalizations in 1998–2000 versus 21% in 2015–2016) and 49,855,225 serious infections in people without vasculitis.

For people with vasculitis with serious infections, the mean age was 67.3 years (35% were ages <65 years), 63% were female, 67% were White, 55% had a Deyo-Charlson comorbidity index score of ≥2, and 70% were covered by a Medicare insurance payer (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24348/abstract). Compared to people with serious infections in the general population, people with vasculitis with serious infections were older and more likely to be White, female, covered by Medicare, or have higher Deyo-Charlson comorbidity index scores. Additionally, people with vasculitis with serious infections had higher health care utilization and in-hospital mortality rates (see Supplementary Table 1, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24348/abstract).

Type of serious infections, rates, and time trends in people with vasculitis. People with OIs were 3–5 years younger than people with other serious infections (Table 1). People with sepsis were more likely to have higher Deyo-Charlson comorbidity index scores compared to those with other serious infections. People with OIs had the highest mean/median length of hospital stay, at 12.7/7.9 days, followed by those with sepsis, at 10.1/6.1 days. OIs and sepsis led all serious infection types in median hospital charges. Our findings revealed that 41% of people with sepsis and vasculitis, and 19–27% of people with other serious infections were discharged to non-home settings (Table 1).

Among people with vasculitis, the overall rate (composite serious infection) and the rates of each serious infection per 100,000 NIS claims significantly increased over time, except for in pneumonia (Table 2 and Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24348/abstract). Raw frequencies and rates in the general NIS population are shown (see Supplementary Tables 2 and 3, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24348/abstract).

The hospitalization rates per 100,000 NIS claims in 1998–2000 versus 2015–2016 (and rates of increase) in people with vasculitis were as follows: OIs, 0.78 versus 0.83 (1.1-fold); SSTI, 1.38 versus 2.52 (1.8-fold); UTI, 0.35 versus 1.48 (4.2-fold); pneumonia, 7.10 versus 6.23 (0.9-fold); and sepsis, 2.53 versus 14.10 (5.6-fold) (Table 2). When vasculitis hospitalization claims were used as the denominator, all trends were significant, but only sepsis, UTI, and SSTI showed the positive trends (Table 3). National estimates for the total number of hospitalization claims

Table 2. Rate of hospitalizations for serious infection in people with vasculitis over time per 100,000 NIS claims*

Time periods	OI	SSTI	UTI	Pneumonia	Sepsis	Composite infection†	Total NIS claims
1998–2000	0.78	1.38	0.35	7.10	2.53	12.14	103,665,051
2001–2002	0.75	1.64	0.32	6.75	2.44	11.89	72,617,381
2003–2004	0.72	1.87	0.28	6.73	2.97	12.57	74,571,583
2005–2006	0.79	1.90	0.36	7.04	3.89	13.98	75,919,595
2007–2008	0.73	2.11	0.34	6.62	4.53	14.34	76,366,797
2009–2010	0.83	2.09	0.31	6.32	5.97	15.52	75,086,597
2011–2012	0.90	2.51	0.33	7.53	8.54	19.81	73,447,261
2013–2014	0.85	2.30	0.39	6.38	11.24	21.15	70,956,610
2015–2016‡	0.83	2.52	1.48	6.23	14.10	25.15	71,445,363
P§	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	
Change from 1998–2000 to 2015–2016¶	1.1	1.8	4.2	0.9	5.6	2.1	

* The first time period, 1998–2000, is a 3-year duration, and the subsequent time periods are a 2-year duration. NIS = US National Inpatient Sample (see Table 1 for other definitions).
† Composite infection indicates any of the 5 hospitalized serious infections as the primary diagnosis.
‡ 2015–2016 was the first year that International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) codes were used; therefore, some rates from this period may reflect transition of the coding system rather than only time trends.
§ *P* by Cochran-Armitage 2-sided test for trend.
¶ Change was calculated by dividing the number from 2015–2016 by the number from 1998–2000.

have remained stable. Vasculitis hospitalization claims have been stable (with slight increase), while claims for OIs and pneumonia have decreased over time.

Outcomes of serious infections in people with vasculitis, rates, and time trends. Overall, both OIs and sepsis were associated with the highest rates of health care utilization and in-hospital mortality in both the first (1998–2000) and the last (2015–2016) periods, with SSTI, UTI, and pneumonia with lower rates (see Supplementary Table 4, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24348/abstract). Over the observation period, the largest increase in median hospital charges posted were seen in OIs and sepsis, with a 250% increase observed in each. Over time, health care utilization and in-hospital mortality rates decreased for each serious infection (with minor exceptions) except for the increasing hospital charges noted for each serious infection. The median hospital charges increased 2.9–3.5-fold from the 1998–2000 range to the 2015–2016 timeframe (see Supplementary Table 4).

Table 3. Rate of hospitalizations for serious infection in people with vasculitis over time per 100,000 vasculitis claims*

Time periods	OI	SSTI	UTI	Pneumonia	Sepsis	Composite infection†	Total vasculitis claims
1998–2000	788	1,393	354	7,173	2,552	12,260	102,674
2001–2002	754	1,654	324	6,820	2,467	12,022	71,837
2003–2004	692	1,792	266	6,457	2,853	12,061	77,732
2005–2006	730	1,757	331	6,500	3,592	12,910	82,183
2007–2008	701	2,016	325	6,322	4,327	13,690	80,006
2009–2010	736	1,856	280	5,618	5,311	13,800	84,420
2011–2012	725	2,031	269	6,098	6,915	16,037	90,739
2013–2014	692	1,880	317	5,224	9,196	17,309	86,720
2015–2016‡	691	2,109	1,236	5,213	11,802	21,051	85,365
P§	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	
Change from 1998–2000 to 2015–2016¶	0.9	1.5	3.5	0.7	4.6	1.7	

* The first time period, 1998–2000, is a 3-year duration, and the subsequent time periods are a 2-year duration. NIS = US National Inpatient Sample (see Table 1 for other definitions).
† Composite infection indicates any of the 5 hospitalized serious infections as the primary diagnosis.
‡ 2015–2016 was the first year that International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) codes were used; therefore, some rates from this period may reflect transition of the coding system rather than only time trends.
§ *P* by Cochran-Armitage 2-sided test for trend.
¶ Change was calculated by dividing the number from 2015–2016 by the number from 1998–2000.

Table 4. Multivariable-adjusted correlates of health care utilization and in-hospital mortality in vasculitis patients hospitalized with serious infection*

	Hospital charges over the median	Discharge to care facility†	Length of hospital stay >3 days	In-hospital mortality
Age category, years				
<50	Ref.	Ref.	Ref.	Ref.
50–64	1.11 (1.00– 1.23)	1.47 (1.29– 1.68)‡	1.24 (1.12–1.38)‡	1.53 (1.27–1.84)‡
65–79	0.98 (0.88–1.10)	2.11 (1.85–2.42)‡	1.28 (1.15–1.43)‡	2.06 (1.70–2.50)‡
≥80	0.82 (0.73–0.93)‡	4.26 (3.71–4.88)‡	1.30 (1.16–1.46)‡	2.27 (1.85–2.77)‡
Sex				
Male	Ref.	Ref.	Ref.	Ref.
Female	0.92 (0.86–0.98)‡	1.09 (1.01–1.16)‡	1.04 (0.97–1.11)	0.88 (0.80–0.98)‡
Race/ethnicity				
White	Ref.	Ref.	Ref.	Ref.
Black	1.12 (0.98–1.27)	1.02 (0.88–1.17)	1.35 (1.18–1.54)‡	0.95 (0.78–1.16)
Hispanic	1.52 (1.32–1.74)‡	0.89 (0.77–1.03)	1.32 (1.16–1.51)‡	1.20 (0.99–1.45)
Other/missing	1.22 (1.12–1.32)‡	1.05 (0.96–1.15)	1.15 (1.06–1.24)‡	1.24 (1.09–1.40)‡
Deyo-Charlson comorbidity index score				
0	Ref.	Ref.	Ref.	Ref.
1	1.15 (1.05–1.25)‡	1.06 (0.96–1.18)	1.12 (1.03–1.23)‡	1.06 (0.91–1.24)
≥2	1.36 (1.25–1.47)‡	1.43 (1.31–1.56)‡	1.39 (1.28–1.50)‡	1.25 (1.10–1.43)‡
Income category, percentile				
0–25th	1.01 (0.92–1.12)	1.19 (1.07–1.31)‡	1.04 (0.95–1.15)	1.02 (0.88–1.19)
25–50th	0.96 (0.88–1.05)	1.08 (0.98–1.18)	1.04 (0.95–1.13)	1.12 (0.97–1.28)
50–75th	0.95 (0.88–1.03)	1.04 (0.95–1.14)	1.01 (0.93–1.10)	1.11 (0.98–1.27)
75–100th	Ref	Ref.	Ref.	Ref.
Primary infection diagnosis				
Sepsis	Ref.	Ref.	Ref.	Ref.
OI	1.18 (1.00–1.39)‡	0.65 (0.56–0.76)‡	1.60 (1.35–1.91)‡	0.68 (0.56–0.83)‡
SSTI	0.40 (0.36–0.44)‡	0.41 (0.37–0.46)‡	0.57 (0.51–0.62)‡	0.07 (0.05–0.10)‡
UTI	0.29 (0.24–0.34)‡	0.33 (0.27–0.41)‡	0.30 (0.25–0.35)‡	0.03 (0.01–0.09)‡
Pneumonia	0.66 (0.62–0.71)‡	0.43 (0.40–0.47)‡	0.70 (0.66–0.76)‡	0.35 (0.31–0.39)‡
Insurance payer				
Medicare	1.12 (1.01–1.23)‡	1.33 (1.19–1.49)‡	1.08 (0.98–1.19)	0.88 (0.76–1.02)
Medicaid	1.27 (1.10–1.46)‡	1.26 (1.06–1.49)‡	1.18 (1.03–1.35)‡	0.94 (0.75–1.19)
Other	1.23 (0.96–1.57)	0.82 (0.60–1.13)	0.96 (0.76–1.21)	1.21 (0.83–1.76)
Private	Ref.	Ref.	Ref.	Ref.
Self	1.04 (0.83–1.31)	0.69 (0.49–0.95)‡	0.80 (0.65–1.00)	1.35 (0.96–1.90)
Hospital region				
Northeast	Ref.	Ref.	Ref.	Ref.
Midwest	0.83 (0.75–0.91)‡	0.89 (0.80–0.98)‡	0.76 (0.69–0.83)‡	0.90 (0.77–1.05)
South	1.01 (0.93–1.11)	0.75 (0.69–0.83)‡	0.94 (0.86–1.03)	1.10 (0.95–1.26)
West	0.72 (0.65–0.79)‡	0.66 (0.59–0.73)‡	0.56 (0.51–0.62)‡	1.01 (0.86–1.17)
Hospital location/teaching				
Rural	Ref.	Ref.	Ref.	Ref.
Urban non-teaching	2.54 (2.30–2.80)‡	0.93 (0.84–1.04)	1.47 (1.33–1.63)‡	1.61 (1.33–1.95)‡
Urban teaching	2.35 (2.13–2.59)‡	0.78 (0.70–0.87)‡	1.39 (1.25–1.53)‡	1.83 (1.51–2.21)‡
Hospital size, by no. of beds				
Small	Ref.	Ref.	Ref.	Ref.
Medium	1.37 (1.24–1.50)‡	0.89 (0.80–0.99)	1.22 (1.10–1.35)‡	1.25 (1.05–1.49)‡
Large	2.12 (1.94–2.31)‡	0.87 (0.79–0.96)‡	1.41 (1.28–1.54)‡	1.52 (1.30–1.78)‡

* Values are the adjusted odds ratio (95% confidence interval [95% CI]) unless indicated otherwise. OI = opportunistic infections; Ref. = reference category; SSTI = skin and soft tissue infections; UTI = urinary tract infection.
† Non-home health care facility, i.e., inpatient, nursing, or rehabilitation facility.
‡ Significant.

Multivariable-adjusted associations with health care utilization and mortality in people with vasculitis. Multivariable-adjusted analyses showed that, compared to sepsis, other serious infections were associated with significantly lower odds of health care utilization (charges above the median, length of stay >3 days, discharged to non-home setting) and in-hospital mortality. OIs, however, were associated with higher hospital charges above the median and higher odds of hospital length of stay of >3 days (Table 4). When compared to sepsis, other serious infections were associated with odds ratios of 0.03–0.68 for in-hospital mortality, depending on the type of serious infection.

Older age (≥80 years), a Deyo-Charlson comorbidity index score of ≥2, urban teaching or non-teaching hospital, or medium or large hospital (by number of beds) were associated with higher health care utilization and in-hospital mortality rates; Northeast region, and Medicare or Medicaid payer type were associated with higher health care utilization (Table 4).

DISCUSSION

The present study is a national study of serious infections in people with vasculitis. We examined the epidemiology and outcomes of serious infections, i.e., OIs, SSTI, UTI, pneumonia, and sepsis, and associated time trends. We identified several factors significantly associated with health care utilization and in-hospital mortality outcomes in people with serious infection and vasculitis. Several findings merit further discussion.

We found increasing rates of serious infections overall during the study period, from 1998 to 2016. Serious infection rates per 100,000 NIS claims increased for SSTI, UTI, and sepsis but remained relatively stable for OIs and pneumonia. In the first study period, 1998–2000, pneumonia was the most common serious infection, with sepsis as the second most common serious infection; in 2015–2016, sepsis was the most common serious infection, followed by pneumonia. We found that sepsis surpassed pneumonia as the most common serious infection in 2011–2012. In 2015–2016, sepsis was 2.3 times more common than pneumonia in people with vasculitis. Some evidence of systematic up-coding of pneumonia to sepsis and possible misclassification errors with sepsis codes (16,17) may explain, at least partially, the increased rate of sepsis diagnosis over time versus pneumonia. To our knowledge, most of the prior studies of serious infections in vasculitis are >10 years old and did not examine time trends in rates of each serious infection (3,4). Our study provides contemporary data on serious infection hospitalizations in people with vasculitis.

In Sweden in 1998–2010 in a population-based sample (4), the rates for sepsis, pneumonia/influenza, and skin infections in people with AAV were 18.2, 50.1, and 7.6 per 1,000 person-years, respectively. Using a different denominator in our hospitalization-based (not population-based) study, the respective rates in our study were 53 for sepsis, 56 for pneumonia, and 18 for SSTI per 1,000 vasculitis hospitalizations in the US in 2009–2010. These rates are not directly comparable due to different denominators, which may underlie the differences; differences in country setting, time period duration, and the definition of infection categories (influenza was not included in our study, and skin versus SSTI) may have also contributed.

Our study showed factors associated with higher health care utilization and/or mortality rates in people with vasculitis hospitalized for serious infections. Among serious infections, sepsis was associated with worse outcomes, except that OIs were associated with higher hospital charges. As an example, sepsis was associated with 34–97% greater odds of in-hospital mortality when compared to other serious infections. Our findings revealed other significant correlates of higher health care utilization and in-hospital mortality rates, namely, older age (≥80 years), Deyo-Charlson comorbidity index score of ≥2, urban teaching or non-teaching hospital, or medium or large hospital (by number of beds). Northeast region and Medicaid payer type were associated with higher health care utilization, but no differences in mortality. Our study identifies important risk factors for outcomes of serious infections in vasculitis and adds to the current knowledge.

Our study findings must be interpreted considering limitations and strengths. Our database study is at risk of misclassification bias because we used the ICD-9-CM or the ICD-10-CM codes to identify vasculitis and serious infections. However, the serious infection (6–10) and vasculitis (11) diagnostic codes were valid in administrative data sets, with high positive predictive values in previous studies. Independent validation in the NIS is not possible because data are deidentified and no medical records are available. Misclassification might have biased our results toward the null. The NIS does not contain vasculitis-specific factors, such as organ involvement, disease severity, and organ damage (disease activity and/or severity measures), as candidate variables that are related to outcomes and therefore their impact could not be analyzed. The NIS does not have data on laboratory test results and medications. Therefore, we are unable to examine the effect of these disease/treatment variables on outcomes of interest. Future studies should examine, in particular, the critical nature of the relative and absolute contribution of glucocorticoids versus immunosuppressive medications versus biologics to the risk of serious infection. The unit of analysis in the NIS is hospitalizations, not people. The NIS does not provide longitudinal data after hospital discharge, which limits the ability to examine the post-discharge outcomes and/or readmission risk. The NIS provides inpatient hospital charges, not actual costs, which are likely much higher than the actual cost. Our study strengths include the use of a national data set, a large sample size, and the inclusion of several potential confounders in our analyses.

In conclusion, among serious infection hospitalizations, people with vasculitis versus those without differed from each other. We found an overall increasing rate of serious infections in vasculitis over time. Time trends for OIs, SSTI, UTI, pneumonia, and sepsis differed. Sepsis surpassed pneumonia as the most common serious infection in 2011–2012. We identified significant correlates of health care utilization and mortality in people with vasculitis hospitalized with serious infections. These findings can better inform patients and providers about the risk and outcomes of serious infections in vasculitis. Interventions are needed in patients with vasculitis to avoid serious infections and to improve its outcome once it occurs.

AUTHOR CONTRIBUTIONS

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

Study conception and design. Singh.
Acquisition of data. Singh, Cleveland.
Analysis and interpretation of data. Singh, Cleveland.

REFERENCES

1. Phillip R, Luqmani R. Mortality in systemic vasculitis: a systematic review. Clin Exp Rheumatol 2008;26 Suppl 51:S94–104.
2. Bosch X, Guilabert A, Espinosa G, Mirapeix E. Treatment of antineutrophil cytoplasmic antibody associated vasculitis: a systematic review. JAMA 2007;298:655–69.
3. Goupil R, Brachemi S, Nadeau-Fredette AC, Déziel C, Troyanov Y, Lavergne V, et al. Lymphopenia and treatment-related infectious complications in ANCA-associated vasculitis. Clin J Am Soc Nephrol 2013;8:416–23.
4. Mohammad AJ, Segelmark M, Smith R, Englund M, Nilsson JÅ, Westman K, et al. Severe infection in antineutrophil cytoplasmic antibody-associated vasculitis. J Rheumatol 2017;44:1468–75.
5. HCUP databases. Healthcare cost and utilization project (HCUP). Overview of the nationwide inpatient sample (NIS). URL: https://www.hcup-us.ahrq.gov/nisoverview.jsp.
6. Jinno S, Lu N, Jafarzadeh SR, Dubreuil M. Trends in hospitalizations for serious infections in patients with rheumatoid arthritis in the US between 1993 and 2013. Arthritis Care Res (Hoboken) 2018;70:652–8.
7. Tektonidou MG, Wang Z, Dasgupta A, Ward MM. Burden of serious infections in adults with systemic lupus erythematosus: a national population-based study, 1996–2011. Arthritis Care Res (Hoboken) 2015;67:1078–85.
8. Schneeweiss S, Robicsek A, Scranton R, Zuckerman D, Solomon DH. Veteran's affairs hospital discharge databases coded serious bacterial infections accurately. J Clin Epidemiol 2007;60:397–409.
9. Grijalva CG, Chung CP, Stein CM, Gideon PS, Dyer SM, Mitchel EF Jr, et al. Computerized definitions showed high positive predictive values for identifying hospitalizations for congestive heart failure and selected infections in Medicaid enrollees with rheumatoid arthritis. Pharmacoepidemiol Drug Saf 2008;17:890–5.
10. Patkar NM, Curtis JR, Teng GG, Allison JJ, Saag M, Martin C, et al. Administrative codes combined with medical records based criteria accurately identified bacterial infections among rheumatoid arthritis patients. J Clin Epidemiol 2009;62:321–7.
11. Bernatsky S, Linehan T, Hanly JG. The accuracy of administrative data diagnoses of systemic autoimmune rheumatic diseases. J Rheumatol 2011;38:1612–6.
12. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol 1992;45:613–9.
13. Medicare eligibility: who may enroll in Medicare. URL: https://www.ehealthmedicare.com/about-medicare/eligibility/.
14. Sabesan VJ, Petersen-Fitts G, Lombardo D, Briggs D, Whaley J. Medicaid payer status is linked to increased rates of complications after treatment of proximal humerus fractures. J Shoulder Elbow Surg 2017;26:948–53.
15. Agency for Healthcare Research and Quality. HCUP NIS description of data elements. ZIPINC_QRTL - Median household income for patient's zip code (based on current year). Healthcare Cost and Utilization Project (HCUP). September 2008. URL: www.hcup-us.ahrq.gov/db/vars/zipinc_qrtl/nisnote.jsp.
16. Lindenauer PK, Lagu T, Shieh MS, Pekow PS, Rothberg MB. Association of diagnostic coding with trends in hospitalizations and mortality of patients with pneumonia, 2003–2009. JAMA 2012;307:1405–13.
17. Rhee C, Dantes R, Epstein L, Murphy DJ, Seymour CW, Iwashyna TJ, et al. Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009–2014. JAMA 2017;318:1241–9.