ARP Announcements

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

Applications Invited for *Arthritis Care & Research* Editor-in-Chief (2021–2026 Term)

The American College of Rheumatology Committee on Journal Publications announces the search for the position of Editor, *Arthritis Care & Research*. The official term of the next *Arthritis Care & Research* editorship is July 1, 2021–June 30, 2026; however, some of the duties of the new Editor will begin during a transition period starting April 1, 2021. ARP/ACR members who are considering applying for this prestigious and rewarding position should submit a nonbinding letter of intent by May 4, 2020 to the Managing Editor, Maggie Parry, at mparry@rheumatology.org, and are also encouraged to contact the current Editor-in-Chief, Dr. Marian Hannan, to discuss details. Initial contact should be made via e-mail to Hannan@hsl.harvard.edu. Applications will be due by June 15, 2020 and will be reviewed during the summer of 2020. Application materials will be available on the ACR web site at https://www.rheumatology.org/Portals/0/Files/ACandR-Editor-Application-Instructions.pdf.

Download the New ACR Publications Mobile App

The brand-new ACR Publications app can be downloaded for free from the Apple store or Google Play. ACR members can log in for full-text access to all articles in A*rthritis Care & Research* and *Arthritis & Rheumatology*. Nonmembers can access abstracts of all *AC&R* and *A&R* articles, the full text of articles published more than one year ago, and select open-access articles published recently, as well as the full text of all articles from *ACR Open Rheumatology* and *The Rheumatologist*.

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 72, No. 5, May 2020

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 72 • May 2020 • NO. 5

Special Articles

Systemic Lupus Erythematosus

Osteoarthritis

Cover image: Cover images (from Khawar et al, pages 602 and 605): The image at left is a hand radiograph at admission showing extensive, prominent periosteal and joint capsular calcifications involving all of the phalanges, metacarpal bones, carpal bones, and the distal radius and ulna of the left hand, with the most prominent calcifications adjacent to the joints. There is marked soft tissue swelling around the metacarpophalangeal, proximal, and distal interphalangeal (IP) joints, as well as periarticular osteopenia. The image at right is a hand radiograph taken 8 months from time of discontinuation of voriconazole showing substantial improvement in the diffuse bilateral proximal interphalangeal (PIP) joints and distal IP joints. A small amount of residual periostitis is seen in the bilateral carpal bones and PIP joints.

Arthritis Care & Research

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

GENERAL INFORMATION

TO SUBSCRIBE

Institutions and Non-Members

Email: wileyonlinelibrary.com
Phone: (201) 748-6645
Write: John Wiley & Sons, Inc.
Attn: Journals Admin Dept UK
111 River Street
Hoboken, NJ 07030

Volumes 72, 2020:

Institutional Print Only:	**Arthritis & Rheumatology and Arthritis Care & Research:** $2,428 in US, Canada, and Mexico $3,097 outside North America
Institutional Online Only:	$2,428 in US, Canada, Mexico, and outside North America
Institutional Print and Online Only:	$3,036 in US, Canada, and Mexico $3,872 outside North America

For submission instructions, subscription, and all other information visit: wileyonlinelibrary.com.

Arthritis Care & Research accepts articles for Open Access publication. Please visit https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/onlineopen.html for further information about OnlineOpen.

Wiley's Corporate Citizenship initiative seeks to address the environmental, social, economic, and ethical challenges faced in our business and which are important to our diverse stakeholder groups. Since launching the initiative, we have focused on sharing our content with those in need, enhancing community philanthropy, reducing our carbon impact, creating global guidelines and best practices for paper use, establishing a vendor code of ethics, and engaging our colleagues and other stakeholders in our efforts.

Follow our progress at www.wiley.com/go/citizenship.

Access to this journal is available free online within institutions in the developing world through the HINARI initiative with the WHO. For information, visit www.healthinternetwork.org.

Disclaimer

The Publisher, the American College of Rheumatology, and Editors cannot be held responsible for errors or any consequences arising from the use of information contained in this journal; the views and opinions expressed do not necessarily reflect those of the Publisher, the American College of Rheumatology and Editors, neither does the publication of advertisements constitute any endorsement by the Publisher, the American College of Rheumatology and Editors of the products advertised.

Members:
American College of Rheumatology/Association of Rheumatology Professionals

For membership rates, journal subscription information, and change of address, please write:
American College of Rheumatology
2200 Lake Boulevard
Atlanta, GA 30319-5312
(404) 633-3777

ADVERTISING SALES AND COMMERCIAL REPRINTS

Sales contacts: Kathleen Malseed, National Account Manager
E-mail: kmalseed@pminy.com
Phone: (212) 852-9824
Pharmaceutical Media, Inc.
30 East 33rd Street, New York, NY 10016

Publisher: Arthritis Care & Research is published byWiley Periodicals, Inc., 101 Station Landing, Suite 300, Medford, MA 02155

Production Editor: Willem Galanis, acrprod@wiley.com

ARTHRITIS CARE & RESEARCH (Print ISSN 2151-464X; Online ISSN 2151-4658 at Wiley Online Library, wileyonlinelibrary.com) is published monthly on behalf of the American College of Rheumatology by Wiley Subscription Services, Inc.; a Wiley Company, 111 River Street, Hoboken, NJ 07030-5774. POSTMASTER: Send all address changes to Arthritis & Rheumatology, John Wiley & Sons, Inc., c/o The Sheridan Press, PO Box 465, Hanover, PA 17331. Send **subscription inquiries care of** John Wiley & Sons, Inc., Attn: Journals Admin Dept UK, 111 River Street, Hoboken, NJ 07030, (201) 748-6645 (nonmember subscribers only; American College of Rheumatology/Association of Rheumatology Professionals members should contact the American Collegeof Rheumatology). Subscription Price: (Volumes 71, 2019: Arthritis & Rheumatology and Arthritis Care & Research) Print only: $2,428.00 in U.S., Canada and Mexico, $3,097.00 rest of world. For all other prices please consult the journal's website at wileyonlinelibrary.com. All subscriptions containing a print element, shipped outside U.S., will be sent by air. Payment must be made in U.S. dollars drawn on U.S. bank. Prices are exclusive of tax. Asia-Pacific GST, Canadian GST and European VAT will be applied at the appropriate rates. For more information on current tax rates, please go to www.wileyonlinelibrary.com/tax-vat. The price includes online access to the current and all online back files to January 1st 2016, where available. For other pricing options including access information and terms and conditions, please visit https://onlinelibrary.wiley.com/library-info/products/price-lists. Terms of use can be found here: https://onlinelibrarywiley.com/terms-and-conditions. **Change of Address:** Please forward to the subscriptions address listed above 6 weeks prior to move; enclose present mailing label with change of address. **Delivery Terms and Legal Title:** Where the subscription price includes print issues and delivery is to the recipient's address, delivery terms are Delivered at Place (DAP); the recipient is responsible for paying any import duty or taxes. Title to all issues transfers Free of Board (FOB) our shipping point, freight prepaid. We will endeavor to fulfill claims for missing or damaged copies within six months of publication, within our reasonable discretion and subject to availability. **Claims** for undelivered copies will be accepted only after the following issue has been received. Please enclose a copy of the mailing label or cite your subscriber reference number in order to expedite handling. Missing copies will be supplied when losses have been sustained in transit and where reserve stock permits. Send claims care of John Wiley & Sons, Inc., Attn: Journals Admin Dept UK, 111 River Street, Hoboken, NJ 07030. If claims are not resolved satisfactorily, please write to Subscription Distribution c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. **Cancellations:** Subscription cancellations will not be accepted after the first issue has been mailed. Printed in U.S.A. **Journal Customer Services:** For ordering information, claims and any enquiry concerning your journal subscription please go to https://hub.wiley.com/community/support/onlinelibrary or contact your nearest office. **Americas:** Email: cs-journals@wiley.com; Tel: 11 781 388 8598 or 11 800 835 6770 (toll free in the USA & Canada). **Europe, Middle East and Africa:** Email: cs-journals@wiley.com; Tel: 144 (0) 1865 778315. **Asia Pacific:** Email: cs-journals@wiley.com; Tel: 165 6511 8000. **Japan:** For Japanese speaking support, Email: cs-japan@wiley.com; Tel: 165 6511 8010 or Tel (toll-free): 005 316 50 480. Visit our Online **Customer Help** available in 7 languages at https://hub.wiley.com/community/support/onlinelibrary. **Back Issues:** Single issues from current and prior year volumes are available at the current single issue price from cs-journals@wiley.com. Earlier issues may be obtained from Periodicals Service Company, 351 Fairview Avenue—Ste 300, Hudson, NY 12534, USA. Tel: 11 518 822-9300, Fax: 11 518 822-9305. Email: psc@periodicals.com. Printed in the USA by The Sheridan Group. Periodical postage paid at Hoboken, NJ and additional offices.

Arthritis Care & Research
Vol. 72, No. 5, May 2020, pp 601–606
DOI 10.1002/acr.23812

American College of Rheumatology
Empowering Rheumatology Professionals

CLINICOPATHOLOGIC CONFERENCE

A 31-Year-Old Man With A Fungal Infection, Elevated Alkaline Phosphatase Level, and Polyarthritis

Talha Khawar,[1] Carsten R. Hamann,[2] Arezoo Haghshenas,[1] Allie Blackburn,[1] and Karina D. Torralba[1]

CASE PRESENTATION

Chief symptoms

A 31-year-old male presented with 4 weeks of progressive joint pain and swelling.

History of present illness

The patient presented with 4 weeks of progressive joint swelling and pain affecting bilateral hands, feet, and ankles. The patient had been nonverbal, hemiplegic, and unable to get out of bed without assistance for the past 6 months. He had not had any past joint swelling or low back pain, and his family could not report any identifiable aggravators or alleviators for his symptoms. He had not lost weight, nor had any fevers, chills, night sweats, cough, hemoptysis or chest pains, vision changes, diarrhea, bloody stools, dysuria, or any rashes.

Medical history

The patient was in normal health until 4 years prior, when he developed chronic sinusitis. He underwent nasal polypectomy and successful removal of a paranasal fungal mass. Four years later, infection recurred with extension into the brain, resulting in loss of consciousness and seizures requiring treatment with levetiracetam. A left frontal/parietal decompressive craniotomy and frontal lobe mass resection were performed; an intracranial hemorrhage during craniotomy resulted in paraplegia and aphasia. The cultures from brain tissue grew *Aspergillus fumigatus*, and treatment with voriconazole was started. Over the next 6 months, additional surgical interventions were performed, including frontal sinusotomy, complete ethmoidectomy, sphenoidotomy, maxillary antrostomy, and ventriculostomy drain placement followed by externalization of the voriconazole periostitis. Voriconazole treatment was continued during this period until the current presentation.

Family and social history

There was no family history of inflammatory arthritis, psoriasis, inflammatory bowel disease, or malignancy. The patient used to be a welder. He is married with 2 children. There was no history of tobacco or illicit drug use.

Physical examination

At the patient's current presentation, he appeared chronically ill and lethargic but arousable. He was unable to move the right side of his body. His right pupil was sluggish but reactive to light. He had tracheostomy and gastrostomy tubes in place. The results of the cardiopulmonary examination were unremarkable. He had soft tissue swelling, bony enlargement, and tenderness and warmth of his metacarpophalangeal joints, as well as of the proximal and distal interphalangeal joints on both hands (Figure 1). He also had tenderness over his distal lower extremities and ankles bilaterally but no swelling or warmth. Digital clubbing, nail pitting, and other dermatologic manifestations of psoriasis were not present.

Laboratory and imaging evaluation

Results for complete blood count, electrolytes (serum sodium, potassium, and chloride), aspartate aminotransferase, alanine aminotransferase, blood urea nitrogen, creatinine, albumin, C-reactive protein, thyroid-stimulating hormone, free T4, total T3, Quantiferon TB Gold, and rapid plasma reagin levels were within normal limits. Results were notable for an erythrocyte sedimentation rate (ESR; Westergren method) of 71 mm/hour (normal 0–20), an alkaline phosphatase level of 1,495 IU/liter (normal 4–147), and a gamma glutamyl transferase (GGT) level of 763 units/liter (normal 0–45). Antinuclear antibodies and other autoantibodies to cyclic citrullinated peptide, antineutrophil cytoplasmic antibody, proteinase-3, myeloperoxidase,

[1]Talha Khawar, MD, Arezoo Haghshenas, MD, Allie Blackburn, MD, Karina D. Torralba, MD: Loma Linda University, Loma Linda, California; [2]Carsten R. Hamann, MD: Herlev and Gentofte Hospital, University of Copenhagen, Hellerup, Denmark.

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

Address correspondence to Karina D. Torralba, MD, 11234 Anderson Street, Room 1521, Loma Linda University Medical Center, Loma Linda, CA 92354. E-mail: Ktorralba@llu.edu.

Submitted for publication June 12, 2018; accepted in revised form November 13, 2018.

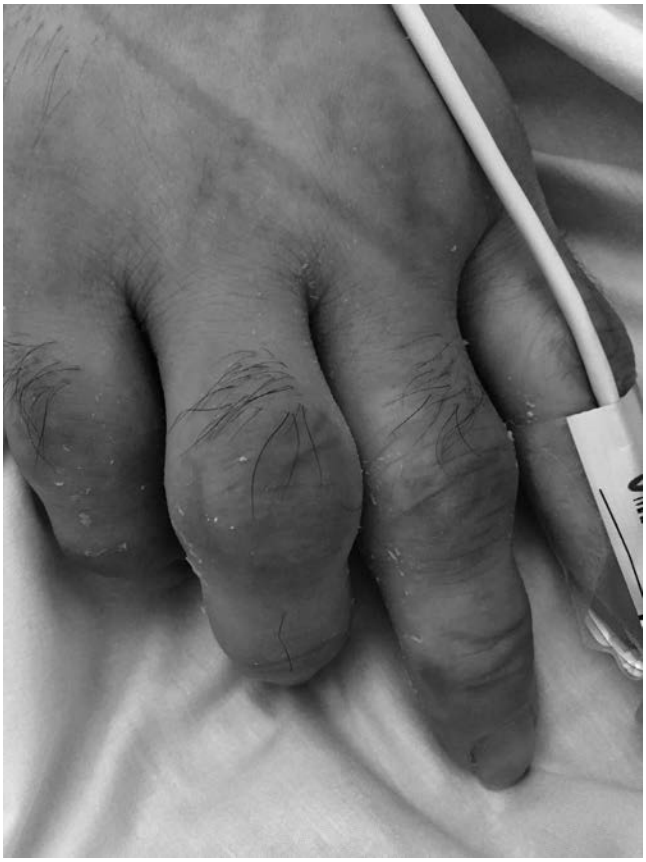

Figure 1. Photograph of right hand showing marked swelling of all the metacarpophalangeal, proximal, and distal interphalangeal joints. There was joint erythema on examination.

double-stranded DNA, Sm, SSA/Ro, SSB/La, RNP, cardiolipin, β_2-glycoprotein, and antithyroid antibodies were negative. Lupus anticoagulant and HLA–B27 test results were negative. The patient's ionized calcium level was 1.18 mmoles/liter (normal 1.00–1.40). The intact parathyroid hormone level was 20.6 pg/ml (normal 15.0–65.0), and total 25-hydroxyvitamin D was 22 ng/ml. The patient's magnesium was 0.8 mmoles/liter (normal 0.7–1.2), and his serum fluoride level was high at 5.3 μmoles/liter (normal 0.0–4.0). A radiograph of the chest showed normal findings. Ankle, foot, and hand radiographs showed extensive periosteal reaction with fluffy periostitis and new bone formation (Figures 2 and 3). A triple-phase bone scan performed to evaluate for complex regional pain syndrome showed prominent uptake adjacent to the joints in both hands, corresponding to areas of periosteal reaction (Figure 4).

CASE SUMMARY

The patient is a 31-year-old male presenting with subacute-onset progressive inflammatory polyarthritis with laboratory and radiologic evidence of inflammation and a bone remodeling reaction based on elevated alkaline phosphatase levels, an elevated ESR, and periosteal reaction as shown on radiographs.

DIFFERENTIAL DIAGNOSIS

Periostitis is an inflammation of the periosteum, the membrane enveloping a bone, and is radiographically characterized by thickening of the periosteum and new bone formation (1). Osteitis deformans is a term used to describe the spectrum of clinical, radiologic, and pathologic features that are associated with periostitis and is typically associated with marked acceleration of the bone remodeling process of breakdown and regrowth (2). The exact mechanism of how periostitis occurs in humans is largely unknown, especially in nonhereditary cases; however, it is theorized to occur as an exaggerated and irregular bone metabolism response to an

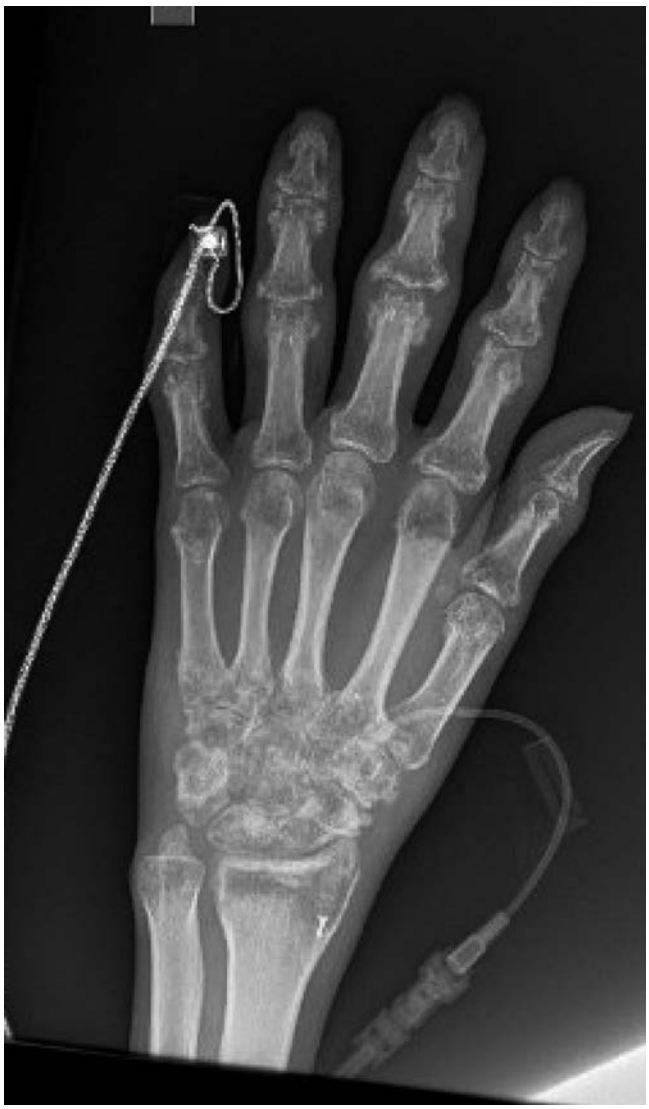

Figure 2. Radiograph of left hand, anteroposterior view, taken at time of admission, showing extensive, prominent periosteal and joint capsular calcifications involving all of the phalanges, metacarpal bones, carpal bones, and the distal radius and ulna, with the most prominent changes adjacent to the joints. There is marked soft tissue swelling around the metacarpophalangeal, proximal, and distal interphalangeal joints, as well as periarticular osteopenia.

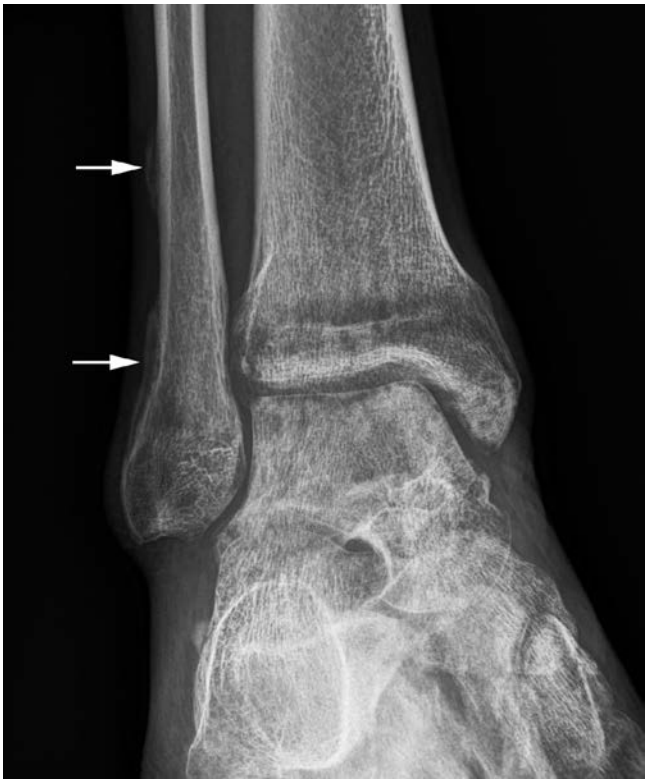

Figure 3. Radiograph of right ankle, anteroposterior view. Bones have a moth-eaten appearance. There is mild soft tissue swelling around the hindfoot. A remodeling/periosteal reaction is noted in the distal fibula (**arrows**).

inflammatory signal, which involves alternating processes of bone resorption and osteoclast invasion of the periosteum's germinative layer and bone formation where the osteoblasts promote apposition of a new layer of bone (2–4). Periostitis can present acutely, usually in the setting of infection, or on a subacute or chronic basis

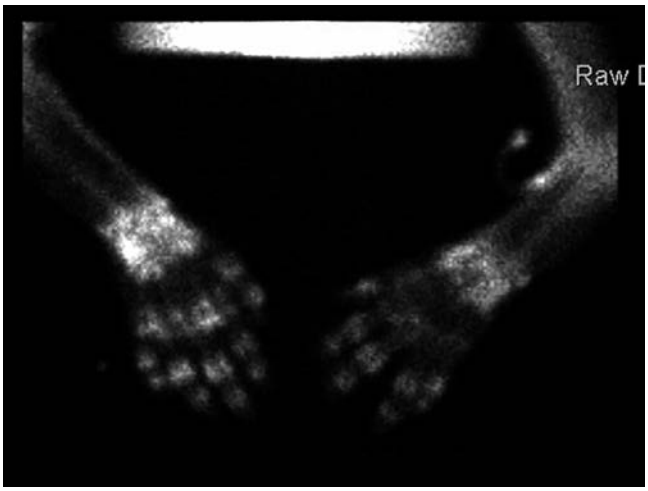

Figure 4. Triple-phase bone scan using 23 mCi ^{99m}Tc-hydroxydiphosphonate to evaluate for complex regional pain syndrome showing prominent uptake adjacent to the joints in both hands, corresponding to areas of periosteal reaction.

Table 1. Causes of periostitis in both acute and chronic forms

Acute periostitis	
Trauma	Excessive physical activity (shin splints, tibial periostalgia, soleus periostalgia)
Infection	Congenital syphilis; *Staphylococcus* species; *Streptococcus* species; acute osteomyelitis
Malignancy	Leukemia
Chronic periostitis	
Hyperosteoarthropathy	Primary hyperosteoarthropathy: pachydermoperiostosis or Touraine-Solente-Gole syndrome; 15-dihydroxyprostaglandin dehydrogenase gene mutation Secondary hyperosteoarthropathy: cardiac, pulmonary, endocrine, gastrointestinal conditions, and malignancies
Drugs and other chemical exposures Fluoride Prostaglandin E_1 Vitamin A Voriconazole	
Infection	Osteomyelitis; syphilis
Malignancy	Eosinophilic granuloma/Langerhans' cell histiocytosis
Benign bone lesions	Osteoid osteoma; aneurysmal bone cyst
Trauma	
Hemophilia	
Thyroid acropachy	
Inflammatory diseases	Psoriatic arthritis; associations with antineutrophil cytoplasmic antibody–associated vasculitis; Takayasu arteritis; sarcoidosis; inflammatory bowel disease; relapsing polychondritis

with varying underlying conditions (Table 1). The main differential diagnoses for this case include the following: hypertrophic osteoarthropathy (HOA), psoriatic arthritis, thyroid acropachy, skeletal fluorosis from fluoride ingestion, and medication-related periostitis from prolonged use of voriconazole.

HOA. Periostitis is often due to HOA, which features the triad of digital clubbing, arthritis, and periostitis of the long bones. HOA comes in both primary and secondary forms. Primary HOA (also known as pachydermoperiostosis or Touraine-Solente-Gole syndrome) comprises 5% of HOA cases and is hereditary due to a 15-dihydroxy prostaglandin dehydrogenase gene mutation (5). Secondary HOA, comprising 95% of HOA cases, is associated with cardiac, pulmonary, endocrine, and gastrointestinal conditions and malignancies (1). HOA, previously known as hypertrophic pulmonary osteoarthropathy because of the wide spectrum of pulmonary conditions, including lung cancer usually associated with HOA, typically also features digital clubbing, periostitis, and excessive proliferation of skin and bone in the extremities (1,6). Our patient did not have any cardiopulmonary conditions or any findings of active malignancy. The

patient's age precludes the diagnosis of primary HOA, which is exclusively seen in childhood. He did have elevated GGT and alkaline phosphatase levels, but this was thought to be due in part to cholestatic hepatic toxicity and in part to bone toxicity as elucidated further in this article.

Psoriatic arthropathy. Psoriatic arthropathy is another diagnosis to consider (7). The patient did have features of synovitis. He had extensive osteoproliferation, which can be seen in psoriatic arthritis usually in association with erosive arthropathy, but the patient lacked other features of psoriatic arthritis, including lack of enthesitis, dactylitis, and a self or family history of psoriasis.

Skeletal fluorosis. Chronic fluoride ingestion, classically described in association with industrial inhalation of fluoride fumes, consumption of water and tea, and cryolite exposure, can lead to skeletal fluorosis, in which there are extensive exostoses of the long bones as well as extensive calcifications of ligaments and cartilage (8). This patient did not have a history suggestive of this condition.

Thyroid acropachy. Thyroid acropachy, which presents with soft tissue swelling of fingers and periosteal reaction of bones, is a manifestation of autoimmune thyroid disorder (9,10). It usually involves the bones of the hand and the feet and accompanies thyroid dermopathy and ophthalmopathy. Thyroid acropachy can occur in patients with either hypothyroidism or euthyroidism. This patient had normal thyroid function as confirmed by test findings and had no features of dermopathy or ophthalmopathy.

Drug-induced periostitis. A number of drugs can cause periostitis, including prostaglandin E_1 (PGE_1) and vitamin A (Table 1). When taken in excessive amounts, vitamin A causes cortical thickening of tubular bones, metaphyseal cupping and fraying, growth plate irregularity, and premature fusion of the ossification centers along with proliferative enthesopathy of the long bones and flaval ligaments (11). Cortical bone thickening has been associated with prolonged use of PGE_1 in conjunction with its intravenous administration to infants with intention to maintain patency of the ductus arteriosus until corrective surgery or transplant can be done (12). This patient had no exposure to vitamin A or PGE_1 but had been taking voriconazole for at least 6 months.

Voriconazole is a broad-spectrum, synthetic, second-generation azole antifungal with activity against a variety of fungal infections, especially particularly invasive aspergillosis and infections caused by *Candida albicans* (13,14). It is a triazole derivative of fluconazole and disrupts fungal cell wall synthesis by inhibiting a cytochrome P450–dependent enzyme, 14-alpha sterole demethylase (13,14). Voriconazole has been described as causing a painful periostitis of the long bones in various immunocompromised patients with associated elevation in alkaline phosphatase levels, with most cases involving lung, heart, liver, stem cell, and bone marrow transplant recipients (15–24). This phenomenon has been reported in patients receiving therapy with this antifungal as either prophylaxis or treatment. Since the current patient had been taking voriconazole for 6 months and had no history to suggest any of the other diagnostic considerations, voriconazole-induced periostitis was highly suspected to be the cause of his skeletal findings.

DISCUSSION

A number of cases, predominantly in the transplant medicine literature, have described the occurrence of periostitis with voriconazole use. However, due to lack of controlled studies, it is not very well understood whether the levels or duration of voriconazole treatment predict the risk of developing periostitis. The duration of therapy for most published reports ranged between 6 weeks to 416 weeks before symptoms of periostitis ensued (15–24). The current patient received voriconazole for 6 months and had significantly more robust radiologic evidence of periostitis compared to other reported cases where the duration of voriconazole use was much longer. All other reported cases involved patients taking dosages of ≥200 mg/day before having symptoms of periostitis. In contrast, the current patient was gradually increased to a dosage of 100 mg/day, which was when he started to show sign of periostitis. Serum fluoride levels in the current patient were slightly elevated but had the lowest value compared to other reported cases (22–24).

Evaluation for voriconazole as a cause of periostitis needs further research. In a way, voriconazole-related periostitis can be considered as a form of fluorosis since part of voriconazole's chemical composition consists of difluorophenyl fluoropyrimidine. Serum fluoride levels tend to be much higher in patients taking voriconazole compared to the general population (16,22–24). Measurement of serum voriconazole levels may be useful in some patients, especially when voriconazole is taken orally due to variable bioavailability, either to evaluate for potential toxicity or to document adequate drug exposure, as was the case in the current patient. The drug levels in the current patient were below the therapeutic range for the majority of the treatment duration, which resulted in a progressive escalation of dose until he started to have symptoms at a dosage of 100 mg/day. The patient's alkaline phosphatase levels were elevated, which is consistent with previously reported cases. However, the patient also had an elevated GGT level, and we hypothesize that this patient's alkaline phosphatase elevation was due in part to cholestatic hepatic toxicity and in part due to bone toxicity. Previous studies have reported levels anywhere from as low as 170 units/liter to as high as 2,548 units/liter. According to the manufacturer's recommendations, these levels are routinely measured, along with other liver function tests, weekly in the first month of treatment and then monthly. There is no recommendation for measuring the bone-specific alkaline

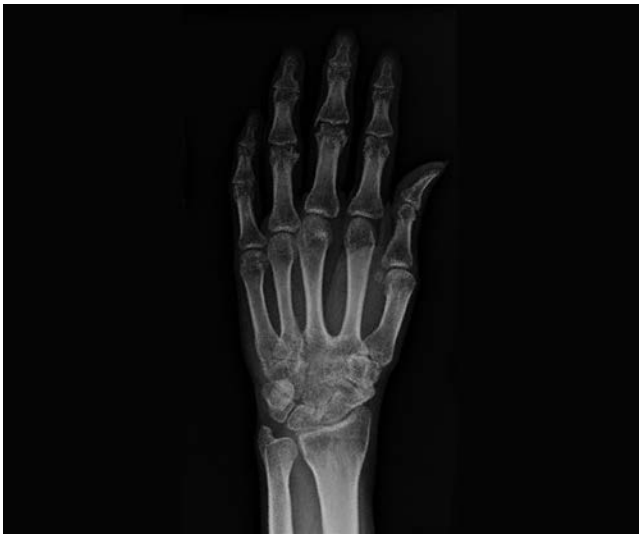

Figure 5. Radiograph of left hand, anteroposterior view, taken 8 months from time of discontinuation of voriconazole. There is substantial improvement of carpal periostitis and soft tissue swelling in the diffuse bilateral proximal interphalangeal (PIP) joints and distal interphalangeal joints. A small amount of residual periostitis is seen in the bilateral carpal bones and PIP joints.

phosphatase level to identify at-risk patients who may develop periostitis. There also is no reported association between periostitis and the levels of bone-specific alkaline phosphatase in patients taking voriconazole (21,25). While serum fluoride levels may be used as surrogate markers for periostitis in symptomatic patients when a bone scan is not readily available, they are not widely available, hence delaying the diagnosis (23). On the other hand, alkaline phosphatase levels, although very high in voriconazole-induced periostitis, are highly variable due to other metabolic factors such as vitamin D, parathyroid hormone levels, calcium intake, and kidney disease, and therefore may not be as reliable (22).

Animal studies have shown that fluoride exposure causes aberrant activation of osteoblasts and inhibits osteoclast-mediated bone resorption (25). Voriconazole promotes osteoblastic differentiation through 2 pathways, either by increasing alkaline phosphatase or by promoting the expression of runx2, a key protein associated with osteoblast differentiation. Additionally, endothelial growth factor and platelet-derived growth factor, which are vital in angiogenesis and osteoblast proliferation and differentiation, are up-regulated with voriconazole exposure (25,26). However, further research is needed to clarify these pathogenic processes.

The cause and effect association between voriconazole use and periostitis development in this patient appears highly probable. Rechallenging the patient with voriconazole and a subsequent redevelopment of periostitis would have definitively proven this association; however, reintroduction of this drug was not acceptable to the next of kin.

PATIENT'S COURSE

Voriconazole treatment was discontinued, and the patient was started on micafungin. After 2 weeks, follow-up radiographs showed improvement of the periarticular soft tissue swelling. The serum alkaline phosphatase levels decreased. A peripherally inserted central catheter was inserted for long-term antimicrobial therapy, and the patient was discharged to a nursing home. Clinic follow-up at 2 months showed improvement in joint pain, swelling, and range of motion. At the 1-year follow-up, radiographs showed substantial improvement of the periostitis (Figure 5).

This case emphasizes the need for keeping voriconazole-induced periostitis as one of the important differentials when encountering patients with symptomatic arthralgias, arthritis, and periostitis while receiving voriconazole treatment.

FINAL DIAGNOSIS

Voriconazole-induced periostitis

REFERENCES

1. Hochberg MC, Silman AJ, Smolan JS, Weinblatt M, Weisnar MH, editors. Rheumatology. New York: Elsevier; 2011.
2. Soriano M. Periostitis deformans. Ann Rheum Dis 1952;11:154–61.
3. Kini U, Nandeesh BM. Pathophysiology of bone formation, remodeling, and metabolism. In: Fogelman I, editor. Radionuclide and hybrid bone imaging. Switzerland: Springer Nature; 2012 p. 29–57.
4. Thrall DE. Principles of radiographic interpretation of the appendicular skeleton. In: Thrall DE, editor. Textbook of veterinary diagnostic radiology. 7th ed. St. Louis: Elsevier; 2018. p. 334–47.
5. Uppal S, Diggle CP, Carr IM, Fishwick CW, Ahmed M, Ibrahim GH, et al. Mutations in 15-hydroxyprostaglandin dehydrogenase cause primary hypertrophic osteoarthropathy. Nat Genet 2008;40:789–93.
6. Qian X, Qin J. Hypertrophic pulmonary osteoarthropathy with primary lung cancer. Oncol Lett 2014;7:2079–82.
7. Gladman DD, Shuckett R, Russell ML, Thorne JC, Schachter RK. Psoriatic arthritis (PSA): an analysis of 220 patients. Q J Med 1987;62:127–41.
8. Yang C, Wang Y, Xu H. Treatment and prevention of skeletal fluorosis. Biomed Environ Sci 2017;30:147–9.
9. Gutch M, Sanjay S, Razi SM, Gupta KK. Thyroid acropachy: frequently overlooked finding. Indian J Endocrinol Metab 2014;18:590–1.
10. Fatourechi V, Ahmed DD, Schwartz KM. Thyroid acropachy: report of 40 patients treated at a single institution in a 26-year period. J Clin Endocrinol Metab 2002;87:5435–41.
11. Rothenberg AB, Berdon WE, Woodard JC, Cowles RA. Hypervitaminosis A-induced premature closure of epiphyses (physeal obliteration) in humans and calves (hyena disease): a historical review of the human and veterinary literature. Pediatr Radiol 2007;37:1264–7.
12. Bhadeka A, Bagalore Prakash P, Allareddy V. Prostaglandin E1-induced periostitis and reversibility with discontinuation. J Pediatr 2017;189:237–37e1.
13. Scott LJ, Simpson D. Voriconazole: a review of its use in the management of invasive fungal infections. Drugs 2007;67:269–98.
14. Johnson LB, Kauffman CA. Voriconazole: a new triazole antifungal agent. Clin Infect Dis 2003;36:630–7.
15. Wang TF, Wang T, Altman R, Eshaghian P, Lynch JP III, Ross DJ, et al. Periostitis secondary to prolonged voriconazole therapy in lung transplant recipients. Am J Transplant 2009;9:2845–50.

16. Chen L, Mulligan ME. Medication-induced periostitis in lung transplant patients: periostitis deformans revisited. Skeletal Radiol 2011;40:143–8.
17. Becce F, Malghem J, Lecouvet FE, Vande Berg BC, Omoumi P. Clinical images: voriconazole-induced periostitis deformans. Arthritis Rheum 2012;64:3490.
18. Bucknor MD, Gross AJ, Link TM. Voriconazole-induced periostitis in two post-transplant patients. J Radiol Case Rep 2013;7:10–7.
19. Baird JH, Birnbaum BK, Porter DL, Frey NV. Voriconazole-induced periostitis after allogeneic stem cell transplantation. Am J Hematol 2015;90:574–5.
20. Skaug M, Spak C, Oza U. Painful periostitis in the setting of chronic voriconazole therapy. Proc (Bayl Univ Med Cent) 2014;27: 350–2.
21. Myint TM, Vucak-Dzumhur M, Ebeling PR, Elder GJ. A case report of disabling bone pain after long-term kidney transplantation. Osteoporos Int 2014;25:769–72.
22. Wermers RA, Cooper K, Razonable RR, Deziel PJ, Whitford GM, Kremers WK, et al. Fluoride excess and periostitis in transplant patients receiving long-term voriconazole therapy. Clin Infect Dis 2011;52:604–11.
23. Gerber B, Guggenberger R, Fasler D, Nair G, Manz MG, Stussi G, et al. Reversible skeletal disease and high fluoride serum levels in hematologic patients receiving voriconazole. Blood 2012;120:2390–4.
24. Moon WJ, Scheller EL, Suneja A, Livermore JA, Malani AN, Moudgal V, et al. Plasma fluoride level as a predictor of voriconazole-induced periostitis in patients with skeletal pain. Clin Infect Dis 2014;59:1237–45.
25. Allen KC, Sanchez CJ Jr, Niece KL, Wenke JC, Akers KS. Voriconazole enhances the osteogenic activity of human osteoblasts in vitro through a fluoride-independent mechanism. Antimicrob Agents Chemother 2015;59:7205–13.
26. Pei J, Li B, Gao Y, Wei Y, Zhou L, Yao H, et al. Fluoride decreased osteoclastic bone resorption through the inhibition of NFATc1 gene expression. Environ Toxicol 2014;29:588–95.

Arthritis Care & Research
Vol. 72, No. 5, May 2020, pp 607–614
DOI 10.1002/acr.23894

American College of Rheumatology
Empowering Rheumatology Professionals

REVIEW

Incomplete Systemic Lupus Erythematosus: What Remains After Application of American College of Rheumatology and Systemic Lupus International Collaborating Clinics Criteria?

Wietske M. Lambers, Johanna Westra, Marcel F. Jonkman, Hendrika Bootsma, and Karina de Leeuw

Incomplete systemic lupus (iSLE) is an acknowledged condition of patients with clinical signs of lupus who do not fulfill classification criteria for SLE. Some patients with iSLE have persistent mild disease, but others have serious organ involvement, and up to 55% progress to established SLE. Research on this subject could reveal predictive or diagnostic biomarkers for SLE. Ideally, it would become possible to discern those patients with critical organ involvement or a high risk for progression to SLE. This high-risk group might benefit from early treatment, which would preferably be confirmed in randomized controlled trials. This process would, however, require agreement on a definition of iSLE. The Systemic Lupus International Collaborating Clinics (SLICC) classification criteria was composed in order to diagnose SLE earlier. The present review outlines the clinical characteristics of iSLE after introduction of SLICC criteria and furthermore proposes a definition of iSLE with the aim of discriminating the high-risk group from those with a lower risk.

Introduction

Systemic lupus erythematosus (SLE) is an autoimmune disease that is characterized by formation of antinuclear autoantibodies (ANAs) and is known to have a wide range of clinical features (1). The judgment of experienced physicians is generally accepted as the gold standard for the diagnosis of SLE; however, especially for research aims, accurate disease classification is important to create comparable, consistent study groups. For that purpose, the American College of Rheumatology (ACR) criteria for SLE were proposed (2,3). A patient is classified as having SLE when 4 of 11 cumulative clinical and immunologic ACR criteria are met (Table 1) (3). In order to increase sensitivity, the Systemic Lupus International Collaborating Clinics (SLICC) more recently composed new criteria that were validated in 2012 (4) (Table 1). The most important differences between the ACR 1997 criteria and the SLICC 2012 criteria include the merging of criteria for subacute or acute cutaneous lupus and photosensitivity and addition of alternative forms of chronic cutaneous lupus; the addition of nonscarring alopecia as a clinical criterion; the redefinition of arthritis; the redefinition of the hematologic criteria; the separation and extension of immunologic criteria; the allowance of biopsy-confirmed lupus nephritis in the presence of ANAs or anti–double-stranded DNA (anti-dsDNA) to be sufficient for classification of SLE; and the requirement of at least 1 immunologic and 1 clinical criterion for SLE classification. Currently, new classification criteria for SLE are under review by a ACR/European League Against Rheumatism (EULAR) collaboration (5).

Some patients with lupus symptoms still do not fulfill any of the current classification criteria for SLE. For example, some patients could have cutaneous lupus and detectable autoantibodies but lack other features. Some patients have gradual disease onset and over time develop serious organ involvement, while others continue to have milder manifestations of the disease. Several terms have been used to qualify this heterogeneous group. The term "undifferentiated connective tissue disease" (UCTD) is used if autoimmunity features do not resemble 1 specific autoimmune disease. However, when patients have typical features of SLE without fulfilling the classification criteria, the terms "latent lupus," "early lupus," "potential lupus," "incomplete lupus," and "incomplete SLE" have all been used (6). The terms "latent lupus" and "early lupus" suggest that there will be progression to SLE, while this is not necessarily the case. "Potential lupus" could be an accurate term in reference to some patients,

Supported by the Dutch Arthritis Foundation.
Wietske M. Lambers, MD, Johanna Westra, PhD, Marcel F. Jonkman, MD, PhD, Hendrika Bootsma, MD, PhD, Karina de Leeuw, MD, PhD: University Medical Center Groningen, University of Groningen, Groningen, The Netherlands.
No potential conflicts of interest relevant to this article were reported.

Address correspondence to Wietske M. Lambers, MD, Department of Rheumatology and Clinical Immunology, University Medical Center Groningen, Huispostcode AA21, Hanzeplein 1, 9700 RB Groningen, The Netherlands. E-mail: w.m.lambers@umcg.nl.
Submitted for publication July 17, 2018; accepted in revised form March 26, 2019.

Table 1. Overview of ACR 1997 criteria and SLICC 2012 classification criteria for SLE*

	Clinical criteria	Immunologic criteria
ACR 1997 criteria†	Malar rash Discoid rash Photosensitivity Oral ulcera Arthritis Serositis Renal involvement Neurologic involvement Hematologic manifestations	Anti-dsDNA, anti-Sm, or anti-phospholipid antibodies ANA
SLICC 2012 criteria‡	Acute or subacute cutaneous lupus Chronic cutaneous lupus Oral or nasal ulcera Alopecia Synovitis Serositis Renal involvement Neurologic involvement Hematologic manifestations	ANA Anti-dsDNA Anti-Sm Antiphospholipid antibodies Low complement Positive Coombs' test

* ACR = American College of Rheumatology; SLICC = Systemic Lupus International Collaborating Clinics; SLE = systemic lupus erythematosus; anti-dsDNA = anti–double-stranded DNA; ANA = antinuclear antibodies.
† Patients were classified as having SLE when 4 of the 11 criteria were met.
‡ Patients were classified as having SLE if 4 criteria with at least 1 clinical and 1 immunologic or biopsy-proven lupus nephritis and ANA or anti-dsDNA were met.

but for patients with a clear form of lupus (for example, cutaneous lupus), this label would not be appropriate, as they would already have lupus of the skin. We decided to use the term "incomplete SLE" (iSLE), as this can include both stable mild disease as well as more severe disease that is still not classifiable as SLE.

Investigating iSLE is of significant relevance, as it could reveal immunologic changes that occur when the disease progresses to SLE. Furthermore, longitudinal follow-up of these patients potentially reveals predictive biomarkers that enable stratification of the risk for progression to established SLE. This investigation of iSLE would improve patient care by allowing high-risk patients to be followed up more intensively and to possibly start treatment earlier; low-risk patients could be exempted from intensive follow-up and be reassured about disease progression. Additionally, patients with iSLE might benefit from inclusion in clinical trials. However, because patients with iSLE are probably an even more heterogeneous group than those with SLE, a consensus definition is of great importance for future research.

In the present review, the characteristics of iSLE before and after introduction of SLICC criteria are outlined. Moreover, clinical and serologic risk factors are described in order to ultimately better define this patient group, especially those patients at high risk of progression to SLE or severe organ involvement.

Characteristics of iSLE and risk factors for progression to SLE

Throughout the past decades, a number of studies investigating iSLE have been published. Most publications on iSLE are based on patients with clinical features of SLE who do not fulfill ACR criteria (3), and only some studies use the SLICC criteria. Table 2 shows an overview of these studies (7–20).

The most commonly occurring features in these patients are mucocutaneous symptoms (up to 46%), arthritis (up to 53%), and hematologic disorders (up to 52%). A significant number of patients with iSLE, however, have serious organ involvement. Up to 36% of patients with iSLE have serositis, up to 27% have renal involvement, and up to 6% have neurologic symptoms. Moreover, 53% of hospitalized iSLE patients have been found to have increased disease damage scores, and in 1 cohort, lupus-associated mortality in patients with iSLE was equal to that in those with SLE (15). Progression to SLE occurs in 5–57% of patients with iSLE within 1–10 years.

Seven studies have reported the progression rate to SLE, each of which are discussed herein. In the first study, by Greer and Panush (7), 38 patients with iSLE who were defined as meeting ≥2 but <4 of the ACR criteria (2) were retrospectively compared to 42 patients with SLE. At inclusion, the median disease duration of the iSLE group was 38 months, and the mean time of consecutive follow-up was 19 months. Compared to patients with iSLE, patients with SLE presented more frequently with malar rash, hematologic features, and organ involvement, whereas discoid rash occurred more often in patients with iSLE. Only 5% of patients (n = 2) developed definite SLE (1 patient at 9 months and 1 patient at 26 months after the first presentation); however, the characterisics of these patients were not published by the authors. At the end of follow-up of the remaining patients with iSLE, 11% were not classified as having a connective tissue disease, 26% had discoid lupus erythematosus, 5% subacute cutaneous lupus erythematosus, and 53% continued to have iSLE.

The second study was a prospective study conducted in Puerto Rico (9), in which 87 patients with iSLE were followed up for a mean of 2.2 years. These iSLE patients were defined as having met ≥1 but <4 criteria according to the ACR criteria for SLE (2)

Table 2. Overview of present studies concerning incomplete SLE*

Study author, year (ref.)	Definition of iSLE	Patients with iSLE, no.	Female, %	Mean age, years	Disease duration at baseline, years	Clinical features of patients (%)	Serology of patients (%)	Follow-up	Progression to SLE	Predictors for progression to SLE
Greer and Panush, 1989 (7)	2–3 ACR 1997 criteria	38	87	37	3.2	Arthritis (47) Discoid rash (34) Photosensitivity (24) Hematologic (18) Serositis (16) Neurologic (3)	ANA (82) Anti-dsDNA (0)	1.6 years (mean)	2 patients (5%)	NA
Calvo-Alén et al, 1995 (8)	Clinical diagnosis SLE, <4 ACR 1997 criteria	22	87	47	2.3	Mucocutaneous (45) Serositis (36) Renal (27) Arthritis (27) Lymphopenia (23)	ANA (95) Anti-dsDNA (22)	NA	NA	NA
Vilá et al, 2000 (9)	1–3 ACR 1997 criteria	79	94	30	4.4	Photosensitivity (25) Lymphopenia (23) Malar rash (11) Arthritis (11) Discoid rash (6) Renal (1)	ANA (97) Anti-Ro/SSA (8) Anti-dsDNA (4)	2.2 years (mean)	9% after mean 4.4 years ± 4 years	Younger age (24.5 vs. 34 years) Photosensitivity Anti-dsDNA Low C3
Swaak et al, 2000 (10)	1 organ system and ANA and clinical suspicion of SLE	100	99	40	4.5	Leucopenia (36) Arthritis (15) Renal (11) Malar rash (4) Discoid rash (4) Pericarditis (4)	ANA (100)	3 years	3% after 3 years	Not shown
Ståhl Hallengren et al, 2004 (11)	4 ACR criteria and 1 organ system	28	93	45	NA	Malar rash (25) Arthritis (32) Renal (18) Hematologic (14) Serositis (14) Discoid rash (7)	ANA (100)	13 years (median, range 10–20 years)	57% after median 5.3 years	Malar rash aCL
Laustrup et al, 2010 (12)	Clinical diagnosis SLE, <4 ACR 1997 criteria	26	Not available	Not available	NA	Photosensitivity (46) Arthritis (31) Hematologic (31) Malar rash (19) Serositis (19) Renal (15) Neurologic (4)	ANA (100)	8 years	27%	None identified
Olsen et al, 2012 (13)	2–3 ACR 1997 criteria	22	86	49	Not available	Photosensitivity (27) Arthritis (23) Hematologic (18) Serositis (9)	ANA (95) Anti-Ro/SSA (9) Anti-dsDNA (14) Anti-Sm (5)	10 years	14% after mean 3.8 years	Younger age; high overall IgG-autoreactivity; high IgM anti-Ro/SSA and IgM Anti-LA/SSB
Al Daabil et al, 2014 (14)	1–3 ACR 1997 criteria	264	94	39	Not available	Arthritis (53) Malar rash (14) Renal (2) Discoid rash (1)	ANA (88) Anti-dsDNA (17)	6.3 years	21% after 6.3 years	Oral ulcers Renal involvement Anti-dsDNA
Chen et al, 2015 (15)	1–3 ACR 1997 criteria and clinically most fitting diagnosis SLE	77[3]	87	34	3.6	Hematologic (52) Arthritis (21) Serositis (21) Renal (17) Malar rash (9) Neurologic (6)	ANA (97) Anti-dsDNA (22)	NA	NA	

(Continued)

Table 2. *(Cont'd)*

Study author, year (ref.)	Definition of iSLE	Patients with iSLE, no.	Female, %	Mean age, years	Disease duration at baseline, years	Clinical features of patients (%)	Serology of patients (%)	Follow-up	Progression to SLE	Predictors for progression to SLE
Rúa-Figueroa et al, 2015 (16)	3 ACR 1997 criteria, clinical diagnosis SLE	345	85	42.9	8.0	Arthritis (44) Hematologic (43) Photosensitivity (20) Malar rash (11) Serositis (9) Renal (4) Neurologic (1)	ANA (95)	5.6 years	NA	
Olsen et al, 2016 (17)	1–3 ACR 1997 criteria	70	94	45	Not available	Photosensitivity (23) Arthritis (16) Malar rash (20) Oral ulcers (13) Serositis (9)	ANA (97)	NA	NA	
Bortoluzzi et al, 2017 (20)	UCTD (ANA + ≥1 autoimmune symptom)	329	97	46	Not available	Malar rash (29) Leukopenia (13) Synovitis (11) Alopecia (11) Ulcers (10) Neurologic (4)	ANA (100) Low complement (14) aPL (14) Anti-dsDNA (2)	5–10 years	44 of 329 (14%) fulfilled SLICC criteria at baseline; 23 of 329 (7%) developed SLE according to SLICC	Not available
Aberle et al, 2017 (18)	Clinical diagnosis SLE, not fulfilling ACR 1997 or SLICC 2012 criteria	291	87	48	Not available	Arthritis (45) ACLE or SCLE (43) Leuko-/lymphopenia (23) Ulcers (11) CDLE (9) Serositis (6) Renal (5) Neurologic (1)	ANA (96) aPL (13) Anti-dsDNA (12)	NA	NA	
Yusof et al, 2018 (19)	ANA plus ≥1 SLICC 2012 criterion	118	88	48	Not available	Synovitis (26) ACLE or SCLE (26) Leuko-/lymphopenia (16) Ulcers (11)	ANA (100) Anti-Ro/SSA (42) Anti-dsDNA (36)	1 year	12%	1 clinical SLICC criterion at baseline; positive family history of autoimmune rheumatic disease

* SLE = systemic lupus erythematosus; ref. = reference; iSLE = incomplete systemic lupus erythematosus; ACR = American College of Rheumatology; ANA = antinuclear antibody; anti–dsDNA = anti–double-stranded DNA; NA = not applicable; aCL = anticardiolipin antibody; UCTD = undifferentiated connective tissue disease; aPL = antiphospholipid antibody; SLICC = Systemic Lupus International Collaborating Clinics; ACLE = acute cutaneous lupus erythematosus; SCLE = subacute cutaneous lupus erythematosus; CDLE = chronic discoid lupus erythematosus.

and had no classification or specific symptoms of other rheumatic diseases. Evolution of iSLE to SLE occurred in 9% of these patients, with a mean interval of 4.4 years between onset of symptoms and diagnosis. These patients in whom iSLE progressed to SLE were younger than those who remained in the iSLE group (ages 24.5 years versus 34 years), but this difference did not reach statistical significance ($P = 0.06$). In terms of clinical manifestations, the patients who developed SLE more frequently had photosensitivity, positive anti-dsDNA, and decreased C3 levels at baseline. Importantly, organ involvement was uncommon in both groups.

The third study was a multicenter study that prospectively evaluated patients with iSLE (10). For these patients, iSLE was defined as the presence of symptoms of 1 organ system, ANA positivity, and clinical suspicion of possibly developing SLE in the future. Although 122 patients were identified using this definition of iSLE, 22 already fulfilled the 1982 ACR criteria of SLE at first evaluation. Of the remaining 100 patients, 3 developed SLE during the next 2 years. Clinical symptoms consisted mainly of fatigue, arthritis, nonhemolytic anemia, and leucopenia, while organ involvement was uncommon. These findings suggest that patients with iSLE whose illness does not progress to SLE during a short term represent a milder disease entity. Unfortunately, no comparison of baseline characteristics was made between the patients who developed SLE and the remaining iSLE group.

The fourth study, by Ståhl Hallengren et al (11), included long-term prospective follow-up of 28 patients with iSLE, which was defined as ANA positivity and symptoms in ≥1 organ. After a median duration of 5.3 years, iSLE in 16 patients (57%) had progressed to SLE according to ACR criteria (2). The iSLE patients whose illness progressed to SLE were all ANA positive (as per protocol for the study), and all but 1 patient had at least 1 clinical symptom at baseline. The progressive patient who did not display clinical sympoms at baseline had a first-degree family member with SLE. All 6 of the patients who had malar rash and all 6 patients who had anticardiolipin antibodies developed established SLE.

In the fifth study, Laustrup et al (12) investigated a cohort of 26 patients with iSLE (clinical diagnosis of SLE, not meeting ACR criteria) (2). All patients had detectable ANA, and the most prevalent clinical symptoms were photosensitivity, malar rash, and hematologic disorders. In 7 of these patients (27%), iSLE transformed into SLE during 8 years of follow-up. No predictive factors could be identified.

In the penultimate study, Al Daabil et al prospectively enrolled 264 patients who fulfilled 1–3 of the ACR classification criteria for SLE (14). Throughout an average follow-up time of 6.3 years, iSLE in 21% of patients evolved to SLE. At baseline, arthritis and presence of anti-dsDNA and anti-Ro/SSA were more frequent in the group that had eventually progressed to SLE. However, after multivariable logistic regression analysis, only oral ulcers, anti-dsDNA, and symptoms of renal involvement were found to be independent risk factors for the development of SLE. During follow-up of the group that did not develop SLE, 61% remained classified as having iSLE, while 18% had another diagnosis (fibromyalgia, autoimmune thyroid disease, mixed connective tissue disease, rheumatoid arthritis, or cutaneous lupus). Importantly, the ANA positivity rate was lower (79%) in this group compared to the remaining SLE group (98%).

Finally, a recent prospective observational study by Yusof et al (19) included 118 subjects with ANA positivity (titer ≥1:80) who fulfilled ≥1 clinical SLICC criteria and had symptom duration of <12 months. Clinical symptoms included mucocutenaous, musculoskeletal, and hematologic features. During the 12 months of follow-up, 19 patients (16%) progressed to a classified connective tissue disease, of which 14 (12%) developed SLE according to SLICC criteria and 5 (4%) developed Sjögren's syndrome. Two patients developed critical organ involment, 1 with serositis and 1 with nephritis. All iSLE patients whose illness progressed to SLE had fulfilled at least 1 clinical SLICC criterion at baseline (not further specified by the authors), indicating that this was an important risk factor of disease progression. Furthermore, after logistic regression analysis, a positive family history of autoimmune rheumatic diseases was associated with a high risk of disease progression. Notably, the authors showed that interferon activity was strongly associated with progression to established SLE.

The above studies are extremely valuable in the underscoring of the variable nature of iSLE. Importantly, most of the aforementioned studies are retrospective in nature, which may result in underestimation of the progression rate. Logic would suggest that patients with a prolonged disease course are more likely to be included than those who quickly progressed to SLE. This is further supported by the fact that the 2 prospective studies (11,19) demonstrated the highest percentage of iSLE patients whose illness progressed to SLE, i.e., 12% of patients after 1 year and 57% after a median of 5.3 years. In summary, clinical symptoms and disease severity are highly variable among patients with iSLE, ranging from persistent mild disease to rapid progression to SLE and/or to critical organ involvement. In regard to clinical features, acute cutaneous lupus erythematosus, photosensitivity, serositis, ulcers, and renal involvement seem to occur more often in patients with iSLE whose illness progresses to SLE. These patients in whom iSLE progressed to SLE were younger. Furthermore, the presence of anti-dsDNA, anticardiolipin antibodies, and hypocomplementemia are all associated with progression to SLE. None of these findings, however, can accurately predict the establishment of SLE.

Consequences for iSLE classification after introduction of SLICC criteria

After the introduction of the SLICC criteria, various researchers have focused on the consequences for classification of clinically diagnosed lupus patients. A recent systematic review and meta-analysis (21) showed that for adult patients with SLE, SLICC cri-

teria increased sensitivity compared to ACR criteria (3) (94.6% versus 89.6%), while specificity decreased only slightly (95.5% versus 98.1%). Unfortunately, most studies on iSLE have noted the number of ACR criteria (3) but have not published individual patient characteristics. Therefore, for the purpose of the present review, retrospective evaluation of the consequences of applying SLICC criteria in these patient cohorts could not be performed.

Four additional studies have applied both ACR and SLICC criteria for the evalution of iSLE. In an observational study, Chen et al (15) included 77 hospitalized iSLE patients (iSLE being defined as fulfilling <4 ACR criteria [3]) in order to analyze the organ damage features of this group. The mean disease duration was 43 months, and the mean Systemic Lupus Erythematosus Disease Activity Index score was 6.6. When the authors applied SLICC criteria in this cohort, 43 patients (56%) who did not meet ACR criteria (3) were classified as having SLE. More than half of the patients (53%) had increased SLICC/ACR Damage Index scores, mostly because of pulmonary arterial hypertension, and renal and neurologic damage. Seventeen of the 41 patients (41%) with increased damage scores did not meet any of the criteria sets.

In the study by Olsen et al (17) (Table 2), none of the identified 70 patients with iSLE (which was defined as fulfilling 1–3 of the ACR criteria for SLE [3]) fulfilled the SLICC criteria for SLE classification. The authors concluded that classification using the SLICC criteria would not change the prevalence of the incomplete lupus designation.

Aberle et al (18) reviewed the medical records of 3,397 patients with a clinical diagnosis of SLE and applied both ACR criteria (3) and SLICC criteria (4) in all patients. They identified 440 subjects who only met 3 ACR criteria (3). One-third of these patients met SLICC criteria (4), resulting in 291 patients (9% of all patients with a clinical SLE diagnosis) who could not be classified by any of the criteria sets. A large proportion of these nonclassifiable patients had organ involvement (6% serositis, 5% renal involvement, and 1% neurologic features) (Table 2). The majority of these patients were being treated with hydroxychloroquine and/or steroids, and 10% required other immunosuppressive drugs.

Bortoluzzi et al (20) retrospectively selected 329 white patients with UCTD (defined as having signs and symptoms suggestive of a connective tissue disease), ANA positivity, and a disease duration of at least 1 year who did not fulfill ACR criteria (3). In retrospect, 44 patients (13%) already fulfilled the SLICC criteria (4) at baseline. The most commonly occurring clinical features in this group were acute cutaneous lupus erythematosus (55%), leukopenia (39%), and synovitis (30%). Regarding critical organ involvement, 7 patients (16%) had neurologic involvement, and 2 (5%) had serositis, while none had renal involvement. Of the remaining 285 patients with UCTD, information regarding 206 could be retrieved from 5 to 10 years follow-up. During this period, 14 patients with UCTD (5%) progressed to SLE according to ACR criteria (3), and 23 patients (8%) according to SLICC criteria (4). Unfortunately, the authors did not show the disease characteristics of these groups.

In summary, more patients were classified as fulfilling SLICC criteria than ACR criteria for SLE (3), but still ~5% of the patients with a clinical diagnosis remained unclassified, as can be expected based on sensitivity. More importantly, a considerable share of these patients had serious organ involvement or required treatment with immunosuppressive drugs and thus might benefit from inclusion in clinical trials.

Requirement of consensus definition for iSLE

Currently, researchers use various definitions for iSLE, which hinders comparability between different studies. Ideally, a classification system would include patients who are at the highest risk of developing SLE or serious organ damage and exclude those who have prolonged mild symptoms or develop other autoimmune diseases. Prospective documentation of a consistent group of patients with iSLE is required in order to better define the high-risk group and to determine predictive biomarkers. We therefore ask for the development of a consensus on the definition of iSLE in order to, ideally, combine forces and start prospective documentation of patients with iSLE.

Definition of iSLE involves a very heterogeneous group of patients and should include patients at the highest risk of developing SLE. Mucocutaneous symptoms, serositis, renal symptoms, anticardiolipin antibodies, low complement, and anti-dsDNA are all associated with progression to SLE. Table 3 shows our proposed

Table 3. Proposed definition of incomplete systemic lupus erythematosus (iSLE)*

Required
ANA at a titer ≥1:80
And ≥1 of the following criteria†
Acute or subacute cutaneous lupus
Chronic cutaneous lupus
Oral or nasal ulcers
Alopecia
Synovitis
Serositis
Neurologic manifestation
Renal manifestation
Or 2 of the following criteria
Hematologic manifestations‡
Immunologic features§
Positive family history of autoimmune rheumatic disease¶
And not meeting ACR 1997 criteria and/or SLICC 2012 criteria for SLE

* ANA = antinuclear antibody; ACR = American College of Rheumatology; SLICC = Systemic Lupus International Collaborating Clinics; SLE = systemic lupus erythematosus.
† As specified in SLICC classification criteria.
‡ As specified in SLICC classification criteria. Hematologic manifestation included hemolytic anemia or leukopenia or lymphopenia (1,000/mm^3 at least once) or thrombocytopenia (100,000/mm^3 at least once).
§ As specified in SLICC classification criteria. Immunologic features included anti–double-stranded DNA or anti-Sm or antiphospholipid antibodies or low complement or direct Coombs' test.
¶ Included first- or second-degree relative with autoimmune rheumatic disease.

definition of iSLE, which aims at including patients with a high risk of developing SLE or serious organ involvement. ANA positivity at a titer ≥1:80 should be present in order to be classified as iSLE, as this is a key feature of SLE. A recent systematic review and meta-regression (22) on the diagnostic value of ANAs reported 97.8% sensitivity and 74.7% specificity for ANA at a titer ≥1:80. Also, in an observational study (23) on 616 patients who were referred due to possible SLE, 99.5% of patients with early SLE were ANA positive. The ACR/EULAR international collaboration on development of new classification criteria for SLE has also reached consensus on using positive ANA at a titer ≥1:80 as entry criterion 5.

Furthermore, a definition of iSLE should include at least 1 clinical symptom. The study by Ståhl Hallengren et al included patients fullfilling at least 1 clinical criterion, and this group had the highest disease progression rate compared to other longitudinal studies on iSLE (11). Moreover, all patients whose illness progressed to iSLE who were included in the study by Yusof et al (19) had fulfilled a clinical criterion. We propose the usage of the clinical criteria as recorded in the SLICC criteria (see Table 1), as these criteria have been demonstrated to be more sensitive than ACR criteria (21).

In the absense of other clinical symptoms, hematologic features have been shown not to be very specific for SLE (23). Therefore, hematologic features should be accompanied by other immunologic features in order to classify iSLE. Having a first- or second-degree relative with an autoimmune disease has also been identified as a risk factor for developing SLE and should therefore be taken into account in the definition of iSLE. Although there is not much literature on this subject, we weighted this factor similarly to an immunologic or hematologic feature. Based on our review, we expect to distinguish a patient group at high risk of progressive disease by using this definition of iSLE.

In summary, there is still a need for better recognition of patients with iSLE, especially those with a high-risk profile for progression to SLE and/or development of organ damage. In the present review, an overview of the current literature was presented in order to clarify the characteristics of high-risk patients. Prospective documentation of a consistent group of patients with iSLE is necessary in order to define the high-risk group and to determine predictive biomarkers. Therefore, it is necessary to reach a widely accepted consensus on a definition for lupus patients who do not fulfill the classification criteria for 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.

REFERENCES

1. Bertsias GK, Pamfil C, Fanouriakis A, Boumpas DT. Diagnostic criteria for systemic lupus erythematosus: has the time come? Nat Rev Rheumatol 2013;9:687–94.
2. Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1982;25:1271–7.
3. Hochberg MC, 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.
4. Petri M, Orbai AM, Alarcón GS, Gordon C, Merrill JT, Fortin PR, et al. Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum 2012;64:2677–86.
5. 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.
6. Costenbader KH, Schur PH. We need better classification and terminology for "people at high risk of or in the process of developing lupus". Arthritis Care Res (Hoboken) 2015;67:593–6.
7. Greer JM, Panush RS. Incomplete lupus erythematosus. Arch Intern Med 1989;149:2473–6.
8. Calvo-Alén J, Bastian HM, Straaton KV, Burgard SL, Mikhail IS, Alarcón GS. Identification of patient subsets among those presumptively diagnosed with, referred, and/or followed up for systemic lupus erythematosus at a large tertiary care center. Arthritis Rheum 1995;38:1475–84.
9. Vilá LM, Mayor AM, Valentin AH, Garcia-Soberal M, Vila S. Clinical outcome and predictors of disease evolution in patients with incomplete lupus erythematosus. Lupus 2000;9:110–5.
10. Swaak AJ, van de Brink H, Smeenk RJ, Manger K, Kalden RJ, Tosi S, et al. Incomplete lupus erythematosus: results of a multicentre study under the supervision of the EULAR Standing Committee on International Clinical Studies Including Therapeutic Trials (ESCISIT). Rheumatology (Oxford) 2001;40:89–94.
11. Ståhl-Hallengren C, Nived O, Sturfelt G. Outcome of incomplete systemic lupus erythematosus after 10 years. Lupus 2004;13:85–8.
12. Laustrup H, Voss A, Green A, Junker P. SLE disease patterns in a Danish population-based lupus cohort: an 8-year prospective study. Lupus 2010;19:239–46.
13. Olsen NJ, Li QZ, Quan J, Wang L, Mutwally A, Karp DR. Autoantibody profiling to follow evolution of lupus syndromes. Arthritis Res Ther 2012;14:R174.
14. Al Daabil M, Massarotti EM, Fine A, Tsao H, Ho P, Schur PH, et al. Development of SLE among "potential SLE" patients seen in consultation: long-term follow-up. Int J Clin Pract 2014;68:1508–13.
15. Chen Z, Li MT, Xu D, Leng XM, Wang Q, Tian XP, et al. Organ damage in patients with incomplete lupus syndromes: from a chinese academic center. Clin Rheumatol 2015;34:1383–9.
16. Rúa-Figueroa I, Richi P, López-Longo FJ, Galindo M, Calvo-Alén J, Olivé-Marqués A, et al. Comprehensive description of clinical characteristics of a large systemic lupus erythematosus cohort from the Spanish Rheumatology Society Lupus Registry (RELESSER) with emphasis on complete versus incomplete lupus differences. Medicine (Baltimore) 2015;94:e267.
17. Olsen NJ, McAloose C, Carter J, Han BK, Raman I, Li QZ, et al. Clinical and immunologic profiles in incomplete lupus erythematosus and improvement with hydroxychloroquine treatment. Autoimmune Dis 2016;2016:8791629.
18. Aberle T, Bourn RL, Munroe ME, Chen H, Roberts VC, Guthridge JM, et al. Clinical and serologic features in patients with incomplete lupus classification versus systemic lupus erythematosus patients and controls. Arthritis Care Res (Hoboken) 2017;69:1780–8.

19. Md Yusof MY, Psarras A, El-Sherbiny YM, Hensor EM, Dutton K, Ul-Hassan S, et al. Prediction of autoimmune connective tissue disease in an at-risk cohort: prognostic value of a novel two-score system for interferon status. Ann Rheum Dis 2018;77:1432–9.
20. Bortoluzzi A, Furini F, Campanaro F, Govoni M. Application of SLICC classification criteria in undifferentiated connective tissue disease and evolution in systemic lupus erythematosus: analysis of a large monocentric cohort with a long-term follow-up. Lupus 2017;26:616–22.
21. Hartman EA, van Royen-Kerkhof A, Jacobs JW, Welsing PM, Fritsch-Stork RD. Performance of the 2012 Systemic Lupus International Collaborating Clinics classification criteria versus the 1997 American College of Rheumatology classification criteria in adult and juvenile systemic lupus erythematosus. A systematic review and meta-analysis. Autoimmun Rev 2018;17:316–22.
22. Leuchten N, Hoyer A, Brinks R, Schoels M, Schneider M, Smolen J, et al. Performance of antinuclear antibodies for classifying systemic lupus erythematosus: a systematic literature review and meta-regression of diagnostic data. Arthritis Care Res (Hoboken) 2018;70:428–38.
23. Mosca M, Costenbader KH, Johnson SR, Lorenzoni V, Sebastiani GD, Hoyer BF, et al. How do patients with newly diagnosed systemic lupus erythematosus present? A multicenter cohort of early systemic lupus erythematosus to inform the development of new classification criteria. Arthritis Rheumatol 2019;71:91–8.

Arthritis Care & Research
Vol. 72, No. 5, May 2020, pp 615–621
DOI 10.1002/acr.23881

American College of Rheumatology
Empowering Rheumatology Professionals

REVIEW

Application of Traditional and Emerging Methods for the Joint Analysis of Repeated Measurements With Time-to-Event Outcomes in Rheumatology

Liubov Arbeeva,[1] Amanda E. Nelson,[1] Carolina Alvarez,[1] Rebecca J. Cleveland,[1] Kelli D. Allen,[2] Yvonne M. Golightly,[1] Joanne M. Jordan,[1] Leigh F. Callahan,[1] and Todd A. Schwartz[1]

Objective. The goal of this paper is to describe approaches for the joint analysis of repeatedly measured data with time-to-event end points, first separately and then in the framework of a single comprehensive model, emphasizing the efficiency of the latter approach. Data from the Johnston County Osteoarthritis (JoCo OA) Project will be used as an example to investigate the relationship between the change in repeatedly measured body mass index (BMI) and the time-to-event end point of incident worsening of radiographic knee OA that was defined as an increased Kellgren/Lawrence grade in at least 1 knee over time.

Methods. First, we provide an overview of the methods for analyzing repeated measurements and time-to-event end points separately. Then, we describe traditional (Cox proportional hazards model [CoxPH]) and emerging (joint model [JM]) approaches, both of which allow combined analysis of repeated measures with a time-to-event end point in the framework of a single statistical model. Finally, we apply the models to JoCo OA data and interpret and compare the results from the different approaches.

Results. Applications of the JM (but not the CoxPH) showed that the risk of worsening radiographic OA is higher when BMI is higher or increasing, thus illustrating the advantages of the JM for analyzing such dynamic measures in a longitudinal study.

Conclusion. Joint models are preferable for simultaneous analyses of repeated measurement and time-to-event outcomes, particularly in the context of chronic disease, where dependency between the time-to-event end point and the longitudinal trajectory of repeated measurements is inherent.

Introduction

Longitudinal studies in which data are collected on participants over years or even decades have become increasingly popular in many epidemiologic fields. Such studies enable the analysis of individual-level changes, represented by repeatedly measured variables, and relate the changing patterns to the development of conditions or diseases causing disability and death. Despite the advantages of having multiple time points, there are several challenges associated with longitudinal data analysis, including non-ignorable missing data and sparse examination times (1–4).

In addition, to monitor risk factors and health outcomes, these studies collect repeated measurements that can encompass different types of variables. Two of these, longitudinally measured variables (e.g., biomarkers, patient-reported outcome measures) and the time to occurrence of an event (e.g., joint replacement, death), are very common in epidemiologic studies. These 2 types of data are often analyzed separately, without considering that longitudinal and survival processes are related (5). However, in a chronic disease context, dependency between time-to-event outcomes and longitudinal trajectories is inherent. The essential characteristic of such chronic conditions is that the course of

Supported by the Centers for Disease Control and Prevention (grants S043, S1734, S3486, U01DP003206, and U01DP006266) and the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases grants P60-AR-30701, P60-AR-049465, and P60-AR-064166).

[1]Liubov Arbeeva, MS, Amanda E. Nelson, MD, MSCR, Carolina Alvarez, MS, Rebecca J. Cleveland, PhD, Yvonne M. Golightly, PT, PhD, Joanne M. Jordan, MD, MPH, Leigh F. Callahan, PhD, Todd A. Schwartz, DrPH: University of North Carolina, Chapel Hill; [2]Kelli D. Allen, PhD: University of North Carolina, Chapel Hill, and Durham VA Medical Center, Durham, North Carolina.

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

Address correspondence to Todd A. Schwartz, DrPH, Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, 3106E McGavran-Greenberg Hall, CB number 7420, Chapel Hill, NC 27599. E-mail: Todd_Schwartz@unc.edu.

Submitted for publication November 30, 2018; accepted in revised form March 19, 2019.

disease is different from one person to another and can change over time for the same person. The repeatedly measured variables can help in understanding the nature of disease progression and provide better estimation of the risk for the event of interest (e.g., death, development or progression of disease, or hospital discharge after surgery).

Investigation of such longitudinal relationships between repeatedly measured variables and the event of interest can provide clinically relevant information about the likely course of disease in a given person. For example, to optimize treatment strategies in early rheumatoid arthritis (RA), it is important to understand the relationship between disease activity over time, represented by longitudinal Disease Activity Score in 28 joints (DAS28) measurements and time to subsequent radiographic joint damage. To evaluate the impact of the longitudinal response trajectory on the time-to-event outcome of interest over time, the data should be analyzed jointly (6). This is because neither the changes in evolution of longitudinal response (e.g., DAS28) nor the risk for event (e.g., radiographic progression of RA) are observable continuously over time, only intermittently during clinic visits. Such analyses require statistical methodology that can relate these unobservable values both to each other and to observable data. However, epidemiologic analyses for various chronic diseases, including rheumatic and musculoskeletal disease (RMD), which might benefit from jointly using these data, often do not utilize this approach. Our didactic study on the use of joint models in rheumatology is therefore designed to provide an example of this methodology in a field where these models are not yet commonly used, despite their appropriateness.

The main goals of this study are to describe mainstream statistical approaches for the analysis of such data, to convince the reader of the advantages of joint analysis of longitudinal measures with time-to-event outcomes, and to demonstrate how to apply these methods in a real and relevant data set using data from the Johnston County Osteoarthritis (JoCo OA) Project. First, we review the methods for analyzing time-to-event and repeated measurements outcomes separately. Then we describe traditional (the Cox proportional hazards model [CoxPH]) and emerging (joint model [JM]) approaches, both of which allow combined analysis of repeated measures with time-to-event outcomes in the framework of a single, comprehensive statistical model. Finally, we apply the models to the JoCo OA data and then interpret and compare the results from the different approaches.

Overview of statistical models

Repeated measurements and linear mixed effects (LME) model. The LME model is a commonly used approach for analysis of repeated measurements (7). The term "mixed" refers to the fact that both fixed and random effects are included in the same model, where fixed effects relate to the mean cohort trajectory, and random effects are individual-specific characteristics that take into account the variability in individual trajectories within a cohort. The LME model is valid under the assumption that the data are missing at random (8), which means that missing data can depend on some baseline characteristics and nonmissing observations of the outcome at previous visits. Since this might be assumed in many situations where missingness is not associated with the outcome of interest, the use of the LME model is justified in such applications. However, when attrition (e.g., due to death, worsening of symptoms, or ineligibility of the participant) depends on missing data (e.g., when it is considered "informative"), the data are called "missing not at random." In this situation, as the data collection is discontinued for such a person, leading to informative missingness, modeling the evolution of the longitudinal response using the LME model may produce biased estimates (2,3,9).

Analysis of time-to-event data. When the main outcome under assessment is the length of the interval from the time origin until the occurrence of the event of interest (e.g., survival time until death), an appropriate methodology is required, as these data have unique properties that cannot be addressed with standard statistical procedures. First, methods based on the normal distribution are not applicable for the analysis of survival times because they tend to be positively skewed, leading to violation of the normality assumption. Second and even more important, it is common that at the end of a study the actual survival times will often be censored, i.e., they are not observed for all individuals. The most common type of censoring, and the focus in this study, is right censoring that occurs when a participant does not experience the event of interest by the end of his or her study follow-up.

The CoxPH (10) is one of the most commonly used approaches for analysis of time-to-event outcomes. In this model, the measure of interest is a hazard, which is the instantaneous risk of the event given that a person has not experienced this event up to a specific time. CoxPH allows the analysis of the effect of 1 or more explanatory variables that may impact the hazard. Predictors that do not change over time are called "time-independent variables" and are among the most commonly analyzed in CoxPH. These predictors often include baseline measurements (e.g., exposure variables, risk factors, covariates, and/or confounders) or variables that do not change over time (e.g., sex, race, or a birth cohort).

The CoxPH model can also be extended to incorporate important explanatory variables that do change over the follow-up time period (11,12). This extension, CoxPH with time-varying covariates (TVCs), offers the opportunity to analyze data collected at different follow-up times for the same individual. In this model, a TVC is assumed to remain constant between 2 observations. Therefore, this model is appropriate for variables that either change in a known way (e.g., age, the dose of an administered drug) or exist independently of individuals (e.g., air pollution levels) (13). These covariates can be associated with the risk for the event but are independent of an individual's time-to-event outcome.

However, the CoxPH with TVCs has limited ability to handle explanatory variables with fluctuation and measurement errors (14). As an illustration, consider how longitudinal measurements of body mass index (BMI) are handled in the CoxPH with TVCs. To obtain the value of BMI, both weight and height need to be measured. Although height is largely stable in the adult population, weight is subject to daily, weekly, and seasonal variability due to fluid balance, food consumption, and other factors such as physical exertion and external temperature (15,16). Measurement errors due to clothing and the calibration of the scale can lead to additional small fluctuations in weight and height. The intervals between visits might be intermittent, spanning a few weeks up to several years. In the CoxPH with TVCs, a value of BMI, observed and recorded only at a specific time, is assumed to remain constant between 2 visits and may be associated with the risk of the event at future time points until the next visit, as shown graphically in Figure 1. The blue dotted line corresponds to the approximation of BMI trajectory in the CoxPH model, which is not a realistic description of the BMI evolution. Application of the CoxPH model to internal TVCs that can be collected only when the individual is available (such as variables measured with errors and not fully observed) can lead to biased results and incorrect inference (14,17). Two approaches, developed in parallel in different scientific areas and for different purposes, can capture the biologic fluctuations and heterogeneity in longitudinal trajectories: stochastic process models (18–20) and JMs (21). In the next sections, we focus on the latter approach, which is more mainstream in biostatistics and thus may be easier to use for those familiar with the LME model and CoxPH. Review and discussion of stochastic process models can be found elsewhere (19).

JM. A JM consists of 2 submodels representing the dynamics of the longitudinal submodel and the time-to-event submodel, as reviewed elsewhere (6,21–23). The fundamental difference between the JM and the CoxPH with TVCs is that, unlike CoxPH, the JM combines the time-to-event model with an appropriate model for the repeated measurements of TVCs to simultaneously make inference on time-to-event and longitudinal processes. The JM technique is more appropriate than the CoxPH with TVCs if there is interest in the effect of a TVC measured with error on the survival process. This is because the CoxPH with TVCs can severely underestimate the association between longitudinal and time-to-event data (17). In a standard specification of the JM, at each time point the risk of event is associated with an unobserved value of the longitudinal outcome at the same time (Figure 1) . These are usually called the true values, as opposed to the observed longitudinal data (collected intermittently and potentially with errors), which are represented in the longitudinal submodel as a sum of such unobserved true values and errors (usually modeled using the LME model). The flexibility in parameterization of the JM (e.g., through various extensions available in the R package JM [24]) allows incorporating not only the current value but also dynamic characteristics of the longitudinal response in the model, e.g., the rate of change, cumulative history, or deviations from population trajectories (5). The survival process can depend on the current slope of the longitudinal trajectory (Figure 1) to capture the situation where 2 individuals have comparable levels of a biomarker but the rate of change is different and affects the risk of an event. We use the term "current" for both slope and value, meaning that the risk for an event at a particular time depends on the concurrent unobserved value of longitudinal outcome as well as the concurrent value of the slope of the true longitudinal trajectory. Alternatively, cumulative effects parameterization allows the entire history of a longitudinal response to be associated with the hazard of event.

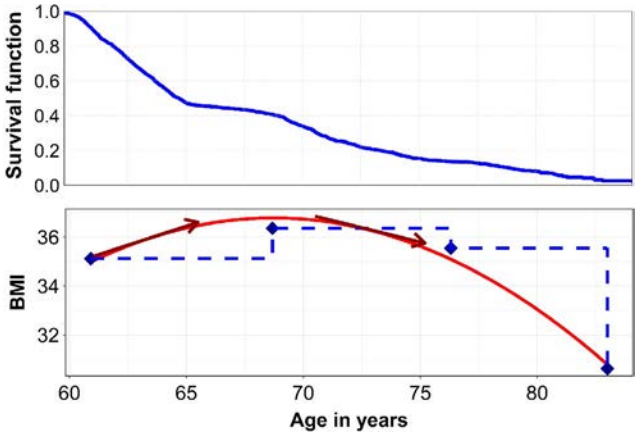

Figure 1. Graphic representation of the features of the joint model (JM). The blue solid line shows the survival function. The diamonds represent individual body mass index (BMI) measurements observed at the baseline and 3 follow-up visits. The blue dotted line corresponds to the approximation of BMI trajectory in the Cox proportional hazards model. A value of BMI, observed and recorded only at a specific time, is assumed to remain constant between 2 visits and may be associated with the risk for event until the next visit. The red line represents the approximation of the longitudinal trajectory in the JM, where the risk for an event is associated with the level of BMI and its change (arrows) at each time point.

Although JMs are becoming increasingly popular in different epidemiologic fields such as oncology (25,26), cardiovascular diseases (27), nephrology (28), and endocrinology (29), these models are still not widely applied in rheumatology. Recently, the JM was applied to a sample of seropositive arthralgia patients to investigate whether a change in individual levels of anti–citrullinated protein antibodies (ACPAs) over time improves the prediction of future RA (30). Higher time-dependent ACPA levels were found to be significantly associated with the development of arthritis, but no difference over baseline measurements of ACPA levels was shown in predictive models.

In our working example, we use repeatedly measured BMI, which is a useful indicator of obesity, to investigate the effect of the longitudinal trajectory of BMI on the time-to-event outcome of worsening Kellgren/Lawrence (K/L) grade in the knee. We chose

this relationship given that 1) obesity is 1 of the most important knee OA risk factors (31), 2) BMI is a good example of a biomarker potentially measured with error, 3) BMI is a potentially modifiable risk factor, and 4) the rate and direction of change in BMI may be as important as the value itself. Various parameterizations of the JM can address known and previously discussed challenges in studying change in BMI and its effect on OA outcomes (32). We emphasize that most (if not all) of the challenges in dealing with repeatedly measured BMI and its change can be applied to other relevant variables in studying their impact on OA and other RMDs.

Working example

Data and measurements. The data used in this study were collected from non-Hispanic African American and Caucasian men and women enrolled in the JoCo OA, which is an ongoing, longitudinal population-based prospective study with clearly defined and repeatedly measured radiographic OA, comorbidities, various biomarkers, sociodemographic, and physiological variables (33). JoCo OA was designed to determine risk factors associated with the occurrence and progression of OA. Our final sample comprised 2,286 participants with 5,325 longitudinal measurements of BMI collected at 4 time points: baseline and 3 follow-up visits (see Supplementary Appendix A, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23881/abstract, for the details on the selection procedure and for the baseline characteristics of this cohort). The time-to-event outcome, worsening radiographic OA of the knee, was defined as an increase of 1 K/L grade or more from any baseline K/L score in at least 1 knee between 2 consecutive or intermittent visits. It is important to note that we include here a working example with a simplified analysis for brevity. In practice, other relevant covariates might be included in the analysis as appropriate.

Statistical analysis. The counting process form of the CoxPH model (12,13) was used to evaluate the association of 2 TVCs, BMI and its change, with worsening knee radiographic OA with adjustment for baseline age and sex. The change in BMI was defined as the percent change in BMI relative to a participant's measurement at the previous visit. We used hazard ratios (HRs) as measures of these associations, and 95% confidence intervals (95% CIs) were used to express the variation around the HRs.

We fitted several JMs using the R package JM (24). A full mathematical description of the models, variables transformation, and interpretation of coefficients are provided in Supplementary Appendix A (available at http://onlinelibrary.wiley.com/doi/10.1002/acr.23881/abstract). In short, the first (basic) model (JM1) consists of the LME model for longitudinal BMI data with normally distributed errors and a survival submodel that specifies the hazard of event as a function of the true longitudinal outcome (see definition in Joint Model section) with adjustment for sex and age at the baseline. In the second joint model (JM2), the risk depends on the slope of the true trajectory at that time. In the third model (JM3), we assumed that the risk depends on both the current true level and the slope of BMI at the same time. JM3 allows us to capture the situations where participants have similar levels of BMI but different rates of change, with this difference affecting the risk of an event. Longitudinal BMI values were logarithmically transformed to satisfy assumptions of normality. In this case, a 1-unit increase of current level of TVC, which is log (BMI) now, corresponds to a 2.7-fold (a mathematical constant, the base of the natural logarithm) difference in BMI. Therefore, the HR quantifies how many times higher the risk of event is for the same participant if his or her BMI at the same time would be 2.7 times higher. To convert it to more interpretable terms, we calculated the HR for a difference of 25% in BMI at the same time for the same participant as follows: as a 25% difference (e.g., 1.25 fold) in BMI level corresponds to log(1.25) = 0.22 in log (BMI), the HR for BMI (Table 1) was calculated relative to 0.22 units of difference in the current level of log (BMI) by taking the exponent of the corresponding coefficient multiplied by 0.22. For the longitudinal change of the BMI, we calculated the HR for BMI slope (Table 1) that compares an increase of 10% over time to an increase of 5% following the procedure previously described (28). The 3 JMs were compared using the Bayesian information criterion (BIC) (34) to select the model with the best fit.

Table 1. Three joint models (JMs) for longitudinal body mass index (BMI) and/or longitudinal change in BMI with risk for incident-worsening radiographic osteoarthritis of the knee fitted to data from the Johnston County Osteoarthritis Project: examples of clinical interpretation*

	HR for BMI†	HR for BMI slope‡
JM1§	1.39 (1.31–1.48)	NA
JM2¶	NA	4.59 (2.14–9.86)
JM3#	1.37 (1.29–1.46)	2.29 (1.20–4.36)

* Values are HR (95% confidence interval). BMI is the concurrent value of body mass index, logarithmically transformed. HR = hazard ratio; NA = not applicable.
† HR for a difference of 25% in BMI at the same time point for the same individual.
‡ HR for increase of 10% vs. increase of 5% at the same time point for the same individual.
§ The survival process depends on the level of BMI at the same time point (concurrent level).
¶ The survival process depends on the slope of BMI at the same time point (concurrent slope).
# The survival process depends on the level of BMI and slope of BMI at the same time point.

Results. In the CoxPH model with TVCs, higher BMI was associated with higher risk of worsening knee radiographic OA (HR per 5 kg/m^2 [1.49, 95% CI 1.42–1.55]). We also found, counterintuitively, that increasing BMI over time was negatively associated with worsening radiographic OA; specifically, the risk decreased by 8% for each 5% increase in BMI over time (HR per 5% 0.92 [95% CI, 0.89–0.95]). The results for JM analysis are

Table 2. Three joint models (JMs) for longitudinal body mass index (BMI) and/or longitudinal change in BMI with risk for incident-worsening radiographic osteoarthritis of the knee fitted to data from the Johnston County Osteoarthritis Project: comparison under different parameterizations*

	JM1† (BIC 2,604.2)	JM2‡ (BIC 2,688.7)	JM3§ (BIC 2,602.8)
Sex, male vs. female	−0.08 (0.06)	−0.09 (0.06)	−0.09 (0.06)
Age at baseline, years¶	0.27 (0.03)#	0.23 (0.03)#	0.29 (0.03)#
Log (BMI)	1.48 (0.14)#	NA	1.42 (0.14)#
Slope of log (BMI)	NA	32.76 (8.38)#	17.77 (7.09)#

* Values are the coefficients with SEs from the time-to-event submodel. BMI is the concurrent value of body mass index. BIC = Bayesian Information Criterion; NA = not applicable.
† The survival process depends on the level of BMI at the same time point (concurrent level).
‡ The survival process depends on the slope of BMI at the same time point (concurrent slope).
§ The survival process depends on the level of BMI and slope of BMI at the same time point.
¶ Variable was standardized to have mean of 0 and standard deviation of 1.
# Significant.

shown in Table 2, representing the coefficients from the survival submodel.

As previously mentioned, the corresponding coefficients can be interpreted in terms of percentage change rather than absolute change (see Supplementary Appendix A, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23881/abstract). As shown in Table 1, JM1 finds the association between the current level of BMI and the risk for increase of K/L grade such that if a participant had a 25% higher BMI at the same time, the K/L grade risk was 1.4 times as high (HR 1.39 [95% CI 1.31–1.48]). In JM2, the slope of BMI trajectory was found to have an association with incident increase in the K/L grade: if BMI increased by 10% each year, the risk for increase of the K/L grade is 4.6 times as high as compared to a 5% increase (HR 4.59 [95% CI 2.14–9.86]). In JM3, both the current level (HR 1.37, 95% CI 1.29–1.46) and the slope of BMI (HR 2.29 [95% CI 1.20–4.36]) were associated with worsening radiographic OA. According to the BIC (Table 1), JM3 provided the best fit to the data compared to JM1 and JM2, providing evidence that the risk for an increase of the K/L grade depends on both the level and the slope of BMI at the current time.

Discussion

JM of longitudinal and time-to-event data continues as an emerging area of statistical research. In the current study, we demonstrated the usefulness and interpretability of the JM approach in rheumatology using OA, which is the most common form of arthritis and a leading cause of disability among adults in the US (35,36) and worldwide, as an exemplar. The association of body mass change over time in relation to OA is especially important because obesity is increasing in prevalence worldwide (37–40). While many studies have provided strong evidence that lowering body mass can reduce risk of OA development and progression (41–43), some have failed to demonstrate this effect potentially due to methodologic difficulties in statistical analyses (32,44). Our results using the JM, but not the CoxPH, show that the risk of increasing K/L grade (i.e., worsening radiographic OA) is higher when BMI is increasing, illustrating the advantages of the JM for such dynamic measures in a longitudinal study. We chose OA as an example to emphasize the importance of detailed modeling of longitudinal trajectories of patient outcomes, in particular in relation to development of an RMD that is strongly associated with older age (45) and frequently is slowly progressive. One can envision such individual trajectories as personal histories of change that led one individual to developing OA and allowed another one to avoid this health problem. Taking advantage of longitudinal design together with this methodology can improve our understanding of the mechanisms of development and progression of such conditions, which in turn can optimize disease prevention and management strategies.

The JM approach can be applied to a very broad family of RMDs that affect people at almost any age. Application of the JM to clinical questions in rheumatology may clarify why the course and the severity of symptoms of RMDs vary from patient to patient and from time to time. In addition, these models provide a natural structure for dynamic individual predictions of longitudinal and time-to-event outcomes (46), which is important both from patient and health provider perspectives. In recent studies (47,48), JMs were used as a tool for optimizing medical screening strategies, in particular the frequency of the screening procedures for people with different stages of disease, which may allow providers to choose the optimal screening schedule for individuals based on their longitudinal history. This approach could maximize benefits and minimize medical costs by avoiding unnecessary screenings and interventions.

Importantly, JMs are also being increasingly used in clinical trials that are crucial to advancements in new drug therapies. In this setting, dropout is a common problem and raises concerns of nonignorable missing data, in particular if a participant leaves the study due to an adverse reaction or a lack of effectiveness of the treatment. As mentioned above, ignoring the mechanism of missingness can cause bias in estimates in the LME model. Perhaps most notably, in the JM framework, dropout time can be considered a survival outcome, while a longitudinal submodel can be used to obtain valid inferences with the correction for nonignorable dropout. Several studies have suggested that JMs of longitudinal data and time to dropout not only provide unbiased estimates (6,25) but also may require smaller sample sizes to achieve

comparable power (49), both critical in driving the field forward to improve knowledge and health.

JMs also have some important limitations. First, JMs are computationally intensive and time consuming, which might pose logistical challenges for researchers working with large data sets. Second, as with any statistical modeling, the LME model and the CoxPH (the 2 submodels of the JM) are based on specific assumptions, which should be properly tested. This prerequisite step becomes more critical when these models are being used jointly and should not be ignored. Our aim in this study was to provide an introduction to the JM approach that is accessible for a clinical audience not necessarily familiar with advanced topics in mixed effects modelling and time-to-event analysis; we emphasize that collaboration with statistical experts in these methods is important in applying JMs in practice.

In conclusion, the potential applications of the JM in RMDs is underappreciated, although these methods provide clear advantages over traditional approaches (while incorporating strengths from these methods). Software is readily available to facilitate applications of the JM to address relevant research and clinical questions in a statistically rigorous and coherent fashion. We hope to stimulate interest in these models among RMD researchers with increased benefits to society through their use.

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.

REFERENCES

1. Gibbons RD, Hedeker D, DuToit S. Advances in analysis of longitudinal data. Annu Rev Clin Psychol 2010;6:79–107.
2. Rubin DB. Inference and missing data. Biometrika 1976;63:581–90.
3. Saha C, Jones MP. Asymptotic bias in the linear mixed effects model under non-ignorable missing data mechanisms. J R Stat Soc Series B Stat Methodol 2005;67:167–82.
4. Carroll RJ. Measurement error in nonlinear models: a modern perspective. 2nd ed. Boca Raton: Chapman & Hall/CRC; 2006.
5. Rizopoulos D. Joint models for longitudinal and time-to-event data: with applications in R. Boca Raton: Taylor & Francis Group/CRC; 2012.
6. Asar O, Ritchie J, Kalra PA, Diggle PJ. Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. Int J Epidemiol 2015;44:334–44.
7. Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics 1982;38:963–74.
8. Verbeke G, Molenberghs G. Linear mixed models for longitudinal data. New York: Springer; 2000.
9. Little RJ, Rubin DB. Statistical analysis with missing data. 2nd ed. Hoboken: Wiley; 2002.
10. Cox DR. Regression models and life-tables. J R Stat Soc Series B Stat Methodol 1972;34:187–220.
11. Cox DR. Partial likelihood. Biometrika 1975;62:269–76.
12. Therneau TM, Grambsch PM. Modeling survival data: extending the Cox model. New York: Springer; 2000.
13. Allison PD, SAS Institute. Survival analysis using the SAS system: a practical guide. Cary (NC): SAS Institute; 1995.
14. Prentice RL. Covariate measurement errors and parameter-estimation in a failure time regression-model. Biometrika 1982;69:331–42.
15. Stevens J, Truesdale KP, McClain JE, Cai J. The definition of weight maintenance. Int J Obes (Lond) 2006;30:391–9.
16. Orsama AL, Mattila E, Ermes M, van Gils M, Wansink B, Korhonen I. Weight rhythms: weight increases during weekends and decreases during weekdays. Obes Facts 2014;7:36–47.
17. Sweeting MJ, Thompson SG. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. Biomed J 2011;53:750–63.
18. Yashin AI, Arbeev KG, Akushevich I, Kulminski A, Akushevich L, Ukraintseva SV. Stochastic model for analysis of longitudinal data on aging and mortality. Math Biosci 2007;208:538–51.
19. Yashin AI, Arbeev KG, Akushevich I, Kulminski A, Ukraintseva SV, Stallard E, et al. The quadratic hazard model for analyzing longitudinal data on aging, health, and the life span. Phys Life Rev 2012;9:177–88.
20. Woodbury MA, Manton KG. Random-walk model of human mortality and aging. Theor Popul Biol 1977;11:37–48.
21. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. Stat Sin 2004;14:809–34.
22. Proust-Lima C, Sene M, Taylor JM, Jacqmin-Gadda H. Joint latent class models for longitudinal and time-to-event data: a review. Stat Methods Med Res 2014;23:74–90.
23. Elashoff RM, Li G, Li N. Joint modeling of longitudinal and time-to-event data. Boca Raton: Chapman and Hall/CRC; 2016.
24. Rizopoulos D. JM: An R package for the joint modelling of longitudinal and time-to-event data. J Stat Softw 2010;35:1–33.
25. Ediebah DE, Galindo-Garre F, Uitdehaag BM, Ringash J, Reljneveld JC, Dirven L, et al. Joint modeling of longitudinal health-related quality of life data and survival. Qual Life Res 2015;24:795–804.
26. Proust-Lima C, Taylor JM. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. Biostatistics 2009;10:535–49.
27. Gilani N, Kazemnejad A, Zayeri F, Hadaegh F, Azizi F, Khalili D. Anthropometric indices as predictors of coronary heart disease risk: joint modeling of longitudinal measurements and time to event. Iran J Public Health 2017;46:1546–54.
28. Fournier MC, Foucher Y, Blanche P, Buron F, Giral M, Dantan E. A joint model for longitudinal and time-to-event data to better assess the specific role of donor and recipient factors on long-term kidney transplantation outcomes. Eur J Epidemiol 2016;31:469–79.
29. Jafari-Koshki T, Mansourian M, Hosseini SM, Amini M. Association of waist and hip circumference and waist-hip ratio with type 2 diabetes risk in first-degree relatives. J Diabetes Complications 2016;30:1050–5.
30. Van Beers-Tas MH, Stuiver MM, de Koning MH, van de Stadt LA, Geskus RB, van Schaardenburg D. Can an increase in autoantibody levels predict arthritis in arthralgia patients? Rheumatology (Oxford) 2018;57:932–4.
31. Manninen P, Riihimaki H, Heliovaara M, Makela P. Overweight, gender and knee osteoarthritis. Int J Obes Relat Metab Disord 1996;20:595–7.
32. Abbate LM, Jordan JM. Weight change in osteoarthritis. Osteoarthritis Cartilage 2012;20:268–70.
33. Jordan JM, Helmick CG, Renner JB, Luta G, Dragomir AD, Woodard J, et al. Prevalence of knee symptoms and radiographic and symptomatic knee osteoarthritis in African Americans and Caucasians: the Johnston County Osteoarthritis project. J Rheumatol 2007;34:172–80.

34. Schwarz G. Estimating dimension of a model. Ann Stat 1978;6: 461–4.
35. Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, et al. The global burden of hip and knee osteoarthritis: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis 2014;73:1323–30.
36. Neogi T. The epidemiology and impact of pain in osteoarthritis. Osteoarthritis Cartilage 2013;21:1145–53.
37. Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the United States, 1999–2004. JAMA 2006;295:1549–55.
38. Ogden CL, Carroll MD, Curtin LR, Lamb MM, Flegal KM. Prevalence of high body mass index in US children and adolescents, 2007–2008. JAMA 2010;303:242–9.
39. Flegal KM, Carroll MD, Ogden CL, Curtin LR. Prevalence and trends in obesity among US adults, 1999–2008. JAMA 2010;303:235–41.
40. Popkin BM, Gordon-Larsen P. The nutrition transition: worldwide obesity dynamics and their determinants. Int J Obes Relat Metab Disord 2004;28:S2–S9.
41. Felson DT, Zhang YQ, Anthony JM, Naimark A, Anderson JJ. Weight-loss reduces the risk for symptomatic knee osteoarthritis in women: the Framingham study. Ann Intern Med 1992;116: 535–9.
42. Anandacoomarasamy A, Leibman S, Smith G, Caterson I, Giuffre B, Fransen M, et al. Weight loss in obese people has structure-modifying effects on medial but not on lateral knee articular cartilage. Ann Rheum Dis 2012;71:26–32.
43. Christensen R, Bartels EM, Astrup A, Bliddal H. Effect of weight reduction in obese patients diagnosed with knee osteoarthritis: a systematic review and meta-analysis. Ann Rheum Dis 2007;66:433–9.
44. Messier SP, Loeser RF, Miller GD, Morgan TM, Rejeski WJ, Sevick MA, et al. Exercise and dietary weight loss in overweight and obese older adults with knee osteoarthritis: the arthritis, diet, and activity promotion trial. Arthritis Rheum 2004;50:1501–10.
45. Murphy L, Schwartz TA, Helmick CG, Renner JB, Tudor G, Koch G, et al. Lifetime risk of symptomatic knee osteoarthritis. Arthritis Rheum 2008;59:1207–13.
46. Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. Biometrics 2011;67:819–29.
47. Rizopoulos D, Taylor JM, Van Rosmalen J, Steyerberg EW, Takkenberg JJ. Personalized screening intervals for biomarkers using joint models for longitudinal and survival data. Biostatistics 2016;17:149–64.
48. Tomer A, Nieboer D, Roobol MJ, Steyerberg EW, Rizopoulos D. Personalized schedules for surveillance of low-risk prostate cancer patients. Biometrics 2019;75:153–62.
49. Chen LM, Ibrahim JG, Chu HT. Sample size and power determination in joint modeling of longitudinal and survival data. Stat Med 2011;30:2295–309.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 72, No. 5, May 2020, pp 622–629
DOI 10.1002/acr.23887

Racial and Ethnic Differences in the Prevalence and Time to Onset of Manifestations of Systemic Lupus Erythematosus: The California Lupus Surveillance Project

Ernest Maningding,[1] Maria Dall'Era,[1] Laura Trupin,[1] Louise B. Murphy,[2] and Jinoos Yazdany[1]

Objective. The California Lupus Surveillance Project (CLSP) is a population-based registry of individuals with systemic lupus erythematosus (SLE) residing in San Francisco County, California from 2007 to 2009, with a special focus on Asian/Pacific Islander and Hispanic patients. We used retrospective CLSP data to analyze racial and ethnic differences in lupus manifestations and in the timing and risk of developing severe manifestations.

Methods. A total of 724 patients with SLE were retrospectively identified. Prevalence ratios (PRs) of SLE manifestations were calculated using Poisson regression models stratified by race/ethnicity and adjusted for sex, age at SLE diagnosis, and disease duration. We studied onset of severe SLE manifestations after SLE diagnosis using Kaplan-Meier methods to examine time-to-event and Cox proportional hazards regression models to estimate hazard ratios (HRs). White patients were the referent group in all analyses.

Results. African Americans, Asian/Pacific Islanders, and Hispanic patients had increased prevalence of renal manifestations (PR 1.74 [95% confidence interval (95% CI) 1.40–2.16], PR 1.68 [95% CI 1.38–2.05], and PR 1.35 [95% CI 1.05–1.74], respectively). Furthermore, African Americans had increased prevalence of neurologic manifestations (PR 1.49 [95% CI 1.12–1.98]), and both African Americans (PR 1.09 [95% CI 1.04–1.15]) and Asian/Pacific Islanders (PR 1.07 [95% CI 1.01–1.13]) had increased prevalence of hematologic manifestations. African Americans, Asian/Pacific Islanders, and Hispanic patients, respectively, had higher risk of developing lupus nephritis (HR 2.4 [95% CI 1.6–3.8], HR 4.3 [95% CI 2.9–6.4], and HR 2.3 [95% CI 1.4–3.8]) and thrombocytopenia (HR 2.3 [95% CI 1.1–4.4], HR 2.3 [95% CI 1.3–4.2], and HR 2.2 [95% CI 1.1–4.7]). Asian/Pacific Islander and Hispanic patients had higher risk of developing antiphospholipid syndrome (HR 2.5 [95% CI 1.4–4.4] and HR 2.6 [95% CI 1.3–5.1], respectively).

Conclusion. This is the first epidemiologic study comparing lupus manifestations among 4 major racial and ethnic groups. We found substantial differences in the prevalence of several clinical SLE manifestations among racial/ethnic groups and discovered that African Americans, Asian/Pacific Islanders, and Hispanic patients are at increased risk of developing several severe manifestations following a diagnosis of SLE.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a chronic, systemic autoimmune disease with a higher prevalence among women and racial/ethnic minority groups in the US. However, despite the growing numbers of Asian/Pacific Islander and Hispanic individuals in the US, little is known about the epidemiology of SLE in these populations. To produce contemporary population-based estimates of incidence and prevalence among various racial/ethnic groups, the Centers for Disease Control and Prevention funded 4 SLE registries across the US, including 2 registries in California and New York, which focused on Hispanic and Asian patients. Estimates from the registries showed, relative to white patients, increased incidence and prevalence of SLE among African

The findings and conclusions herein are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention or the National Institutes of Health.

Supported by the Centers for Disease Control and Prevention (grants A114297 and U01DP005120) and the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases grant P30-AR-070155). Dr. Yazdany's work was supported by the University of California, San Francisco (Robert L. Kroc Chair in Rheumatic and Connective Tissue Diseases). Drs. Yazdany and Dall'Era's work was supported by the Russell/Engleman Medical Research Center for Arthritis.

[1]Ernest Maningding, MD, MPH, Maria Dall'Era, MD, Laura Trupin, MPH, Jinoos Yazdany, MD, MPH: University of California, San Francisco; [2]Louise B. Murphy, PhD: Centers for Disease Control and Prevention, Atlanta, Georgia.

Dr. Dall'Era has received consulting fees from AstraZeneca, Genentech, Janssen, and Biogen (less than $10,000 each). No other disclosures relevant to this article were reported.

Address correspondence to Ernest Maningding, MD, MPH, Department of Medicine, University of California, San Francisco, Box 0920, San Francisco, CA 94143. E-mail: ejmaningding@gmail.com.

Submitted for publication October 22, 2018; accepted in revised form March 26, 2019.

SIGNIFICANCE & INNOVATIONS

- Relative to white patients, racial/ethnic minority groups (African Americans, Asian/Pacific Islanders, and Hispanics) develop more renal, neurologic, and hematologic manifestations. They also develop lupus nephritis, thrombocytopenia, and antiphospholipid syndrome sooner.
- To our knowledge, this is the first study to use rigorous epidemiologic methods to compare systemic lupus erythematosus (SLE) manifestations across 4 racial/ethnic groups, including Asian/Pacific Islanders and Hispanics, 2 understudied populations.
- Data collected in this study support the importance of increased clinician awareness of SLE and its accelerated progression in these racial/ethnic groups.

Americans (1–4), Asian/Pacific Islanders (2,3), Hispanics (2,3), and American Indian/Alaskan Natives (5). Specifically, the California Lupus Surveillance Project (CLSP) reported a higher age-standardized incidence and prevalence of SLE among African American (15.5 per 100,000 person-years; 241.0 per 100,000 persons), Asian/Pacific Islander (4.1 per 100,000 person-years; 90.5 per 100,000 persons), and Hispanic (4.2 per 100,000 person-years; 94.7 per 100,000 persons) populations relative to the white population (2.8 per 100,000 person-years; 55.2 per 100,000 persons) in San Francisco County during the period 2007–2009 (3).

There are currently few studies exploring racial/ethnic differences in the clinical presentation of SLE or in the development of severe disease manifestations subsequent to SLE diagnosis. Previous epidemiologic studies suggest that in comparison to white patients, African Americans have a more severe presentation of symptoms at the time of diagnosis of SLE and a worse overall prognosis (6–8). However, no studies to date have analyzed racial/ethnic differences in manifestations of SLE across the 4 major racial/ethnic populations in the US, including Asian/Pacific Islander and Hispanic groups. To address this gap, we examined data gathered in the CLSP to investigate racial/ethnic differences in the prevalence of SLE manifestations and in the risk and timing of development of severe SLE manifestations.

PATIENTS AND METHODS

CLSP. The California Department of Public Health collaborated with the University of California, San Francisco (UCSF) to conduct the CLSP, as described elsewhere (3). The State of California's institutional review board granted a waiver for this public health surveillance activity. The project was reviewed and approved by the UCSF Committee on Human Research.

Source population/catchment area criteria. The CLSP includes residents of San Francisco County within the period 2007–2009. According to US census estimates during this time period, San Francisco County averaged 790,582 residents, with the following racial/ethnic composition: 56% white, 35% Asian/Pacific Islander, 7% African American, and 1% American Indian/Alaskan Native (9). At total of 15% of the San Francisco population identified as Hispanic, which is reported separately from race in the census.

Case definition and case ascertainment. In this analysis, patients were defined as having SLE if they met at least 4 of the 11 American College of Rheumatology revised classification criteria defined in 1982 and updated in 1997 (10,11). Three primary sources within the catchment area provided possible cases of SLE: 1) community rheumatology and nephrology clinics, 2) community hospitals, and 3) integrated health care systems including UCSF, Kaiser Permanente, and the San Francisco Veterans Administration Medical Center. Potential cases were identified using International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnostic codes 710.0 (SLE), 695.4 (discoid lupus), 710.8 (other specified connective tissue disease), and 710.9 (unspecified connective tissue disease). Secondary sources of possible cases of SLE included a commercial laboratory, which was queried for a comprehensive panel of SLE-related serologic tests, as described elsewhere (3), and the California Office of Statewide Health Planning and Development hospital discharge database, which was queried for patient discharges using similar ICD-9-CM codes to those listed above.

Clinical manifestations of SLE. A defined number of SLE manifestations for each patient was ascertained from the medical records and grouped within the following 10 categories: mucocutaneous, serositis, cardiovascular, pulmonary, gastrointestinal, renal, musculoskeletal, neurologic, hematologic, and serologic. The manifestations included under each of these categories are defined in Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23887/abstract.

Severe SLE manifestations. We evaluated the following 4 severe SLE manifestations: lupus nephritis, thrombocytopenia, neuropsychiatric lupus, and antiphospholipid syndrome (APS). Because the initial chart abstraction did not include the date of first appearance of lupus nephritis, we created a surrogate variable to define nephritis. Any 1 of the following 3 laboratory measurements were used to define onset of lupus nephritis: 24-hour proteinuria >500 mg, 24-hour urine protein-to-creatinine ratio >0.5, or spot protein-to-creatinine ratio >0.5. We chose this approach to increase the sensitivity of our definition and to ensure that barriers to access and timely diagnosis did not bias the results. A positive laboratory result nearest to the date of SLE diagnosis was used to define the time of initial lupus nephritis. Thrombocytopenia was defined as a documented platelet count below 100,000/mm^3 or physician documentation of thrombocytopenia that was

unexplained by medication effect or other causes. Neuropsychiatric manifestations included seizures, psychosis, and acute confusional state, as documented in the medical record. APS was diagnosed when there was physician documentation of APS. We also evaluated a combined outcome, i.e., development of any of lupus nephritis, thrombocytopenia, neuropsychiatric lupus, or APS.

Independent variables. Information regarding sex and race/ethnicity was abstracted from each medical chart. For race/ethnicity, patients were classified as non-Hispanic white, non-Hispanic African American, Asian/Pacific Islander, or Hispanic (any race). Non-Hispanic whites and non-Hispanic African Americans will be referred to as whites and African Americans from this point forward. American Indian/Alaskan natives were identified but excluded from this analysis due to small sample size (n = 4). Physician-documented age at SLE diagnosis was categorized into discrete age groups (≤18, 19–29, 30–39, 40–49, ≥50 years) to account for the nonlinear relationship between age of diagnosis and disease manifestations. At the time of chart abstraction, the number of years since physician-documented date of SLE diagnosis for each subject was calculated from the year of last reported clinic visit date. Years elapsed since diagnosis were categorized as ≤5, 6–10, 11–15, and ≥16 years, which approximated quartiles of the distribution.

Statistical analyses. Baseline racial/ethnic differences in age at SLE diagnosis and years since SLE diagnosis were examined using analysis of variance after determining that the assumptions of normally distributed residuals and homoscedasticity were not violated. Racial/ethnic differences in sex and SLE manifestation were examined using a chi-square test. Race/ethnicity-stratified prevalence ratios (PRs) for clinical manifestations of SLE were calculated using a Poisson regression model with robust error variances, including the covariates sex, age at SLE diagnosis, and years since SLE diagnosis. A Poisson regression model was chosen due to its appropriateness for analyzing count data and the rarity of individual manifestations. Kaplan-Meier survival methods were used to examine the time to onset of severe SLE manifestations. Individual survival curves representing separate race/ethnicities were compared using the log rank test. For each severe SLE manifestation, we estimated the risk of manifestation onset for race/ethnicity, sex, and age at SLE diagnosis using multivariable Cox proportional hazards regression models that modeled the 3 characteristics simultaneously. Higher hazard ratios (HRs) indicated a greater risk of developing specific manifestations over time. The proportional-hazards assumption based on Schoenfeld residuals was validated for appropriateness of use.

For the survival analysis, the baseline was defined as the physician-documented date of SLE diagnosis. Subjects who already had evidence of the reported outcome(s) at the time of SLE diagnosis were treated as having developed the outcome on day 1. Patients were included in follow-up until they developed the outcome of interest or were censored at the date of the last recorded clinical visit through 2009. A chi-square test was used

Table 1. Baseline characteristics of prevalent cases of systemic lupus erythematosus (SLE) in San Francisco County, 2007–2009*

Variable	Total (n = 724)	Non-Hispanic white (n = 189)	Non-Hispanic African American (n = 135)	Asian/Pacific Islander (n = 265)	Hispanic (n = 109)	Missing† (n = 26)	*P*
Age at SLE diagnosis, years							0.341
≤18	134	44 (23)	18 (13)	49 (18)	17 (16)	6 (23)	
19–29	205	52 (28)	35 (26)	76 (29)	39 (36)	3 (12)	
30–39	143	37 (20)	32 (24)	50 (19)	18 (17)	6 (23)	
40–49	106	19 (10)	24 (18)	38 (14)	21 (19)	4 (15)	
≥50	136	37 (20)	26 (19)	52 (20)	14 (13)	7 (27)	
Years since SLE diagnosis							<0.001
≤5	231	47 (25)	45 (33)	78 (29)	44 (41)	17 (65)	
6–10	141	31 (16)	19 (14)	66 (25)	21 (19)	4 (15)	
11–15	125	28 (15)	29 (21)	45 (17)	22 (20)	1 (4)	
≥16	227	83 (44)	40 (30)	71 (27)	20 (20)	4 (15)	
Sex							0.735
Male	76	17 (9)	13 (10)	32 (12)	12 (11)	2 (8)	
Female	648	172 (91)	122 (90)	233 (88)	97 (89)	24 (92)	
SLE manifestation							
Lupus nephritis‡	256	37 (14)	52 (20)	134 (52)	33 (13)	3 (1)	<0.001
Thrombocytopenia	210	40 (19)	51 (24)	81 (39)	36 (17)	2 (1)	0.009
Neuropsychiatric§	102	29 (28)	28 (27)	30 (29)	15 (15)	0 (0)	0.089
APS	111	22 (20)	20 (18)	48 (43)	21 (19)	0 (0)	0.205

* Values are the number (%) unless indicated otherwise. *P* values were calculated by analysis of variance (age at SLE diagnosis, years since SLE diagnosis) and chi-square test (sex, SLE manifestation). APS = antiphospholipid antibody syndrome.
† Observations without recorded race/ethnicity or date of last clinic visit are grouped in the missing category.
‡ Any of the following surrogate laboratory results were used to define lupus nephritis: 24-hour urine for protein >500 mg, 24-hour urine protein-to-creatinine ratio >0.5, or spot protein-to-creatinine ratio >0.5.
§ Neuropsychiatric lupus manifestations studied include seizures, psychosis, and acute confusional state.

to evaluate racial/ethnic differences in the presence of severe SLE manifestations identified at the date of SLE diagnosis.

Given the use of a surrogate variable to define lupus nephritis onset, a sensitivity analysis was performed in which only those with subsequently physician-diagnosed lupus nephritis were analyzed. All statistical analyses were performed using STATA, version 13 (12).

RESULTS

Study population characteristics. A total of 724 patients with SLE residing in San Francisco County between 2007 and 2009 were retrospectively identified (Table 1). The distribution by race/ethnicity was as follows: white (26.2%), African American (18.8%), Asian/Pacific Islander (36.9%), and Hispanic (15.5%). Asian/Pacific Islander patients were predominantly Chinese (49.4%), followed by Filipino (15.4%) and Vietnamese (6.7%). Most Hispanic patients had no ethnic origin specified (50.9%), followed by South or Central American (except Brazilian) (31.3%) and Mexican (13.4%). Nineteen patients had missing race/ethnicity information (2.6%) and 7 patients (1.0%) had a missing date of last clinic visit. All 26 (3.6%) of these patients were excluded from further analysis. Females comprised 89.5% of the patients. No statistically significant differences in age at SLE diagnosis (P = 0.341) or sex (P = 0.735) were observed by race/ethnicity. Conversely, there was a statistically significant difference in the duration of SLE among race/ethnicities, with white patients more likely to have ≥16 years of follow-up from diagnosis (44% compared to 15–30% for other groups; $P < 0.001$). With respect to the severe SLE manifestations studied, African Americans and Asian/Pacific Islanders had higher prevalence of lupus nephritis (20% and 52%, respectively, compared to 13–14% among other groups; $P < 0.001$) and thrombocytopenia (24% and 39%, respectively, compared to 17–19% among other groups; P = 0.009). Neuropsychiatric lupus was less common among Hispanic patients (15% compared to 27–29% among other groups; P = 0.089), and APS was more common among Asian/Pacific Islanders (43% compared to 18–20% among other groups; P = 0.205), although these differences did not meet statistical significance.

Clinical symptoms of SLE. In comparison to white patients, African American, Asian/Pacific Islander, and Hispanic patients had increased prevalence of renal manifestations (PR 1.74, PR 1.68, and PR 1.35, respectively) (Table 2). African Americans had increased prevalence of neurologic manifestations (PR 1.49), and both African Americans and Asian/Pacific Islanders had increased prevalence of hematologic manifestations (PR 1.09 and PR 1.07, respectively). Because we were performing multiple statistical comparisons, we applied a Bonferroni correction ($P < 0.005$) when interpreting our P values to reduce the risk of Type I errors. After applying this correction, the higher prevalence of renal manifestations among Hispanic patients (P = 0.006), neurologic manifestations among African Americans (P = 0.007), and hematologic manifestations among Asian/Pacific Islanders (P = 0.015) lost their statistical significance. There were no statistically significant differences in prevalence between racial/ethnic minority groups and white patients for the mucocutaneous, serositis, cardiovascular, pulmonary, gastrointestinal, musculoskeletal, or serologic manifestation categories. Supplementary Table 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23887/abstract, shows the prevalence of individual manifestations within each category, stratified by race/ethnicity.

Time to development of severe manifestations of SLE. Figure 1 displays Kaplan-Meier curves illustrating time to incident severe SLE manifestations following SLE diagnosis, stratified by race/ethnicity. Mean and median follow-up times were 12.5 and 9.8 years, respectively. Log rank tests demonstrated statistically significant differences between race/ethnicities in the time to development of lupus nephritis ($P < 0.01$), thrombocytopenia (P = 0.04), APS ($P <$

Table 2. Prevalence ratios of systemic lupus erythematosus (SLE) manifestations stratified by race/ethnicity among prevalent cases of SLE in San Francisco County, 2007–2009*

Characteristic	Non-Hispanic white (n = 189)	Non-Hispanic African American (n = 135)	Asian/Pacific Islander (n = 265)	Hispanic (n = 109)
Mucocutaneous	Ref.	0.98 (0.89–1.07)	1.04 (0.97–1.11)	1.00 (0.91–1.10)
Serositis	Ref.	1.13 (0.92–1.38)	0.92 (0.76–1.12)	1.09 (0.87–1.37)
Cardiovascular	Ref.	1.38 (0.95–2.01)	1.09 (0.77–1.55)	1.34 (0.88–2.04)
Pulmonary	Ref.	1.09 (0.87–1.36)	0.96 (0.78–1.17)	1.13 (0.89–1.44)
Gastrointestinal	Ref.	1.64 (0.84–3.18)	1.30 (0.71–2.37)	1.71 (0.85–3.44)
Renal	Ref.	1.74 (1.40–2.16)†	1.68 (1.38–2.05)†	1.35 (1.05–1.74)‡
Musculoskeletal	Ref.	1.00 (0.91–1.11)	0.95 (0.87–1.04)	1.07 (0.97–1.18)
Neurologic	Ref.	1.49 (1.12–1.98)‡	0.76 (0.56–1.04)	0.98 (0.67–1.44)
Hematologic	Ref.	1.09 (1.04–1.15)†	1.07 (1.01–1.13)‡	1.06 (1.00–1.13)
Serologic	Ref.	1.00 (1.00–1.01)	1.00 (1.00–1.01)	1.00 (1.00–1.01)

* Values are prevalence ratio (95% confidence interval). Calculations based on Poisson regression model with robust error variances adjusting for sex, age at SLE diagnosis, and years since SLE diagnosis. Ref. = reference.
† $\alpha < 0.001$.
‡ $\alpha < 0.05$.

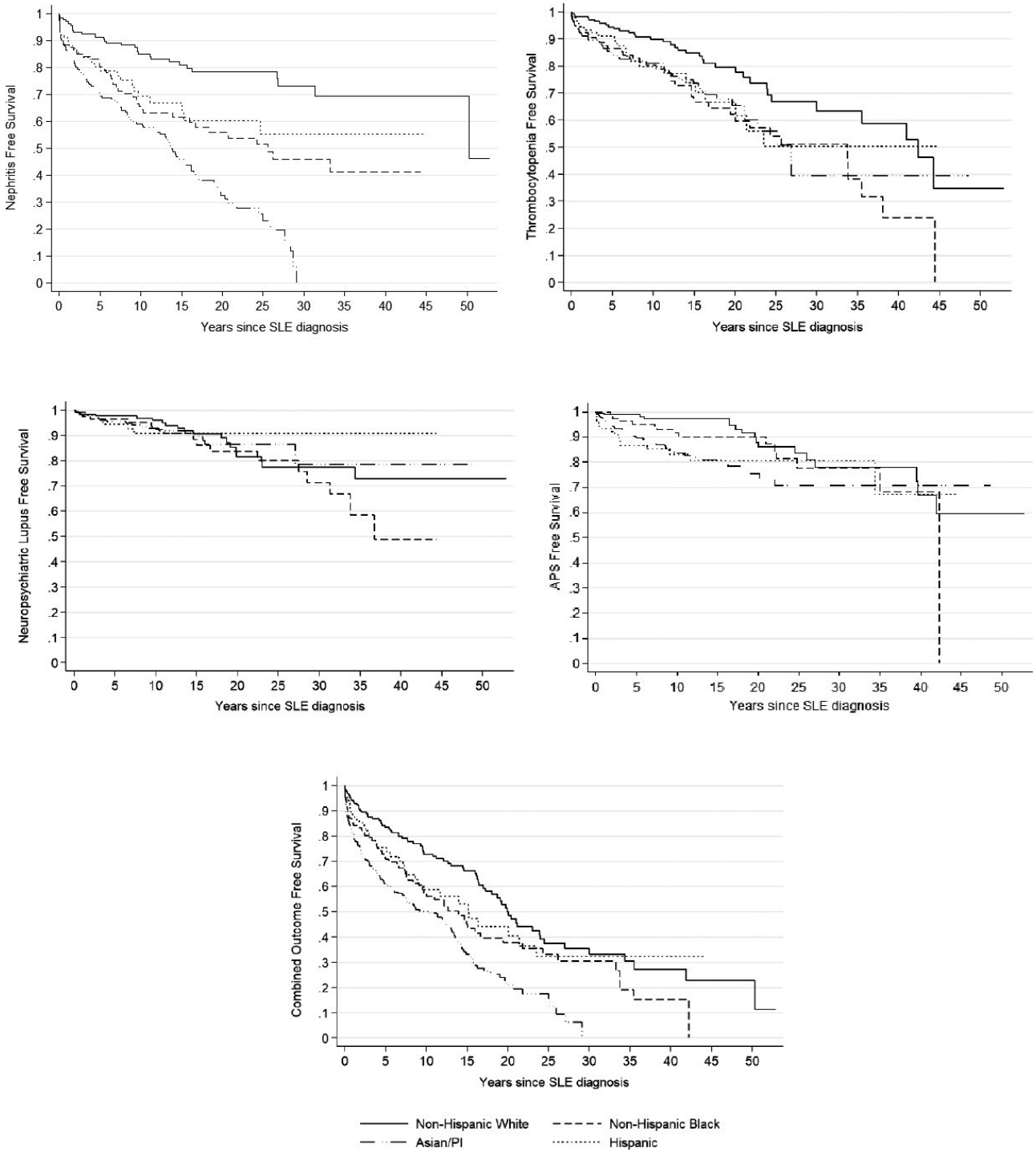

Figure 1. Time to incident severe systemic lupus erythematosus (SLE) manifestation following SLE diagnosis, stratified by race/ethnicity, among prevalent SLE cases in San Francisco County, 2007–2009. Combined outcome (bottom center) includes any of lupus nephritis (top left), thrombocytopenia (top right), neuropsychiatric lupus (bottom left), or antiphospholipid antibody syndrome (APS) (bottom right). Any of the following surrogate laboratory results were used to define lupus nephritis: 24-hour urine for protein >500 mg, 24-hour urine protein-to-creatinine ratio >0.5, or spot protein-to-creatinine ratio >0.5. Neuropsychiatric lupus manifestations studied include seizures, psychosis, and acute confusional state. PI = Pacific Islander.

0.01), and the combined outcome ($P < 0.01$), but not for neuropsychiatric lupus ($P = 0.59$). For all racial/ethnic minorities, the risk for development of lupus nephritis, thrombocytopenia, APS, and the combined outcome was greatest in the first year after disease onset.

Results from the Cox proportional hazard regression models are displayed in Table 3. All racial/ethnic minorities with SLE had statistically significant increased HRs relative to white patients for lupus nephritis and thrombocytopenia. Asian/Pacific Islanders

Table 3. Factors associated with severe manifestations of systemic lupus erythematosus (SLE) in multivariable Cox proportional hazards regression model among prevalent cases of SLE in San Francisco County, 2007–2009*

Variables	Lupus nephritis†	Thrombocytopenia	Neuropsychiatric‡	APS	Combined§
Race/ethnicity					
Non-Hispanic white (ref.)					
Non-Hispanic African American	2.4 (1.6–3.8)	2.3 (1.1–4.4)	1.5 (0.8–2.9)	1.5 (0.7–3.0)	1.6 (1.1–2.2)
Asian/Pacific Islander	4.3 (2.9–6.4)	2.3 (1.3–4.2)	1.1 (0.6–2.0)	2.5 (1.4–4.4)	2.4 (1.8–3.2)
Hispanic	2.3 (1.4–3.8)	2.2 (1.1–4.7)	1.0 (0.4–2.5)	2.6 (1.3–5.1)	1.4 (0.9–2.1)
Sex					
Female (ref.)					
Male	1.6 (1.1–2.4)	2.1 (1.2–3.6)	1.8 (0.9–3.6)	1.4 (0.7–2.6)	1.8 (1.3–2.5)
Age at diagnosis, years					
≤18	1.1 (0.8–1.5)	1.0 (0.6–1.8)	1.1 (0.6–2.2)	1.1 (0.6–1.8)	0.9 (0.7–1.2)
19–29 (ref.)					
30–39	0.8 (0.5–1.2)	0.7 (0.4–1.4)	0.8 (0.4–1.8)	0.7 (0.4–1.4)	0.8 (0.5–1.1)
40–49	1.0 (0.6–1.5)	0.5 (0.2–1.1)	1.2 (0.5–2.8)	0.7 (0.4–1.6)	0.9 (0.7–1.4)
≥50	1.0 (0.6–1.5)	1.6 (0.9–2.8)	1.7 (0.8–4.0)	1.0 (0.5–2.0)	1.4 (1.0–2.0)

* Values are hazard ratio (95% confidence interval). Calculations based on Cox proportional hazards model that contained each of race/ethnicity, sex, and age at diagnosis. Higher hazards indicate a shorter time to development of specific manifestations. APS = antiphospholipid antibody syndrome; ref. = reference.
† Any of the following surrogate laboratory results were used to define lupus nephritis: 24-hour urine for protein >500 mg, 24-hour urine protein-to-creatinine ratio >0.5, or spot protein-to-creatinine ratio >0.5.
‡ Neuropsychiatric lupus manifestations studied include seizures, psychosis, and acute confusional state.
§ Combined outcome is any of lupus nephritis, thrombocytopenia, neuropsychiatric, or APS.

and Hispanic patients had increased HRs for APS, and African Americans and Asian/Pacific Islanders had increased HRs for the combined outcome relative to white patients. There were no statistically significant differences in HRs for neuropsychiatric lupus among the racial/ethnic groups. Relative to women, men with SLE had between 1.4 and 2.1 times the HR of lupus nephritis, thrombocytopenia, and the combined outcome. There were no statistically significant differences in HRs for the severe SLE manifestations among the categories for age at SLE diagnosis. Furthermore, there were no significant differences in the proportion of severe SLE manifestations identified at SLE diagnosis among race/ethnicities (data not shown).

Of the patients identified with lupus nephritis based on the primary variable definition, 76% had a diagnosis of lupus nephritis indicated in their medical charts by the treating physician. A sensitivity analysis examining only individuals diagnosed with lupus nephritis by the treating physician revealed slightly reduced HRs with preserved statistical significance for all racial/ethnic minority groups, except African Americans (P = 0.054) (see Supplementary Table 3, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23887/abstract).

DISCUSSION

Using a large, racially/ethnically diverse population-based registry, we identified racial/ethnic differences in the prevalence of SLE manifestations and in the risk and timing of development of severe SLE manifestations. This analysis demonstrates substantial differences in the prevalence of several clinical SLE manifestations among race/ethnicities. African American, Asian/Pacific Islander, and Hispanic patients had a greater prevalence of renal abnormalities in comparison to white patients. In addition, African Americans had increased neurologic manifestations, and both African Americans and Asian/Pacific Islanders had increased hematologic manifestations in comparison to white patients. These findings are not explained by racial/ethnic differences in sex, age at SLE diagnosis, or duration of SLE disease, as all of these risk factors were accounted for in this analysis. We also found that African American, Asian/Pacific Islander, and Hispanic patients are at increased risk of developing a number of severe manifestations (lupus nephritis, thrombocytopenia, and APS) earlier than white patients following SLE diagnosis. Men also developed lupus nephritis and thrombocytopenia earlier than women. Our study represents the first comprehensive examination of differences in SLE manifestations in Asian/Pacific Islander and Hispanic patients in the US, 2 racial/ethnic groups that have been understudied in population-based epidemiologic investigations.

We found that for racial/ethnic minorities, the risk of lupus nephritis was greatest during the first year following diagnosis of SLE. Subsequently, the annual risk of lupus nephritis remained approximately constant, with the highest burden of risk experienced by Asian/Pacific Islanders. White patients, on the other hand, had a nearly constant incidence rate. The HRs calculated in our analyses are consistent with those of at least 1 previous study, which also found increased risks for African Americans (HR 1.5), Asians (HR 1.8), and Hispanics (HR 1.5) in comparison to white patients (6), although the HRs in that study were not as large as discovered here.

There are several possible explanations for why racial/ethnic minority groups with SLE are at greater risk of developing lupus nephritis. Genetic factors have been proposed to explain why African Americans with SLE have a greater likelihood of renal disease,

more severe disease presentation, and poorer prognosis relative to white patients (13–16). A number of studies have also attributed worse outcomes among African American and Hispanic patients to socioeconomic factors, although no analysis has looked specifically at the association between socioeconomic factors and the onset of lupus nephritis (16–18). Surprisingly, the highest risk for lupus nephritis was observed among Asian/Pacific Islanders, nearly double that of African Americans and Hispanics. This is despite the fact that Asian/Pacific Islanders have the highest average levels of income and education and the best access to health care among racial/ethnic minority groups in San Francisco County (19), suggesting that there may be additional underlying factors that increase risk in this population. In African Americans, certain genotypes increase the risk of lupus nephritis, and although this is a potential mechanism among Asians, it remains understudied (20).

There are no previous studies describing the time to development of thrombocytopenia, neuropsychiatric lupus, or APS in patients with SLE. All racial/ethnic groups, including white patients, were at greatest risk of thrombocytopenia and APS during the first year following SLE diagnosis and continued to develop these manifestations throughout the follow-up period. It is plausible that genetic variation partially explains the observed racial/ethnic differences in the risk of APS. No defined racial/ethnic predominance for primary APS has been documented, but several studies support the increased risk conferred by various genetic variants, particularly human leukocyte antigen associations (21,22). Future studies investigating racial/ethnic group-level differences in genetic variability are needed.

Numerous studies demonstrate that men with SLE are more likely to have organ damage, including renal disease and neuropsychiatric abnormalities, and have an increased mortality rate 1 year following initial SLE hospitalization (23–27). A previous analysis has shown that men develop lupus nephritis earlier than women, with a reported relative hazard of 1.7 (6). Our study has yielded similar findings and shows that men had the greatest risk of lupus nephritis during the first year following diagnosis of SLE (data not shown). Furthermore, our results show that men develop thrombocytopenia earlier following SLE diagnosis, a result not previously demonstrated. Although our analysis does not readily identify causes for these differences, several theories exist to explain differences in sex in SLE presentation, including differences in sex hormones and decreased medical-seeking behavior among men, possibly leading to their delayed diagnosis (28).

A potential contributor to the more severe progression of lupus identified in men relative to women and in the studied minority groups relative to white patients could be that SLE is diagnosed at a more advanced stage in these populations. This would explain both the relatively accelerated appearance of lupus nephritis and thrombocytopenia occurring within the first year and why racial/ethnic minorities appear to have greater risk for development of several manifestations of SLE after controlling for duration of disease. There is some preliminary support for this hypothesis (28,29). In one study, compared with women, men had higher risk of severe disease activity at the time of SLE diagnosis as determined by a Systemic Lupus Erythematosus Disease Activity Index score of ≥12, independent of age, racial/ethnic group, anti-Ro positivity, or time to criteria accrual (odds ratio 3.11 [95% confidence interval 1.09–8.92]) (29). Further work is needed to unravel the contribution of delayed access to diagnosis and treatment.

One of the major strengths of the CLSP is the careful and systematic attention to case ascertainment using a variety of sources: university and community clinics, hospitals, regional laboratories, and state administrative databases. Asian and Hispanic patients were further identified through physicians who focused on these populations (e.g., Chinese patients at San Francisco's Chinese hospital) and through multilingual abstractors.

Several limitations to this study exist. There is always a potential for incomplete case ascertainment despite the efforts described above. Each clinic and hospital had to voluntarily agree to participate in the CLSP. Unfortunately, 2 community hospitals in San Francisco chose not to participate in the program. Given the small number of cases identified solely through community-based hospitals, <5 cases were expected to be missing from this data (3). Incomplete case ascertainment might also have occurred because surveillance efforts were focused on rheumatology clinics and did not include primary care clinics. Capture–recapture analysis reported in a prior publication estimated an additional 147 patients, although this estimation had a wide confidence interval (3). A second limitation is that the quality of medical record documentation of SLE manifestations and criteria varied widely depending on the clinic or hospital setting. Older charts that may have documented early manifestations of disease, particularly serologic laboratory results, were difficult to obtain and may have been inadequately captured. Third, race and ethnicity were determined from the medical record, not through patient self-report. Race and ethnicity were sometimes poorly documented in the medical records, leading to missing data for race and ethnicity. Importantly, this study also did not account for variables that historically have been associated with severe disease manifestations, including socioeconomic status, medications, access to care, and coexisting medical conditions. The analysis of genetic and other biologic data would have been useful if collected previously for this study population. A longitudinal cohort study called the California Lupus Epidemiology Surveillance Study has emerged from analysis of biologic specimens voluntarily provided by members of the CLSP cohort.

The current study found important differences in the characteristics and progression of SLE between racial/ethnic minority groups and white patients. To our knowledge, it is the first study to use rigorous epidemiologic methods to compare SLE manifestations across 4 racial/ethnic groups, including Asian/Pacific Islanders and Hispanics, 2 understudied populations. Data collected in this study support the importance of increased clinician awareness of SLE and its accelerated progression in these racial/ethnic groups. These data also advocate for greater resource

allocation on early diagnosis and treatment in these populations. Future studies should attempt to collect and analyze data on additional risk factors, including socioeconomic status, access to care, medication and appointment adherence, coexisting medical conditions, and genetic variation and the relationship of these variables to the onset of disease manifestations.

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. Maningding 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. Maningding, Dall'Era, Trupin, Yazdany.
Acquisition of data. Dall'Era, Yazdany.
Analysis and interpretation of data. Maningding, Dall'Era, Trupin, Murphy, Yazdany.

REFERENCES

1. 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.
2. Izmirly PM, Wan I, Sahl S, Buyon JP, Belmont HM, Salmon JE, et al. The incidence and prevalence of systemic lupus erythematosus in New York County (Manhattan), New York: the Manhattan Lupus Surveillance Program. Arthritis Rheumatol 2017;69:2006–17.
3. Dall'Era M, Cisternas MG, Snipes K, Herrinton LJ, Gordon C, Helmick CG. The incidence and prevalence of systemic lupus erythematosus in San Francisco County, California: the California Lupus Surveillance Project. Arthritis Rheumatol 2017;69:1996–2005.
4. Lim SS, Bayakly AR, Helmick CG, Gordon C, Easley KA, Drenkard C. The incidence and prevalence of systemic lupus erythematosus, 2002–2004: the Georgia Lupus Registry. Arthritis Rheumatol 2014;66:357–68.
5. Ferucci ED, Johnston JM, Gaddy JR, Sumner L, Posever JO, Choromanski TL, et al. Prevalence and incidence of systemic lupus erythematosus in a population-based registry of American Indian and Alaska Native people, 2007–2009. Arthritis Rheumatol 2014;66:2494–502.
6. Seligman VA, Lum RF, Olson JL, Li H, Criswell LA. Demographic differences in the development of lupus nephritis: a retrospective analysis. Am J Med 2002;112:726–9.
7. Alarcón GS, McGwin GJ Jr, Bartolucci AA, Roseman J, Lisse J, Fessler BJ, et al. Systemic lupus erythematosus in three ethnic groups. Part IX. Differences in damage accrual. Arthritis Rheum 2001;44:2797–806.
8. Contreras G, Lenz O, Pardo V, Borja E, Cely C, Iqbal K, et al. Outcomes in African Americans and Hispanics with lupus nephritis. Kidney Int 2006;69:1846–51.
9. CDC/National Center for Health Statistics. Bridged-race population estimates: data files and documentation. URL: http://www.cdc.gov/nchs/nvss/bridged_race/data_documentation.htm.
10. 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.
11. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1997;40:1725.
12. StataCorp. Stata statistical software: release 13. College Station (TX): StataCorp; 2013.
13. Reveille JD, Moulds JM, Ahn C, Friedman AW, Baethge B, Roseman J, et al, for the LUMINA Study Group. Systemic lupus erythematosus in three ethnic groups. Part I. The effects of HLA class II, C4, and CR1 alleles, socioeconomic factors, and ethnicity at disease onset. Arthritis Rheum 1998;41:1161–72.
14. Alarcón GS, Bastian HM, Beasley TM, Roseman JM, Tan FK, Fessler BJ, et al. Systemic lupus erythematosus in a multi-ethnic cohort (LUMINA) XXXII: [corrected] contributions of admixture and socioeconomic status to renal involvement. Lupus 2006;15:26–31.
15. Salmon JE, Millard S, Schachter LA, Arnett FC, Ginzler EM, Gourley MF, et al. Fc gamma RIIA alleles are heritable risk factors for lupus nephritis in African Americans. J Clin Invest 1996;97:1348–54.
16. Korbet SM, Schwartz MM, Evans J, Lewis EJ. Severe lupus nephritis: racial differences in presentation and outcome. J Am Soc Nephrol 2007;18:244–54.
17. Reveille JD, Bartolucci A, Alarcón GS. Prognosis in systemic lupus erythematosus: negative impact of increasing age at onset, black race, and thrombocytopenia, as well as causes of death. Arthritis Rheum 1990;33:37–48.
18. Pons-Estel BA, Catoggio LJ, Cardiel MH, Soriano ER, Gentiletti S, Villa AR, et al. The GLADEL multinational Latin American prospective inception cohort of 1,214 patients with systemic lupus erythematosus: ethnic and disease heterogeneity among "Hispanics." Medicine (Baltimore) 2004;83:1–17.
19. Tseng W, Cook WK, Chung C. Demographic and socioeconomic profiles of Asian Americans, Native Hawaiians, and Pacific Islanders in the United States. 2011. URL: http://www.apiahf.org/sites/default/files/Demographic_Socioeconomic_Profiles_AANHPI.pdf.
20. Lanata CM, Nititham J, Taylor KE, Chung SA, Torgerson G, Seldin MF, et al. Genetic contributions to lupus nephritis in a multi-ethnic cohort of systemic lupus erythematous patients. PLoS One 2018;13:e0199003.
21. Horita T, Merrill JT. Genetics of antiphospholipid syndrome. Curr Rheumatol Rep 2004;6:458–62.
22. Namjou B. Antiphospholipid syndrome: genetic review. Curr Rheumatol Rep 2003;5:391–4.
23. Prete PE, Majlessi A, Gilman S, Hamideh F. Systemic lupus erythematosus in men: a retrospective analysis in a Veterans Administration Healthcare System population. J Clin Rheumatol 2001;7:142–50.
24. Tan TC, Fang H, Magder LS, Petri MA. Differences between male and female systemic lupus erythematosus in a multiethnic population. J Rheumatol 2012;39:759–69.
25. Kaufman LD, Gomez-Reino JJ, Heinicke MH, Gorevic PD. Male lupus: retrospective analysis of the clinical and laboratory features of 52 patients, with a review of the literature. Semin Arthritis Rheum 1989;18:189–97.
26. Ward MM, Studenski S. Systemic lupus erythematosus in men: a multivariate analysis of gender differences in clinical manifestations. J Rheumatol 1990;17:220–4.
27. Resende AL, Titan SM, Barros RT, Woronik V. Worse renal outcome of lupus nephritis in male patients: a case-control study. Lupus 2011;20:561–7.
28. Lu LJ, Wallace DJ, Ishimori ML, Scofield RH, Weisman MH. Male systemic lupus erythematosus: a review of sex disparities in this disease [review]. Lupus 2010;19:119–29.
29. Muoz-Grajales C, Gonzalez LA, Alarcon GS, Acosta-Reyes J. Gender differences in disease activity and clinical features in newly diagnosed systemic lupus erythematosus patients. Lupus 2016;25:1217–23.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 72, No. 5, May 2020, pp 630–640
DOI 10.1002/acr.23891

Early Magnetic Resonance Imaging–Based Changes in Patients With Meniscal Tear and Osteoarthritis: Eighteen-Month Data From a Randomized Controlled Trial of Arthroscopic Partial Meniscectomy Versus Physical Therapy

Jamie E. Collins,[1] Elena Losina,[1] Robert G. Marx,[2] Ali Guermazi,[3] Mohamed Jarraya,[4] Morgan H. Jones,[5] Bruce A. Levy,[6] Lisa A. Mandl,[2] Scott D. Martin,[7] Rick W. Wright,[8] Kurt P. Spindler,[5] and Jeffrey N. Katz,[1] for the MeTeOR Investigator Group

Objective. The present study was undertaken to evaluate changes in knee magnetic resonance imaging (MRI) findings over the course of 18 months in subjects with osteoarthritic change and meniscal tear treated with arthroscopic partial meniscectomy (APM) or nonoperatively with physical therapy (PT).

Methods. We used 18-month follow-up data from the Meniscal Tear in Osteoarthritis Research Trial. MRI results were read with reference to the MRI Osteoarthritis Knee Score. We focused on 18-month change in bone marrow lesions (BMLs), cartilage thickness, cartilage surface area, osteophyte size, effusion-synovitis, and Hoffa-synovitis. We used multinomial logistic regression to assess associations between MRI-based changes in each feature and treatment type.

Results. A total of 351 subjects were randomized, and 225 had both baseline and 18-month MRI results. In both treatment groups, patients experienced substantial changes in several MRI-based markers. In 60% of the APM group, versus 33% of the PT group, cartilage surface area damage advanced in ≥2 subregions (adjusted odds ratio 4.2 [95% confidence interval 2.0–9.0). Patients who underwent APM also had greater advancement in scores for osteophytes and effusion-synovitis. We did not find significant associations between treatment type and change in cartilage thickness, BMLs, or Hoffa-synovitis.

Conclusion. This cohort of patients with meniscal tear and osteoarthritis showed marked advancement in MRI-based features over 18 months. Patients treated with APM showed more advancement in some features compared to those treated nonoperatively. The clinical relevance of these early findings is unknown and requires further study.

INTRODUCTION

Recent estimates suggest that 14 million adults in the US have knee osteoarthritis (OA), including 8 million individuals under 65 years of age (1). Of these, ~80% have a concomitant meniscal tear (2). Nonoperative treatment of symptomatic meniscal tear typically includes a physical therapy (PT) regimen (muscle strengthening, endurance, flexibility, and balance training). Surgical treatment

ClinicalTrials.gov identifier: NCT00597012.

Supported by the Rheumatology Research Foundation (Investigator Award) and the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases grant R01-AR-05557 and grants K24-AR-057827 and P30-AR-072577).

[1]Jamie E. Collins, PhD, Elena Losina, PhD, Jeffrey N. Katz, MD, MSc: Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; [2]Robert G. Marx, MD, MSc, Lisa A. Mandl, MD: Hospital for Special Surgery, Weill Cornell Medicine, New York, New York; [3]Ali Guermazi, MD, PhD: Boston University School of Medicine, Boston, Massachusetts; [4]Mohamed Jarraya, MD: Boston University School of Medicine, Boston, Massachusetts, and Mercy Catholic Medical Center, Darby, Pennsylvania; [5]Morgan H. Jones, MD, MPH, Kurt P. Spindler, MD: Cleveland Clinic, Cleveland, Ohio; [6]Bruce A. Levy, MD: Mayo Clinic, Rochester, Minnesota; [7]Scott D. Martin, MD: Massachusetts General Hospital, Boston; [8]Rick W. Wright, MD: Washington University School of Medicine, St. Louis, Missouri. See Appendix A for members of the MeTeOR Investigator Group.

Dr. Collins has received consulting fees from Boston Imaging Core Labs (less than $10,000). Dr. Losina has received consulting fees from Regeneron and Samumed (less than $10,000 each). Dr. Marx has received consulting fees from Mend (less than $10,000). Dr. Guermazi has received consulting fees from Galapagos, Roche, AstraZeneca (less than $10,000 each), Pfizer, Merck Serono, and TissueGene (more than $10,000 each) and owns stock or stock options in Boston Imaging Core Labs. Dr. Jones has received consulting fees from Samumed (less than $10,000). Dr. Spindler has received consulting fees from Cytori Therapeutics and DePuy Mitek (less than $10,000 each). Dr. Katz has received research grants from Samumed and Flexion Therapeutics. No other disclosures relevant to this article were reported.

Address correspondence to Jamie E. Collins, PhD, Orthopedics and Arthritis Center for Outcomes Research, Department of Orthopedic Surgery, Brigham and Women's Hospital, 75 Francis Street, BTM 5016, Boston, MA 02115. E-mail: Jcollins13@bwh.harvard.edu.

Submitted for publication November 2, 2018; accepted in revised form March 26, 2019.

SIGNIFICANCE & INNOVATIONS

- To our knowledge, this is the first study to use data from a randomized controlled trial investigating surgical versus nonoperative treatment for patients with knee osteoarthritis and meniscal tear to evaluate the early magnetic resonance imaging (MRI)–based changes over an 18-month period for patients undergoing arthroscopic partial meniscectomy (APM) or physical therapy.
- We found marked MRI-based advancement in both groups. In addition, we found that patients who were treated with APM had higher odds of advancement in cartilage surface area, osteophytes, and effusion-synovitis, although the data did not provide sufficient evidence to establish an association between treatment type and change in cartilage thickness, bone marrow lesions, or Hoffa-synovitis.
- The clinical relevance of these findings requires further study and should be considered a research priority.

typically consists of arthroscopic partial meniscectomy (APM). Data from several large clinical trials have suggested that patients with meniscal tear and osteoarthritic changes treated with APM plus PT experience similar pain relief compared to patients treated with PT alone, although crossover from PT to APM makes interpretation challenging in several of these trials (3–7).

In patients with OA, meniscal tear has been shown to be an independent risk factor for progression of cartilage damage (8). Observational studies have suggested that a history of APM may be associated with a higher risk of incident OA (9,10). It is unclear in these studies whether the risk of progression is attributable to the initial meniscal damage or to the surgical treatment. This question can best be addressed in a clinical trial setting, in which all subjects have knee pain and meniscal tear and are deemed surgical candidates by their orthopedic surgeons. The aim of our study was to evaluate early magnetic resonance imaging (MRI)–based changes in patients with meniscal tear and OA treated with APM and those treated nonoperatively.

PATIENTS AND METHODS

Study sample. We used data from the Meniscal Tear in Osteoarthritis Research (MeTeOR) Trial, a multicenter randomized controlled trial (RCT) investigating APM plus PT versus PT alone to treat symptomatic meniscal tear in OA (4). Subjects were recruited from orthopedic surgery clinics in 7 US referral centers. Eligible subjects were ages ≥45 years old, had evidence of meniscal tear on MRI, evidence of OA changes on MRI or radiography, and knee symptoms. The full inclusion and exclusion criteria have been published elsewhere (11).

Subjects were randomized to APM with PT or PT alone. For subjects undergoing APM, the surgeon used standard arthroscopic portals and trimmed the damaged meniscus back to a stable rim (11). Surgeons also trimmed loose fragments of cartilage and bone but did not penetrate the subchondral bone. Subjects randomized to the PT arm followed a standardized, strengthening-based PT protocol, including weekly sessions with a therapist and home-based exercises; generally the program lasted 6 weeks (4,11). Subjects were permitted to see their orthopedic surgeons throughout the study and could discuss with the surgeon the option of crossing over to receive APM if symptoms persisted despite PT.

Outcome. Subjects underwent MRI at baseline as part of routine clinical care. Each of the centers performed cartilage-sensitive sequences, permitting us to assess the MRI results with semiquantitative methods. At 18 months, subjects underwent MRI using the same sequences as performed at baseline. Baseline and 18-month MRI results were read using the MRI OA Knee Score (MOAKS) in pairs, unblinded to time by an experienced musculoskeletal radiologist (AG) who is an expert in semiquantitative MRI analysis of knee OA (12). In a sample of 10 subjects, the MOAKS total OA scores for this reader were closely associated to the total OA scores of another highly experienced reader, with an interclass correlation of 0.98 (13). The reader was blinded to the subject treatment and all other demographic information. We focused on 18-month change in several joint features: bone marrow lesions (BMLs), cartilage surface area, cartilage thickness, osteophyte size, effusion-synovitis, and Hoffa-synovitis. Given the biomechanical models demonstrating greater contact pressures associated with APM, we envisioned that the most striking effects would be observed in cartilage damage (with contact pressures transmitted directly to cartilage) and osteophytes (bony enlargement in response to greater contact pressure) (14,15). In the MOAKS system, each joint feature is divided into subregions, and each subregion is scored on an ordinal scale from 0–3. We assessed change in each feature as described below.

BMLs. BML size is assessed in 14 subregions, and thus the change in number of subregions affected has a theoretical range of –14 to 14 because BMLs can both develop and resolve. We assessed the change in the number of subregions affected by any BML (i.e., with a score >0). We assessed the maximum advancement in BML size score across all subregions, which was grouped into "no change," "advancement by 1 grade," and "advancement by 2+ grades." We also assessed whether there were any subregions with improvement in score and whether there were any subregions with advancement in score.

Cartilage. Cartilage surface area and thickness were analyzed separately. We assessed the number of subregions with advancement, the number of subregions with new cartilage damage (i.e., a score of 0 at baseline and >0 at follow-up), and the maximum advancement across all subregions. The number of subregions with advancement and the number of subregions with new damage have a possible range of 0–14; based on distribution these were grouped into 0, 1, 2+ subregions. Maximum

advancement was grouped into "no change," "advancement by 1 grade," and "advancement by 2+ grades."

Osteophytes. We assessed the number of locations with advancement, the number of locations with new osteophytes, and the maximum advancement across all locations. Osteophytes are scored in 12 locations, and thus the number of locations with advancement and the number of locations with new osteophyte has a possible range of 0–12. Based on distribution, the number of locations with advancement was grouped into 0, 1, 2+ subregions, and the number of locations with new osteophytes and the maximum advancement across all subregions were grouped into "no advancement" versus "any advancement."

Synovitis. Effusion-synovitis represents a combination of effusion and synovial thickening, and Hoffa-synovitis is seen as hyperintensity on fat-suppressed water sensitive sequences and is a sensitive but not specific surrogate marker for the true synovitis. They are each rated on an ordinal scale from 0–3. Changes were classified as "improvement," "no change," and "advancement."

Statistical analyses. We first evaluated the association between baseline characteristics and treatment group to ensure that the groups were balanced after excluding crossovers. Baseline Kellgren/Lawrence (K/L) grade was imbalanced between the treatment groups and was thus adjusted in multivariable models. For each joint feature considered, we used multinomial logistic regression with structural advancement of that feature as the dependent variable and treatment group as the independent variable. We calculated odds ratios (ORs) with associated 95% confidence intervals (95% CIs), where the OR represents the increased odds of experiencing structural advancement for subjects receiving APM versus PT. To adjust for multiple testing, we used the Holm step-down procedure (16,17).

The primary analysis compared subjects randomized to and receiving APM with subjects randomized to and receiving PT without crossover to APM. The analysis is mechanistic in focus; consequently, it examines how the treatment received (surgical versus nonoperative) affects progression. Subjects who crossed over from PT to APM had higher baseline pain and slower initial clinical improvement. Since these factors may be associated with more rapid structural progression, we did not include the crossovers in the primary analysis (18). We performed a secondary as-treated analysis, also mechanistic in focus, in which subjects crossing over from PT to APM within 6 months of randomization were analyzed in the surgical group. Finally, we performed an intent-to-treat (ITT) analysis, in which subjects were analyzed according to randomization group irrespective of treatment received. This analysis was considered important because baseline factors, known and unknown, could be more important and have a bigger impact than the intervention. Subjects crossing over from PT to APM between 6 and 18 months were excluded from all analyses to ensure that subjects analyzed in the surgical arm in as-treated analyses were exposed to surgery for at least 12 months. Subjects randomized to the APM arm who did not receive surgery were also excluded. This group was very small, and unlike the PT-to-APM crossovers, where we have information on treatment received (number of PT visits, date of surgery), we do not know what other treatment

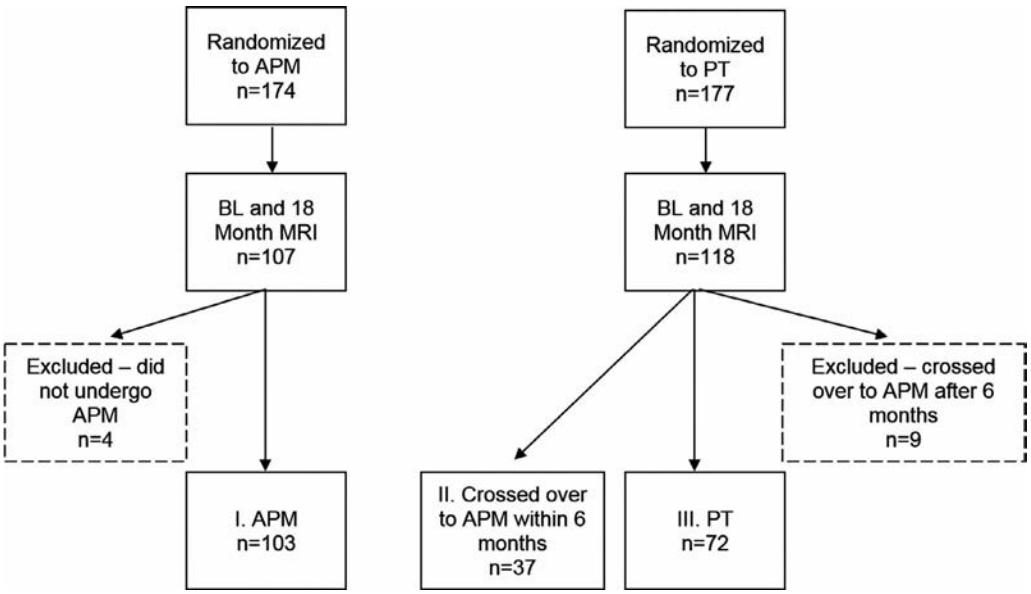

Figure 1. Sample details and analytic cohorts. A total of 351 subjects were enrolled and randomized in the Meniscal Tear in Osteoarthritis Research Trial, and 225 had both baseline and 18-month results from magnetic resonance imaging (MRI). A total of 13 were excluded from all analyses, leaving 103 in the arthroscopic partial meniscectomy (APM) group, 37 in the APM to physical therapy (PT) crossover group, and 72 in the PT group. The primary analysis is APM (bottom left) versus PT (bottom right). The first secondary analysis is as-treated: APM plus APM to PT crossover (bottom left plus bottom middle) versus PT (bottom right). The second secondary analysis is intent-to-treat: APM (bottom left) versus APM to PT crossover plus PT (bottom middle plus bottom right). BL = baseline.

courses, if any, this group pursued. When these subjects were included in the analyses, the results did not differ meaningfully.

Sensitivity analyses to examine the impact of missing data were conducted. For each outcome, we used multiple imputation (MI) to impute an outcome for those subjects missing baseline and/or 18-month MRI data (19,20). We did this under 2 different assumptions: first, we assumed that the missing data were associated with observed covariates (treatment group, K/L grade, sex, race, and baseline MOAKS if available [missing at random (MAR)]). Then, we assumed that the missing data were associated both with observed covariates and with unobserved outcomes; that is, that structural progression may be better or worse than expected based on observed covariates alone (missing not at random [MNAR]). We took a so-called tipping-point approach, asking how severe the missing data mechanism must be in order to change the study's conclusions (21). To do this, we imputed data under various not-at-random mechanisms ranging from more (MNAR1) to less (MNAR5) plausible. Details of each mechanism are described in Supplementary Appendix A, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23891/abstract. All analyses were conducted using SAS, version 9.4.

RESULTS

Cohort characteristics. A total of 351 subjects were randomized, and 225 subjects had both baseline and 18-month MRI results. Of the 225 with paired MRI data, 9 subjects crossed over from PT to APM between 6 and 18 months, and 4 subjects were randomized to APM but did not undergo surgery (Figure 1). These 13 subjects (5.8%) were excluded from all analyses. A total of 175 subjects were included in the primary analysis: 103 were randomized to and underwent APM, and 72 were randomized to PT and did not cross over. An additional 37 patients were randomized to PT and crossed over to APM in the first 6 months. Thus, the secondary as-treated analysis consisted of 212 subjects (140 subjects in the APM group and 72 in the PT group), and the ITT analysis consisted of 212 subjects (103 in the APM group and 109 in the PT group). The included subjects did not differ on baseline characteristics compared to the subjects excluded (see Supplementary Tables 1 and 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23891/abstract).

The primary analytic sample was 56% female sex and 89% white race. The mean ± SD age was 59 ± 7 years, and the mean ± SD baseline score for the Knee Injury and Osteoarthritis Outcome Score pain subscale (0–100 scale; 100 indicates worst pain) was 45 ± 16. The treatment groups were balanced on baseline demographics and clinical characteristics, with the exception of K/L grade. The APM group had a higher percentage of patients with K/L grade 3 and a lower percentage of patients with K/L grade 2 compared to the PT group (Table 1). The treatment groups were balanced on baseline MOAKS (see Supplementary Table 3, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.23891/abstract).

Table 1. Cohort characteristics*

	Treatment	
	APM plus PT (n = 103)	PT alone (n = 72)
Sex		
Male	44 (43)	33 (46)
Female	59 (57)	39 (54)
Race		
Nonwhite	12 (12)	8 (11)
White	91 (88)	64 (89)
Age, mean ± SD years	58.9 ± 7.9	58.4 ± 6.1
BMI, mean ± SD, kg/m^2	29.8 ± 6.1	30.0 ± 5.3
Baseline KOOS pain score, mean ± SD†	44.7 ± 15.4	46.1 ± 17.2
Baseline WOMAC pain score, mean ± SD†	37.8 ± 17.2	40.7 ± 17.8
Baseline WOMAC function score, mean ± SD†	35.7 ± 17.5	38.0 ± 19.5
Baseline K/L grade		
0	23 (22)	14 (19)
1	26 (25)	19 (26)
2	23 (22)	22 (31)
3	31 (30)	17 (24)
Meniscal tear category‡		
None or signal abnormality on meniscus	0 (0)	2 (3)
Nondegenerative simple tear	13 (13)	9 (13)
Short degenerative complex tear	42 (41)	24 (33)
Long degenerative complex tear	33 (32)	19 (26)
Meniscal root tear	15 (15)	18 (25)

* Values are the frequency (%) unless indicated otherwise. APM = arthroscopic partial meniscectomy; PT = physical therapy; BMI = body mass index; KOOS = Knee Injury and Osteoarthritis Outcome Score; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index; K/L = Kellgren/Lawrence.
† Range 0–100; 100 indicates worst pain.
‡ Based on central readings; patients were enrolled on the basis of readings at local centers.

The median number of days between randomization and intervention start (surgery or first PT visit) was 21 for the APM group and 9 for the PT group.The median time between baseline MRI and 18-month MRI was 579 days (19.1 months); the median time between randomization and 18-month MRI was 542 days (17.9 months).

Change in joint features. *Cartilage*. The number of subregions with advancement in cartilage surface area score ranged 0–7 with a mean ± SD score of 1.9 ± 1.9. The mean ± SD number of subregions with advancement in cartilage surface area score was 2.3 ± 1.9 in the APM group compared to 1.3 ± 1.6 in the PT group (Figure 2A). Among subjects undergoing APM, 19% had 0 subregions with advancement in cartilage surface area score, 21% had 1 subregion with advancement, and 60%

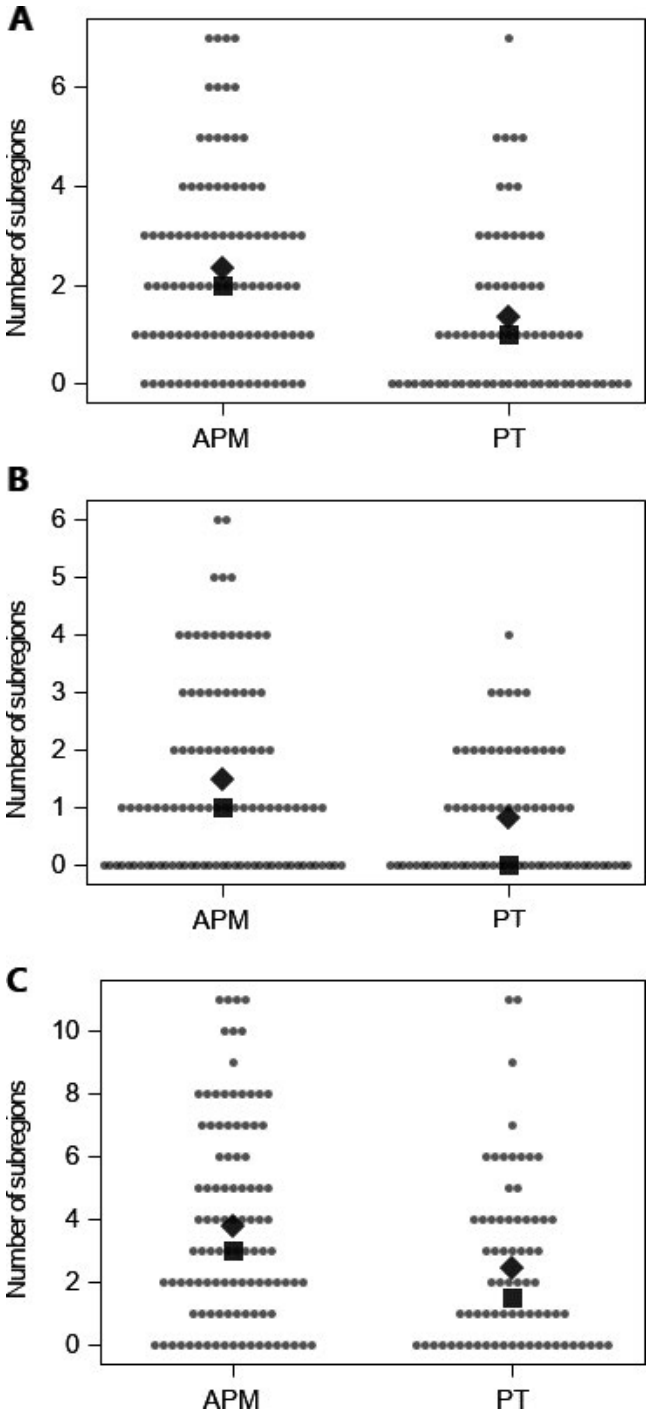

Figure 2. Early magnetic resonance imaging (MRI)–based advancement in cartilage and osteophytes by treatment group (primary analysis). Each panel shows the distribution of MRI-based advancement by treatment group for **A**, cartilage surface area, **B**, cartilage thickness, and **C**, osteophytes. Number of subregions with advancement is along the y-axis, and treatment group (arthroscopic partial meniscectomy [APM] versus physical therapy [PT]) is along the x-axis. Circles represent individual participants, diamonds the mean, and squares the median.

had 2+ subregions with advancement. Among subjects in the PT arm, 43% had 0 subregions with advancement in cartilage surface area score, 24% had 1 subregion with advancement, and 33% had 2+ subregions with advancement. This translates to a 2-fold increased odds of 1 subregion with advancement (OR 2.0 [95% CI 0.9–4.8]) and a 4.2-fold increased odds of 2+ subregions with advancement for APM versus PT (OR 4.2 [95% CI 2.0–9.0]) (Table 2). We also found significantly increased odds of advancement for APM versus PT when evaluating the number of subregions affected by cartilage surface area damage and maximum advancement in damage score (Table 2).

The number of subregions with advancement in cartilage thickness score ranged 0–6 with a mean ± SD of 1.2 ± 1.4. The mean ± SD number of subregions with advancement in cartilage thickness score was 1.5 ± 1.6 in the APM group and 0.8 ± 1.0 in the PT group (Figure 2B). Of subjects undergoing APM, 38% had 2+ subregions with advancement compared to 26% of those receiving PT. Compared to subjects receiving PT, subjects receiving APM had approximately 2-fold elevated odds of having 2+ subregions with advancement, but these associations did not reach statistical significance (OR 1.98 [95% CI 0.96–4.10]) (Table 2). Similarly, patients who had APM had 2-fold greater odds of advancement in the number of subregions affected by reduced cartilage thickness scores and maximum advancement in cartilage thickness score; these associations did not reach statistical significance (Table 2).

Osteophytes. The number of locations with advancement in osteophyte score ranged 0–11 with a mean ± SD of 3.2 ± 3.0. The mean ± SD number of locations with advancement in osteophyte score was 3.8 ± 3.2 in the APM group compared to 2.4 ± 2.7 in the PT group (Figure 2C). Osteophyte advancement was frequent: 82% of subjects in the APM group and 68% of subjects in the PT group experienced advancement in MRI-based osteophyte score in at least 1 subregion. Subjects undergoing APM had a 2.6-fold increased odds of having 2+ subregions with advancement in osteophyte score compared to 0 subregions (OR 2.6 [95% CI 1.3–5.6]) (Table 2).

BMLs. The change in the number of subregions affected by BMLs ranged from –4 to 7 with a mean ± SD of 0.4 ± 1.4. The mean ± SD change in the number of subregions with any BML was 0.6 ± 1.6 subregions in the APM group and 0.2 ± 1.1 in the PT group (Figure 3A). Treatment was not significantly associated with change in BMLs (Table 2).

Synovitis. Of subjects undergoing APM, 14% experienced advancement in Hoffa-synovitis compared to 10% of subjects receiving PT (Figure 3B). This difference was not statistically significant. Of subjects undergoing APM, 24% experienced advancement in effusion-synovitis, 45% had no change, and 31% improved (Figure 3C). Of subjects receiving PT, 8% experienced advancement, 40% no change, and 51% improvement. The adjusted odds of advancement versus improvement associated with APM was 5.0 (95% CI 1.8–13.8).

Secondary analysis. The subjects included in the secondary analysis did not differ on baseline characteristics compared to the subjects who were excluded (see Supplementary Tables 4 and 5,

Table 2. Association between treatment group and advancement, primary analysis*

	APM (n = 103)	PT (n = 72)	*P*	APM vs. PT, OR (95% CI)
Cartilage surface area				
Number of SRs with advancement in cartilage surface area score			0.0008†	
0 SRs with advancement	19 (19)	31 (43)		
1 SR with advancement	21 (21)	17 (24)		2.03 (0.85–4.82)
2+ SRs with advancement	61 (60)	24 (33)		4.22 (1.99–8.96)
Number of additional SRs affected by any cartilage surface area damage			0.0075†	
0 additional SRs affected	35 (35)	42 (58)		
1 additional SR affected	35 (35)	16 (22)		2.82 (1.32–6.01)
2+ additional SRs affected	31 (31)	14 (19)		2.67 (1.21–5.87)
Maximum advancement in cartilage surface area score across all SRs			0.0015†	
No change	19 (19)	31 (43)		
Advance by 1 grade	33 (33)	22 (31)		2.50 (1.13–5.53)
Advance by 2+ grades	49 (49)	19 (26)		4.19 (1.91–9.20)
Cartilage thickness				
Number of SRs with advancement in cartilage thickness score			0.1666	
0 SRs with advancement	39 (39)	38 (53)		
1 SR with advancement	24 (24)	15 (21)		1.55 (0.70–3.45)
2+ SRs with advancement	38 (38)	19 (26)		1.98 (0.96–4.10)
Number of additional SRs affected by any cartilage thickness damage			0.1896	
0 additional SRs affected	52 (51)	43 (60)		
1 additional SR affected	21 (21)	18 (25)		1.01 (0.48–2.16)
2+ additional SRs affected	28 (28)	11 (15)		2.09 (0.92–4.75)
Maximum advancement in cartilage thickness score across all SRs			0.1787	
No change	39 (39)	38 (53)		
Advance by 1 grade	32 (32)	19 (26)		1.66 (0.80–3.45)
Advance by 2+ grades	30 (30)	15 (21)		1.97 (0.90–4.32)
Osteophytes				
Number of locations with advancement in osteophyte score			0.0097†	
0 locations with advancement	19 (18)	23 (32)		
1 location with advancement	10 (10)	13 (18)		0.89 (0.31–2.51)
2+ locations with advancement	74 (72)	36 (50)		2.64 (1.25–5.58)
Any additional locations affected by any osteophyte			0.0230	
No	30 (29)	33 (46)		
Yes	73 (71)	39 (54)		2.10 (1.11–3.99)
Any advancement in osteophytes score across all locations			0.0388	
No	19 (18)	23 (32)		
Yes	84 (82)	49 (68)		2.13 (1.04–4.35)
BMLs				
Change in number of SRs affected by any BML			0.3595	
Improvement	19 (19)	14 (20)		
No change	38 (37)	33 (46)		0.90 (0.38–2.09)
1 additional SR affected	23 (23)	16 (23)		1.10 (0.42–2.87)
2+ additional SRs affected	22 (22)	8 (11)		2.10 (0.71–6.23)
Maximum advancement in BML size score across all SRs			0.2543	
No change	39 (38)	36 (51)		
Advance by 1 grade	31 (30)	19 (27)		1.53 (0.73–3.22)
Advance by 2+ grades	32 (31)	16 (23)		1.85 (0.85–4.02)
Any SRs with improvement in BML size score			0.9042	
No	53 (52)	38 (54)		
Yes	49 (48)	33 (46)		1.04 (0.56–1.92)
Any of SRs with advancement in BML size score			0.1102	
No	39 (38)	36 (51)		
Yes	63 (62)	35 (49)		1.68 (0.89–3.16)
Hoffa-synovitis and effusion-synovitis				
Change in Hoffa-synovitis			0.6610	
Improvement	27 (26)	17 (24)		
No change	62 (60)	47 (66)		0.80 (0.39–1.64)
Advance	14 (14)	7 (10)		1.19 (0.39–3.62)
Change in effusion-synovitis			0.0063†	
Improvement	32 (31)	37 (51)		
No change	46 (45)	29 (40)		1.84 (0.94–3.59)
Advance	25 (24)	6 (8)		4.99 (1.80–13.85)

* Values are the number (%) unless indicated otherwise. Analysis adjusted for baseline Kellgren/Lawrence grade. APM = arthroscopic partial meniscectomy; PT = physical therapy; SR = subregion; BML = bone marrow lesion.
† Statistically significant after Holm correction.

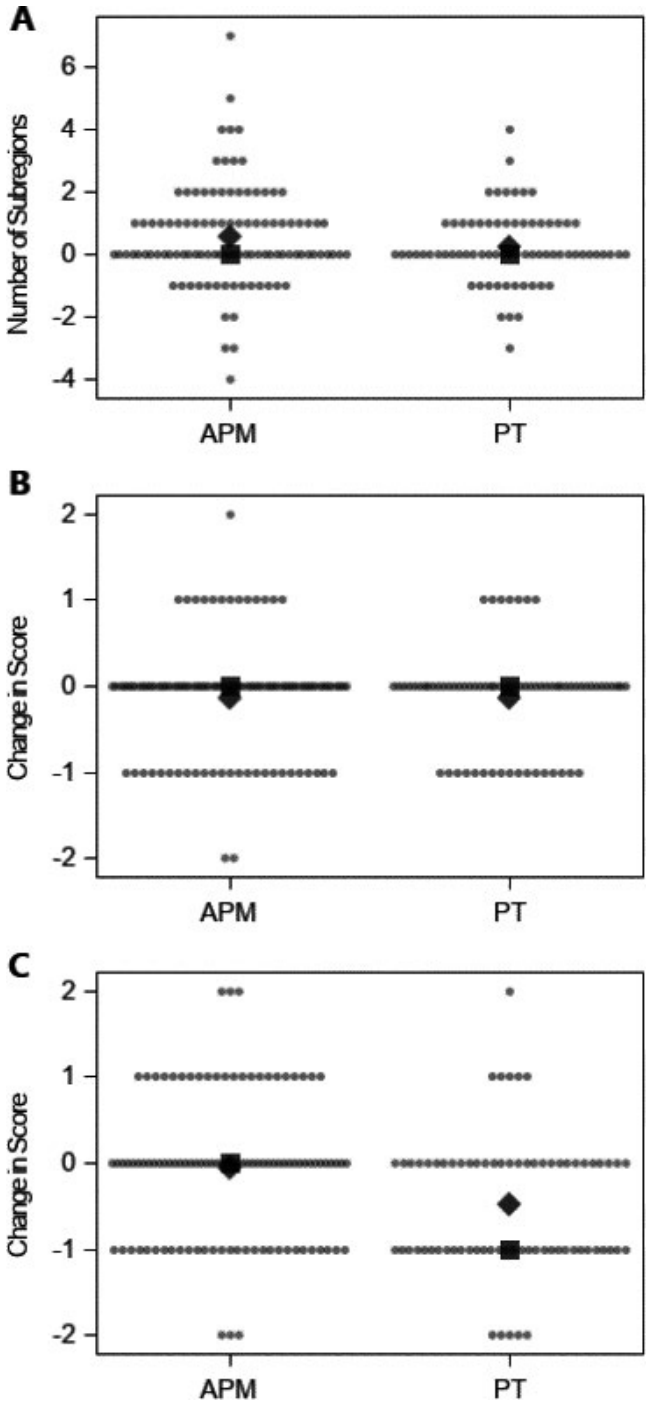

Figure 3. Early magnetic resonance imaging (MRI)–based advancement in bone marrow lesion (BML), Hoffa-synovitis, and effusion-synovitis by treatment group (primary analysis). Each panel shows the distribution of MRI-based advancement by treatment group for **A**, BML, **B**, Hoffa-synovitis, and **C**, effusion-synovitis. Advancement is along the y-axis, and treatment group (arthroscopic partial meniscectomy [APM] versus physical therapy [PT]) is along the x-axis. Advancement is measured in number of subregions for BML and in change in score for Hoffa-synovitis and effusion-synovitis. Circles represent individual participants, diamonds the mean, and squares the median.

available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23891/abstract).

As treated. Secondary analysis was performed for the as-treated sample, which included the 37 subjects who crossed over from PT to APM in the APM group. Results were similar to the main analysis, with statistically significant differences between treatment arms in cartilage surface area, osteophytes, and effusion-synovitis (see Supplementary Table 6, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23891/abstract). As in the primary analysis, we did not find significant associations between treatment group and changes in BML, cartilage thickness, or Hoffa-synovitis.

ITT. The ITT analysis included the 37 subjects who crossed over from PT to APM in the PT group. Results were similar to the main analysis, with increased odds of advancement of cartilage surface area and advancement in effusion-synovitis in the APM versus PT groups. The odds of advancement in osteophytes were in the same direction but attenuated in this analysis compared to the primary and as-treated analyses and did not reach statistical significance (see Supplementary Table 7, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23891/abstract). As in the primary and secondary analyses, we did not find significant associations between treatment group and changes in BML, cartilage thickness, or Hoffa-synovitis.

Sensitivity analysis for missing data. Sensitivity analysis with MI for missing data demonstrated similar associations as the main analysis under a MAR mechanism. That is, if we assume that patients missing 18-month cartilage-change data are similar to those with data or that we can reasonably impute change from subject characteristics (age, sex, K/L grade, baseline MOAKS if available), then our conclusions do not change (Table 3). As we change the missing data mechanism and assume that subjects in the APM group with missing data are progressing less than observed APM subjects, and/or PT subjects with missing data are progressing more than observed PT subjects, the associations are attenuated. Generally, in order to change the conclusion about the association between treatment group and progression, we would have to assume an extreme missing data mechanism: namely, that PT subjects with missing data progress in the same fashion as observed APM subjects and APM subjects with missing data progress in the same fashion as observed PT subjects.

DISCUSSION

We evaluated data from an RCT of APM with PT versus PT alone and found that both treatment groups had substantial early

Table 3. Association between treatment group and progression, sensitivity analysis with multiple imputation for missing data*

	Primary analysis	Secondary as-treated analysis	MAR	MNAR1	MNAR2	MNAR3	MNAR4	MNAR5
Cartilage surface area								
Number of SRs with advancement in cartilage surface area score								
0 SRs with advancement (Ref.)								
1 SR with advancement	2.0 (0.9–4.8)	1.7 (0.8–3.8)	1.7 (0.7–4.0)	1.2 (0.5–2.7)	1.4 (0.6–3.0)	1.3 (0.5–3.2)	1.4 (0.7–2.9)	1.0 (0.4–2.2)
2+ SRs with advancement	4.2 (2.0–9.0)	3.3 (1.7–6.5)	3.3 (1.6–6.5)	2.5 (1.2–5.2)	2.0 (1.01–3.9)	2.0 (0.97–4.2)	2.2 (1.2–4.0)	1.4 (0.7–2.8)
Number of additional SRs affected by any cartilage surface area damage								
0 additional SRs affected (Ref.)								
1 additional SR affected	2.8 (1.3–6.0)	2.4 (1.2–4.8)	2.6 (1.3–5.2)	2.0 (0.98–4.1)	1.9 (0.9–4.0)	1.8 (0.9–3.5)	1.7 (0.9–3.4)	1.4 (0.7–2.6)
2+ additional SRs affected	2.7 (1.2–5.9)	2.1 (1.0–4.5)	2.3 (1.1–5.0)	1.8 (0.8–3.7)	1.8 (0.8–4.0)	1.5 (0.7–3.3)	1.7 (0.8–3.6)	1.3 (0.6–2.6)
Maximum advancement in cartilage surface area score across all SRs								
No change (Ref.)								
Advance by 1 grade	2.5 (1.1–5.5)	2.0 (1.0–4.2)	1.9 (1.0–3.9)	1.3 (0.6–2.6)	1.6 (0.8–3.2)	1.5 (0.7–3.4)	1.6 (0.8–3.1)	1.1 (0.5–2.1)
Advance by 2+ grades	4.2 (1.9–9.2)	3.3 (1.6–6.7)	3.2 (1.6–6.5)	2.4 (1.2–4.9)	2.1 (1.1–4.2)	2.1 (1.0–4.5)	2.1 (1.1–4.0)	1.4 (0.7–2.8)
Cartilage thickness								
Number of SRs with advancement in cartilage thickness score								
0 SRs with advancement (Ref.)								
1 SR with advancement	1.6 (0.7–3.4)	1.4 (0.7–3.1)	1.4 (0.7–2.9)	1.0 (0.4–2.2)	1.2 (0.6–2.6)	1.4 (0.6–3.1)	1.3 (0.6–2.8)	1.1 (0.5–2.4)
2+ SRs with advancement	2.0 (1.0–4.1)	1.9 (1.0–3.8)	1.9 (1.0–3.6)	1.5 (0.7–3.0)	1.4 (0.7–2.9)	1.4 (0.7–3.0)	1.5 (0.8–3.0)	1.1 (0.6–2.3)
Number of additional SRs affected by any cartilage thickness damage								
0 additional SRs affected (Ref.)								
1 additional SR affected	1.0 (0.5–2.2)	0.9 (0.5–1.9)	0.8 (0.4–1.6)	0.7 (0.3–1.6)	0.7 (0.3–1.6)	0.9 (0.4–1.7)	0.9 (0.4–2.0)	0.9 (0.4–1.7)
2+ additional SRs affected	2.1 (0.9–4.8)	1.9 (0.9–4.1)	1.9 (0.9–4.1)	1.7 (0.8–3.5)	1.7 (0.8–3.7)	1.5 (0.7–3.4)	1.5 (0.7–3.1)	1.2 (0.6–2.6)
Maximum advancement in cartilage thickness score across all SRs								
No change (Ref.)								
Advance by 1 grade	1.7 (0.8–3.4)	1.7 (0.8–3.3)	1.6 (0.8–3.3)	1.2 (0.6–2.5)	1.5 (0.7–3.2)	1.5 (0.7–3.0)	1.3 (0.7–2.7)	1.2 (0.6–2.3)
Advance by 2+ grades	2.0 (0.9–4.3)	1.8 (0.8–3.7)	1.8 (0.9–3.7)	1.4 (0.7–2.8)	1.4 (0.7–2.9)	1.5 (0.7–3.2)	1.4 (0.7–2.7)	1.3 (0.6–2.6)
Osteophytes								
Number of locations with advancement in osteophyte score								
0 locations with advancement (Ref.)								
1 location with advancement	0.9 (0.3–2.5)	0.8 (0.3–2.0)	0.7 (0.3–1.8)	0.7 (0.3–1.6)	0.6 (0.2–1.7)	0.7 (0.3–1.9)	0.8 (0.3–2.1)	0.7 (0.3–1.9)
2+ locations with advancement	2.6 (1.3–5.6)	2.5 (1.2–5.0)	2.4 (1.1–5.0)	1.9 (1.0–3.7)	1.7 (0.9–3.4)	1.7 (0.9–3.3)	1.9 (1.0–3.6)	1.2 (0.6–2.3)
Any additional locations affected by any osteophyte								
No (Ref.)								
Yes	2.1 (1.1–4.0)	2.0 (1.1–3.6)	1.9 (1.1–3.4)	1.6 (0.8–3.0)	1.3 (0.7–2.4)	1.5 (0.8–2.8)	1.5 (0.8–2.6)	1.2 (0.7–2.1)
Any advancement in osteophytes score across all locations								
No (Ref.)								
Yes	2.1 (1.0–4.4)	2.0 (1.0–3.9)	2.0 (1.0–3.9)	1.8 (0.8–3.7)	1.6 (0.8–3.1)	1.5 (0.8–2.9)	1.5 (0.8–3.0)	1.3 (0.7–2.5)

(Continued)

Table 3. *(Cont'd)*

	Primary analysis	Secondary as-treated analysis	MAR	MNAR1	MNAR2	MNAR3	MNAR4	MNAR5
Bone marrow lesions								
Change in number of subregions affected by any BML								
Improvement (Ref.)								
No change	0.9 (0.4–2.1)	1.2 (0.6–2.8)	1.2 (0.6–2.6)	0.9 (0.4–1.9)	0.9 (0.4–2.1)	1.0 (0.4–2.3)	0.9 (0.4–1.9)	0.8 (0.4–1.8)
1 additional SR affected	1.1 (0.4–2.9)	1.3 (0.5–3.3)	1.1 (0.4–2.9)	0.9 (0.3–2.2)	1.0 (0.4–2.2)	1.0 (0.5–2.4)	0.9 (0.4–2.2)	0.9 (0.4–2.1)
2+ additional SR affected	2.1 (0.7–6.2)	2.1 (0.7–6.0)	2.0 (0.7–5.6)	1.4 (0.5–4.4)	1.6 (0.5–5.0)	1.6 (0.5–5.2)	1.5 (0.5–4.3)	1.3 (0.5–3.5)
Maximum advancement in BML size score across all SRs								
No change (Ref.)								
Advance by 1 grade	1.5 (0.7–3.2)	1.5 (0.7–2.9)	1.4 (0.7–2.9)	1.1 (0.6–2.2)	1.3 (0.6–2.8)	1.3 (0.6–2.7)	1.3 (0.6–2.6)	1.1 (0.5–2.2)
Advance by 2+ grades	1.9 (0.9–4.0)	1.3 (0.6–2.8)	1.3 (0.6–2.9)	1.0 (0.5–1.9)	1.1 (0.6–2.3)	1.1 (0.5–2.4)	1.1 (0.6–2.1)	1.1 (0.5–2.2)
Any SRs with improvement in BML size score								
No (Ref.)								
Yes	1.0 (0.6–1.9)	0.8 (0.4–1.4)	0.8 (0.5–1.5)	0.7 (0.4–1.3)	1.0 (0.5–1.8)	0.9 (0.5–1.7)	0.9 (0.5–1.6)	1.0 (0.5–1.9)
Any of SRs with advancement in BML size score								
No (Ref.)								
Yes	1.7 (0.9–3.2)	1.4 (0.8–2.5)	1.5 (0.8–2.6)	1.1 (0.6–2.0)	1.3 (0.7–2.3)	1.3 (0.7–2.3)	1.2 (0.7–2.3)	1.2 (0.6–2.2)
Hoffa-synovitis and effusion-synovitis								
Change in Hoffa-synovitis								
Improvement (Ref.)								
No change	0.8 (0.4–1.6)	0.8 (0.4–1.6)	0.8 (0.4–1.7)	0.7 (0.3–1.5)	0.7 (0.4–1.4)	0.8 (0.4–1.6)	0.8 (0.4–1.6)	0.8 (0.4–1.6)
Advance	1.2 (0.4–3.6)	1.7 (0.6–4.7)	1.8 (0.6–5.5)	1.3 (0.5–3.5)	1.5 (0.5–4.4)	1.5 (0.5–4.2)	1.2 (0.4–3.6)	1.1 (0.4–3.1)
Change in effusion-synovitis								
Improvement (Ref.)								
No change	1.8 (0.9–3.6)	1.9 (1.0–3.6)	1.9 (1.02–3.4)	1.6 (0.9–3.0)	1.6 (0.9–2.9)	1.4 (0.8–2.5)	1.6 (0.8–3.0)	1.2 (0.6–2.1)
Advance	5.0 (1.8–13.8)	4.1 (1.5–10.9)	4.1 (1.5–11.4)	3.7 (1.3–10.6)	2.6 (1.0–6.5)	2.7 (1.0–7.4	2.5 (1.0–6.2)	1.7 (0.7–4.2)

* Values are the odds ratio (95% confidence interval) for progression for arthroscopic partial meniscectomy vs. physical therapy presented in cells. MAR = missing at random; MNAR = missing not at random; SR = subregion; Ref. = reference; BML = bone marrow lesion.

advancement of MRI-based biomarkers of each of the structural joint features examined. Patients undergoing APM had greater early advancement in MRI-based markers over 18 months than those treated nonoperatively for cartilage surface area, osteophytes, and effusion-synovitis.

Two RCTs found no differences in radiographic advancement between subjects treated with APM and those treated nonoperatively for degenerative meniscal tear (5,6). However, radiographic OA grade is an insensitive marker of structural change (22,23); both studies found radiographic advancement rates of <5%. Roemer et al found that both meniscal damage and partial meniscectomy were associated with incident OA (K/L grade 2) over 4 years in a nested case-control sample from the Osteoarthritis Initiative (10). In addition, the authors used MOAKS to evaluate MRI-based cartilage progression, defining progression as any increase in either size or thickness of cartilage damage. In the incident OA cases, partial meniscectomy was associated with worsening cartilage damage compared to knees with meniscal damage and without meniscectomy, and to knees without meniscal damage. However, only 26 knees underwent meniscal surgery and had MRI results. Our analysis builds on the work of Roemer et al (10) by taking advantage of the large MeTeOR Trial cohort. In MeTeOR, all subjects had documented meniscal tear, and all had substantial enough pain and functional limitation that subjects and their enrolling surgeons were prepared to proceed to APM. This balancing of structural and symptom severity between treatment groups is difficult to achieve in observational studies. The surgeries in MeTeOR were done in a uniform manner, and follow-ups were at regular intervals post randomization. Like Roemer et al (10), we found associations between APM and subsequent cartilage advancement.

To our knowledge, this is the first study to evaluate early MRI-based changes in a follow-up evaluation of data from an RCT of APM versus nonoperative therapy. We found that MeTeOR participants who had APM had higher likelihood of MRI-based advancement in cartilage surface area, osteophytes, and effusion-synovitis. We did not find significant associations between treatment type and advancement in BMLs or Hoffa-synovitis. The lack of association between treatment and BMLs and Hoffa synovitis may reflect the transient nature of BMLs and synovitis; these features do not reflect cumulative damage as do cartilage damage and osteophytosis. The results were similar in the main per-protocol analysis and in the secondary as-treated and ITT analyses.

The clinical relevance of these early MRI findings remains uncertain. It will be important to determine whether subjects who demonstrate these changes in imaging findings over 18 months are at higher risk of worsening in symptom severity, functional limitation, and total joint replacement over subsequent follow-up. The findings underscore the importance of clinical follow-up of this cohort and, more generally, of individuals with meniscal tear treated either operatively or nonoperatively.

Only 225 of 351 randomized patients (64%) had MRI data available both at baseline and 18 months. Although our analyses found associations between some aspects of structural disease progression and treatment group, we were concerned about the amount of missing data. Our tipping-point sensitivity analysis with MI suggested that it would take an extreme missing data mechanism to change the conclusions. We would have to assume that PT patients with missing data are actually advancing at rates similar to the APM patients, and that APM subjects with missing data actually progress at rates similar to PT subjects. Although this extreme scenario does not seem plausible, we can never rule out a missing data mechanism with 100% certainty.

These results should be interpreted within the context of the study limitations. The primary analytic sample was limited to those subjects undergoing MRI at 18 months. Of patients randomized to PT, 31% crossed over to APM within 6 months of randomization. Subjects crossing over to APM had shorter symptom duration and greater baseline pain; thus, the balancing of confounders inherent to randomization may have been disrupted. We evaluated differences in known potential confounders between the groups and adjusted where necessary. However, we cannot be certain that the groups were balanced on unknown confounders. To further minimize risk of bias, we conducted 3 sets of analyses, all adjusted for K/L grade and factors imbalanced at baseline: the primary analysis that excluded crossovers, a sensitivity as-treated analysis that included crossovers in the APM group, and an ITT analysis that included crossovers in the PT group. All 3 of these analyses yielded similar conclusions. Due to the multinomial nature of many of the outcomes variables, we used logistic regression and present ORs. ORs overstate relative risks, especially when the prevalence of the outcome is high, as in this analysis (24). Thus, these ORs should not be interpreted as relative risks. Caution should be taken in generalizing these results to a more general knee OA cohort. First, inclusion criteria for the MeTeOR Trial included evidence of meniscal tear on MRI results and symptoms consistent with torn meniscus (i.e., clicking, catching, popping). Each patient had to be willing to undergo APM if randomized to the APM group (11). Thus, the MeTeOR Trial may be more generalizable to patients with knee OA and meniscal tear with symptoms. Patients were recruited from academic medical centers, and only 26% of eligible subjects agreed to participate in the MeTeOR RCT (4). Finally, this analysis is a secondary analysis of an RCT, and as such, we did not conduct a formal power analysis (25). To address the uncertainty in our parameter estimates, we included 95% CIs (26).

In discussing treatment options for symptomatic meniscal tear, patients and providers must weigh the potential benefits and risks of treatment options, including these findings on structural advancement. Future work will assess the association between early structural advancement and subsequent pain, function, and risk of total knee replacement. Clinicians should be aware that regardless of treatment, there was MRI evidence of progression. At this point, the clinical meaning of the MRI-based changes doc-

umented in this study is unknown. Assessing the relevance of these MRI-based changes is an important research priority.

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. Collins 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. Collins, Losina, Marx, Guermazi, Jarraya, Jones, Levy, Mandl, Martin, Wright, Spindler, Katz.

Acquisition of data. Losina, Guermazi, Jarraya, Katz.

Analysis and interpretation of data. Collins, Losina, Katz.

REFERENCES

1. Deshpande BR, Katz JN, Solomon DH, Yelin EH, Hunter DJ, Messier SP, et al. Number of persons with symptomatic knee osteoarthritis in the US: impact of race and ethnicity, age, sex, and obesity. Arthritis Care Res (Hoboken) 2016;68:1743–50.
2. Bhattacharyya T, Gale D, Dewire P, Totterman S, Gale ME, McLaughlin S, et al. The clinical importance of meniscal tears demonstrated by magnetic resonance imaging in osteoarthritis of the knee. J Bone Joint Surg Am 2003;85:4–9.
3. Thorlund JB, Juhl CB, Roos EM, Lohmander LS. Arthroscopic surgery for degenerative knee: systematic review and meta-analysis of benefits and harms. Br J Sports Med 2015;49:1229–35.
4. Katz JN, Brophy RH, Chaisson CE, De Chaves L, Cole BJ, Dahm DL, et al. Surgery versus physical therapy for a meniscal tear and osteoarthritis. N Engl J Med 2013;368:1675–84.
5. Herrlin SV, Wange PO, Lapidus G, Hallander M, Werner S, Weidenhielm L. Is arthroscopic surgery beneficial in treating non-traumatic, degenerative medial meniscal tears? A five year follow-up. Knee Surg Sports Traumatol Arthrosc 2013;21:358–64.
6. Yim JH, Seon JK, Song EK, Choi JI, Kim MC, Lee KB, et al. A comparative study of meniscectomy and nonoperative treatment for degenerative horizontal tears of the medial meniscus. Am J Sports Med 2013;41:1565–70.
7. Kise NJ, Risberg MA, Stensrud S, Ranstam J, Engebretsen L, Roos EM. Exercise therapy versus arthroscopic partial meniscectomy for degenerative meniscal tear in middle aged patients: randomised controlled trial with two year follow-up. BMJ 2016;354:i3740.
8. Englund M, Guermazi A, Roemer FW, Aliabadi P, Yang M, Lewis CE, et al. Meniscal tear in knees without surgery and the development of radiographic osteoarthritis among middle-aged and elderly persons: the multicenter osteoarthritis study. Arthritis Rheum 2009;60:831–9.
9. Englund M, Roos EM, Lohmander LS. Impact of type of meniscal tear on radiographic and symptomatic knee osteoarthritis: a sixteen-year followup of meniscectomy with matched controls. Arthritis Rheum 2003;48:2178–87.
10. Roemer FW, Kwoh CK, Hannon MJ, Hunter DJ, Eckstein F, Grago J, et al. Partial meniscectomy is associated with increased risk of incident radiographic osteoarthritis and worsening cartilage damage in the following year. Eur Radiol 2017;27:404–13.
11. Katz JN, Chaisson CE, Cole B, Guermazi A, Hunter DJ, Jones M, et al. The meteor trial (meniscal tear in osteoarthritis research): rationale and design features. Contemp Clin Trials 2012;33:1189–96.
12. Hunter DJ, Guermazi A, Lo GH, Grainger AJ, Conaghan PG, Boudreau RM, et al. Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). Osteoarthritis Cartilage 2011;19:990–1002.
13. MacFarlane LA, Yang H, Collins JE, Guermazi A, Jones MH, Teeple E, et al. Associations among meniscal damage, meniscal symptoms and knee pain severity. Osteoarthritis Cartilage 2017;25:850–57.
14. Bedi A, Kelly NH, Baad M, Fox AJ, Brophy RH, Warren RF, et al. Dynamic contact mechanics of the medial meniscus as a function of radial tear, repair, and partial meniscectomy. J Bone Joint Surg Am 2010;92:1398–408.
15. Zhang AL, Miller SL, Coughlin DG, Lotz JC, Feeley BT. Tibiofemoral contact pressures in radial tears of the meniscus treated with all-inside repair, inside-out repair and partial meniscectomy. Knee 2015;22:400–4.
16. Aickin M, Gensler H. Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm methods. Am J Public Health 1996;86:726–8.
17. Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat 1979;6:65–70.
18. Katz JN, Wright J, Spindler KP, Mandl LA, Safran-Norton CE, Reinke EK, et al. Predictors and outcomes of crossover to surgery from physical therapy for meniscal tear and osteoarthritis: a randomized trial comparing physical therapy and surgery. J Bone Joint Surg Am 2016;98:1890–96.
19. Rubin D. Multiple imputation for nonresponse in surveys. New York: John Wiley & Sons; 1987.
20. Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. Stat Methods Med Res 2014;23:440–59.
21. Liublinska V, Rubin DB. Sensitivity analysis for a partially missing binary outcome in a two-arm randomized clinical trial. Stat Med 2014;33:4170–85.
22. Hunter DJ, Zhang W, Conaghan PG, Hirko K, Menashe L, Reichmann WM, et al. Responsiveness and reliability of MRI in knee osteoarthritis: a meta-analysis of published evidence. Osteoarthritis Cartilage 2011;19:589–605.
23. Wirth W, Duryea J, Hellio Le Graverand MP, John MR, Nevitt M, Buck RJ, et al. Direct comparison of fixed flexion, radiography and MRI in knee osteoarthritis: responsiveness data from the osteoarthritis initiative. Osteoarthritis Cartilage 2013;21:117–25.
24. Davies HT, Crombie IK, Tavakoli M. When can odds ratios mislead? BMJ 1998;316:989–91.
25. Hoenig JM, Heisey DM. The abuse of power: the pervasive fallacy of power calculations for data analysis. Am Stat 2001;55:19–24.
26. Ranstam J. Why the P-value culture is bad and confidence intervals a better alternative. Osteoarthritis Cartilage 2012;20:805–8.

APPENDIX A

Members of the MeTeOR Investigator Group are as follows: Robert H. Brophy, MD, Brian J. Cole, MD, MBA, Diane L. Dahm, MD, Lindsey A. MacFarlane, MD, MPH, Mathew J. Matava, MD, Clare E. Safran-Norton, PT, PhD, Faith Selzer, PhD, Matthew Smith, MD, Michael J. Stuart MD, and John Wright, MD.

Arthritis Care & Research
Vol. 72, No. 5, May 2020, pp 641–651
DOI 10.1002/acr.23884

American College of Rheumatology
Empowering Rheumatology Professionals

Duration of Symptom Relief and Early Trajectory of Adverse Events for Oral Nonsteroidal Antiinflammatory Drugs in Knee Osteoarthritis: A Systematic Review and Meta-Analysis

Mikala C. Osani, Elizaveta E. Vaysbrot, Mengyu Zhou, Timothy E. McAlindon, and Raveendhara R. Bannuru

Objective. Despite an extensive body of research on nonsteroidal antiinflammatory drugs (NSAIDs) in osteoarthritis, the duration of their efficacy and timeline of adverse event (AE) onset have been understudied. We conducted a systematic review and meta-analyses from 2 to 26 weeks to characterize the efficacy and AE trajectories of oral NSAIDs in knee osteoarthritis.

Methods. We searched MEDLINE, Embase, Web of Science, Google Scholar, and the Cochrane Database from inception to May 2018. Randomized controlled trials assessing the efficacy and/or safety of Federal Drug Administration–approved NSAIDs in knee osteoarthritis patients were included. Two independent reviewers assessed quality and extracted data. We calculated standardized mean differences (SMDs) and risk ratios (RRs) with 95% confidence intervals (95% CIs).

Results. We included 72 randomized controlled trials (26,424 participants). NSAIDs demonstrated moderate, statistically significant effects on pain that peaked at 2 weeks (SMD –0.43 [95% CI –0.48, –0.38]), but the magnitude of the effects decreased over time. The results for function were similar. The incidence of gastrointestinal (GI) AEs was significantly higher in NSAID users than placebo users as early as 4 weeks (RR 1.38 [95% CI 1.21, 1.57]). The incidence of cardiovascular (CV) AEs in NSAID users was not significantly different from placebo. Most GI and CV AEs were transient and of minor severity.

Conclusion. NSAIDs produced significant pain and function improvements that peaked at 2 weeks but decreased over time. The incidence of minor GI and CV AEs consistently rose, reaching significance as early as 4 weeks. Clinicians should weigh the durability of efficacy with the early onset of minor AEs along with patient tolerability and preferences when formulating an NSAID regimen.

INTRODUCTION

Osteoarthritis (OA) is a leading cause of pain and disability among adults in the US (1), with involvement of the knee joint accounting for >80% of the disease's disability burden. The prevalence of the disease is rising, and ~14 million adults in the US are now experiencing symptomatic knee OA (2–4). Since the natural history of OA is long, patients may need therapy for many years, even after arthroplasty.

Oral nonsteroidal antiinflammatory drugs (NSAIDs) are the pharmaceuticals used most frequently for pain control and are routinely recommended in OA clinical practice guidelines (5,6). In the US, 65% of patients with OA are prescribed NSAIDs (7). Given the current need to limit the use of opioid medications, NSAIDs can be expected to play an even larger role in clinical practice (8). Despite widespread use, a gap currently exists in our knowledge regarding the consistency and duration of the beneficial effects of NSAIDs on pain and functional outcomes in patients with knee

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

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

Mikala C. Osani, BA, Elizaveta E. Vaysbrot, MD, MS, Mengyu Zhou, MD, Timothy E. McAlindon, MD, MPH, Raveendhara R. Bannuru, MD, PhD: Tufts Medical Center, Boston, Massachusetts.

Dr. McAlindon has received consulting fees from Flexion, Astellas, Pfizer, Samumed, Seikugaku, and Regeneron (less than $10,000 each). Dr. Bannuru has received consulting fees from Fidia (less than $10,000). 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. E-mail: rbannuru@tuftsmedicalcenter.org.

Submitted for publication October 23, 2018; accepted in revised form March 19, 2019.

[Correction added on 22 April 2020, after first online publication: Dr. McAlindon's disclosure statement has been updated to reflect that fees received were less than $10,000 each.]

SIGNIFICANCE & INNOVATIONS

- Current research on oral nonsteroidal antiinflammatory drugs (NSAIDs) in osteoarthritis does not provide information about the durability of efficacy or the onset of early adverse events.
- We conducted meta-analyses of efficacy and safety at 2, 4, 8, 12, and 26 weeks to characterize the trajectory of efficacy and early adverse events for oral NSAIDs in knee osteoarthritis.
- Our results suggest that the beneficial effects of NSAIDs peak at 2 weeks and begin to decline by 8 weeks, whereas minor gastrointestinal and cardiovascular adverse events begin to manifest as early as 4 weeks.
- Information on the efficacy and safety trajectory of oral NSAIDs can guide clinicians and patients in selecting an appropriate NSAID treatment regimen.

OA, and the efficacy of these drugs has largely been tested in randomized controlled trials (RCTs) of short-term duration. For clinicians treating chronic conditions like OA that involve long-term management, the degree of superiority of a treatment over placebo is often of equal importance to its duration of efficacy.

In addition to the uncertainty in efficacy trajectory, there is a lack of research on the timing and evolution of adverse reactions to NSAIDs used in OA. NSAIDs are associated with cardiovascular (CV) adverse events (AEs), kidney injury, and gastrointestinal (GI) toxicity; the latter is shown to be likelier for nonselective NSAIDs than for selective cyclooxygenase 2 (COX-2) inhibitors (9–12). Furthermore, the general OA population, which is characterized by older age and a more frequent use of concomitant medications, could be at a higher risk for NSAID-associated complications. Although serious GI and CV risks such as GI bleed or myocardial infarction are associated with prolonged NSAID use, minor AEs contributing to patient discomfort may begin to manifest even when the treatment duration is relatively short (10,13).

Given the chronic nature of knee OA symptoms and the resulting need for long-term therapeutic solutions, assessing the benefits and risks of any drug, including NSAIDs, in its temporal context is important. Understanding the durability of the efficacy of oral NSAIDs, as well as the time course of onset of minor AEs, is key to this decision-making process. Therefore, we conducted a systematic review and meta-analysis to comprehensively and quantitatively characterize the efficacy trajectory of oral NSAIDs on pain and functional improvement and to summarize the timing of onset and the subsequent progression of minor GI and CV AEs in patients with knee OA.

MATERIALS AND METHODS

Data sources/searches. We searched MEDLINE, Embase, Web of Science, Google Scholar, and the Cochrane Central Register of Controlled Trials from inception to May 3, 2018 (see Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23884/abstract). We hand-searched reference lists of relevant systematic reviews and meta-analyses and within supplements of conference proceedings that had been published up to May 2018. We limited our search to randomized placebo-controlled trials involving NSAIDs in human subjects with knee OA. No restrictions were placed on publication date, status, or language. We adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines, but we elected not to register our study protocol in the PROSPERO database (see Supplementary Table 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23884/abstract).

Study selection. We included RCTs that assessed the efficacy and/or safety of Federal Drug Administration (FDA)–approved NSAIDs versus placebo in patients with knee OA. We included studies that involved multiple treatment arms, as well as studies that involved treatment modalities other than NSAIDs, as long as the study compared ≥1 FDA-approved NSAID at an approved dosage against a placebo arm. We included combined knee and hip studies if they either reported separate results for the knee or if they had included >70% knee OA patients. Nonrandomized studies and studies in which the location of OA was undefined were excluded. Each abstract recovered by the search was screened by 2 independent reviewers (MCO, RRB), in line with the preestablished inclusion and exclusion criteria. Full manuscripts of abstracts that were included after initial screening were subsequently gathered and assessed for eligibility in further detail by the same 2 reviewers (MCO, RRB). Discordant results in inclusion or exclusion that resulted during either screening stage were adjudicated by a third reviewer (EEV).

Data extraction and quality assessment. Data from each RCT were independently extracted by 2 reviewers (MCO, RRB). We drafted a data extraction form in Microsoft Excel to gather information on study and population characteristics, NSAID classification, dosage and frequency, rescue medication protocol, pain and functional outcomes, discontinuation rates and reasons, and relevant safety outcomes. We collected pain and functional outcomes that were reported by any validated scale; in the event that >1 scale was reported, results for all scales were collected. Based on their mechanisms of action, we determined 3 overarching NSAID classes: traditional (nonselective) NSAIDs, an older group of NSAIDs without a strong COX-2 selectivity (e.g., diclofenac, ibuprofen, indomethacin, naproxen, and piroxicam); selective COX-2 inhibitors, a newer class of coxib NSAIDs developed specifically for COX-2 selectivity (celecoxib was the only representative treatment); and intermediate COX inhibitors, those from the traditional cohort of NSAIDs demonstrating relative COX-2 selectivity but with a chemical structure different from coxibs (e.g., etodolac, meloxicam, and nabumetone) (14–16).

In studies assessing multiple doses of NSAIDs, we only collected data on the dose that most closely matched the recommended dosing range for OA. Information regarding the recommended dosing for this indication was obtained directly from the FDA website and/or from package inserts. In studies assessing multiple interventions against a common placebo group, we evenly divided the shared group into 2 or more smaller groups and included them as independent comparators, as referenced in the Cochrane Handbook, section 16.5.4 (17). To comprehensively assess the efficacy trajectories of NSAIDs while maintaining the robustness of our analyses, we collected pain and functional data at all reported time points and grouped the data into the following time point categories: 2 weeks (0–2 weeks), 4 weeks (3–6 weeks), 8 weeks (7–10 weeks), 12 weeks (11–16 weeks), and 26 weeks (17–29 weeks). In all circumstances, we prioritized data that were presented in manuscript text or tables over graphical data. Data that were only presented in figures or graphs were recovered using Engauge Digitizer and double-checked by a second reviewer (MCO, EEV, or RRB) (18). We transferred the outcome data from Excel into RevMan software, and study quality was assessed in RevMan using the Cochrane Risk of Bias Tool (19,20). Data extraction and quality ratings were reviewed in their entirety for consistency. Discrepancies were arbitrated by a third reviewer (EEV).

Outcome definitions. We selected the following outcomes of interest: pain, function, rate of discontinuation due to lack of efficacy, rate of discontinuation due to AEs, incidence of treatment-related AEs and serious AEs, and incidence of GI and CV AEs. Pain and functional outcomes were reported as the mean change from baseline to follow-up; in our primary analyses of all pain and all function, Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) scales were prioritized (21). If no other scales were available, nonstandard Likert scales were included in analyses of all pain and all function but were not used in any other analyses. Rates of discontinuation were reported as the number of participants who discontinued treatment or withdrew from the study due to lack of efficacy or due to any AE. Discontinuation rates were collected for active treatment periods only; we did not collect discontinuation data that were reported after treatment had been stopped or changed, or after randomization had been broken. Serious AEs (SAEs) were defined as those explicitly designated by the outcome assessors as SAEs within the study period. The criteria for SAEs have been delineated by the FDA and include AEs that are potentially life threatening or result in hospitalization, disability or permanent damage, congenital anomalies or birth defects, or death, or events that may jeopardize the patient and may require medical intervention to prevent one of the above outcomes. Treatment-related AEs were specifically collected so as to better highlight the differences between treatment and placebo groups and were defined as any AEs (or "side effects") that were described by the study investigators as treatment-related or drug-related, or were determined to be of probable, possible, and/or certain relationship to the study treatment. We excluded studies that only reported the incidence of "any adverse event" or "treatment-emergent adverse events" from our analysis of treatment-related AEs (22). We collected the incidence of GI and CV AEs as the sum total of the respective AEs at the study end points and at separate time points falling within the preestablished follow-up categories of 2, 4, 8, 12, and 26 weeks, as available. Although we anticipated that the majority of AEs would be minor due to limitations of follow-up time, all GI and CV events were counted regardless of severity. The AEs that were most commonly observed were summarized. All safety data were reported as the number of patients experiencing ≥1 event.

Statistical analysis. For continuous outcomes, we calculated standardized mean differences (SMDs) and 95% confidence intervals (95% CIs) using the mean change from baseline to follow-up. We conducted meta-analyses using random-effects models to account for methodologic and clinical heterogeneity. To allow for direct comparability of effect sizes across different outcomes and subgroups, SMDs were used for all analyses of continuous outcome measures regardless of the variation in their scales. We analyzed dichotomous outcomes using the Mantel-Haenszel method and reported the effects as risk ratios (RRs) and 95% CIs (23). Heterogeneity was assessed using the I^2 statistic (24). Analyses were conducted using RevMan software (20). Funnel plots were visually inspected for asymmetry as a means of assessing publication bias. To aid in the clinical interpretation of SMDs, we used the benchmark of 0.37 units for clinical significance (or importance) per the definition published by Wandel et al (25).

We planned the following a priori subgroup and sensitivity analyses, all of which were contingent upon the availability of data: analyses based on NSAID classification (selective COX-2 inhibitor versus intermediate COX inhibitor versus traditional NSAID), analyses limiting by pain scale (WOMAC versus visual analog scale), or functional scale (WOMAC versus any other functional scale), analyses limited to knee OA patients, analyses with potential outliers removed (conducted in the event that I^2 was ≥75%, as referenced in the Cochrane Handbook, sections 9.5.2 and 9.5.3), and analyses limiting by study quality (17). In sensitivity analyses limiting by study quality, we chose to eliminate studies of very low quality. Very low-quality studies were defined a priori as those that received ≥2 high risk-of-bias ratings OR 1 specific high-risk rating in the "other" category in addition to ≥2 unclear risk ratings OR ≥3 unclear risk-of-bias ratings in dimensions other than the "other" category using the Cochrane Risk of Bias tool (19).

RESULTS

The initial systematic search returned 1,607 potentially relevant abstracts, of which 191 were eligible for full text review.

Table 1. Effects of NSAIDs on pain and function*

Outcome	All NSAIDs	All NSAIDs, very low-quality removed†	All NSAIDs, knee OA only	Celecoxib	Intermediate COX inhibitors	Traditional NSAIDs
Pain, all time points						
RCTs; patients, no.	48; 17,861	29; 11,741	32; 12,875	20; 7,996	9; 2,902	25; 7,303
2 weeks	−0.43 (−0.48, −0.38), I^2 = 55%	−0.48 (−0.55, −0.41), I^2 = 64%	−0.40 (−0.46, −0.35), I^2 = 44%	−0.41 (−0.48, −0.34), I^2 = 54%	−0.31 (−0.41, −0.21), I^2 = 40%	−0.51 (−0.59, −0.44), I^2 = 55%
RCTs; patients, no.	59; 22,911	37; 15,287	41; 16,931	31; 11,699	9; 2,902‡	29; 8,896
4 weeks	−0.40 (−0.46, −0.34), I^2 = 76%	−0.40 (−0.44, −0.36), I^2 = 21%	−0.38 (−0.45, −0.30), I^2 = 80%	−0.34 (−0.39, −0.29), I^2 = 32%	−0.31 (−0.38, −0.23), I^2 = 0‡	−0.44 (−0.54, −0.34), I^2 = 77%
RCTs; patients, no.	13; 6,341	9; 4,648	6; 3,849	9; 4,970	1; 308	4; 1,218
8 weeks	−0.36 (-0.43, −0.30), I^2 = 41%	−0.42 (−0.49, −0.35), I^2 = 23%	-0.37 (−0.49, −0.26), I^2 = 69%	−0.37 (−0.46, −0.28), I^2 = 56%	−0.26 (−0.49, −0.04), I^2 = NA	−0.37 (−0.49, −0.25), I^2 = 0
RCTs; patients, no.	24; 11,096	17; 7,925	15; 7,762	13; 6,472	2; 571	13; 4,657
12 weeks	−0.30 (−0.34, −0.26), I^2 = 0	−0.31 (−0.36, −0.27), I^2 = 0	−0.27 (−0.32, −0.23), I^2 = 0	−0.27 (−0.32, −0.22), I^2 = 0	−0.25 (−0.41, −0.08), I^2 = 0	−0.36 (−0.42, −0.30), I^2 = 0
RCTs; patients, no.	2; 976	2; 976	2; 976	2; 976	ND	ND
26 weeks	−0.21 (−0.39, −0.03), I^2 = 48%	−0.21 (−0.39, −0.03), I^2 = 48%	−0.21 (−0.39, −0.03), I^2 = 48%	−0.21 (−0.39, −0.03), I^2 = 48%	NA	NA
Function, all time points						
RCTs; patients, no.	28; 9,595	21; 7,317	15; 5,619	13; 5,261	1; 263	16; 4,551
2 weeks	−0.45 (−0.52, −0.38), I^2 = 59%	−0.47 (−0.56, −0.37), I^2 = 69%	−0.44 (−0.50, −0.37), I^2 = 23%	−0.40 (−0.46, −0.35), I^2 = 0	−0.15 (−0.40, 0.09), I^2 = NA§	−0.51 (−0.63, −0.39), I^2 = 69%
RCTs; patients, no.	35; 11,979	26; 8,966	22; 8,002	20; 6,219	1; 263	21; 6,282
4 weeks	−0.38 (−0.43, −0.33), I^2 = 45%	−0.40 (−0.46, −0.33), I^2 = 49%	−0.34 (−0.40, −0.28), I^2 = 31%	−0.32 (−0.37, −0.26), I^2 = 19%	−0.31 (−0.56, −0.07), I^2 = NA	−0.43 (−0.52, −0.35), I^2 = 57%
RCTs; patients, no.	7; 2,492	5; 1,630	1; 460	4; 1,581	ND	3; 911
8 weeks	−0.37 (−0.45, −0.29), I^2 = 0	−0.41 (−0.52, −0.30), I^2 = 0	−0.33 (−0.52, −0.15), I^2 = NA	−0.35 (−0.45, −0.25), I^2 = 0	NA	−0.40 (−0.61, −0.20), I^2 = 48%
RCTs; patients, no.	23; 10,118	17; 7,320	14; 6,784	13; 6,395	2; 571	12; 4,165
12 weeks	−0.34 (−0.39, −0.29), I^2 = 28%	−0.35 (−0.42, −0.28), I^2 = 44%	−0.31 (−0.36, −0.25), I^2 = 17%	−0.29 (−0.34, −0.24), I^2 = 0	−0.26 (−0.43, −0.10), I^2 = 0	−0.40 (−0.48, −0.31), I^2 = 39%
RCTs; patients, no.	2; 976	2; 976	2; 976	2; 976	ND	ND
26 weeks	−0.19 (−0.32, −0.07), I^2 = 0	−0.19 (−0.32, −0.07), I^2 = 0	−0.19 (−0.32, −0.07), I^2 = 0	−0.19 (−0.32, −0.07), I^2 = 0	NA	NA

* Values are the standardized mean difference (SMD) (95% confidence interval [95% CI]) unless indicated otherwise. All SMDs (95% CIs) were statistically significant, except for Function, 2 weeks, with intermediate cyclooxygenase (COX) inhibitors, as indicated with a footnote. NSAIDs = nonsteroidal antiinflammatory drugs; OA = osteoarthritis; RCT = randomized controlled trial; I^2 = measure of heterogeneity, with 100% being the maximum possible heterogeneity; ND = no data; NA = not applicable.
† Very low quality was defined as studies that received ≥2 high risk-of-bias ratings OR 1 specific high risk rating in the "other" category in addition to ≥2 unclear risk ratings OR ≥3 unclear risk-of-bias ratings in dimensions other than the "other" category. Negative standardized mean differences favor treatment, and positive standardized mean differences favor placebo.
‡ Paul et al (27) was removed due to I^2 value of 93% and extremely large effect.
§ SMD (95% CI) was not statistically significant.

Of the 191 articles that underwent full text review, 72 RCTs were eligible for our analyses (see Supplementary Figure 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23884/abstract). The efficacy and/or safety of the following oral NSAIDs were assessed by the included RCTs: celecoxib (35 RCTs), naproxen (18 RCTs), diclofenac (11 RCTs), nabumetone (7 RCTs), ibuprofen (6 RCTs), meloxicam (3 RCTs), etodolac (2 RCTs), indomethacin (1 RCT), and piroxicam (1 RCT).

Baseline characteristics of the included RCTs are reported in Supplementary Table 3, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23884/abstract. A supplementary table of studies that were excluded due to inappropriate population characteristics can be found in Supplementary Table 4, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.23884/abstract. The publication dates of included RCTs ranged from 1976 to 2017, and the sample sizes comprising eligible treatment arms in included RCTs ranged from 47 to 844 (median 323). The follow-up duration ranged from 1 week to 2 years, but 96% of the trials had a duration of 13 weeks or less (median 6 weeks). The mean age of included participants ranged from 53 to 69 years (median 62 years), and the mean body mass index of patients ranged from 27 to 34 kg/m^2 (median 31.5 kg/m^2). The percentage of females ranged from 49% to 85% (median 68%). Limited use of acetaminophen as rescue medication was permitted in 69% of the included RCTs.

A summary of study quality assessment is shown in Supplementary Table 5, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23884/abstract. Supplementary Figure 2, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.23884/abstract, shows the overall risk of bias distribution. The majority of studies were of moderate quality; potential attrition bias and reporting bias were the most common reasons for high risk-of-bias ratings. The majority of RCTs (80%) reported industry sponsorship and/or direct industry involvement of 1 or more investigators.

Overall effects of NSAIDs on pain and function. Our primary analyses of pain and function combined all oral NSAIDs, regardless of classification. Results of all analyses of pain and functional outcomes are shown in Table 1.

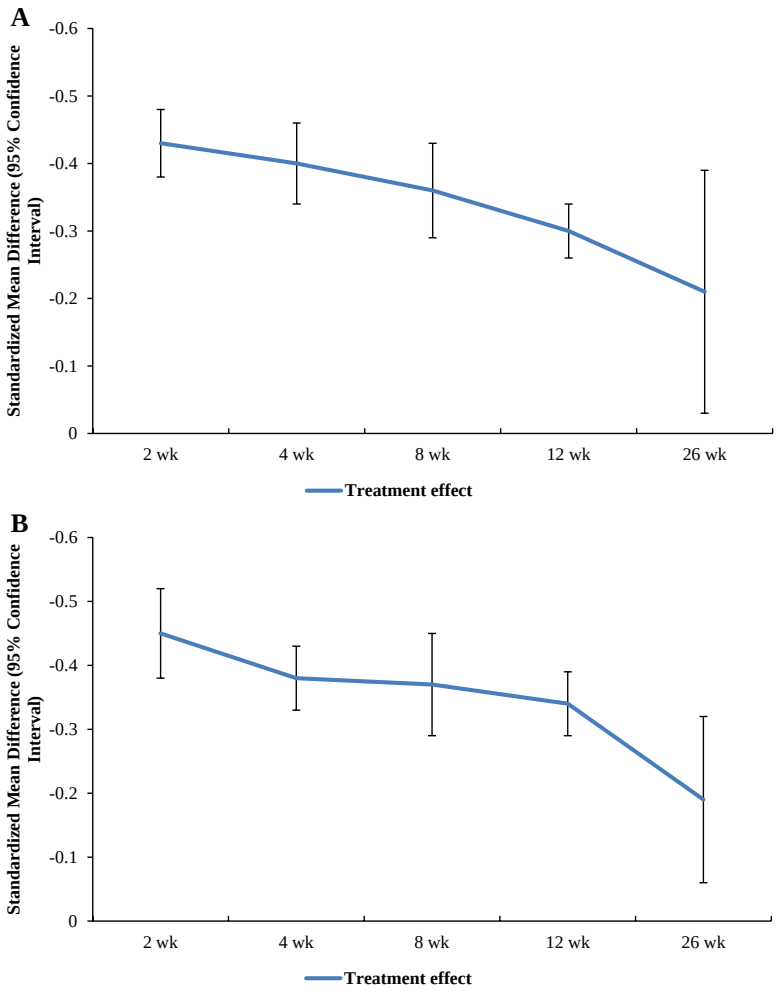

Figure 1. Trajectory of overall effects of NSAIDs on pain (**A**) and function (**B**). wk = week. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.23884/abstract.

NSAIDs showed statistically significant, clinically important effects on pain as early as 2 weeks from baseline, with an SMD of –0.43 (95% CI –0.48, –0.38). This treatment effect remained statistically significant up to 26 weeks (SMD –0.21 [95% CI –0.39, –0.03]), although the effects attenuated progressively over time and lost clinical significance (Figure 1A). The analysis of pain at 4 weeks demonstrated high heterogeneity (I^2 = 76%), prompting a sensitivity analysis excluding outliers (26,27). This analysis reduced the I^2 value to 29%, and the treatment effect decreased (SMD –0.36 [95% CI –0.40, –0.33]). Sensitivity analyses restricted to populations with knee OA only were not notably different from those observed in the primary analysis at any time point. Sensitivity analyses by study quality also showed results similar to the primary analysis, but there was a trend for effect sizes to increase slightly with the removal of very low-quality studies (Table 1). There was no notable asymmetry in our visual inspection of the funnel plots at any time point.

With respect to functional improvement, NSAIDs again showed consistent statistically significant benefits compared with placebo, from 2 weeks (SMD –0.45 [95% CI –0.52, –0.38]) to 26 weeks (SMD –0.19 [95% CI –0.32, –0.07]) (Figure 1B). None of the analyses of functional improvement demonstrated I^2 values necessitating sensitivity analysis. Sensitivity analyses restricted to populations with knee OA only showed beneficial effects on functional outcomes that were not notably different from the primary analysis at any time point. Sensitivity analyses limiting by study quality produced results similar to the main analysis, with a tendency for effect sizes to increase (Table 1). Again, there was no notable asymmetry in our visual inspection of the funnel plots at any time point.

Sensitivity analyses restricting by pain or functional assessment scale demonstrated that functional measurements obtained using scales other than the WOMAC (primarily, the Lequesne Algofunctional Index) tended to result in slightly smaller effect sizes (see Supplementary Figure 3, parts A and B, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23884/abstract).

Overall safety of NSAIDs. Patients receiving oral NSAIDs were more likely to withdraw due to an AE during a study's treatment period (RR 1.16 [95% CI 1.02, 1.32]) but were less likely to withdraw due to a lack of efficacy (RR 0.38 [95% CI 0.34, 0.43]) (Table 2). Patients receiving oral NSAIDs experienced a higher incidence of treatment-related AEs (RR 1.21 [95% CI 1.04, 1.40]), CV AEs (RR 1.37 [95% CI 1.05, 1.77]), and GI AEs (RR 1.36 [95% CI 1.25, 1.49]) during the study follow-up period. As

Table 2. Results for discontinuations and safety*

Outcome	RCTs, no.	Patients, no.	Effect estimate, RR (95% CI)	Follow-up range, weeks
All NSAIDs vs. placebo				
Withdrawals due to adverse events	60	22,993	1.16 (1.02, 1.32), I^2 = 22%†	1–104 (median 6)
Withdrawals due to lack of efficacy	50	18,309	0.38 (0.34, 0.43), I^2 = 37%†	1–104 (median 6)
Treatment-related adverse events	24	9,548	1.21 (1.04, 1.40), I^2 = 54%†	1–13 (median 6)
Gastrointestinal adverse events	59	23,026	1.36 (1.25, 1.49), I^2 = 38%†	1–26 (median 6)
Cardiovascular adverse events	36	14,654	1.37 (1.05, 1.77), I^2 = 0†	1–13 (median 6)
Serious adverse events	40	17,278	0.90 (0.68, 1.19), I^2 = 0	1–13 (median 12)
Celecoxib vs. placebo				
Withdrawals due to adverse events	28	11,177	1.01 (0.84, 1.22), I^2 = 23%	1–26 (median 12)
Withdrawals due to lack of efficacy	23	9,084	0.40 (0.34, 0.48), I^2 = 31%†	1–26 (median 12)
Treatment-related adverse events	13	4,722	0.99 (0.86, 1.13), I^2 = 7%	1–13 (median 6)
Gastrointestinal adverse events	27	10,984	1.14 (1.03, 1.27), I^2 = 0†	1–26 (median 12)
Cardiovascular adverse events	18	7,732	1.24 (0.86, 1.80), I^2 = 0	1–13 (median 12)
Serious adverse events	23	9,723	0.89 (0.60, 1.32), I^2 = 0	1–13 (median 12)
Intermediate COX inhibitors vs. placebo				
Withdrawals due to adverse events	11	3,419	1.11 (0.78, 1.57), I^2 = 29%	4–12 (median 6)
Withdrawals due to lack of efficacy	9	2,906	0.49 (0.39, 0.62), I^2 = 25%†	4–12 (median 6)
Treatment-related adverse events	3	1,045	1.05 (0.82, 1.33), I^2 = 0	4–6 (median 4)
Gastrointestinal adverse events	11	3,419	1.40 (1.06, 1.86), I^2 = 59%†	4–12 (median 6)
Cardiovascular adverse events	5	2,029	1.29 (0.63, 2.63), I^2 = 0	4–12 (median 6)
Serious adverse events	5	1,829	1.98 (0.57, 6.93), I^2 = 0	4–12 (median 6)
Traditional NSAIDs vs. placebo				
Withdrawals due to adverse events	33	10,302	1.36 (1.16, 1.59), I^2 = 1%†	2–104 (median 6)
Withdrawals due to lack of efficacy	25	7,066	0.31 (0.25, 0.39), I^2 = 43%†	2–104 (median 6)
Treatment-related adverse events	13	4,263	1.42 (1.13, 1.78), I^2 = 61%†	4–13 (median 6)
Gastrointestinal adverse events	32	9,892	1.49 (1.31, 1.68), I^2 = 45%†	1–13 (median 6)
Cardiovascular adverse events	17	5,542	1.92 (1.17, 3.16), I^2 = 29%†	1–13 (median 6)
Serious adverse events	20	6,573	0.93 (0.63, 1.38), I^2 = 0	1–13 (median 6)

* Risk ratios (RRs) <1 favor treatment, and RRs >1 favor placebo. RCT = randomized controlled trial; 95% CI = 95% confidence interval; NSAIDs = nonsteroidal antiinflammatory drugs; I^2 = measure of heterogeneity, with 100% being the maximum possible heterogeneity; COX = cyclooxygenase.
† RR (95% CI) was statistically significant.

we had expected, the most commonly reported GI AEs were transient and mild and included upper abdominal pain, diarrhea, dyspepsia, and nausea. Edema and hypertension were the most commonly reported CV AEs, and they were mild in severity and duration. The incidence of serious AEs (SAEs) over the duration of study did not differ between groups.

GI and CV safety trajectory of NSAIDs (all classes combined). We assessed the likelihood of experiencing GI AEs at 2, 4, 12, and 26 weeks and CV AEs at 2, 4, and 12 weeks (Figure 2, parts A and B). No RCT reported data on GI or CV safety at 8 weeks. Patients receiving NSAIDs were more likely to experience a minor GI AE as early as 4 weeks after initiating treatment (n = 31 studies; RR 1.38 [95% CI 1.21, 1.57]), at 12 weeks (n = 22 studies; RR 1.36 [95% CI 1.18, 1.56]), and at 26 weeks (n = 1 study; RR 1.55 [95% CI 0.52, 4.62]) (Figure 2A). Although the overall risk of developing a minor CV AE was higher in patients using NSAIDs versus placebo, it did not reach statistical significance at any individual time point in the analysis of all NSAID classes combined.

Comparing the efficacy and safety of different classes of NSAIDs. Traditional NSAIDs performed consistently better than the other classes (Table 1 and Figure 3A). At 2 weeks, traditional NSAIDs demonstrated effects on pain that were 24% and 64% greater than those of celecoxib and intermediate COX inhibitors, respectively; at 12 weeks, the effects of traditional NSAIDs on pain were 33% and 44% greater than those of celecoxib and intermediate COX inhibitors, respectively. Only studies assessing the efficacy of celecoxib extended to 26 weeks; so a comparison of the different classes could not be undertaken at this time point.

Traditional NSAIDs also outperformed the other classes with regard to functional improvement, demonstrating effects that ranged from 14% to 42% better than those of celecoxib (Table 1 and Figure 3B). Interestingly, for both pain and functional efficacy outcomes, celecoxib outperformed intermediate COX inhibitors at most time points. Once again, due to a lack of data, a comparison of the different NSAID classes was not possible at 26 weeks.

Traditional NSAIDs demonstrated the largest effects with regard to efficacy outcomes and also demonstrated the least favorable safety profile of all the classes (Table 2). Patients

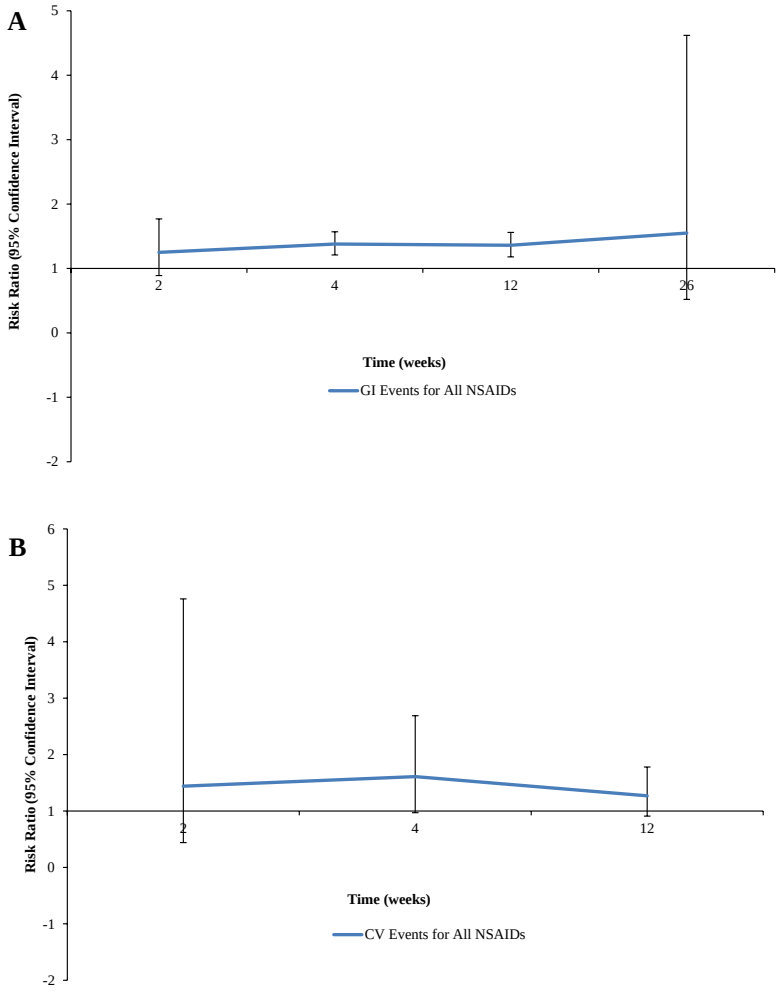

Figure 2. Trajectories of the gastrointestinal (GI) (**A**) and cardiovascular (CV) (**B**) safety of nonsteroidal antiinflammatory drugs (NSAIDs). Color figure can be viewed in the online issue, which is available athttp://onlinelibrary.wiley.com/doi/10.1002/acr.23884/abstract.

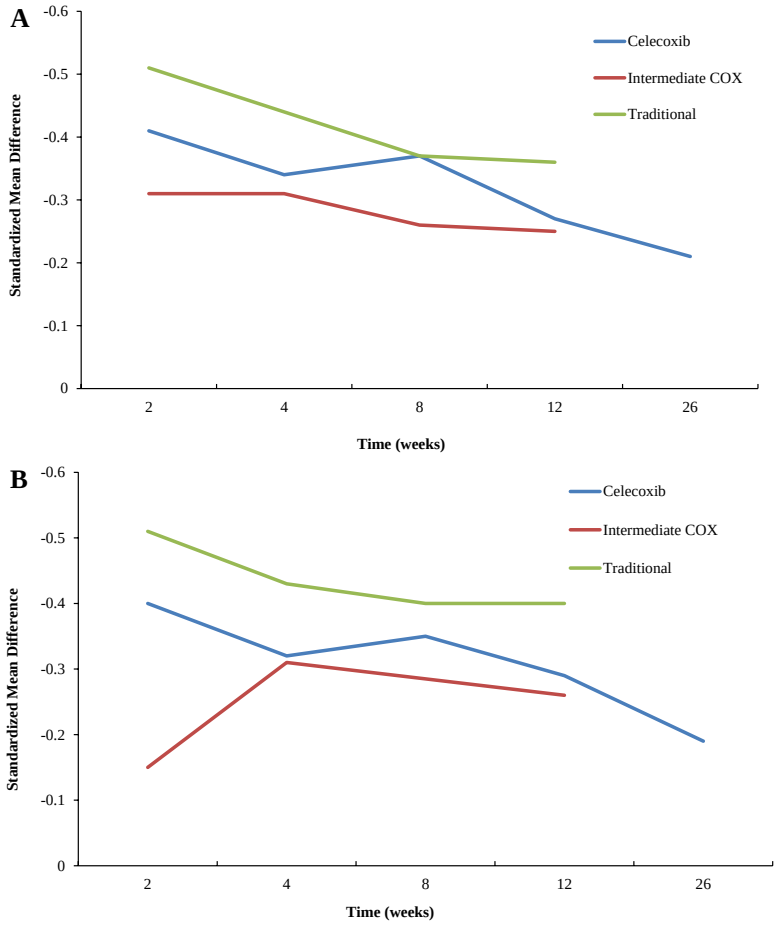

Figure 3. Trajectory of effects of different classes of nonsteroidal antiinflammatory drugs on pain (**A**) and function (**B**). COX = cyclooxygenase.

receiving traditional NSAIDs were significantly more likely to withdraw due to an AE during the study period (median 6 weeks; range 2–104). These patients were also 42% more likely than patients receiving placebo to report treatment-related AEs (RR 1.42 [95% CI 1.13, 1.78]), ~50% more likely to report GI AEs (RR 1.49 [95% CI 1.31, 1.68]), and 92% more likely to report CV AEs (RR 1.92 [95% CI 1.17, 3.16]) over a median follow-up time of only 6 weeks.

Patients receiving celecoxib had a statistically significantly higher risk of experiencing a GI AE (RR 1.14 [95% CI 1.03, 1.27]) than patients receiving placebo over the course of the study period (median 12 weeks; range 1–26), but the effect on CV risk was not significant (RR 1.24 [95% CI 0.86, 1.80]). Patients receiving intermediate COX inhibitors were also significantly more likely than those receiving placebo to report GI AEs (RR 1.40 [95% CI 1.06, 1.86]) over the course of the treatment period (median 6 weeks; range 4–12) but not CV AEs (RR 1.29 [95% CI 0.63, 2.63]).

Comparing the safety trajectories of different classes of NSAIDs. The trajectory of GI AEs differed based on NSAID classification. Patients receiving traditional NSAIDs experienced the highest likelihood of GI AEs at most time points, and the risk ratios were statistically significant at 4 (RR 1.54 [95% CI 1.23, 1.93]) and 12 weeks (RR 1.52 [1.31, 1.77]). For patients using intermediate COX inhibitors, the likelihood of developing GI AEs was higher than in the placebo group at most time points, but it was statistically significant only at 4 weeks (RR 1.37 [95% CI 1.02, 1.84]). Patients receiving celecoxib showed the lowest likelihood of GI AEs among NSAID classes, not reaching statistical significance at any individual time point.

The rates of CV AEs observed in patients receiving celecoxib and intermediate COX inhibitors were not statistically significantly different from their respective placebo groups at any time point. Patients receiving traditional NSAIDs were statistically significantly more likely to report a CV AE at 4 weeks. There were no significant differences between the traditional NSAIDs and placebo groups at 2 or 12 weeks. The majority of GI and CV AEs collected at any time point for all NSAID classifications were transient and of minor severity.

DISCUSSION

Our meta-analysis showed that, while NSAIDs demonstrated rapid benefits for pain and functional outcomes, the effects

attenuated and lost clinical significance by 8 weeks. Although the magnitude of the effect differed between the 3 different classes of NSAIDs, the effect consistently waned over time across all classes. Meanwhile, the incidence of minor GI and CV AEs in NSAID users rose as early as 2 weeks after treatment initiation and remained elevated thereafter. The use of traditional NSAIDs was associated with the least favorable safety profile.

Our study expands upon the oral NSAIDs-related findings from another meta-analysis that was conducted over a decade ago (28). At the time, the study assembled pain outcomes from 25 RCTs of oral NSAIDs, which included some treatments we considered ineligible for our review (e.g., valdecoxib and lumiracoxib) and double-counted the results from 1 RCT published as 2 reports (29,30). Our study included a much larger pool of 72 RCTs and evaluated both pain and functional outcomes in addition to the timeline of safety measurements. Nevertheless, the results of the prior meta-analysis were consistent with ours with regard to pain relief that peaked at ~2 weeks and steadily declined thereafter. Based on our findings coupled with the current literature, clinicians should weigh the likelihood of a decline in symptomatic benefit against the risk of early-onset minor AEs, along with the patient perceptions, tolerability, and preferences when extending an NSAID treatment regimen beyond 12 weeks (31–33).

The results of our subgroup analyses contingent upon NSAID classification revealed that although all NSAIDs shared a similar trend of efficacy, traditional NSAIDs as a group demonstrated the largest effects on pain and function. This finding could be explained by the fact that the group of selective COX-2 inhibitors in our review was populated by only 1 drug, celecoxib, since other coxibs that might belong to the group did not satisfy our selection criteria (i.e., FDA-approval for use in the US). Three recent network meta-analyses ranking the efficacy of individual NSAIDs of all classes demonstrated very modest clinical effects of celecoxib, even at its maximum approved daily dose, relative to other coxibs or traditional NSAIDs (34–36); our findings thus corroborate those results. An important takeaway from the aforementioned network meta-analyses was that the effects of NSAIDs were dose-dependent and varied among individual drugs within and even between NSAID classes. While these meta-analyses provided detailed treatment rankings, a major strength of our study is that we have examined the therapeutic trajectory instead of single time points targeted by network meta-analyses and answered a broader question regarding the expected duration of beneficial effects of NSAIDs. The information from our study and from previous network meta-analyses provides clinicians with a strong background of evidence by which to establish the optimal treatment regimen.

The distinction between NSAIDs based on COX-2 selectivity was made primarily to examine their safety profile, with the focus on GI and CV AEs. In our study, all classes of NSAIDs demonstrated a greater probability of GI AEs. The incidence of minor GI AEs rose with the decline of COX-2 selectivity among NSAID classes, reaching the highest point with the traditional NSAIDs group. This group (and, to a lesser extent, the intermediate NSAIDs group) demonstrated statistically significantly more minor GI AEs at 4 weeks; the relative risk of minor GI events remained elevated at 12 weeks in patients taking traditional NSAIDs. We did not observe the similarity in GI tolerability between the selective and intermediate COX inhibitors noted in a 2015 network meta-analysis on NSAID-induced GI injury, possibly owing to the difference in the selection of coxibs under review. Our analyses of CV AE risks indicated that the incidence of minor CV events also rose with the reduction in COX-2 selectivity; they were the lowest for celecoxib and the highest for the traditional NSAIDs (37).

Although the follow-up time of our study limited our safety analyses to the observation of minor AEs, the overall trends we observed in our results align with the findings from a 2004 study by Richy et al (13). This study assessed a more heterogeneous population of NSAID users but demonstrated an early development of GI complications after initiation of treatment with NSAIDs, with timing that varied from 1 week for indomethacin to over 1 month for other NSAIDs.

Current clinical practice guidelines recommend the use of traditional NSAIDs with a proton-pump inhibitor (PPI), or use of celecoxib with or without a PPI, to minimize the risk of GI toxicity in patients with moderate or high comorbidity risk (5,6). For patients with CV comorbidities, naproxen or celecoxib have been suggested to minimize the risk of a CV AE (38). Clinical practice guidelines have also indicated that NSAIDs should be used at the lowest effective dose and for the shortest duration (5). The results of our study support these recommendations, demonstrating the rapidity with which minor negative reactions can occur during NSAID treatment.

The results of our study should be interpreted in light of certain limitations. First, we did not perform separate analyses of the studies using the individual drugs' highest recommended dose because the data would be too scarce to derive meaningful trajectories of their effects on pain and function. Thus, our results may not have pinpointed the absolute measure of potency for the included NSAIDs. Second, the quality of our study was limited by the quality of the underlying data. One of the primary concerns among the included studies was the potential for attrition bias. There was a tendency for attrition rates in both treatment and placebo study arms to be high, but the reasons for discontinuation were unbalanced: a larger share of patients withdrew from the placebo group due to lack of efficacy, whereas more patients from treatment groups withdrew due to AEs. In the context of our results, a higher withdrawal rate due to AEs in the intervention group could skew treatment effects toward the null because patients who discontinued may have been experiencing pain relief or functional improvement despite any adverse experience; conversely, a higher withdrawal rate due to lack of efficacy in the placebo group could inflate the placebo effects because the participants who experienced the least benefit have discontinued the study. In a majority

of the studies, withdrawals were tallied at the end of the study and details were not provided for each time point. Therefore, we could not quantify the effect of attrition on the treatment effect at specific time points. Overall, due to the attrition imbalance we observed across many of the RCTs, our results may have ultimately understated the overall treatment effects of NSAIDs.

Another limitation of our study is the lack of data at and beyond the 26-week time point. Only 2 studies reported efficacy results at 26 weeks, and celecoxib was the only treatment represented at this time point. Consequently, the overall estimates for treatment effects at 26 weeks are less precise due to a lower number of participants, and we were unable to conduct subgroup analyses restricting by NSAID classification at this time point. Given that celecoxib was found to be less effective than traditional NSAIDs overall, our efficacy estimates for pain and functional outcomes at 26 weeks could be an underestimation of the longer-term effects of NSAIDs overall. The scarcity of longer-term follow-up data means that our results may not be generalizable beyond 12 weeks.

Our analyses of safety outcomes were limited by several factors. First, the risk estimates from our study might be smaller than those observed in clinical practice because the knee OA population that was included is more restricted and less representative of the general OA population, and because patients with previous GI or CV issues were most likely excluded from the enrollment. However, our risk estimates are less biased compared with the observational studies because the randomized nature of our data more accurately controls for confounding factors and other biases that limit the interpretation of non-RCT data. Second, we were unable to evaluate the risks for major vascular events (such as myocardial infarction, stroke, or coronary death) or serious GI complications (e.g., GI bleed, perforation, or obstruction) because very few of these events were observed during the study periods in the included pool of studies. Therefore, our safety analyses incorporated more commonly reported minor events, such as the symptoms of GI upset or edema or hypertension, and the resulting RRs may inadequately reflect the risks of major GI and CV AEs. Furthermore, some of the AEs assembled within the CV AEs group, such as edema and hypertension, may have been at least partly effected by prostaglandin-mediated effects on renal physiology (39).

Third, the median follow-up time for included studies was 6 weeks, which prevented us from analyzing data on AEs that may arise from extended NSAID use. The shorter follow-up times of the included studies may have also introduced bias with regard to the types of AEs we collected. For example, minor GI events are common and may manifest in numbers sufficient for an analysis in short-term usage periods of an oral treatment, but very few major CV events may be detected within such a brief follow-up period. Finally, in order to maximize the use of available data, we collected composite rates, and in situations where individual events were reported separately (e.g., diarrhea, dyspepsia, and nausea), we summed the number of participants experiencing each event to mimic a composite rate for the organ system. We considered this approach to be justified by the fact that summation of event rates occurred in both treatment and placebo groups; however, the raw event rates may be a slight overestimation of the actual number of patients who experienced GI and/or CV AEs. Despite the above limitations, we detected a statistically significantly heightened risk of minor GI AEs and (in the case of traditional NSAIDs) minor CV AEs as early as 4 weeks after treatment initiation in the knee OA RCT population. Considering that this estimate is coming from a relatively low-risk population, these values may be a conservative estimate of those observed in the general OA population.

Our results should be interpreted with caution because they focus on the trajectory of response to single regimens in contrast to the dose adjustments and switching that happen in clinical practice (40). Repeated cycles of continuous NSAID use of longer duration have been suggested both as an alternative to intermittent as-needed use and as a replacement for chronic use (41). However, the results of our study suggest that such a treatment regimen may lack long-term efficacy while increasing the risks for adverse treatment effects. Even though repeated NSAID cycles are used by some clinicians, their efficacy trajectory is unknown because long-term clinical trial data on this treatment regimen are lacking. Future research should focus on incorporating study designs that mimic real-world clinical practice to better characterize the efficacy trajectory in these scenarios.

In conclusion, this study described the efficacy and safety trajectories of oral NSAIDs for knee OA over a 26-week period and showed that oral NSAIDs provide statistically significant pain reduction and functional improvement from as early as 2 weeks and up to 26 weeks of use, with the magnitude of the effect decreasing over this time period and no longer attaining clinical significance after 8 weeks. At the same time, a statistically significant risk of minor GI AEs was evident from 4 weeks of exposure. This information should be taken into account together with patient-specific safety profiles and preferences, comorbid conditions, and concomitant medications to aid clinicians in their decisions on the prescription of oral NSAIDs.

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, Vaysbrot, McAlindon, Bannuru.
Acquisition of data. Osani, Vaysbrot, Zhou, Bannuru.
Analysis and interpretation of data. Osani, Bannuru.

REFERENCES

1. Lawrence RC, Felson DT, Helmick CG, Arnold LM, Choi H, Deyo RA, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States: part II. Arthritis Rheum 2008;58: 26–35.
2. Deshpande BR, Katz JN, Solomon DH, Yelin EH, Hunter DJ, Messier SP, et al. Number of persons with symptomatic knee osteoarthritis in

the US: impact of race and ethnicity, age, sex, and obesity. Arthritis Care Res (Hoboken) 2016;68:1743–50.
3. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380:2163–96.
4. Felson DT, Lawrence RC, Dieppe PA, Hirsch R, Helmick CG, Jordan JM, et al. Osteoarthritis: new insights. Part 1: the disease and its risk factors. Ann Intern Med 2000;133:635–46.
5. McAlindon TE, Bannuru RR, Sullivan MC, Arden NK, Berenbaum F, Bierma-Zeinstra SM, et al. OARSI guidelines for the non-surgical management of knee osteoarthritis. Osteoarthritis Cartilage 2014; 22:363–88.
6. National Institute for Health and Clinical Excellence: Guidance. In: Osteoarthritis: care and management in adults. London: National Institute for Health and Care Excellence; 2014.
7. Gore M, Tai KS, Sadosky A, Leslie D, Stacey BR. Use and costs of prescription medications and alternative treatments in patients with osteoarthritis and chronic low back pain in community-based settings. Pain Pract 2012;12:550–60.
8. Deveza LA, Hunter DJ, Van Spil WE. Too much opioid, too much harm. Osteoarthritis Cartilage 2018;26:293–5.
9. Nissen SE, Yeomans ND, Solomon DH, Luscher TF, Libby P, Husni ME, et al. Cardiovascular safety of celecoxib, naproxen, or ibuprofen for arthritis. N Engl J Med 2016;375:2519–29.
10. Schneider V, Levesque LE, Zhang B, Hutchinson T, Brophy JM. Association of selective and conventional nonsteroidal antiinflammatory drugs with acute renal failure: a population-based, nested case-control analysis. Am J Epidemiol 2006;164:881–9.
11. Sostres C, Gargallo CJ, Lanas A. Nonsteroidal anti-inflammatory drugs and upper and lower gastrointestinal mucosal damage. Arthritis Res Ther 2013;15 Suppl 3:S3.
12. Fendrick AM, Greenberg BP. A review of the benefits and risks of nonsteroidal anti-inflammatory drugs in the management of mild-to-moderate osteoarthritis. Osteopath Med Prim Care 2009;3:1.
13. Richy F, Bruyere O, Ethgen O, Rabenda V, Bouvenot G, Audran M, et al. Time dependent risk of gastrointestinal complications induced by non-steroidal anti-inflammatory drug use: a consensus statement using a meta-analytic approach. Ann Rheum Dis 2004;63:759–66.
14. Yang M, Wang HT, Zhao M, Meng WB, Ou JQ, He JH, et al. Network meta-analysis comparing relatively selective COX-2 inhibitors versus coxibs for the prevention of NSAID-induced gastrointestinal injury. Medicine (Baltimore) 2015;94:e1592.
15. Hawkey CJ. COX-1 and COX-2 inhibitors. Best Pract Res Clin Gastroenterol 2001;15:801–20.
16. Warner TD, Giuliano F, Vojnovic I, Bukasa A, Mitchell JA, Vane JR. Nonsteroid drug selectivities for cyclo-oxygenase-1 rather than cyclo-oxygenase-2 are associated with human gastrointestinal toxicity: a full in vitro analysis. Proc Natl Acad Sci U S A 1999;96:7563–8.
17. Higgins J, Green S, eds. Cochrane handbook for systematic reviews of interventions, version 5.1.0. URL: www.handbook.cochrane.org.
18. Markum Mitchell. Engauge digitizer. URL: https://github.com/markummitchell/engauge-digitizer.
19. Cochrane risk of bias tool. Copenhagen: Nordic Cochrane Centre; 2014.
20. Review manager (RevMan). Version 5.3. Copenhagen: Nordic Cochrane Centre; 2014.
21. Cochrane Musculoskeletal Group. Proposed outcomes. URL: http://musculoskeletal.cochrane.org/proposed-outcomes.
22. ICH E9 statistical principles for clinical trials. European Medicines Agency;1998. URL: https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials.
23. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959;22:719–48.
24. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21:1539–58.
25. Wandel S, Juni P, Tendal B, Nuesch E, Villiger PM, Welton NJ, et al. Effects of glucosamine, chondroitin, or placebo in patients with osteoarthritis of hip or knee: network meta-analysis. BMJ 2010; 341:c4675.
26. Ghosh S, Paul S, Das N, Bhattacharyya TK. A study on the effects of diclofenac sodium and etoricoxib in the treatment of osteoarthritis. J Indian Med Assoc 2007;105:260–2.
27. Paul S, Das N, Ghosh S. The effects of aceclofenac and nabumetone in osteoarthritis. JNMA J Nepal Med Assoc 2009;48:121–5.
28. Bjordal JM, Klovning A, Ljunggren AE, Slordal L. Short-term efficacy of pharmacotherapeutic interventions in osteoarthritic knee pain: a meta-analysis of randomised placebo-controlled trials. Eur J Pain 2007;11:125–38.
29. Bensen WG, Fiechtner JJ, McMillen JI, Zhao WW, Yu SS, Woods EM, et al. Treatment of osteoarthritis with celecoxib, a cyclooxygenase-2 inhibitor: a randomized controlled trial. Mayo Clin Proc 1999;74:1095–105.
30. Zhao SZ, McMillen JI, Markenson JA, Dedhiya SD, Zhao WW, Osterhaus JT, et al. Evaluation of the functional status aspects of health-related quality of life of patients with osteoarthritis treated with celecoxib. Pharmacotherapy 1999;19:1269–78.
31. Huerta C, Castellsague J, Varas-Lorenzo C, Garcia Rodriguez LA. Nonsteroidal anti-inflammatory drugs and risk of ARF in the general population. Am J Kidney Dis 2005;45:531–9.
32. Helin-Salmivaara A, Saarelainen S, Gronroos JM, Vesalainen R, Klaukka T, Huupponen R. Risk of upper gastrointestinal events with the use of various NSAIDs: a case-control study in a general population. Scand J Gastroenterol 2007;42:923–32.
33. Helin-Salmivaara A, Virtanen A, Vesalainen R, Gronroos JM, Klaukka T, Idanpaan-Heikkila JE, et al. NSAID use and the risk of hospitalization for first myocardial infarction in the general population: a nationwide case-control study from Finland. Eur Heart J 2006;27:1657–63.
34. Da Costa BR, Reichenbach S, Keller N, Nartey L, Wandel S, Juni P, et al. Effectiveness of non-steroidal anti-inflammatory drugs for the treatment of pain in knee and hip osteoarthritis: a network meta-analysis. Lancet 2017;390:e21–33.
35. Van Walsem A, Pandhi S, Nixon RM, Guyot P, Karabis A, Moore RA. Relative benefit-risk comparing diclofenac to other traditional non-steroidal anti-inflammatory drugs and cyclooxygenase-2 inhibitors in patients with osteoarthritis or rheumatoid arthritis: a network meta-analysis. Arthritis Res Ther 2015;17:66.
36. Bannuru RR, Schmid CH, Kent DM, Vaysbrot EE, Wong JB, McAlindon TE. Comparative effectiveness of pharmacologic interventions for knee osteoarthritis: a systematic review and network meta-analysis. Ann Intern Med 2015;162:46–54.
37. Conaghan PG. A turbulent decade for NSAIDs: update on current concepts of classification, epidemiology, comparative efficacy, and toxicity. Rheumatol Int 2012;32:1491–502.
38. Trelle S, Reichenbach S, Wandel S, Hildebrand P, Tschannen B, Villiger PM, et al. Cardiovascular safety of non-steroidal anti-inflammatory drugs: network meta-analysis. BMJ 2011;342:c7086.
39. Horl WH. Nonsteroidal anti-inflammatory drugs and the kidney. Pharmaceuticals (Basel) 2010;3:2291–321.
40. Gore M, Sadosky A, Leslie D, Tai KS, Seleznick M. Patterns of therapy switching, augmentation, and discontinuation after initiation of treatment with select medications in patients with osteoarthritis. Clin Ther 2011;33:1914–31.
41. Bruyere O, Cooper C, Pelletier JP, Branco J, Luisa Brandi M, Guillemin F, et al. An algorithm recommendation for the management of knee osteoarthritis in Europe and internationally: a report from a task force of the European Society for Clinical and Economic Aspects of Osteoporosis and Osteoarthritis (ESCEO). Semin Arthritis Rheum 2014;44:253–63.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 72, No. 5, May 2020, pp 652–660
DOI 10.1002/acr.23890

Therapeutic Alliance Between Physical Therapists and Patients With Knee Osteoarthritis Consulting Via Telephone: A Longitudinal Study

Belinda J. Lawford,[1] Kim L. Bennell,[1] Penny K. Campbell,[1] Jessica Kasza,[2] and Rana S. Hinman[1]

Objective. To explore therapeutic alliance between physical therapists and patients with knee osteoarthritis during telephone consultations. Specifically, to describe and compare physical therapist and patient ratings, to determine whether alliance changes over time, and to evaluate whether individual characteristics are associated with alliance.

Methods. We performed a secondary analysis of 84 patients in the intervention arm of a randomized controlled trial who completed 5–10 consultations with 1 of 8 physical therapists via telephone over 26 weeks, involving education, advice, and prescription of a strengthening and physical activity program. Therapeutic alliance was measured after the second (week 4) and final consultations (week 26), using the Working Alliance Inventory–Short Form.

Results. Patient and physical therapist ratings of the alliance were high. At week 4, patients rated the overall alliance, and all 3 subscales, higher than therapists. At 26 weeks, patients rated the Task subscale higher than therapists. Patient ratings for the Goal subscale decreased over time, while physical therapist ratings for total alliance and the Bond subscale increased. For patients, the variables of living with others, consulting with a therapist with no previous experience delivering care remotely, having more telephone consultations, and having a higher self-efficacy were associated with greater alliance ratings. Therapists were more likely to perceive a stronger alliance if they had less clinical experience and when treating patients who were younger and had higher self-efficacy.

Conclusion. Physical therapist perceptions of the therapeutic alliance tended to be lower than those of patients early in treatment, but differences were small and of unclear clinical significance. Some subgroups of patients rated the alliance more strongly than others.

INTRODUCTION

Knee osteoarthritis (OA) is a common and debilitating condition affecting approximately 24% of the adult population (1), causing pain, impaired function, and reduced quality of life. There is no cure for OA, and lifestyle management remains the cornerstone of evidence-based care. Recommended strategies include education, exercise, and, if appropriate, weight loss (2,3). However, effect sizes for these interventions for improving pain and function are modest at best (2,3); thus there is much research interest in understanding how treatment effectiveness could be improved.

An emerging area of interest is the relationship between clinicians and patients, known as the therapeutic alliance. Therapeutic alliance is conceptualized as the sense of collaboration, warmth, and support between a patient and clinician (4). Bordin's model of therapeutic alliance (4) focuses on 3 elements of this relationship: agreement on goals, agreement on tasks, and personal bond. Therapeutic alliance has been studied extensively in patients with psychological conditions receiving psychotherapy, with research showing it is correlated with positive outcomes following therapy (5), such as recovery from schizophrenia, improved global mental health, reduced drug use, better adherence to therapy, and a reduction in depressive symptoms. Similarly, there is

Supported by the National Health and Medical Research Council (Partnership Project grant 1112133) the Medibank Better Health Foundation, with in-kind support from Musculoskeletal Australia (formerly Arthritis and Osteoporosis Victoria), HealthChange Australia, and the Australian Physiotherapy Association. Dr. Lawford's work was supported by the National Health and Medical Research Council Centre of Research Excellence (grant 1079078). Dr. Bennell's work was supported by a National Health and Medical Research Council Fellowship (grant 1058440). Dr. Hinman's work was supported by a National Health and Medical Research Council Fellowship (grant 1154217).

[1]Belinda J. Lawford, BMedSci, BHlthSci(Hons), PhD, Kim L. Bennell, BAppSci(Physio), PhD, Penny K. Campbell, BAppSci(FoodSci&Nutr), Rana S. Hinman, BPhysio(Hons), PhD: University of Melbourne, Parkville, Victoria, Australia; [2]Jessica Kasza, BSc(Hons), PhD: Monash University, Clayton, Victoria, Australia.

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

Address correspondence to Rana S. Hinman, BPhysio(Hons), PhD, Centre for Health Exercise and Sports Medicine, Department of Physiotherapy, School of Health Sciences, University of Melbourne, Parkville, Victoria, Australia 3010. E-mail: ranash@unimelb.edu.au.

Submitted for publication January 8, 2019; accepted in revised form March 26, 2019.

SIGNIFICANCE & INNOVATIONS

- The therapeutic alliance between patients and clinicians has an important influence on clinical outcomes, yet no previous studies have investigated the alliance between patients with knee osteoarthritis and physical therapists who consult via telephone.
- We found that, on average, patient and physical therapist ratings of the alliance were high. Patients rated the alliance significantly higher than physical therapists early in treatment, but scores generally converged by the end of treatment. However, differences were small and of unclear clinical significance.
- For patients, the variables of living with others, consulting with a therapist who had no previous experience delivering care remotely, having more consultations, and having higher self-efficacy were associated with greater alliance ratings. Therapists were more likely to perceive a stronger alliance if they had less clinical experience and when treating patients who were younger and with higher self-efficacy.
- Findings suggest that physical therapists should not assume that their perceptions of the alliance match those of their patients, and that some subgroups of patients tend to rate the alliance higher. However, further research is required to investigate whether differences in ratings have a meaningful impact on clinical outcomes.

increasing evidence that therapeutic alliance is also important in other fields of health care (6). For example, a stronger patient-physician alliance has been correlated with better emotional acceptance of terminal disease and less time spent in intensive care in patients with cancer (7), and with more positive rehabilitation outcomes in patients with multiple sclerosis (8).

There is emerging evidence that the therapeutic alliance is important in musculoskeletal rehabilitation. A systematic review found that the alliance between patients with chronic musculoskeletal pain (including chronic low back pain or general musculoskeletal pain) and their treating therapists (including physical therapists, occupational therapists, psychologists, and sports therapists) was positively associated with improvements in pain severity and physical function in all 5 included studies (9). Two recent systematic reviews have investigated the impact of the therapeutic alliance between physical therapists and patients on outcomes of treatment for chronic musculoskeletal pain (10). These reviews have concluded that a stronger therapeutic alliance was associated with significant improvements in pain (11,12), as well as improvements in disability and function and greater global perceived effect (11). Furthermore, a recent scoping review of therapeutic alliance in musculoskeletal physical therapy and occupational therapy practice also found some evidence that an enhanced therapeutic alliance has beneficial effects on adherence to interventions (13). Unfortunately, the few studies eligible for inclusion in these reviews all involved patients who had chronic low back pain, and thus research investigating the therapeutic alliance between physical therapists and patients with knee OA is required.

All of the literature investigating the therapeutic alliance in chronic musculoskeletal pain has focused on face-to-face, in-person consultations between patients and clinicians. However, interest in (14,15), and evidence to support (16,17), remotely delivered health care (i.e., telerehabilitation) for chronic musculoskeletal conditions is growing. Given that physical touch and nonverbal cues (e.g., gestures, head nodding, and eye contact) play an important role in communication between physical therapists and their patients (18,19), the delivery of care via telerehabilitation has been regarded as possibly having a negative impact on the therapeutic alliance (20,21). Although there is qualitative evidence that both physical therapists and patients with OA who have received, or delivered, care via telephone (20,22) and Skype video-conferencing (23) believe that they are able to develop strong relationships, no previous studies have quantitively investigated the therapeutic alliance between physical therapists and patients with OA consulting via telerehabilitation. Thus, the goal of this study was to explore the therapeutic alliance developed between physical therapists and patients with knee OA during telephone-delivered consultations for exercise advice and support. Specific aims were to describe and compare physical therapist and patient ratings of the therapeutic alliance, to determine whether the therapeutic alliance changes over time, and to evaluate whether individual patient or physical therapist characteristics are associated with therapeutic alliance.

MATERIALS AND METHODS

Design. This study used data collected concurrently from the intervention arm of a randomized controlled trial (RCT; Australian New Zealand Clinical Trials Registry ANZCTRN 12616000054415) evaluating the effectiveness of incorporating exercise advice and support by physical therapists for adults with knee OA into an existing musculoskeletal telephone service delivered by nurses (24).

Patients. Eighty-four patients with knee OA were randomized to the intervention arm of the RCT. Briefly, inclusion criteria for involvement in the trial included meeting the National Institute for Health and Care Excellence OA clinical criteria (age ≥45 years, activity-related joint pain, and morning stiffness ≤30 minutes) (2), having average knee pain of ≥4 on an 11-point numeric rating scale, and having a history of knee pain for at least 3 months. Exclusion criteria have been published elsewhere (24). Patients were recruited from metropolitan, regional, and rural areas across Australia using advertisements on social media, through

community organizations, in medical clinics, on the radio, in newspapers, and using previous volunteer databases.

Patient demographic variables were collected at baseline. These included: 1) age; 2) sex; 3) living arrangement (alone or with others); 4) employment status (currently employed part- or full-time or unemployed/retired); 5) education level (<3 years of high school, ≥3 years of high school, some tertiary training, graduated from university or polytechnic, or any postgraduate study); 6) self-efficacy, measured using the Arthritis Self-Efficacy Scale (25); overall scores range 3–30, with higher scores indicating greater self-efficacy, and scores for subscales (pain, function, other symptoms) range 1–10; 7) physical function, measured using the Western Ontario and McMaster Universities Osteoarthritis Index (26); scores range 0–68, where lower scores indicate better function; and 8) overall average knee pain in the past week, measured with a numeric rating scale (range 0–10, where 0 = no pain and 10 = worst pain possible).

Physical therapists. Eight physical therapists were recruited in Victoria, Australia, to deliver the intervention for the trial. Selection criteria included a physical therapy qualification, at least 2 years of musculoskeletal professional experience, and current Australian registration to practice. Demographic variables were collected from each physical therapist, including years of clinical experience, work setting (private and/or public), previous experience delivering care remotely (yes or no), and previous training in behavior change (yes or no).

Intervention. The intervention has been described in detail elsewhere (24). Briefly, all patients (including those in the control arm) received an initial telephone call from a nurse who staffed the existing musculoskeletal help-line, where the patients received general information and advice about OA. Patients who were allocated to the intervention group additionally received between 5 and 10 telephone consultations from 1 of the 8 physical therapists over a 26-week period, which are the focus of this study. The number of consultations was negotiated between patients and physical therapists based on individual progress and goals. Initial consultations were approximately 40 minutes in length, and follow-up consultations were approximately 20 minutes. Physical therapists worked with patients to devise goals and an action plan that involved a home-based structured strengthening exercise program and/or a physical activity plan.

Over the 26 weeks, physical therapists adjusted the program as necessary, while also providing support by working to increase patient knowledge and understanding of knee OA and the benefits of exercise. Physical therapists used person-centered practice principles and behavior-change techniques to help increase the patient's motivation to exercise and build confidence in their ability to undertake, and adhere to, an exercise program. These principles and skills were taught to the physical therapists during a training program by HealthChange Australia (http://www.healthchange.com/) prior to trial commencement (27). This training involved an initial 2-day workshop, a period of practice consultations with patients with knee OA, and a final follow-up training day (27). Briefly, the methodology involved a set of practice principles to foster effective communication, techniques to identify and address barriers to behavior change, and a framework to guide decision-making.

Patients were provided with a study folder containing information about OA and its effective management, as well as exercise instructions. Patients were also mailed 3 exercise resistance bands for home exercises and provided with access to a study website that contained video demonstrations of each exercise.

Therapeutic alliance measures. Therapeutic alliance was measured using the Working Alliance Inventory–Short Form (WAI-SF) (28,29), a commonly used valid and reliable measure (29). The WAI-SF contains 12 statements relating to the perceived trust and agreement between the therapist and client (e.g., "My patient/physiotherapist and I agree about the things they/I will need to do in therapy to help improve my situation"). Responses to each item range across on a 7-point scale from never feeling that way to always feeling that way. Overall scores range 12–84 (with higher scores indicating a stronger therapeutic alliance) (28,29). The WAI-SF has 3 subscales: Task (agreement on management methods being used; items 1, 2, 8, and 12), Bond (feelings of appreciation and trust; items 3, 5, 7, and 9), and Goal (agreement on aims and objectives of treatment; items 4, 6, 10, and 11). Subscale scores range 4–28, with higher scores indicating a stronger alliance.

Patients with knee OA and physical therapists completed the WAI-SF after their second consultation (at approximately week 4 of the 26-week intervention), as recommended (29). Patients then completed the WAI-SF again as part of their 26-week follow-up questionnaire at the end of the intervention period. Physical therapists completed the WAI-SF immediately after their final consultation with the patient, which occurred anywhere between weeks 21 and 26 of the 26-week intervention, depending on the individual patient. For ease of reporting, the initial completion of the WAI-SF will be referred to as week 4 and the final as week 26.

Statistical analysis. Analysis was undertaken using Stata software, version 15.1. Mean ± SDs of patient and physical therapist characteristics, and of therapeutic alliance ratings, were calculated. To address each of the aims, mixed linear regression models were fit for each of the subscale and total therapeutic alliance scores, including data from both time points and from physical therapists and patients. Random intercepts for physical therapists and for patients were included to account for the clustering of measurements within patients and patients within physical therapists. Restricted maximum likelihood was used to estimate parameters, and the Kenward and Roger small-sample

Table 1. Characteristics of patients with knee osteoarthritis (n = 84)*

Characteristics	Values
Male, no. (%)	30 (36)
Age, years	62.3 ± 9.3
Body mass index, kg/m^2	30.9 ± 6.8
Location, no. (%)†	
Metropolitan	45 (54)
Nonmetropolitan	39 (46)
Living arrangements, no. (%)	
Living with others	71 (85)
Living alone	13 (15)
Employment status, no. (%)	
Work full- or part-time	35 (42)
Unemployed or retired	49 (58)
Education, no. (%)	
<3 years high school	5 (6)
≥3 years high school	19 (23)
Some tertiary training	21 (24)
Graduated university/polytechnic	24 (29)
Any postgraduate study	15 (18)
Knee pain (NRS)	6.0 ± 1.5
Physical function (WOMAC)	29.4 ± 10.3
No. of calls with physical therapist during trial	6.3 ± 1.8
Self-efficacy (ASES)	
Pain	6.0 ± 1.7
Function	7.5 ± 1.6
Other symptoms	6.7 ± 7.8
Total	20.2 ± 4.0

* Values are the mean ± SD unless indicated otherwise. NRS = numeric rating scale (range 0–10, where lower scores indicate less pain); WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index (range 0–68, where lower scores indicate better function); ASES = Arthritis Self-Efficacy Scale (subscales range 1–10, where higher scores indicate greater self-efficacy; total scores range 3–30).
† Defined according to The Australian Statistical Geography Standard Remoteness Structure (http://www.abs.gov.au/websitedbs/d3310114.nsf/home/remoteness+structure).

method was used to adjust for the small numbers of physical therapist clusters.

To address the first and second aims, fixed effects for time (week 4 or 26) and an indicator for whether a physical therapist or a patient provided the measurement, as well as the interaction between time and this factor, were included. These fixed effects were used to compare changes of ratings from week 4 to week 26 within physical therapists and patients, as well as between physical therapists and patients at weeks 4 and 26. For the third aim, patient/physical therapist characteristics were additionally included in the mixed linear regression models, including the main effect and all 2- and 3-way interactions with time and with physical therapist/patient. Linear regression assumptions (linearity, normality, and heteroscedasticity of residuals) were assessed with standard diagnostic plots. Results in the article are based on complete case analyses. In sensitivity analyses, missing alliance measures were imputed 25 times using chained equations with predictive mean matching, with missing values imputed from among the 3 nearest neighbors (30). Imputation models included all demographic characteristics in Table 1.

RESULTS

Characteristics of patients with knee OA. Approximately one-third of patients in the sample (Table 1) were male (36%) and the mean ± SD age of the patients was 62.3 ± 9.3 years. Just over half of patients (54%) lived in metropolitan areas of Australia. At baseline, the mean ± SD knee pain was 6.0 ± 1.5 on an 11-point numeric rating scale. Patients completed a mean ± SD 6.3 ± 1.8 of a maximum of 10 telephone consultations with physical therapists.

Characteristics of physical therapists. Four of the 8 physical therapists were male. Half of the physical therapists worked exclusively in private physical therapy settings (Table 2). Physical therapists had a mean ± SD 13.8 ± 8.2 years of clinical experience, and none had previous experience delivering care via telephone, but 2 had previously delivered care via video-conferencing. Some physical therapists had completed postgraduate training in knee OA or behavior-change techniques, and most had completed postgraduate training in exercise therapy.

Therapeutic alliance ratings. Table 3 shows mean therapeutic alliance ratings by patients and physical therapists at week 4 and week 26 for each of the 3 subscales of the WAI-SF (Goal, Task, and Bond) and overall total scores. Overall, both patient and physical therapist scores were high. At weeks 4 and 26, the mean ± SD total patient ratings were 75.3 ± 7.4 and 73.3 ± 9.7 of 84, respectively, with mean ± SD physical therapist scores of 71.0 ±

Table 2. Characteristics of physical therapists (n = 8)*

Characteristics	Values
Male	4 (50)
Age, mean ± SD years	35.4 ± 8.2
Clinical experience, mean ± SD years	13.8 ± 8.2
No. of patients consulted with during the trial, mean ± SD	10.5 ± 2.1
Work setting	
Mixed private and public	2 (25)
Private	5 (63)
Public	1 (12)
Previous experience delivering care remotely via technology	
Yes	2 (25)
No	6 (75)
Postgraduate training in knee osteoarthritis	
Yes	3 (37)
No	5 (63)
Postgraduate training in exercise	
Yes	7 (88)
No	1 (12)
Postgraduate training in behavior change†	
Yes	3 (37)
No	5 (63)

* Values are the number (%) unless indicated otherwise.
† Excluding the trial-specific training in person-centered principles and behavior change techniques.

Table 3. Ratings of therapeutic alliance by patients with knee osteoarthritis and their treating physical therapists*

	Patient ratings (n = 84)			Physical therapist ratings (n = 8)			Difference between patient and physical therapist ratings†	
Subscales‡	Week 4 (n = 70)	Week 26 (n = 81)	Change over time, mean (95% CI)	Week 4 (n = 82)	Week 26 (n = 76)	Change over time, mean (95% CI)	Week 4 mean (95% CI)	Week 26 mean (95% CI)
Goal	25.7 ± 2.8	24.6 ± 3.9	−1.1 (−1.9, −0.2)	24.3 ± 1.8	24.7 ± 2.2	0.4 (−0.4, 1.2)	1.4 (0.5, 2.2)	−0.1 (−0.9, 0.8)
Task	25.1 ± 2.5	24.7 ± 3.5	−0.5 (−1.3, 0.3)	23.3 ± 2.4	23.9 ± 3.3	0.5 (−0.3, 1.2)	1.8 (1.1, 2.6)	0.9 (0.1, 1.7)
Bond	24.5 ± 3.2	24.0 ± 3.8	−0.5 (−1.3, 0.2)	23.4 ± 2.1	24.5 ± 2.4	1.0 (0.2, 1.8)	1.1 (0.3, 1.9)	−0.4 (−1.2, 0.3)
Total§	75.3 ± 7.4	73.3 ± 9.7	−2.1 (−4.1, 0.0)	71.0 ± 5.5	73.1 ± 7.2	1.8 (−0.2, 3.8)	4.3 (2.2, 6.4)	0.4 (−1.6, 2.4)

* Values are the mean ± SD unless indicated otherwise. Differences between the number of consultations at weeks 4 and 26 are due to missing data. 95% CI = 95% confidence interval.
† Calculated as (patient score) – (physical therapist score). Positive values indicate a higher patient rating.
‡ Scores range 4–28.
§ Total scores range 12–84.

5.5 at week 4 and 73.1 ± 7.2 at week 26. Over weeks 4 and 26, subscale scores ranged from 24.0 to 25.7 of 28 for patients, and from 23.3 to 24.7 for physical therapists.

Differences between patient and physical therapist ratings. At week 4, patients rated the overall alliance, as well as all subscales, significantly higher than physical therapists (Table 3). At 26 weeks, patients rated only the Task subscale significantly higher than physical therapists (mean difference 0.9 [95% confidence interval (95% CI) 0.1, 1.7]), with no significant differences between ratings for the Goal or Bond subscales or total scores.

Change in alliance over time. For patients, scores for the Goal subscale decreased significantly over time (–1.1 [95% CI –1.9, –0.2]). Scores for the total, as well as Task and Bond subscales, did not change significantly. For physical therapists, ratings for the Bond subscale of the alliance improved significantly from week 4 to 26 (1.0 [95% CI 0.2, 1.8]). Scores for the total, as well as Goal and Task subscales, did not change significantly.

Characteristics associated with alliance ratings. Table 4 shows patient and physical therapist characteristics that were significantly associated with total therapeutic alliance scores. Scatter plots of each continuous variable against alliance ratings are provided in Supplementary Figures 1–8, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23890/abstract. Patients who lived with others rated the alliance significantly higher than those who lived alone at both week 4 and week 26 (regression coefficient 7.6 [95% CI 2.8, 12.4] and 5.7 [95% CI 1.0, 10.5], respectively). Patients who were treated by physical therapists who had prior experience delivering care remotely rated the alliance significantly lower at week 26 than those consulting physical therapists who had no such experience (regression coefficient –4.6 [95% CI –8.5, –0.7]). Higher ratings of self-efficacy for "other" arthritis symptoms were associated with higher ratings of therapeutic alliance at week 4 (for each 1-unit increase in self-efficacy, there was an estimated 1.2-unit increase in alliance [95% CI 0.1, 2.2]). Each 1-unit increase in self-efficacy for pain was associated with an increase in alliance at week 26 of 1.0 units (95% CI 0.0, 2.0). Each additional telephone consultation during the trial was associated with an increase in alliance ratings at week 26 of 1.0 units (95% CI 0.1, 1.9).

Patient age was inversely associated with physical therapists' alliance ratings at 26 weeks (–0.3 [95% CI –0.4, –0.1]), as physical therapists' alliance ratings tended to decrease with increasing age of the patient. Similarly, physical therapists' years of experience were inversely associated with their ratings of alliance at 26 weeks (–0.2 [95% CI –0.5, –0.0]), with more experienced physical therapists tending to rate the alliance lower. Patient self-efficacy (total scores) was positively associated with physical therapists' alliance ratings at week 26 (0.4 [95% CI 0.0, 0.9]) and self-efficacy for managing pain was positively associated with physical therapists' alliance scores at week 26 (1.1 [95% CI 0.0, 2.2]), as physical therapists rated the alliance higher for patients with higher self-efficacy.

DISCUSSION

The aim of this study was to describe and explore the therapeutic alliance between physical therapists and patients with knee OA during telephone consultations for exercise advice and support. We found that both patient and physical therapist ratings of the alliance were high. Patients rated the alliance significantly higher than physical therapists early in treatment, with scores generally converging after their final consultation. However, differences in ratings were generally small (e.g., up to 6.4 of 84 units) and of uncertain clinical relevance. A number of baseline characteristics were associated with higher alliance ratings both by patients (living with others, being treated by a physical therapist with no prior experience delivering care remotely, having higher self-efficacy, and having more telephone consultations) and physical therapists (having less clinical experience and treating patients who were younger and had higher self-efficacy). However, given the large number of characteristics explored and the subsequent risk of Type I error, these findings must be interpreted with caution.

Our findings reflect research in patients with psychological conditions, who also tend to rate the alliance higher than their

Table 4. Association between baseline variables and ratings of therapeutic alliance*

	Patient rating		Physical therapist rating	
	Total week 4	Total week 26	Total week 4	Total week 26
Binary variables, mean ± SD				
Patient sex				
Male	74.2 ± 8.3	71.5 ± 10.2	70.2 ± 5.1	72.3 ± 7.2
Female	76.0 ± 6.9	74.3 ± 9.4	71.4 ± 5.7	73.6 ± 7.2
Regression coefficient (95% CI)	–1.9 (–5.5, 1.8)	–2.6 (–6.1, 0.9)	–1.2 (–4.7, 2.3)	–1.3 (–4.9, 2.3)
Patient living arrangements				
Living with others	76.5 ± 6.6	74.0 ± 9.3	71.4 ± 4.9	72.9 ± 7.1
Living alone	69.0 ± 8.9	68.6 ± 11.4	68.5 ± 7.6	74.7 ± 7.6
Regression coefficient (95% CI)	7.6 (2.8, 12.4)	5.7 (1.0, 10.5)	3.1 (–1.3, 7.6)	–1.6 (–6.5, 3.3)
Patient employment status				
Work full- or part-time	77.0 ± 6.5	73.4 ± 10.7	71.6 ± 4.4	75.0 ± 6.3
Unemployed/retired	74.2 ± 7.8	73.2 ± 9.1	70.5 ± 6.1	71.7 ± 7.5
Regression coefficient (95% CI)	2.1 (–1.6, 5.9)	0.4 (–3.1, 3.8)	1.1 (–2.4, 4.5)	3.2 (–0.3, 6.8)
Physical therapist previous experience delivering care remotely				
Yes	75.8 ± 8.5	70.0 ± 12.3	70.6 ± 4.8	72.1 ± 6.5
No	75.1 ± 7.0	74.6 ± 8.3	71.1 ± 5.7	73.5 ± 7.4
Regression coefficient (95% CI)	0.1 (–4.0, 4.2)	–4.6 (–8.5, –0.7)	–0.5 (–4.4, 3.3)	–1.6 (–5.6, 2.4)
Physical therapist training in behavior change†				
Yes	77.9 ± 6.3	73.1 ± 12.1	69.9 ± 4.6	71.4 ± 6.2
No	74.0 ± 7.7	73.4 ± 8.4	71.5 ± 5.9	73.9 ± 7.5
Regression coefficient (95% CI)	3.9 (0.0, 7.9)	–0.1 (–3.8, 3.7)	–1.6 (–5.4, 2.1)	–3.4 (–7.3, 0.4)
Continuous variables, regression coefficient (95% CI)				
Patient age	0.0 (–0.2, 0.2)	0.0 (–0.2, 0.1)	–0.1 (–0.3, 0.1)	–0.3 (–0.4, –0.1)
Patient self-efficacy				
Pain	0.5 (–0.6, 1.6)	1.0 (0.0, 2.0)	0.7 (–0.3, 1.7)	1.1 (0.0, 2.2)
Function	0.6 (–0.6, 1.7)	0.3 (–0.8, 1.4)	0.2 (–0.8, 1.3)	0.8 (–0.3, 1.9)
Other symptoms	1.2 (0.1, 2.2)	0.9 (–0.1, 1.8)	0.6 (–0.4, 1.6)	0.7 (–0.3, 1.7)
Total	0.4 (–0.1, 0.9)	0.4 (0.0, 0.8)	0.3 (–0.1, 0.7)	0.4 (0.0, 0.9)
Physical function (WOMAC)	0.0 (–0.2, 0.1)	0.0 (–0.2, 0.1)	0.0 (–0.2, 0.1)	–0.1 (–0.3, 0.0)
Knee pain (NRS)	–0.1 (–1.4, 1.1)	–0.9 (–2.0, 0.3)	–0.1 (–1.3, 1.1)	0.4 (0.0, 1.9)
Physical therapist years of experience	0.0 (–0.3, 0.2)	0.0 (–0.2, 0.3)	–0.1 (–0.3, 0.1)	–0.2 (–0.5, –0.0)
Number of calls during the trial	–0.6 (–1.6, 0.4)	1.0 (0.1, 1.9)	–0.2 (–1.1, 0.8)	–0.5 (–1.6, 0.6)
Ordinal variables, mean ± SD, regression coefficient (95% CI)				
Education level				
<3 years high school	73.6 ± 6.0, –2.9 (–10.8, 5.1)	73.8 ± 5.5, –2.1 (–9.9, 5.8)	74.6 ± 3.3, 6.1 (–1.7, 14.0)	75.0 ± 7.1, 2.5 (–5.3, 10.4)
≥3 years high school	78.8 ± 5.9, 2.6 (–3.0, 8.1)	72.5 ± 2.6, –3.3 (–8.6, 2.0)	70.7 ± 5.7, 2.2 (–3.1, 7.5)	72.1 ± 7.6, –1.2 (–6.6, 4.1)
Some tertiary training	73.6 ± 7.7, –2.6 (–8.2, 2.9)	74.3 ± 9.4, –1.7 (–6.8, 3.5)	71.9 ± 4.5, 3.4 (–1.7, 8.6)	75.1 ± 6.7, 2.4 (–2.8, 7.6)
Graduated university/polytechnic	73.1 ± 7.9, –3.5 (–8.9, 1.8)	71.3 ± 10.3, –4.8 (–9.8, 0.2)	71.2 ± 5.7, 3.0 (–2.1, 8.0)	72.0 ± 6.9, –0.7 (–5.9, 4.6)
Any postgraduate study	76.8 ± 7.8, (ref.)	75.9 ± 9.0, (ref.)	68.5 ± 6.1, (ref.)	72.5 ± 8.0, (ref.)
Physical therapist work setting				
Private	74.9 ± 7.6, 1.6 (–4.5, 7.7)	72.8 ± 9.0, –3.9 (–9.8, 2.0)	71.2 ± 6.0, –0.1 (–6.1, 5.8)	74.2 ± 7.0, 2.8 (–3.1, 8.8)
Private and public	78.4 ± 4.6, 5.3 (–1.8, 12.4)	72.8 ± 12.5, –3.6 (–10.4, 3.2)	70.2 ± 4.0, –0.9 (–7.7, 5.9)	71.3 ± 6.8, –0.4 (–7.3, 6.4)
Public	72.9 ± 9.1 (ref.)	76.6 ± 8.1 (ref.)	71.2 ± 1.7 (ref.)	71.0 ± 7.9 (ref.)

* 95% CI = 95% confidence interval; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index (range 0–68, where lower scores

treating psychologists (31–35), and with differences being greater early in treatment (34). Although comparison across studies is difficult, 1 study also used the WAI-SF and reported a 6-point difference between client and therapist scores (31). Our observed patient-physical therapist differences were smaller (4.3 at week 4 and 0.4 at week 26), which may reflect differences in the sample characteristics and/or nature of the interventions. Patients may rate the alliance higher than clinicians because patients have a different role in therapy (36) and likely commence treatment with expectations of improvement and an optimistic outlook, which can influence the strength of the alliance (37). Clinicians, on the other hand, may not necessarily have the same treatment expectations, because they have a broader reference group when judging the strength of the alliance with any 1 patient (34,35). Although our findings suggest that physical therapists' perceptions of the therapeutic alliance do not necessarily reflect those of their patient, the

clinical importance of such small differences in perceptions of the alliance remains unclear. Commonly used measures of alliance, such as the WAI-SF, do not have standardized scores or cutoffs to define a score representative of a good alliance (38). Thus, determining whether the magnitude of the differences between patients and clinicians that we observed in our study has a meaningful impact on therapy is difficult. Further research is required to determine the consequences, if any, of differences in patient and clinician perceptions of therapeutic alliance.

This is the first study to investigate the therapeutic alliance between patients with knee OA and physical therapists who consult via telephone. Given the importance of physical touch and nonverbal gestures (e.g., head nodding, eye contact) in physical therapy (18,19), the therapeutic alliance developed during telephone consultations may be weaker than that of traditional face-to-face consultations. Unfortunately, the design of our trial did not include a face-to-face treatment arm to test this hypothesis. Furthermore, the few previous studies investigating therapeutic alliance in face-to-face physical therapy consultations have used different measurement tools (11,12), making direct comparisons with our findings difficult. One study (11), which used a different version of the WAI, investigated the strength of the alliance between physical therapists and patients with chronic low back pain during 8 weeks of exercise and spinal manipulative therapy. After the second consultation, therapeutic alliance scores ranged from 97 to 99 of 112 (i.e., 86–88% of maximum), which are similar to what we observed in our telephone consultations (71–75 of 84, i.e., 85–89% of maximum). Consulting via telephone may not significantly compromise the strength of the alliance between patients and clinicians, but further research comparing telephone and face-to-face consultations using the same physical therapists, and the same intervention, is required. In other populations, there is evidence that the alliance between nonprofessional health coaches who consult via telephone with adolescents with arthritis (39) and parents of children with psychosocial and behavioral issues (40) is similar to the alliance when face-to-face, which also suggests that a strong alliance can exist without in-person face-to-face contact.

An unexpected finding of our study was that patients rated the alliance higher at week 26 if their physical therapist did not have prior experience delivering care remotely. Among our small sample of 8 physical therapists, only 2 had prior experience delivering care remotely, both of whom had done so via video-conferencing. To our knowledge, no previous studies have investigated the influence of prior telerehabilitation experience on therapeutic alliance. Given that remote consultations require strong communication skills to compensate for lack of physical or visual cues (41), possibly those with prior experience consulting remotely would be more skilled at communicating effectively via the telephone in this study, and hence have an enhanced ability to develop a stronger alliance with their patients. Why patients in our study perceived a lower alliance with physical therapists who had prior experience delivering care remotely is not clear. Possibly the inexperienced physical therapists were cognizant of their inexperience and actively compensated for this gap in ways that led to their patients perceiving a greater alliance at week 26. Due to the large numbers of patient and physical therapist characteristics that were considered, we emphasize that this finding is inconclusive.

We also found that higher patient self-efficacy was associated with a stronger alliance, rated by both patients and physical therapists. To our knowledge, no previous studies have explored the relationship between self-efficacy and therapeutic alliance. Intuitively, those with greater self-efficacy managing their condition may be more willing and confident to engage in an active self-management exercise approach like that used in this study, and therefore may perceive a greater alliance with their physical therapist. Similarly, physical therapists may find it easier to build a positive alliance with patients who have higher self-efficacy, because the physical therapists may experience less resistance when prescribing a management program. These findings provide further evidence of the important role of self-efficacy in knee OA, and further reinforce the idea that self-efficacy is an important parameter to consider in clinical consultations (42).

Physical therapists in our study tended to rate the alliance lower for older patients at 26 weeks. This finding contrasts with previous research in psychotherapy, where patient age has been found to be unrelated to therapist-rated or patient-rated alliance (36,43). In fact, there is some evidence that older patients are more satisfied with physical therapy (44) and medical care (45), and that medical physicians are more likely to have person-centered encounters (which is linked to greater therapeutic alliance [46]) with patients ages >65 years (45). In physical therapy, there is some evidence that patient age is related to adherence to sports rehabilitation and that age moderates the relationship between therapeutic alliance and adherence, with younger patients having greater adherence (47). Our study does not make clear why physical therapists tended to rate therapeutic alliance lower with older patients. However, given that older patients in our study did not perceive a lower alliance themselves, the clinical implications of these findings are unclear and further investigation is required.

In our study, patients who lived alone tended to rate therapeutic alliance lower than those who lived with others at both time points. This finding may reflect differences in social support, which has previously been linked to the development of therapeutic alliance. For example, patients with psychological conditions who have better social functioning have been found to develop a stronger alliance with their psychologist (32,34,48), and, conversely, patients who have difficulty maintaining social relationships are more likely to have difficulty forming a strong alliance (49). Collectively, these findings suggest that clinicians should be aware of their patient's social and/or living arrangements and may need to specifically focus on improving the strength of their therapeutic relationship for those who are living alone and/or socially isolated.

Our study has a number of strengths and limitations. Strengths include the use of a valid and reliable measure of

therapeutic alliance and the collection of data at 2 time points from both physical therapists and patients. Our study also had a relatively large sample size, and uniquely, investigated telephone calls as the medium for consultations. Our study also had some limitations. We were unable to evaluate whether therapeutic alliance with the physical therapist was a treatment effect moderator in our clinical trial, because the control arm did not undergo consultations with a physical therapist, and thus we could only collect therapeutic alliance data from within the intervention arm. Therefore, we cannot make inferences about the effect of therapeutic alliance on clinical outcomes or treatment adherence. We included a large number of patient/physical therapist characteristics as independent variables for potential association with therapeutic alliance, and thus our findings should be interpreted with caution. The precise timing, and number, of consultations varied among individual patients, which may have contributed to some variation in alliance scores. In addition, patients and physical therapists assessed the alliance at different time points, with patients completing the WAI-SF during their 26-week follow-up questionnaire, and physical therapists completing it immediately following their final consultation with the patient. Therefore, physical therapists may have been able to recall the alliance more accurately than patients, which may have contributed to some of the observed differences in physical therapist and patient scores. Our measure of therapeutic alliance, the WAI-SF, was not developed specifically for use in musculoskeletal physical therapy and has been found to exhibit a ceiling effect in these populations (50). Finally, our physical therapists underwent an intensive training program in person-centered principles before the trial, and therefore our findings may not be generalizable to other populations of physical therapists who have not undergone such training.

In conclusion, patients and physical therapists who consulted via telephone both rated the alliance highly. Physical therapist perceptions of the therapeutic alliance tend to be lower than those of patients early in treatment and some subgroups of patients rate the alliance less strongly than others. Our findings suggest that physical therapists should not assume that their perceptions of the alliance reflect those of their patients, and that physical therapists may benefit by actively working to strengthen the alliance with patients who live alone and/or have low self-efficacy. However, differences in alliance ratings between groups and over time were small and of unclear clinical importance. Further research is required to determine the clinical relevance of differences between patient and therapist ratings of the alliance and to investigate differences between the alliance in remotely delivered and face-to-face consultations.

AUTHOR CONTRIBUTIONS

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

Study conception and design. Lawford, Bennell, Campbell, Kasza, Hinman.

Acquisition of data. Campbell.

Analysis and interpretation of data. Lawford, Bennell, Kasza, Hinman.

REFERENCES

1. Pereira D, Peleteiro B, Araujo J, Branco J, Santos RA, Ramoa E. The effect of osteoarthritis definition on prevalence and incidence estimates: a systematic review. Osteoarthritis Cartilage 2011;19:1270–85.
2. National Institute for Health and Care Excellence. Osteoarthritis: care and management, clinical guideline CG177. London: National Institute for Health and Care Excellence; 2014.
3. Royal Australian College of General Practitioners. Guideline for the management of knee and hip osteoarthritis. East Melbourne (Australia): Royal Australian College of General Practitioners; 2018.
4. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. Psychother Theory Res Pract 1979;16:252–60.
5. Flückiger C, Del Re AC, Wampold BE, Horvath AO. The alliance in adult psychotherapy: a meta-analytic synthesis. Psychotherapy (Chic) 2018;3:316–40.
6. Mistiaen P, van Osch M, van Vliet L, Howick J, Bishop FL, Di Blasi Z, et al. The effect of patient–practitioner communication on pain: a systematic review. Eur J Pain 2015;20:675–88.
7. Mack JW, Block SD, Nilsson M, Wright A, Trice E, Friedlander R. Measuring therapeutic alliance between oncologists and patients with advanced cancer: the Human Connection Scale. Cancer 2009;115:3302–11.
8. Rosti-Otajarvi E, Mantynen A, Koivisto K, Huhtala H, Hamalainen P. Predictors and impact of the working alliance in the neuropsychological rehabilitation of patients with multiple sclerosis. J Neurol Sci 2014;338:156–61.
9. Lakke SE, Meerman S. Does working alliance have an influence on pain and physical functioning in patients with chronic musculoskeletal pain: a systematic review. J Compassionate Health Care 2016;3:1.
10. Taccolini Manzoni AC, Bastos de Oliveira NT, Nunes Cabral CM, Aquaroni Ricci N. The role of the therapeutic alliance on pain relief in musculoskeletal rehabilitation: a systematic review. Physiother Theory Pract 2018;34:901–15.
11. Ferreira PH, Ferreira ML, Maher CG, Refshauge KM, Latimer J, Adams RD. The therapeutic alliance between clinicians and patients predicts outcome in chronic low back pain. Phys Ther 2013;93:470–8.
12. Fuentes J, Armijo-Olivo S, Funabashi M, Miciak M, Dick B, Warren S, et al. Enhanced therapeutic alliance modulates pain intensity and muscle pain sensitivity in patients with chronic low back pain: an experimental controlled study. Phys Ther 2014;94:477–89.
13. Babatunde F, MacDermid J, MacIntyre N. Characteristics of therapeutic alliance in musculoskeletal physical therapy and occupational therapy practice: a scoping review of the literature. BMC Health Serv Res 2017;17:375–98.
14. Lawford BJ, Bennell KL, Hinman RS. Consumer perceptions of and willingness to use remotely delivered service models for exercise management of knee and hip osteoarthritis: a cross-sectional survey. Arthritis Care Res (Hoboken) 2017;69:667–76.
15. Lawford BJ, Bennell KL, Kasza J, Hinman RS. Physical therapists' perceptions of telephone- and internet video–mediated service models for exercise management of people with osteoarthritis. Arthritis Care Res (Hoboken) 2018;70:398–408.

16. Shukla H, Nair S, Thakker D. Role of telerehabilitation in patients following total knee arthroplasty: evidence from a systematic literature review and meta-analysis. J Telemed Telecare 2016;23: 339–46.
17. Cottrell MA, Galea OA, O'Leary SP, Hill AJ, Russell TG. Real-time telerehabilitation for the treatment of musculoskeletal conditions is effective and comparable to standard practice: a systematic review and meta-analysis. Clin Rehabil 2016;1:625–38.
18. Hiller A, Guillemin AM, Delany C. Exploring healthcare communication models in private physical therapy practice. Patient Educ Couns 2015;98:1222–8.
19. Bjorbaekm WS, Mengshoel AM. "A touch of physiotherapy": the significance and meaning of touch in the practice of physiotherapy. Physiother Theory Pract 2016;32:10–9.
20. Lawford BJ, Delany C, Bennell KL, Hinman RS. "I was really pleasantly surprised": first-hand experience and shifts in physical therapist perceptions of telephone-delivered exercise therapy for knee osteoarthritis: a qualitative study. Arthritis Care Res (Hoboken) 2019;71:545–57.
21. Cottrell MA, Hill AJ, O'Leary SP, Raymer ME, Russell TG. Service provider perceptions of telerehabilitation as an additional service delivery option within an Australian neurosurgical and orthopaedic physical therapy screening clinic: a qualitative study. Musculoskelet Sci Pract 2017;32:7–16.
22. Lawford BJ, Delany C, Bennell KL, Hinman RS. "I was really sceptical...But it worked really well": a qualitative study of patient perceptions of telephone-delivered exercise therapy by physical therapists for people with knee osteoarthritis. Osteoarthritis Cartilage 2018;26:741–50.
23. Hinman R, Nelligan RK, Bennell KL, Delany C. "Sounds a bit crazy, but it was almost more personal:" a qualitative study of patient and clinician experiences of physical therapist–prescribed exercise for knee osteoarthritis via Skype. Arthritis Care Res (Hoboken) 2017;69:1834–44.
24. Hinman RS, Lawford BJ, Campbell PK, Briggs AM, Gale J, Bills C. Telephone-delivered exercise advice and behavior change support by physical therapists for people with knee osteoarthritis: protocol for the telecare randomized controlled trial. Phys Ther 2017;97: 524–36.
25. 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.
26. 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.
27. Lawford BJ, Delany C, Bennell KL, Bills C, Gale J, Hinman RS. Training physical therapists in person-centered practice for people with osteoarthritis: a qualitative case study. Arthritis Care Res (Hoboken) 2018;70:558–70.
28. Tracey TJ, Kokotovic AM. Factor structure of the working alliance inventory. Psychol Assess 1989;1:207–10.
29. Horvath AO, Greenberg LS. Development and validation of the Working Alliance Inventory. J Couns Psychol 1989;36:223–33.
30. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med 2011;30: 377–99.
31. Lysake PH, Davis LW, Buck KD, Outcalt S, Ringer JM. Negative symptoms and poor insight as predictors of the similarity between client and therapist ratings of therapeutic alliance in cognitive behavior therapy for patients with schizophrenia. J Nerv Ment Dis 2011;199:191–5.
32. Couture SM, Roberts DL, Penn DL, Cather C, Otto MW, Goff D. Do baseline client characteristics predict the therapeutic alliance in the treatment of schizophrenia? J Nerv Ment Dis 2006;194:10–4.
33. Schönberger M, Humle F, Teasdale TW. The development of the therapeutic working alliance, patients' awareness and their compliance during the process of brain injury rehabilitation. Brain Inj 2006;20:445–54.
34. Shick Tryon G, Collins Blackwell S, Felleman Hammel E. A meta-analytic examination of client–therapist perspectives of the working alliance. Psychother Res 2007;17:629–42.
35. Hartmann A, Joos A, Orlinsky DE, Zeeck A. Accuracy of therapist perceptions of patients' alliance: exploring the divergence. Psychother Res 2015;25:408–19.
36. Bachelor A. Clients' and therapists' views of the therapeutic alliance: similarities, differences and relationship to therapy outcome. Clin Psychol Psychother 2013;20:118–35.
37. Constantino MJ, Arnow BA, Blasey C, Agras WS. The association between patient characteristics and the therapeutic alliance in cognitive-behavioral and interpersonal therapy for bulimia nervosa. J Consult Clin Psychol 2005;73:203–11.
38. Working Alliance Inventory. Criteria for "good alliance." URL: http://wai.profhorvath.com/criteria.
39. White M, Stinson JN, Lingley-Pottie P, McGrath PJ, Gill N, Vijenthira A. Exploring therapeutic alliance with an internet-based self-management program with brief telephone support for youth with arthritis: a pilot study. Telemed J E Health 2012;18:271–6.
40. Lingely-Pottie P, McGrath PJ. A therapeutic alliance can exist without face-to-face contact. J Telemed Telecare 2006;12:396–9.
41. Car J, Freeman GK, Partridge MR, Sheikh A. Improving quality and safety of telephone based delivery of care: teaching telephone consultation skills. Qual Saf Health Care 2004;13:2–3.
42. Somers TJ, Wren AA, Shelby RA. The context of pain in arthritis: self-efficacy for managing pain and other symptoms. Curr Pain Headache Rep 2012;16:502–8.
43. Hersoug AG, Hoglend P, Monsen JT, Havik OE. Quality of working alliance in psychotherapy: therapist variables and patient/therapist similarity as predictors. J Psychother Pract Res 2001;10:205–16.
44. Hush JM, Cameron K, Mackey M. Patient satisfaction with musculoskeletal physical therapy care: a systematic review. Phys Ther 2011;91:25–36.
45. Peck BM. Age-related differences in doctor-patient interaction and patient satisfaction. Curr Gerontol Geriatr Res 2011;2011:137492.
46. Pinto RZ, Ferreira ML, Oliveira VC, Franco MR, Adams R, Maher CG, et al. Patient-centred communication is associated with positive therapeutic alliance: a systematic review. J Physiother 2012;58:77–87.
47. Levy AR, Polman RC, Borkoles E. Examining the relationship between perceived autonomy support and age in the context of rehabilitation adherence in sport. Rehabil Psychol 2008;53: 224–30.
48. Keller SM, Zoellner LA, Feeny NC. Understanding factors associated with early therapeutic alliance in PTSD treatment: adherence, childhood sexual abuse history, and social support. J Consult Clin Psychol 2010;78:974–9.
49. Constantino M, Castonguay L, Schut A. The working alliance: a flagship for the scientist-practitioner model in psychotherapy. Counseling based on process research: applying what we know. Cambridge: Cambridge University Press; 2000.
50. Hall AM, Ferreira ML, Clemson L, Ferreira P, Latimer J, Maher CG. Assessment of the therapeutic alliance in physical rehabilitation: a RASCH analysis. Disabil Rehabil 2012;34:257–66.

Arthritis Care & Research
Vol. 72, No. 5, May 2020, pp 661–668
DOI 10.1002/acr.23882

American College of Rheumatology
Empowering Rheumatology Professionals

Feasibility and Preliminary Outcomes of a Physical Therapist–Administered Physical Activity Intervention After Total Knee Replacement

Meredith B. Christiansen,[1] Louise M. Thoma,[1] Hiral Master,[1] Dana Voinier,[1] Laura A. Schmitt,[1] Melissa L. Ziegler,[1] Michael P. LaValley,[2] and Daniel K. White[1]

Objective. To explore the feasibility, fidelity, safety, and preliminary outcomes of a physical therapist–administered physical activity (PA) intervention after total knee replacement (TKR).

Methods. People who had undergone a unilateral TKR and were receiving outpatient physical therapy (PT) were randomized to a control or intervention group. Both groups received standard PT for TKR. The intervention included being provided with a Fitbit Zip, step goals, and 1 phone call a month for 6 months after discharge from PT. Feasibility was measured by rates of recruitment and retention, safety was measured by the frequency of adverse events, and fidelity was measured by adherence to the weekly steps/day goal created by the physical therapist and participant monitoring of steps/day. An Actigraph GT3X measured PA, which was quantified as steps/day and minutes/week of engaging in moderate-to-vigorous PA. Our preliminary outcome was the difference in PA 6 months after discharge from PT between the control and intervention groups.

Results. Of the 43 individuals who were enrolled, 53.4% were women, the mean ± SD age was 67.0 ± 7.0 years, and the mean ± SD body mass index was 31.5 ± 5.9 kg/m^2. For both the control and intervention groups, the recruitment and retention rates were 64% and 83.7%, respectively, and adherence to the intervention ranged from 45% to 60%. No study-related adverse events occurred. The patients in the intervention group accumulated a mean 1,798 more steps/day (95% confidence interval [95% CI] 240, 3,355) and spent 73.4 more minutes/week (95% CI –14.1, 160.9) engaging in moderate-to-vigorous PA at 6 months than those in the control group.

Conclusion. A physical therapist–administered PA intervention is feasible and safe, demonstrates treatment fidelity, and may increase PA after TKR. Future research is needed to establish the effectiveness of the intervention.

INTRODUCTION

Knee osteoarthritis (OA) is a leading cause of pain and disability in older adults (1). Total knee replacement (TKR) is the definitive treatment for knee OA that resolves most knee pain and limitations in physical function (1). However, physical activity (PA), which is defined as any body movement that results in energy expenditure above a resting level (2), remains mostly unchanged after TKR (3,4) and leaves individuals who have undergone TKR at risk of inactivity-related health problems after surgery, such as weight gain (5,6), cardiovascular disease (7), diabetes mellitus (8,9), and premature death (10,11). The number of TKR surgeries, which have doubled over the past 15 years in the US (12), is expected to increase substantially; more than 3.5 million TKRs are expected to be performed every year by 2030 (13). Thus, there is a critical need to improve PA in individuals after TKR.

Outpatient physical therapy (PT) is an optimal setting to deliver a PA intervention. Physical therapists are experts in prescribing and tailoring therapeutic exercise programs to promote PA. After TKR, physical therapists often treat patients in an outpatient setting 2–3 times a week for 6 to 8 weeks as part of standard postoperative care, which aligns with the recommended frequency of one-on-one visits that are needed to promote behavioral change to increase PA (14,15). Furthermore, >90% of patients in PT agree that physical therapists should discuss PA as a part of care (16).

ClinicalTrials.gov identifier: NCT02724137.
Supported by the NIH (grants R21-AR-071079-01A1, U54-GM-104941, and K12-HD-055931-01).
[1]Meredith B. Christiansen, DPT, Louise M. Thoma, DPT, PhD, Hiral Master, MPT, MPH, Dana Voinier, DPT, Laura A. Schmitt, DPT, Melissa L. Ziegler, PhD, Daniel K. White, MSPT, ScD, MSc: University of Delaware, Newark; [2]Michael P. LaValley, PhD: Boston University, Boston, Massachusetts.

No potential conflicts of interest relevant to this article were reported.
Address correspondence to Daniel K. White, MSPT, ScD, MSc, STAR Health Sciences Complex, 540 South College Avenue, Newark, DE 19713. E-mail: dkw@udel.edu.
Submitted for publication November 20, 2018; accepted in revised form March 19, 2019.

SIGNIFICANCE & INNOVATIONS

- Outpatient physical therapy (PT) is an ideal setting to increase physical activity (PA) because physical therapists are experts in exercise and commonly see patients after total knee replacement (TKR) for multiple visits; however, it is unclear if a PA intervention in PT is feasible and safe, demonstrates fidelity, and can increase PA in individuals after TKR.
- We found the physical therapist–administered PA intervention to be feasible and safe and to have modest fidelity. Moreover, we observed clinically meaningful increases in PA among those in the intervention group.
- Study participants receiving the intervention reached levels of PA that were consistent with the 2018 Physical Activity Guidelines for Americans, and the majority of those in the intervention walked >6,000 steps/day (a level needed to prevent the development of functional limitation) 6 months after discharge from PT.

Recording steps/day with an activity tracker, e.g., a Fitbit Zip (Fitbit, Inc.), and receiving weekly steps/day goals from a health care professional has been shown to increase PA (17–19). This type of intervention increases PA through a behavioral change technique that includes feedback from the activity tracker and self-monitoring of the goal (20,21). At present, it is not known if a physical therapist–administered PA intervention is feasible and safe and demonstrates fidelity. Additionally, it is not known if such an intervention can increase PA in people after TKR. The purpose of this study was to explore the feasibility, safety, fidelity, and preliminary outcomes of a physical therapist–administered PA intervention after TKR.

PATIENTS AND METHODS

Design. This was a single-center, randomized, controlled pilot study. This study was registered at and approved by the University of Delaware Institutional Review Board, and informed consent was obtained from all study participants.

Study participants. We recruited individuals who were receiving outpatient PT for a unilateral TKR at the Delaware Physical Therapy Clinic at the University of Delaware (UDPT) in Newark, Delaware. Potential study participants were informed by their treating physical therapist about our study during the initial PT appointment at UDPT. If the person was interested in participating in the study, a research assistant screened them for eligibility at their next PT appointment. Patients were eligible to participate if they were >45 years of age and had self-reported "yes" when asked if they were interested in increasing PA. Participants were excluded if they had any additional comorbidities that would prevent them from participating in a PA intervention (e.g., unstable angina), had another lower extremity surgery in the previous 6 months, or had another lower extremity surgery planned within 6 months after enrolling in the study.

Study procedures. Once consent was obtained and the patients were enrolled, participants were randomized into the control or intervention group. A research assistant randomized each participant using a manila envelope with notecards labeled "A" for intervention and "B" for control. The research assistant who analyzed the PA data was blinded to group assignment.

Control group. Participants in the control group received standard outpatient PT provided by a licensed physical therapist using the Delaware Physical Therapy Clinic–Rehab Practice Guidelines for Unilateral TKR (see Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23882/abstract). Standard PT also included a printed home exercise program with an exercise log that was updated weekly by the physical therapist (see Supplementary Table 2, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.23882/abstract). Participants in the control group also received a monthly phone call for 6 months after discharge from PT to discuss their overall health and to serve as a reminder for the PA data collection at 6 months and 12 months. These calls also served to match the number of points of contact between the intervention and control groups. Weekly steps/day goals and feedback on PA was not provided to control group participants.

Intervention group. The intervention group received a Fitbit Zip, weekly steps/day goal from a physical therapist, and monthly follow-up phone calls from a research assistant (for 6 months) to promote PA. In addition, the intervention group was provided with the same standard outpatient PT of the control group.

Fitbit Zip (activity tracker). A Fitbit Zip was provided within 1 week of enrolling in the study. Participants were given written and face-to-face instructions on how to set up, use, and sync the Fitbit Zip to their smartphone, tablet, or home computer using the app provided by Fitbit. If the participant did not have a smartphone, tablet, or home computer, they were instructed to use the Fitbit Zip as a pedometer. We asked participants to wear the Fitbit Zip around their waist at their right anterior superior iliac crest daily (during waking hours) and to monitor their steps/day count with the Fitbit Zip and record steps/day count in their home exercise program. Extra batteries and instructions on how to install the batteries were provided as needed. After the 6-month follow-up, participants returned their Fitbit Zip to the research team.

Weekly steps/day goal setting. Participants in the intervention group jointly set weekly steps/day goals with the physical therapist, starting at least 3 weeks after TKR surgery. Several factors were considered to progress the steps/day goal, including if the weekly steps/day goal was achieved in at least 4 of the last 7 days, the participant's current health status, the physical therapist's clinical judgment, and the participant's personal PA goal. The end goal was

to walk at least 6,000 steps/day until discharge. We used this as a goal because this threshold is known to protect against development of functional limitation in people with knee OA (22). If the participant achieved 6,000 steps/day, they were encouraged to continue to increase their steps/day since health benefits persist with more PA. Weekly steps/day goals were recorded on a standardized goal-setting form and a home exercise program log (see Supplementary Table 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23882/abstract).

Monthly follow-up phone calls. A research assistant contacted participants in the intervention group once a month for 6 months and continued jointly setting steps/day goals with the study participant after discharge from PT. The purpose of the phone calls once a month for 6 months was to provide additional PA goal setting after discharge from PT. A research assistant, instead of a physical therapist, called participants because this model of delivery is more feasible in the clinical practice.

Baseline participant characteristics. Using electronic medical records, research assistants extracted self-reported health history, demographic characteristics, worst knee pain on a visual analog scale (VAS), and objective knee range of motion (ROM) measurements that were collected by a licensed physical therapist at the initial PT evaluation (i.e., our study baseline). Quantifying worst knee pain using a VAS is a valid and reliable method to assess pain (23). ROM was measured with the participant in supine position using a standardized goniometer, with the axis positioned at the lateral epicondyle, the stationary arm in line with the greater trochanter, and the moveable arm aligned with the lateral malleolus.

Feasibility, safety, and fidelity. We assessed feasibility by evaluating recruitment and retention rates. The recruitment rate was calculated as the number of participants who enrolled in the study divided by the number of people who were screened for eligibility at UDPT. The retention rate was calculated as the number of participants who completed the study from baseline to 6 months divided by the total number of enrolled participants. We measured fidelity 2 ways, including the physical therapist's adherence to administering the intervention (establishing a weekly steps/day goal with the participant) and the participant's ability to adhere to the intervention (monitoring steps/day). Adherence to administering the weekly steps/day goal by the physical therapist was measured by a research assistant counting the number of goals documented in the home exercise program log from baseline to discharge from PT. We classified adherence as "achieved" for participants who had ≥80% of the weekly steps/day goal recorded by the physical therapist and "not achieved" for those with <80% of the weekly steps/day goals recorded, which is

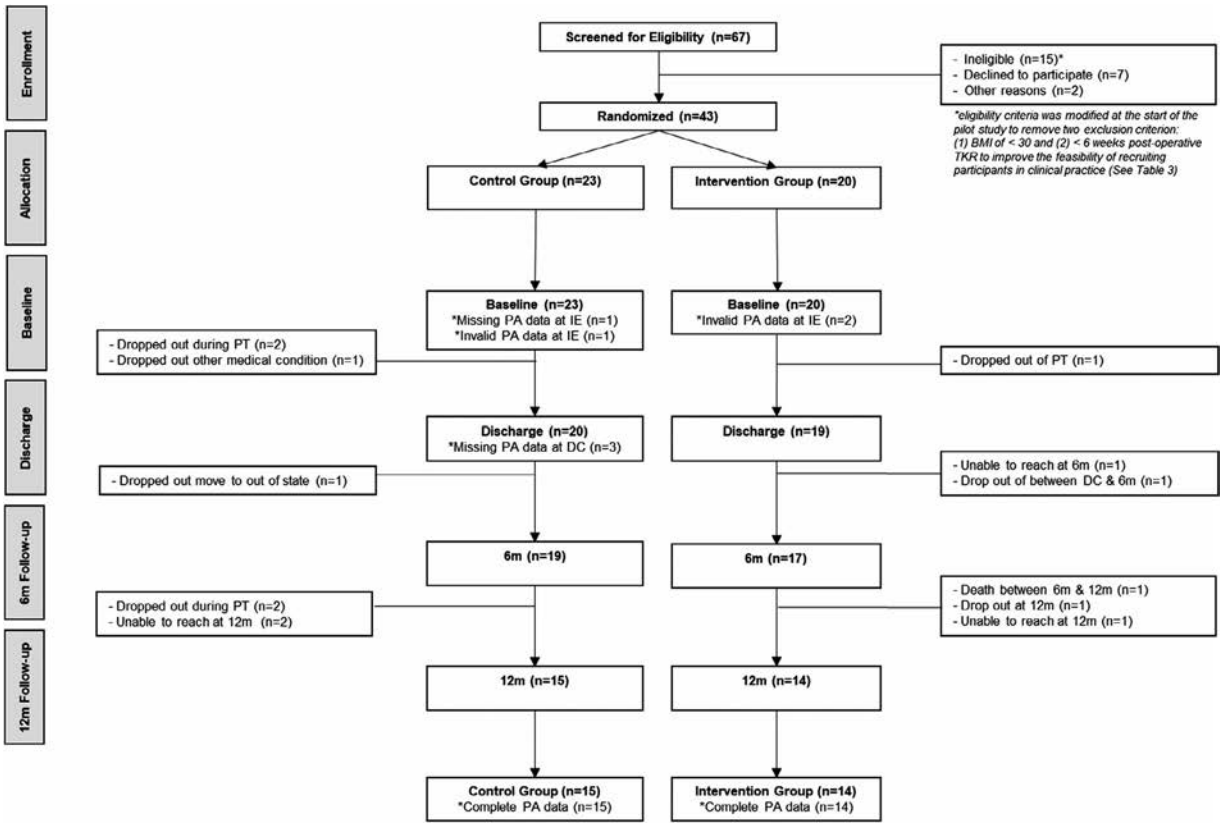

Figure 1. Consolidated Standards of Reporting Trials (CONSORT) diagram demonstrating the flow of enrolled participants. BMI = body mass index; TKR = total knee replacement; PT = physical therapy; PA = physical activity; DC = discharge.

consistent with the definition of adherence from a pharmacologic perspective (24). Participants documented the number of days they recorded steps/day using the Fitbit Zip, and those with ≥80% of documented steps/day in the home exercise program log from baseline to discharge were classified as "achieved" adherence and those with <80% as "not achieved." Lastly, we assessed safety by reporting the number of adverse events recorded from baseline to 12 months. An adverse event was defined as any unfavorable or unintended diagnosis, sign, symptom, or disease associated with the study, which may or may not be related to the intervention.

PA measures. PA was objectively measured using the Actigraph GT3X, which is a reliable and valid measurement of PA in older adults after TKR (25). PA was quantified in units of steps/day and time spent engaging in moderate-to-vigorous PA/week. Participants wore the monitor on the waist, positioned at the right anterior superior iliac spine, from the time they got up in the morning until they went to sleep. The participants were also instructed to remove the monitor when the device could get wet (e.g., showering, swimming). Participants wore the monitor for 1 week at baseline, discharge, 6-month follow-up, and 12-month follow-up (Figure 1). Data from the Actigraph GT3X were downloaded and analyzed following a standardized protocol (https://epi.grants.cancer.gov/nhanes_pam/) reported by Troiano et al (26). Briefly, we defined a valid wear day as ≥10 hours of wear time, excluding time with ≥90 consecutive minutes of <100 activity counts, and we only included PA data with ≥4 valid wear days (26).

Sample size calculation. The sample size for this study was based on the general notion for pilot studies to recruit at least 30 participants for each parameter and expect 20% to drop out (27,28). At the start of the study, we intended to recruit 72 participants, with 36 participants in each group to successfully retain a total of 60 participants in the control and intervention groups at 12 months. We stopped enrollment at 43 participants as we received funding for a larger trial to investigate the effectiveness of our intervention.

Statistical analysis. We calculated the mean and 95% confidence intervals (95% CIs) for continuous variables and frequency counts for categorical variables to describe participant characteristics. Independent *t*-tests and chi-square tests were used to evaluate baseline differences between the intervention and control groups for continuous and categorical variables, respectively. We described safety ("adverse event"/"no event") and intervention fidelity ("achieved"/"not achieved") in terms of frequencies. We calculated the mean ± SDs and 95% CIs between groups at baseline, discharge, 6-month follow-up and 12-month follow-up for PA outcomes, which included steps/day and minutes/week in moderate-to-vigorous PA. We also calculated within-group differences from baseline to 6 months and 12-months using the mean ± SD and 95% CI. For all statistical analyses, a *P* value of less than 0.05 was considered statistically significant. Data were analyzed using SAS, version 9.4.

RESULTS

Study sample. Of the enrolled participants, 53.4% were women, the mean ± SD age was 67.0 ± 7.0 years, the mean ± SD body mass index was 31.5 ± 5.9 kg/m, and the mean ± SD time from TKR surgery to the first PT appointment was 13.8 ± 21.3 days (Table 1). On average, participants in the intervention group attended a mean ± SD 20 ± 8 PT sessions for a mean ± SD

Table 1. Participant characteristics at baseline*

	All participants (n = 43)	Intervention (n = 20)	Control (n = 23)	*P*
Age, years	67.0 ± 7.0	66.5 ± 6.9	67.5 ± 7.2	0.5
Women, no. (%)	23 (53.4)	8 (40.0)	15 (65.2)	0.1
BMI (kg/m^2)†	31.5 ± 5.9	31.1 ± 5.6	32.0 ± 6.3	0.6
Education (≥ college), no. (%)	20 (51.3)	10 (55.0)	10 (47.6)	0.6
White (other), no. (%)	39 (91)	19 (95)	20 (87)	0.7
Time from TKR to PT, days	13.8 ± 21.3	9.6 ± 7.0	18.0 ± 35.9‡	0.3
Total number of PT visits§	19 ± 8	20 ± 8	18 ± 7	0.4
Duration of PT in weeks§	9.7 ± 3.4	10.4 ± 5.5	9.0 ± 2.7	0.3
Unilateral TKR side (right), no. (%)	22 (52.7)	15 (75.0)	7 (30.4)	<0.004
Comorbidity (≥1), no. (%)¶	21 (50.3)	11 (55.0)	10 (45.5)	0.4
Knee pain at worst (VAS)	6.5 ± 2.9	6.5 ± 2.8	6.6 ± 3.0	0.9
Knee flexion (degrees)¶	85.4 ± 31.4	76.3 ± 36.2	94.5 ± 26.6	0.07
Knee extension (degrees)#	5.7 ± 4.4	5.5 ± 3.8	5.9 ± 5.0	0.9

* Values are the mean ± SD unless indicated otherwise. Comorbidities included cardiovascular disease, chronic obstructive pulmonary disease, stroke, cancer, history of falls, diabetes mellitus, or depression. BMI = body mass index; TKR = total knee replacement; PT = physical therapy; VAS = visual analog scale.
† Data missing: n = 40.
‡ One participant in the control group started PT 6 months after unilateral TKR.
§ Data missing: n = 39.
¶ Data missing: n = 42.
# Knee flexion and extension passive range of motion on unilateral TKR side, lacking knee extension. Data missing: n = 42.

Table 2. Reasons patients did not enroll in the study*

Ineligible (n = 15)	
Not meeting BMI criteria†	1
Not meeting number of days postoperative TKR†	2
Bilateral TKR	1
Lower extremity surgery within 6 months	7
Not interested in increasing PA	3
Medical reason (orthostatic hypotension)	1
Decline to participate (n = 7)	
Not interested in participating in research	7
Other reasons (n = 2)	
Unable to consent due to clinic logistics	1
Previously enrolled in a PA study	1

* PA = physical activity.
† To improve the feasibility of recruiting participants in clinical practice, eligibility criteria were modified at the start of the pilot study to remove 2 exclusion criterion, including participants with body mass index (BMI) of <30 and <6 weeks postoperative total knee replacement (TKR).

10.4 ± 5.5 weeks, and participants in the control group attended a mean ± SD 18 ± 7 PT sessions for mean ± SD 9.0 ± 2.7 weeks. We did not adjust for any baseline characteristics in our statistical model (Table 1) because there were no differences between control and intervention groups (except for the side of the TKR [P = 0.004], which we did not consider to be a potential confounder).

Feasibility, safety, and fidelity. Between March 2016 and June 2017, 67 individuals were screened after having unilateral TKR, of whom 43 were eligible for the study and were enrolled and randomized (Table 2). The recruitment rate was 64% (43 of 67 participants) (Figure 2). The overall retention rate was 83.7% (36 of 43 participants) at 6 months and was 67% (29 of 43 participants) at 12 months (Figure 2). During the study, 3 adverse events occurred with participants in the control group. One participant had 2 adverse events by injuring the contralateral knee and bruising a tendon in the back of the ipsilateral knee between the 6-month and 12-month follow-up visits. Another participant was diagnosed with gout in the first metatarsal joint during PT. All adverse events were unrelated to the PA intervention. Sixty percent (12 of 20 participants) of those in the intervention group monitored steps/day at least 80% of the time while in PT, while 45% of physical therapists (9 of 20) were in adherence with the administration of the intervention by documenting weekly steps/day goals ≥80% of the time.

Physical activity. Participants in the control group (n = 23) had similar counts of daily steps at baseline as the intervention group (n = 20) (a mean of 2,214 steps/day [95% CI 1,573, 2,855] for the control group and a mean of 2,494 steps/day [95% CI 1,803, 3,168] for the intervention group) (Table 3). However, participants in the control group spent less minutes/week engaging in moderate-to-vigorous PA than in the intervention group, although this difference did not meet statistical significance (for the control group, a mean 19.4 minutes/week [95% CI 9.9, 28.9] of moderate-to-vigorous PA and for the intervention group, a mean 35.6 minutes/week [95% CI 16.7, 54.5]) (Table 3). Of the 43 participants, complete PA data was available for 15 in the control group and 14 in the intervention group from baseline to 12 months (Figure 1).

At 6 months, the intervention group accumulated a mean 1,798 (95% CI 240, 3,355) more steps/day and spent a mean 73.4 (95% CI −14.1, 160.9) more minutes/week engaging in moderate-to-vigorous PA than the control group (Table 3). In particular, the control group walked 3,941 (95% CI 3,021, 4,863) steps/day and spent 77.2 (95% CI 33.3, 121.2) minutes/week in moderate-to-vigorous PA, while the intervention group walked 5,739 (95% CI 4,369, 7,109) steps/day and

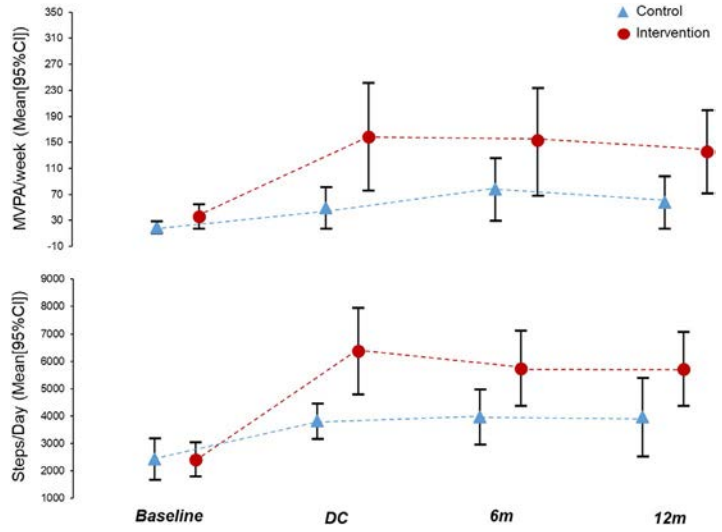

Figure 2. Physical activity between groups at baseline, discharge (DC), 6 months, and 12 months. For the control group, the sample size was n = 21 at baseline, n = 17 at discharge, n = 19 at 6 months, and n = 15 at 12 months. For the intervention group, the sample size was n = 18 at baseline, n = 19 at discharge, n = 17 at 6 months, and n = 14 at 12 months. MVPA = moderate-to-vigorous physical activity, 95% CI = 95% confidence interval.

Table 3. Physical activity between groups at baseline, discharge, 6 months, 12 months*

	Control group	Intervention group	Difference between groups
Steps/day			
Baseline	2,214 ± 1,407 (1,573, 2,855)	2,494 ± 1,391 (1,803, 3,186)	280 ± 1,340 (–631, 1,191)
Discharge	3,823 ± 1,356 (3,126, 4,520)	6,389 ± 3,279 (4,808, 7,969)	2,566 ± 2,561 (828, 4,303)
6 months	3,941 ± 1,910 (3,021, 4,863)	5,739 ± 2,665 (4,369, 7,109)	1,798 ± 2,296 (240, 3,355)
12 months	4,169 ± 1,890 (3,123, 5,217)	6,114 ± 1,989 (4,966, 7,262)	1,945 ± 1,938 (466, 3,422)
MVPA/week, minutes			
Baseline	19.4 ± 20.8 (9.9, 28.9)	35.6 ± 37.9 (16.7, 54.5)	16.2 ± 29.9 (–3.7, 35.7)
Discharge	46.4 ± 65.4 (12.8, 80.0)	158.4 ± 172.3 (75.2, 241.3)	112.0 ± 133.1 (21.5, 202.2)
6-month follow-up	77.2 ± 91.3 (33.3, 121.2)	150.6 ± 161.2 (67.7, 233.5)	73.4 ± 129.0 (–14.1, 160.9)
12-month follow-up	57.7 ± 72.7 (17.5, 98.0)	133.8 ± 98.1 (77.1, 190.4)	76.1 ± 85.9 (10.5, 141.5)

* Values are the mean ± SD (95% confidence interval). MVPA = moderate-to-vigorous physical activity.

spent 150.6 (95% CI 67.7, 233.5) minutes/week engaging in moderate-to-vigorous PA at 6 months (Table 3). At 12 months, the intervention group accumulated 1,945 more steps/day (95% CI 466, 3,422) and spent 76.1 more minutes/week (95% CI 10.5, 141.5) in moderate-to-vigorous PA than the control group (Table 3). The control group walked 4,169 (95% CI 3,123, 5,217) steps/day and spent 57.7 (95% CI 17.5, 98.0) minutes/week in moderate-to-vigorous PA, while the intervention group walked 6,114 (95% CI 4,966, 7276) steps/day and spent 133.8 (95% CI 77.1, 190.4) minutes/week in moderate-to-vigorous PA at 12 months (Table 3).

DISCUSSION

We found that a physical therapist–administered PA intervention was feasible and safe and had modest fidelity. Also, the intervention was potentially effective for people after TKR, which was demonstrated by our findings of improvements in PA that met clinically meaningful levels. For instance, the intervention group walked about 6,000 steps/day after discharge from outpatient PT, which is a meaningful threshold for reducing the risk of functional limitation in people with knee OA (22). In contrast, the control group walked about the same amount as the general population, i.e., ~4,000 steps/day (29). Moreover, the intervention group spent >150 minutes/week engaging in moderate-to-vigorous PA at discharge and at the 6-month follow-up visit, meeting the 2018 Department of Health and Human Services PA guidelines for aerobic activity (10). Our results show promise in informing ways to change behavior that leads to inactivity and subsequent weight gain that are common after TKR.

Despite increases in PA, adherence to the intervention was modest from both participants and treating physical therapists. Less than half of the physical therapists (45%) documented that they consistently provided weekly steps/day goals, and 60% of participants monitored their steps/day. While we were unable to track how often patients looked at their Fitbit Zip, it is possible that self-monitoring steps/day from a device with a daily goal may be more useful to increase PA than a weekly goal given by a PT. The intervention was safe, as no participants in the intervention group reported an adverse event. We also noted that the monthly phone calls made by a research assistant from the 6-month to the 12-month follow-up did not result in a noticeable change in PA.

Our findings, that use of a Fitbit Zip monitor in conjunction with step goals was feasible and safe, showed good fidelity, and could increase PA after TKR, are consistent with other studies (30–32). A study by Losina et al demonstrated that financial incentives combined with telephone health coaching and a Fitbit led to a clinically meaningful increase in steps/day within the first 6 months after TKR (33). Furthermore, our finding that PA did not change from 6 to 12 months after surgery is consistent with the literature (3,34,35). In a meta-analysis, Hammett et al observed no change in PA at 6 months after TKR, with only a small change in PA at 12 months (3). Both the control and intervention groups in our study had no significant differences in PA from discharge to 6-month and 12-month follow-up visits, indicating that providing a PA intervention during outpatient PT may be an optimal time to improve PA.

Strengths of our study included measuring PA using an accelerometer-enabled device and a brief PA intervention. Our intervention took roughly 5 minutes/week to set and progress weekly steps/day goals, which improved the feasibility for clinical implementation. However, our study was not without limitations. We had a small sample size with a moderate dropout rate at 12 months. The dropout rate was equal between the control and intervention groups, with most participants dropping out after the 6-month follow-up visit. This rate indicates that a study follow-up of more than 6 months may be challenging in this patient population. Since this was a pilot study (i.e., we sought to determine feasibility, safety, and fidelity), we felt it was acceptable to have a small and underpowered sample size. There was high variability around the number of days after surgery that participants started outpatient PT and enrolled in our study; the mean ± SD days after TKR that participants enrolled in the study was 14 ± 21 days, and the range was 2–175 days. Outpatient PT typically commences between 2 and 12 weeks after TKR, and the time postsurgery when participants enrolled in our study was within this range. Therefore, this variability is expected (36). Lastly, we only included people after TKR who were interested in increasing their PA, which may have limited the generalizability of our sample. We did so because those who are not interested in PA

may require additional behavioral change techniques (e.g., motivational interviewing) to increase PA, which was not feasible in our study.

Overall, a physical therapist–administered PA intervention was feasible and safe, demonstrated modest fidelity, and promoted an increase in PA after TKR that appeared to remain for up to 12 months after discharge from PT. Our preliminary findings suggest that PT may indeed be a practical delivery model for a PA intervention after TKR. Given these study findings, further research is needed, with appropriate power, to establish the effectiveness of this intervention.

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. White 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. Christiansen, Thoma, White.
Acquisition of data. Christiansen, Master, Voinier, Schmitt.
Analysis and interpretation of data. Christiansen, Thoma, Master, Ziegler, LaValley, White.

REFERENCES

1. Weinstein AM, Rome BN, Reichmann WM, Collins JE, Burbine SA, Thornhill TS, et al. Estimating the burden of total knee replacement in the United States. J Bone Joint Surg Am 2013;95:385–92.
2. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep 1985;100:126–31.
3. Hammett T, Simonian A, Austin M, Butler R, Allen KD, Ledbetter L, et al. Changes in physical activity after total hip or knee arthroplasty: a systematic review and meta-analysis of six-and twelve-month outcomes. Arthritis Care Res (Hoboken) 2018;70:892–901.
4. Booth FW, Roberts CK, Laye MJ. Lack of exercise is a major cause of chronic diseases. Compr Physiol 2012;2:1143–211.
5. Zeni JA Jr, Snyder-Mackler L. Most patients gain weight in the 2 years after total knee arthroplasty: comparison to a healthy control group. Osteoarthritis Cartilage 2010;18:510–4.
6. Riddle DL, Singh JA, Harmsen WS, Schleck CD, Lewallen DG. Clinically important body weight gain following knee arthroplasty: a five-year comparative cohort study. Arthritis Care Res (Hoboken) 2013;65:669–77.
7. Shortreed SM, Peeters A, Forbes AB. Estimating the effect of long-term physical activity on cardiovascular disease and mortality: evidence from the Framingham Heart Study. Heart 2013;99:649–54.
8. Sowers MR, Karvonen-Gutierrez CA. The evolving role of obesity in knee osteoarthritis. Curr Opin Rheumatol 2010;22:533–7.
9. Piva SR, Susko AM, Khoja SS, Josbeno DA, Fitzgerald GK, Toledo FG. Links between osteoarthritis and diabetes: implications for management from a physical activity perspective. Clin Geriatr Med 2015;31:67–87.
10. US Department of Health and Human Services: 2018 physical activity guidelines for Americans. 2018. URL: https://health.gov/paguidelines/second-edition/pdf/Physical_Activity_Guidelines_2nd_edition.pdf.
11. Hawker GA, Croxford R, Bierman AS, Harvey PJ, Ravi B, Stanaitis I. All-cause mortality and serious cardiovascular events in people with hip and knee osteoarthritis: a population based cohort study. PLoS One 2014;9:e91286.
12. Losina E, Thornhill TS, Rome BN, Wright J, Katz JN. The dramatic increase in total knee replacement utilization rates in the United States cannot be fully explained by growth in population size and the obesity epidemic. J Bone Joint Surg Am 2012;94:201–7.
13. Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. J Bone Joint Surg 2007;89:780–5.
14. Guide to physical therapist practice. Second Edition. American Physical Therapy Association. Phys Ther 2001;81:9–746.
15. Brawley LR, Rejeski WJ, King A. Promoting physical activity for older adults: the challenges for changing behavior. Am J Prev Med 2003;3 Suppl 2:172–83.
16. Black B, Ingman M, Janes J. Physical therapists' role in health promotion as perceived by the patient: descriptive survey. Phys Ther 2016;96:1588–96.
17. Chase JA. Interventions to increase physical activity among older adults: a meta-analysis. Gerontologist 2015;55:706–18.
18. Furber S, Monger C, Franco L, Mayne D, Jones LA, Laws R. The effectiveness of a brief intervention using a pedometer and step-recording diary in promoting physical activity in people diagnosed with type 2 diabetes or impaired glucose tolerance. Health Promot J Austr 2008;19:189–95.
19. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. JAMA 2007;298: 2296–304.
20. Lyons EJ, Lewis ZH. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. J Med Internet Res 2014;16:e192.
21. Conroy DE, Dubansky A, Remillard J, Murray R, Pellegrini CA, Phillips SM, et al. Using behavior change techniques to guide selections of mobile applications to promote fluid consumption. Urology 2017;99:33–7.
22. White DK, Tudor-Locke C, Zhang Y, Fielding R, LaValley M, Felson DT, et al. Daily walking and the risk of incident functional limitation in knee OA: an observational study. Arthritis Care Res (Hoboken) 2014;66:1328–36.
23. Hjermstad MJ, Fayers PM, Haugen DF, Caraceni A, Hanks GW, Loge JH. Studies comparing numerical rating scales, verbal rating scales, and visual analogue scales for assessment of pain intensity in adults: a systematic literature review. J Pain Symptom Manage 2011;41:1073–93.
24. Morrison A, Stauffer ME, Kaufman AS. Defining medication adherence in individual patients. Patient Prefer Adherence 2015;9: 893–7.
25. Almeida GJ, Irrgang JJ, Fitzgerald GK, Jakicic JM, Piva SR. Reliability of physical activity measures during free-living activities in people after total knee arthroplasty. Phys Ther 2016;96:898–907.
26. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, Mcdowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008;40:181–8.
27. Browne RH. On the use of a pilot sample for sample size determination. Stat Med 1995;14:1933–40.
28. Lancaster GA, Dodd S, Williamson PR. Design and analysis of pilot studies: recommendations for good practice. J Eval Clinical Pract 2004;10:307–12.
29. Thoma LM, Dunlop D, Song J, Lee J, Tudor-Locke C, Aguiar EJ, et al. Are older adults with symptomatic knee osteoarthritis less active than the general population? Analysis from the Osteoarthritis Initiative and the National Health and Nutrition Examination Survey. Arthritis Care Res (Hoboken) 2018;70:1448–54.
30. Paxton RJ, Forster JE, Miller MJ, Gerron KL, Stevens-Lapsley JE, Christiansen CL. A feasibility study for improved physical activity after total knee arthroplasty. J Aging Phys Act 2018;26:7–13.

31. Wang JB, Cadmus-Bertram LA, Natarajan L, White MM, Madanat H, Nichols JF, et al. Wearable sensor/device (Fitbit One) and SMS text-messaging prompts to increase physical activity in overweight and obese adults: a randomized controlled trial. Telemed J E Health 2015;21:782–92.
32. Cadmus-Bertram LA, Marcus BH, Patterson RE, Parker BA, Morey BL. Randomized trial of a fitbit-based physical activity intervention for women. Am J Prev Med 2015;49:414–8.
33. Losina E, Collins JE, Deshpande BR, Smith SR, Michl GL, Usiskin IM, et al. Financial incentives and health coaching to improve physical activity following total knee replacement: a randomized controlled trial. Arthritis Care Res (Hoboken) 2018;70:732–40.
34. Smith TO, Mansfield M, Dainty J, Hilton G, Mann CJ, Sackley CM. Does physical activity change following hip and knee replacement? Matched case-control study evaluating Physical Activity Scale for the Elderly data from the Osteoarthritis Initiative. Physiotherapy 2018;104:80–90.
35. Kahn TL, Schwarzkopf R. Does total knee arthroplasty affect physical activity levels? Data from the Osteoarthritis Initiative. J Arthroplasty 2015;30:1521–5.
36. Artz N, Elvers KT, Lowe CM, Sackley C, Jepson P, Beswick AD. Effectiveness of physiotherapy exercise following total knee replacement: systematic review and meta-analysis. BMC Musculoskelet Disord 2015;16:15.

Arthritis Care & Research
Vol. 72, No. 5, May 2020, pp 669–678
DOI 10.1002/acr.23898

American College of Rheumatology
Empowering Rheumatology Professionals

Depression Subtypes in Individuals With or at Risk for Symptomatic Knee Osteoarthritis

Alan M. Rathbun,[1] Megan S. Schuler,[2] Elizabeth A. Stuart,[3] Michelle D. Shardell,[4] Michelle S. Yau,[5] Joseph J. Gallo,[3] Alice S. Ryan,[1] and Marc C. Hochberg[1]

Objective. The present study was undertaken to identify depression subtypes in individuals with or at risk for symptomatic knee osteoarthritis (OA) and to evaluate differences in pain and disability trajectories between groups.

Methods. Participants (n = 4,486) were enrolled in the Osteoarthritis Initiative. Latent class analysis was applied to the 20-item Center for Epidemiologic Studies Depression Scale measured at baseline to identify groups with similar patterns of depressive symptoms, and subtypes were assigned using posterior probability estimates. The relationships between depression subtypes and Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain and disability subscales were modeled over 4 years and stratified by baseline knee OA status (symptomatic [n = 1,626] or at risk [n = 2,860]).

Results. Four subtypes were identified: asymptomatic (80.6%), catatonic (5.3%), anhedonic (10.6%), and melancholic (3.5%). Catatonic and anhedonic subtypes were differentiated by symptoms corresponding to psychomotor agitation and the inability to experience pleasure, respectively. The melancholic subtype expressed symptoms related to reduced energy and movement, anhedonia, and other somatic symptoms. Detectable mean differences in pain and disability compared to the asymptomatic group were observed for the anhedonic (1.5–2.3 WOMAC units) and melancholic (4.8–6.6 WOMAC units) subtypes, and associations were generally larger in individuals with symptomatic knee OA relative to those at risk.

Conclusion. Among individuals with or at risk for symptomatic knee OA, there is evidence of depression subtypes characterized by distinct clusters of depressive symptoms that have differential effects on reports of pain and disability over time. Our findings thus imply that depression interventions could be optimized by targeting the specific symptomology that these subtypes exhibit.

INTRODUCTION

Osteoarthritis (OA) is the most common joint disorder in the US; symptomatic knee OA affects ~9.3 million American adults and 10% of men and 13% of women 60 years of age or older (1,2). Knee OA represents a failure of normal joint repair and is often accompanied by symptoms of pain and disability (3). Worsening OA disease severity may lead to the development of psychiatric comorbidity, particularly depressive symptoms, which can exacerbate the course of pain, disability, and disease progression (4–8). The bidirectional relationship between OA disease severity and depressive symptoms may influence the severity of both, thus complicating medical management and contributing to higher health care costs, decreased quality of life, and greater mortality (9–11).

Depression presenting in chronic diseases is difficult to recognize; indeed, depressive symptoms in arthritis patients are underdiagnosed (12,13). A contributing factor to the underrecognition

This article was prepared using an Osteoarthritis Initiative (OAI) public-use data set, and its contents do not necessarily reflect the opinions or views of the OAI Study Investigators, the NIH, or the private funding partners of the OAI. The OAI is a public–private partnership between the NIH (contracts N01-AR-2-2258, N01-AR-2-2259, N01-AR-2-2260, N01-AR-2-2261, and N01-AR-2-2262) and private funding partners (Merck Research Laboratories, Novartis Pharmaceuticals, GlaxoSmithKline, and Pfizer, Inc.) and is conducted by the OAI Study Investigators. Private sector funding for the OAI is managed by the Foundation for the NIH. Drs. Hochberg and Rathbun are part of the OAI investigative team.

Supported by the Rheumatology Research Foundation (Scientist Development Award to Dr. Rathbun).

[1]Alan M. Rathbun, PhD, MPH, Alice S. Ryan, PhD, Marc C. Hochberg, MD, MPH: VA Maryland Health Care System and University of Maryland School of Medicine, Baltimore; [2]Megan S. Schuler, PhD: RAND Corporation, Boston, Massachusetts; [3]Elizabeth A. Stuart, PhD, Joseph J. Gallo, MD, MPH: Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland; [4]Michelle D. Shardell, PhD: University of Maryland School of Medicine, Baltimore; [5]Michelle S. Yau, PhD, MPH: Institute for Aging Research, Hebrew SeniorLife, Boston, Massachusetts.

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

Address correspondence to Alan M. Rathbun, PhD, MPH, Howard Hall Suite 200, 660 West Redwood Street, University of Maryland School of Medicine, Baltimore, MD 21201. E-mail: arathbun@som.umaryland.edu.

Submitted for publication November 30, 2018; accepted in revised form April 2, 2019.

SIGNIFICANCE & INNOVATIONS

- Depressive symptoms in individuals with or at risk for symptomatic knee osteoarthritis present as 1 of 4 unique subtypes.
- Depression subtypes are differentiated primarily by psychomotor agitation, anhedonia, and other somatic symptoms.
- Anhedonic and melancholic depression subtypes may be risk factors for increased pain and disability.
- Results highlight the need for protocols designed to address the specific symptomatology of different depression subtypes.

of depression in OA patients is the overlapping somatic symptomology between the conditions; as a result, many individuals are only treated for their chronic disease and not depressive symptoms (12). A meta-analysis estimated that depressive symptoms are present in 18.5% of adults with knee OA, a prevalence that is more than double that in the US general population and has remained unchanged over time, despite increases in depression treatment rates (14–16). Depression treatment in American adults predominantly consists of pharmacotherapy using antidepressant medications (15). However, many patients do not achieve symptomatic remission, and nonresponse is even more pervasive in those with chronic diseases (10,17).

Increasingly, major depressive disorder is recognized as heterogeneous with respect to clinical presentation (12). Official classifications and corresponding treatments for major depressive disorder have traditionally used a "one size fits all" approach, yet it is becoming more accepted that such definitions and management strategies do not accurately reflect the immense heterogeneity of depressive symptomology (12,18). In the research and clinical setting, depressive symptoms are generally evaluated in terms of symptom count or a dichotomous indicator, which does not differentiate patients with disparate symptomology (18). By contrast, nascent research has begun utilizing the depression symptomics framework, a methodology that evaluates how different constellations of depressive symptoms cluster and differentially impact health, well-being, and medical management (19).

Currently, variability in depressive symptoms among OA patients and the implications of distinct depression profiles on OA disease severity is poorly understood. Understanding heterogeneity in depressive symptoms among individuals with or at risk for symptomatic knee OA could lead to more personalized interventions that target individuals' symptomatic profiles in order to improve clinical outcomes for both conditions. The current study's objectives were to identify depression subtypes in individuals with or at risk for symptomatic knee OA based on patterns of depressive symptoms and to examine the impact of depression subtypes on pain and disability stratified by individuals with or at risk symptomatic knee OA.

MATERIALS AND METHODS

Study data and sample. The sample included participants from the Osteoarthritis Initiative (OAI), an observational cohort study designed to identify risk factors and biomarkers for the onset and progression of knee OA, and methodologic details have been published previously (20). Institutional review boards at each site approved the OAI study, and all participants provided informed consent. The OAI cohort (n = 4,796) was restricted to participants (n = 4,486) with baseline symptomatic knee OA data and baseline radiographs that were read centrally at Boston University by trained, certified, radiologic technicians (21). This sample (Figure 1) was used to examine heterogeneity in baseline depressive symptoms, and there were 1,626 and 2,860 individuals with or at risk (respectively) for symptomatic knee OA at baseline, defined as "pain most days of a month in past 12 months" (20). Only 1 knee per participant selected by random sampling was included in the analysis. Participants with complete baseline data and at least 1 follow-up observation were included in longitudinal analyses. Complete data on covariates and outcomes at study enrollment were available for 4,300 participants, and information on pain and disability during follow-up were available for 4,204, 4,043, 3,990, and 3,970 participants at the first, second, third, and fourth annual follow-up visits, respectively, in individuals with fully observed baseline data. Depressive episodes last between 6 to 12 months, and a 4-year follow-up period provided sufficient time to examine associations between baseline depression subtypes and pain and disability (22).

Depressive symptoms. The 20-item Center for Epidemiologic Studies Depression Scale (CES-D) was used to assess an array of depressive symptoms, such as psychomotor agitation/retardation, poor appetite, restless sleep, sadness, feelings of loneliness, social interactions, impaired concentration, and anhedonia (23). CES-D items have a reference timeframe corresponding to the occurrence of depressive symptoms in the prior week and response options ranging from 0–3, where increasing score is representative of greater symptomatic frequency (23). Binary indicators were created for each symptom, classifying individuals who responded with "2" (i.e., occasionally or a moderate amount of time) or "3" (i.e., most or all of the time) as having a given symptom, an approach that has been used in prior research (24). These 20 binary symptom indicators were used to identify depression subtypes.

Pain and disability. The Western Ontario and McMaster Universities Arthritis Index (WOMAC) (Likert version 3.1) was used to assess knee OA disease severity. The WOMAC assessed 3 distinct domains: stiffness (2 items), pain (5 items), and disability (17 items) (25). WOMAC item responses are on a Likert scale, ranging from 0 (none) to 4 (extreme), with higher scores indicating greater severity. Item scores are summed across subscales to calculate summary scores for each domain. In the current study,

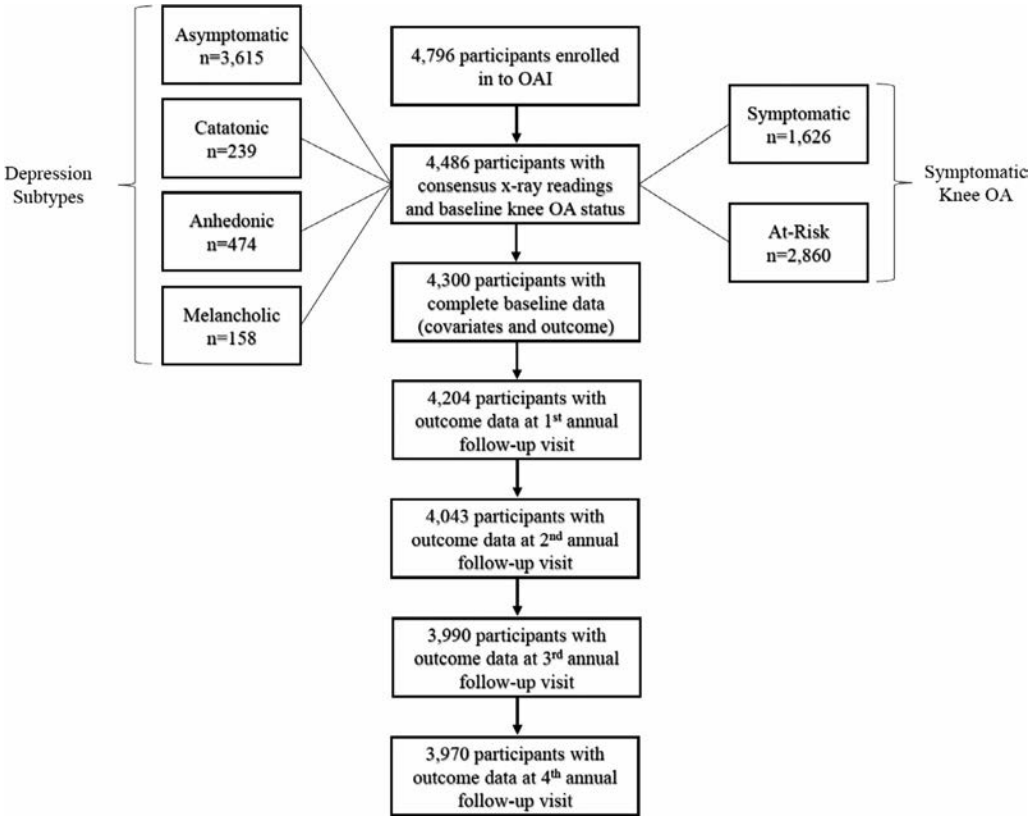

Figure 1. Study sample flow diagram. OAI = Osteoarthritis Initiative; OA = osteoarthritis.

pain and disability subscales were used as outcome measures and assessed at baseline and 4 annual follow-up visits. Pain and disability scores were rescaled to range 0–100 so that estimates can be compared to minimal perceptible clinical differences. Clinically significant differences on rescaled WOMAC pain and disability scores are ~9.7 and 9.3 units, respectively (26).

Confounders. Potential confounders measured at study baseline were selected a priori based on review of the research literature. Sociodemographic and behavioral measures were age (years), sex, race (white or nonwhite), marital status (married, widowed, divorced, separated, or never married), educational attainment (high school, college graduate, or graduate degree), health insurance (insured or uninsured), employment (employed or unemployed), alcohol consumption (none, minimal, or moderate), and smoking (never or former). Clinical characteristics included body mass index (BMI) (kg/m^2), comorbidity, history of knee injuries, treatment with analgesic medications, OA disease severity (Kellgren-Lawrence [K/L] grade), and total WOMAC score. Comorbidity was measured using the Charlson Comorbidity Index, a composite scale comprising 22 different comorbid conditions that does not incorporate major depression (27). History of knee injuries was assessed as "ever injured badly enough to limit ability to walk for at least 2 days." Analgesic medications were operationalized as treatment with acetaminophen, nonsteroidal antiinflammatory drugs, or opioids within the previous 30 days. K/L grade is an ordinal scale (range 0–4), where higher values represent worse structural disease (28).

Latent class analysis (LCA). LCA is an approach that identifies subgroups (i.e., classes) based on multiple indicators of a given construct (e.g., depression) and was used to identify classes of individuals with similar patterns of depressive symptoms (29). LCA assumes mutually exclusive and exhaustive classes of individuals that are differentiated within a population by values of observed indicators (29). The LCA model estimates the prevalence of each class in the overall sample and item-response probabilities within each class; namely, the probabilities of endorsing each indicator given membership in a specific class (29). For each individual, posterior class probabilities are estimated to provide a patient's likelihood of membership in each subtype given their indicator response pattern (29).

LCA models were implemented sequentially with 1 to 6 classes and were evaluated regarding fit, parsimony, and clinical interpretability. Akaike's information criterion and Bayesian information criterion fit statistics were used to compare LCA models with different numbers of classes (30,31). Model uncertainty was assessed using relative entropy, with values ranging 0–1, where higher estimates indicate greater classification certainty. Class prevalence estimates and item-response probabilities were used to qualitatively examine and describe the different classes and select the optimal number of depression subtypes. LCA model

fitting was conducted in the overall study sample (n = 4,486) and then stratified by participants with (n = 1,626) or at risk (n = 2,860) for symptomatic knee OA to assess measurement invariance. LCA estimation allows missing values in response variables, and missing values were assumed to be missing at random. After identifying the optimal LCA model, posterior probability estimates were used to assign individuals to the subtype for which they had the highest probability of membership. Chi-square tests and analysis of variance were used to evaluate between-group differences in baseline covariates by depression subtype. Analyses were conducted using R statistical software, version 3.4.1.

Propensity scores. Multiple-group propensity score weights were used to balance between-group differences in potential baseline confounders by depression subtype in order to promote causal interpretations regarding their effect on pain and disability (32). Propensity score weights were estimated using boosted regression, which has been shown to outperform other methods of estimation concerning bias reduction (32). The generalized boosted model is a flexible machine learning estimation routine that fits multiple regression trees to account for potentially complex and nonlinear relationships between exposure and covariates without overfitting data (32). The boosting algorithm optimizes balance on covariates across groups; indicators for missing covariates were automatically included in the propensity model such that depression subtypes were balanced regarding observed covariates as well as missing data patterns (32). Given that longitudinal analyses were stratified by symptomatic knee OA status at baseline, weights were estimated separately in at risk and symptomatic OIA participants. Propensity score weights in the subsamples were stabilized with the marginal probability for each depression subtype. Standardized covariate differences were used to assess balance in the weighted and unweighted samples, and differences of ≥0.2 SD were considered evidence of imbalance.

Weighted estimating equations (WEEs). WEEs were used to assess the relationship between depression subtype and pain and disability of >4 years stratified by participants with and at risk for symptomatic knee OA. WEEs account for missing data using weighting, where weights are the inverse probability of observation conditional on predictors of missing data and were stabilized using the time-specific marginal probabilities for response. Final weights were the product of the time-invariant propensity score weights and time-specific nonresponse weights. WEEs were implemented using survey analysis methods that are appropriate for clustered data and can be used to estimate population-average exposure effects while accounting for between-person heterogeneity in SE estimates. Models included categorical indicators for depression subtype, follow-up time, and their interaction to determine whether there were differences in pain and disability across subtypes over time. Differences in pain and disability by depression subtype at each time point were estimated with 95% confidence intervals (95% CIs), and an α level of 0.05 was used to define statistical significance.

RESULTS

Depression subtypes. A 4-class LCA model was chosen based on clinical interpretability of the subtypes, prevalence estimates, and fit statistics. Indicators of model fit implied that a 5-class model provided a more optimal solution; however, the fifth class was not uniquely distinct when compared to 1 of the

Table 1. Baseline Center for Epidemiologic Studies Depression scale (CES-D) item-response probabilities by depressive symptom subtype among participants who had or were at risk for knee osteoarthritis

CES-D item	Asymptomatic	Catatonic	Anhedonic	Melancholic
I was bothered by things that don't usually bother me.	0.005	0.126	0.019	0.392
I did not feel like eating; my appetite was poor.	0.002	0.085	0.015	0.158
I felt that I could not shake off the blues even with help from my family and friends.	0.004	0.076	0.014	0.529
I felt that I was just as good as other people.*	0.085	0.125	0.500	0.425
I had trouble keeping my mind on what I was doing.	0.018	0.268	0.057	0.502
I felt depressed.	0.002	0.078	0.024	0.742
I felt that everything I did was an effort.	0.009	0.319	0.049	0.641
I felt hopeful about the future.*	0.073	0.257	0.752	0.847
I thought my life had been a failure.	0.002	0.026	0.024	0.298
I felt fearful.	0.004	0.052	0.009	0.290
My sleep was restless.	0.116	0.460	0.206	0.661
I was happy.*	0.033	0.274	0.722	0.882
I talked less than usual.	0.011	0.190	0.056	0.390
I felt lonely.	0.013	0.127	0.050	0.537
People were unfriendly.	0.004	0.033	0.015	0.119
I enjoyed life.*	0.011	0.141	0.617	0.773
I had crying spells.	0.001	0.038	0.011	0.192
I felt sad.	0.002	0.054	0.023	0.585
I felt that people disliked me.	0.005	0.037	0.015	0.141
I could not get going.	0.005	0.264	0.066	0.592

* CES-D item was reverse scored.

other identified subtypes. Moreover, class prevalence structure and item-response probabilities were consistent by symptomatic knee OA status at baseline, and the relative model entropy of 0.86 indicated a high degree of classification certainty. Four depression subtypes were identified (Table 1): asymptomatic (80.6%), catatonic (5.3%), anhedonic (10.6%), and melancholic (3.5%). CES-D score was lowest in asymptomatic participants (Table 2), comparable between the catatonic and anhedonic subtypes, and highest in the melancholic group. Similarly, the proportions of individuals meeting CES-D screening criteria for probable depression were 0.4%, 36.7%, 36.2%, and 100% for the asymptomatic, catatonic, anhedonic, and melancholic subtypes, respectively.

The asymptomatic group was the most prevalent subtype and had low item-response probabilities for every depressive symptom. The second most common subtype was the anhedonic group, which was characterized by high item-response probabilities for symptoms corresponding to the inability to experience pleasure and happiness. The catatonic subtype was the third highest in prevalence and was more likely to endorse symptoms related to psychomotor agitation and somatic symptoms; in particular, decreased energy and movement, difficultly concentrating, and restless sleep. The least common subtype was the melancholic group, which had high item-response probabilities for the widest spectrum of symptoms, including sadness, loneliness, anhedonia, psychomotor agitation, and other somatic symptoms.

Subtype characteristics. When compared to those classified as asymptomatic, other subtypes, particularly catatonic and melancholic, were more likely to be female, nonwhite, not married, and of lower socioeconomic status as measured by educational attainment and employment and health insurance status. Age was comparable across the asymptomatic, catatonic, and

Table 2. Baseline sample characteristics by depressive symptom subtype before propensity score weighting*

Variable	Asymptomatic (n = 3,615)	Catatonic (n = 239)	Anhedonic (n = 474)	Melancholic (n = 158)	*P*
Age, mean ± SD years	61.24 ± 9.09	61.06 ± 9.37	61.87 ± 9.81	57.44 ± 8.27	<0.001
Female sex	2,054 (56.8)	157 (65.7)	283 (59.7)	113 (71.5)	<0.001
White race	630 (17.4)	82 (34.6)	122 (25.7)	53 (33.5)	<0.001
Marital status					<0.001
Married	2,537 (70.8)	124 (52.1)	267 (56.3)	75 (47.8)	
Widowed	250 (7.0)	30 (12.6)	59 (12.4)	13 (8.3)	
Divorced	467 (13.0)	41 (17.2)	80 (16.9)	34 (21.7)	
Separated	45 (1.3)	11 (4.6)	11 (2.3)	8 (5.1)	
Never married	285 (8.0)	32 (13.4)	57 (12.0)	27 (17.2)	
Education					<0.001
No degree	1,265 (35.3)	137 (57.6)	252 (53.2)	97 (61.4)	
College degree	1,111 (31.0)	53 (22.3)	123 (25.9)	37 (23.4)	
Graduate degree	1,208 (33.7)	48 (20.2)	99 (20.9)	24 (15.2)	
Employment	2,272 (62.9)	133 (55.6)	279 (59.0)	88 (55.7)	0.022
Health insurance	3,507 (97.9)	219 (92.0)	442 (94.0)	138 (87.3)	<0.001
Smoking status					<0.001
Never	1,944 (54.6)	104 (43.9)	232 (49.5)	78 (50.0)	
Current	177 (5.0)	29 (12.2)	41 (8.7)	30 (19.2)	
Former	1,438 (40.4)	104 (43.9)	196 (41.8)	48 (30.8)	
Alcohol consumption					0.015
None	641 (17.9)	55 (23.1)	104 (22.0)	42 (26.6)	
Minimal	2,634 (73.5)	163 (68.5)	338 (71.5)	105 (66.5)	
Moderate	307 (8.6)	20 (8.4)	31 (6.6)	11 (7.0)	
Charlson Comorbidity Index score, mean ± SD	0.34 ± 0.78	0.54 ± 1.04	0.54 ± 1.04	0.68 ± 1.03	<0.001
BMI, mean ± SD	28.40 ± 4.68	29.33 ± 5.27	29.17 ± 5.14	29.94 ± 5.79	<0.001
Symptomatic knee OA	1,224 (33.9)	116 (48.5)	202 (42.6)	84 (53.2)	<0.001
K/L grade					<0.001
0	1,455 (40.2)	74 (31.0)	163 (34.4)	52 (32.9)	
1	650 (18.0)	42 (17.6)	81 (17.1)	32 (20.3)	
2	916 (25.3)	61 (25.5)	127 (26.8)	52 (32.9)	
3	484 (13.4)	54 (22.6)	79 (16.7)	19 (12.0)	
4	110 (3.0)	8 (3.3)	24 (5.1)	3 (1.9)	
History of knee injury	938 (26.2)	69 (29.0)	133 (28.4)	48 (30.4)	0.419
Treated with analgesics	1,238 (34.3)	108 (45.6)	196 (41.4)	95 (60.5)	<0.001
WOMAC score, mean ± SD	10.18 ± 13.17	17.40 ± 18.81	15.51 ± 17.64	21.97 ± 20.28	<0.001
CES-D score, mean ± SD	4.03 ± 3.56	13.92 ± 4.59	13.85 ± 4.47	28.97 ± 7.2	<0.001
CES-D score ≥16	15 (0.4)	87 (36.7)	170 (36.2)	158 (100.0)	<0.001

* Values are the number (%) unless indicated otherwise. Some frequency data may not add up to the column total because of missing data. BMI = body mass index; OA = osteoarthritis; K/L = Kellgren-Lawrence (grade); WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index; CES-D = Center for Epidemiologic Studies Depression scale.

anhedonic subtypes but lower in the melancholic group (57 years versus 61 years). The catatonic, anhedonic, and melancholic subtypes had a higher likelihood of being current smokers but were less likely to consume alcohol than the asymptomatic group. In addition, asymptomatic participants had fewer comorbid conditions, lower BMI, lower probability of symptomatic knee OA, and less treatment with analgesic medication compared to every other subtype. Similarly, K/L grade and total WOMAC score were worse in the catatonic, anhedonic, and melancholic subtypes compared to the asymptomatic group. Subtype characteristics and covariate trends of the overall sample were similar to the stratified samples (see Supplementary Tables 1 and 2, respectively, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23898/abstract). In general, propensity score weights reduced the magnitude of between-subtype standardized covariate differences at baseline below 0.2 SD (see Supplementary Tables 3 and 4, respectively, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.23898/abstract).

OA pain and disability. In individuals at risk for symptomatic knee OA, the asymptomatic group had almost no change in pain and disability, while other subtypes generally sustained small increases (Figures 2B and 2D). Time-specific differences in outcomes were smallest in magnitude between the catatonic and asymptomatic groups, ranging over 4 years from $\beta = -0.20$ (95% CI 2.23, 1.84) to $\beta = 1.66$ (95% CI –0.59, 3.91) for pain and $\beta = -1.06$ (95% CI –2.69, 0.57) to $\beta = 0.70$ (95% CI –1.07, 2.48) for disability (Table 3). However, the between-subtype differences in pain and disability for the anhedonic and asymptomatic groups increased in magnitude from baseline ($\beta = 0.45$ [95% CI –0.73, 1.64] and $\beta = -0.17$ [95% CI –1.17, 0.82], respectively) and were as high as $\beta = 2.28$ (95% CI 0.33, 4.22) and $\beta = 2.63$ (95% CI 0.92, 4.35) during follow-up, respectively. Similarly, the melancholic subtype had worse outcomes across all 4 annual follow-up visits, but the time-specific differences in pain and disability were not statistically significant.

Among participants with symptomatic knee OA, every group experienced improvement from baseline except the melancholic subtype, which had persistently greater pain and disability during the follow-up period (Figures 2A and 2C). When compared to the asymptomatic group, time-specific differences in pain and disability among the catatonic and anhedonic subtypes were small in magnitude and generally ≤1 rescaled WOMAC units (Table 4). By contrast, differences in pain between the melancholic and

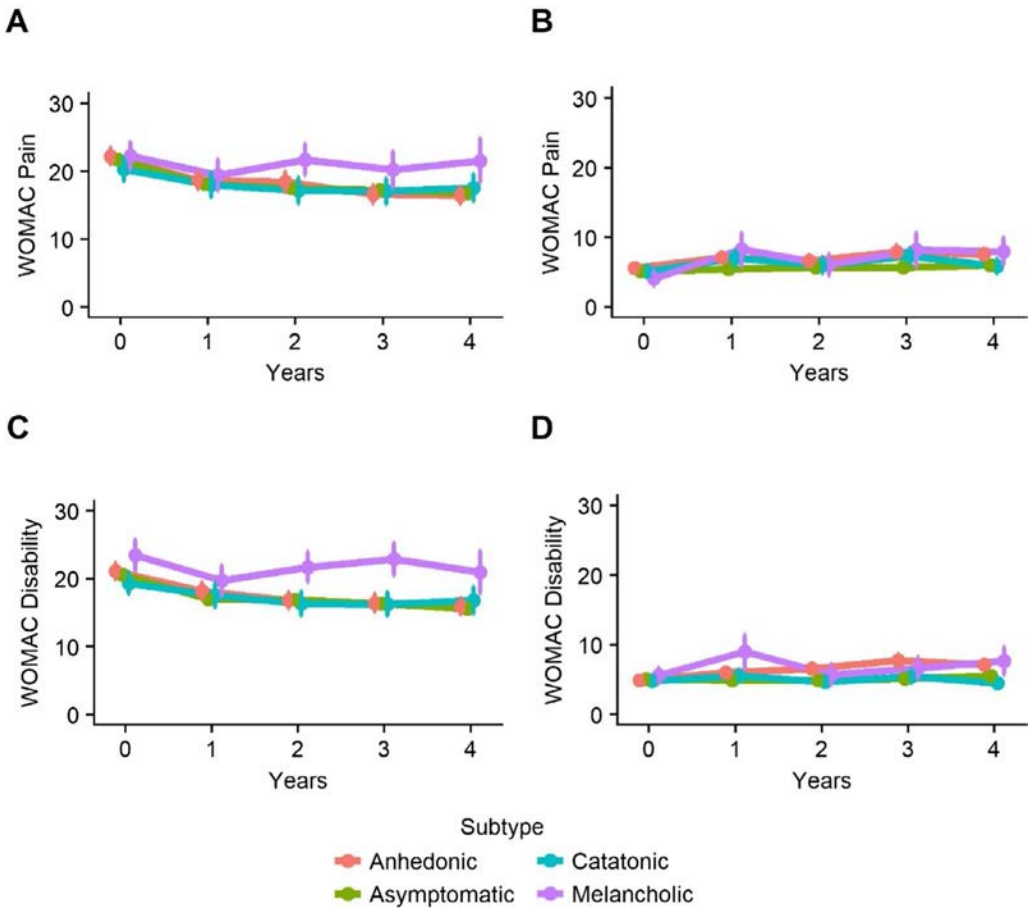

Figure 2. Pain (**A** and **B**) and disability (**C** and **D**) trajectories by baseline depression subtypes among participants with (**A** and **C**) or at risk for (**B** and **D**) symptomatic knee osteoarthritis. WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index.

Table 3. Time-specific differences in pain and disability by baseline depression subtype among participants at risk for symptomatic knee osteoarthritis*

	Catatonic			Anhedonic			Melancholic		
Time point	β	95% CI	*P*	β	95% CI	*P*	β	95% CI	*P*
Pain									
Baseline	0.01	–1.74, 1.76	0.994	0.45	–0.73, 1.64	0.451	–1.02	–3.17, 1.12	0.349
Year 1	1.54	–0.91, 4.00	0.219	1.59	0.07, 3.12	0.041	2.77	–1.99, 7.53	0.254
Year 2	0.32	–2.02, 2.66	0.790	0.97	–0.62, 2.57	0.231	0.40	–2.73, 3.53	0.802
Year 3	1.66	–0.59, 3.91	0.148	2.28	0.33, 4.22	0.022	2.59	–2.23, 7.42	0.292
Year 4	–0.20	–2.23, 1.84	0.850	1.56	–0.12, 3.23	0.069	1.95	–1.87, 5.77	0.316
Disability									
Baseline	–0.26	–1.79, 1.26	0.734	–0.17	–1.17, 0.82	0.737	0.50	–1.78, 2.78	0.667
Year 1	0.70	–1.07, 2.48	0.437	1.12	–0.16, 2.40	0.087	4.17	–0.52, 8.86	0.082
Year 2	–0.20	–1.75, 1.34	0.798	1.70	0.17, 3.23	0.030	0.80	–2.19, 3.80	0.599
Year 3	0.32	–1.57, 2.21	0.742	2.63	0.92, 4.35	0.003	1.55	–1.55, 4.64	0.327
Year 4	–1.06	–2.69, 0.57	0.204	1.57	0.00, 3.14	0.050	2.12	–1.73, 5.96	0.281

* Reference is asymptomatic. 95% CI = 95% confidence interval.

asymptomatic groups increased from β = 0.47 (95% CI –3.68, 4.62) at baseline to β = 4.79 (95% CI –1.77, 11.35) at the fourth annual follow-up visit (Table 4). Similarly, differences in disability increased from β = 2.80 (95% CI –1.84, 7.44) at baseline to time-specific differences as large as β = 6.56 (95% CI 1.72, 11.40) rescaled WOMAC units in the melancholic subtype during follow-up.

DISCUSSION

The current study identified 4 distinct depression subtypes based on patterns of depressive symptoms in individuals with or at risk for symptomatic knee OA. Consistent with a previous meta-analysis (16), our findings indicate that ~80% of OAI participants expressed few symptoms of depression. However, our findings demonstrated moderate heterogeneity in the 20% of participants reporting more depressive symptoms at baseline, and these subtypes were qualified as catatonic, anhedonic, and melancholic. Moreover, detectable effects on pain and disability across 4 years of follow-up were limited to the anhedonic and melancholic subtypes and were largest in individuals with symptomatic knee OA who exhibited both somatic and cognitive symptomology. These results imply that there is variability in both the expression and severity of depressive symptoms among participants with or at risk for symptomatic knee OA that may lead to differences in knee OA outcomes.

Few studies have examined depressive symptom heterogeneity in patients with chronic physical diseases, and the current study highlights 3 symptomatic depression subtypes identified in the OAI cohort (33). The catatonic and anhedonic subtypes were differentiated by somatic and cognitive symptoms, while the melancholic group had a broader constellation of symptomology. Prior work in cancer patients identified a mild depression subtype presenting with concentration and sleep problems and psychomotor agitation that is similar to the catatonic group reported in the current study (34). Prior research also provides evidence for a nondysphoric (i.e., without sadness) depression subtype in older adults typified by slowness of movement and, unlike in the catatonic group, other cognitive symptoms (35–37). Anhedonic subtypes epitomized solely by the absence of happiness have not been widely reported (38). Research conducted in the general

Table 4. Time-specific differences in pain and disability by baseline depression subtype among participants with symptomatic knee osteoarthritis*

	Catatonic			Anhedonic			Melancholic		
Time point	β	95% CI	*P*	β	95% CI	*P*	β	95% CI	*P*
Pain									
Baseline	–1.51	–5.06, 2.04	0.404	0.39	–2.31, 3.09	0.779	0.47	–3.68, 4.62	0.824
Year 1	–0.14	–3.96, 3.68	0.942	0.57	–2.32, 3.46	0.701	1.33	–3.40, 6.05	0.582
Year 2	–0.33	–4.34, 3.69	0.873	0.95	–2.18, 4.08	0.551	4.25	–0.45, 8.95	0.077
Year 3	–0.17	–4.07, 3.72	0.932	–0.64	–3.58, 2.30	0.668	2.98	–2.49, 8.44	0.286
Year 4	0.82	–3.34, 4.98	0.699	–0.31	–3.18, 2.55	0.830	4.79	–1.77, 11.35	0.152
Disability									
Baseline	1.27	–2.07, 4.60	0.457	0.48	–2.17, 3.12	0.723	2.80	–1.84, 7.44	0.237
Year 1	0.58	–3.12, 4.27	0.760	1.24	–1.60, 4.08	0.393	2.76	–1.74, 7.27	0.230
Year 2	–0.54	–4.26, 3.19	0.778	–0.08	–2.80, 2.64	0.954	4.82	0.31, 9.33	0.037
Year 3	–0.01	–3.57, 3.54	0.994	0.14	–2.68, 2.96	0.921	6.56	1.72, 11.40	0.008
Year 4	1.21	–2.92, 5.34	0.567	0.31	–2.32, 2.94	0.817	5.35	–0.90, 11.60	0.094

* Reference is asymptomatic. 95% CI = 95% confidence interval.

population derived an anhedonic subtype with similar characteristics to current results: older age and more proportionate sex distribution (39). However, that study suggested that the anhedonic class had fewer future depressive episodes and stressful events than other subtypes, which is surprising given that anhedonia is associated with chronic stress and is a risk factor for major depression (39,40). Subtypes encompassing cognitive and somatic symptoms are consistently reported, and melancholic depression characterized by sadness, anhedonia, decreased energy and movement, difficulty concentrating, restless sleep, and other physical and emotional problems has been identified in the general population and the elderly (33,41,42). Melancholic depression is a more severe phenotype, evidenced by higher CES-D scores satisfying screening criteria for every participant in this group, and is associated with an almost 2-fold response time to pharmacologic treatment compared to other subtypes (43). Nonetheless, one-third of individuals in the catatonic and anhedonic groups screened positive for probable depression, and these subtypes not only represent differences in severity but illustrate different patterns of symptomology and may explain (in part) why ~60% of depressed OA patients do not receive care for mental health (44).

Differences in pain and disability by depression subtype further highlight the difficulty in managing psychosomatic factors in knee OA patients. The catatonic subtype was the only group that did not experience significantly greater pain and disability than asymptomatic participants. This finding is contrary to research showing that older adults who report somatic symptoms of depression but deny feelings of sadness are at an increased risk for functional impairment (35). Perhaps the catatonic subtype represents a mild phenotype, but it may also indicate somatic symptom overlap, where CES-D items detect symptomology of knee OA or another unrelated medical condition (e.g., fatigue) that is not predictive of pain and disability (34,45,46). By contrast, participants at risk for symptomatic knee OA in the anhedonic subtype had statistically greater pain and disability (1.5–2.3 WOMAC units) compared to asymptomatic individuals that was small in magnitude and not clinically significant. However, prior research indicates that OAI participants experience minimal changes in their knee OA symptoms; therefore, psychosomatic factors that contribute to persistently higher disease severity may be relevant at the population level (47). The cognitive symptoms of the anhedonic subtype may represent perseverative thought, a process of repetitive pessimistic thinking associated with negative affect and other traits closely related to depression, which could act as a response shift and lead to reports of worse pain and disability (48). Alternatively, the anhedonic subtype may constitute a mild mood disorder which, following the onset of a chronic physical disease, decompensates into melancholic depression in some individuals with symptomatic knee OA. This premise is supported by persistently higher disease severity (4.8–6.6 WOMAC units) among symptomatic knee OA participants in the melancholic subtype, perhaps indicative of substantive increases in pain severity and decreased physical performance associated with a greater number and spectrum of depressive symptoms (49). Unfortunately, it is not possible to determine causality regarding knee OA and depression in the current study, and the melancholic subtype may represent more severe depressive symptoms that are a consequence of pain, disability, or other related factors. Notwithstanding, anhedonia, and to a larger extent, melancholic depression, may be a modifiable risk factor for worsening knee OA disease severity and a potential target for intervention.

There are limitations that should be considered when interpreting this study's results. First, the 20-item CES-D lacks assessments of symptoms (e.g., hallucinations, risk taking behaviors, etc.) commonly found in psychotic and atypical depression subtypes (23). Nevertheless, the CES-D is a valid and reliable measure of depressive symptoms in individuals with knee OA that has been used in prior OAI studies of depression (6,7,49). Second, a 3-step design was used, where the measurement model was estimated, participants were assigned to a subtype, and a structural model of outcomes was fit; an approach that may lead to misclassification of class assignment. However, relative model entropy suggested a minimal level of class uncertainty, and any resulting bias would attenuate associations to the null. Last, the potential for confounding by unmeasured factors, such as depression treatments, cannot be eliminated here, as is the case for all observational studies. These limitations are mitigated by the current study's strengths. First, this is one of the largest LCA studies on depression heterogeneity in individuals with or at risk for chronic musculoskeletal disorders. Second, modern statistical techniques that leveraged machine learning and inverse probability weighting methods were used to overcome model misspecification when adjusting for potential confounders and missing data. Finally, the OAI is a well-documented, prospective cohort that measures comprehensive sociodemographic and clinical characteristics relevant to assessments of depression subtypes and their influence on OA pain and disability.

In conclusion, our study findings indicate that depression subtypes among individuals with or at risk for symptomatic knee OA are differentiated primarily by psychomotor agitation, anhedonia, and other somatic symptoms. Moreover, anhedonic and melancholic depression subtypes may be risk factors for increased pain and disability in individuals who develop and experience symptomatic knee OA. These results demonstrate the advantages of using a data-driven approach to identify distinct depression subtypes that represent unique clinical phenotypes presenting with different spectrums of symptoms. Consistent with clinical recommendations, our findings support the need for depression screening in knee OA patients during primary care and rheumatology encounters with simple, reliable, and valid instruments (e.g., the 2-item Patient Health Questionnaire) (50). If depression screening becomes routine in patients with musculoskeletal disorders, more comprehensive symptomatic assessments could be conducted after referral to mental health professionals, where patients

then would receive treatments tailored to their specific depressive symptomology. However, the feasibility and cost-effectiveness of more complex assessment methods in the secondary care setting that are required for depression subtype identification are unknown. Ultimately, our findings highlight the difficulty of using standard depression treatments in individuals with musculoskeletal disorders, and protocols that address the specific symptomatology of different subtypes of depression that may present in patients with knee OA are needed.

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. Rathbun 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. Rathbun, Hochberg.

Acquisition of data. Rathbun.

Analysis and interpretation of data. Rathbun, Schuler, Stuart, Shardell, Yau, Gallo, Ryan, Hochberg.

ADDITIONAL DISCLOSURES

Author Schuler is an employee of the RAND Corporation.

REFERENCES

1. Lawrence RC, Felson DT, Helmick CG, Arnold LM, Choi H, Deyo RA, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. Arthritis Rheum 2008;58:26–35.
2. Zhang Y, Jordan JM. Epidemiology of osteoarthritis. Clin Geriatr Med 2010;26:355–69.
3. Lane NE, Brandt K, Hawker G, Peeva E, Schreyer E, Tsuji W, et al. OARSI-FDA initiative: defining the disease state of osteoarthritis. Osteoarthritis Cartilage 2011;19:478–82.
4. Rathbun AM, Harrold LR, Reed GW. Temporal associations between the different domains of rheumatoid arthritis disease activity and the onset of patient-reported depressive symptoms. Clin Rheumatol 2015;34:653–63.
5. Rathbun AM, Harrold LR, Reed GW. Temporal effect of depressive symptoms on the longitudinal evolution of rheumatoid arthritis disease activity. Arthritis Care Res (Hoboken) 2015;67:765–75.
6. Rathbun AM, Stuart EA, Shardell M, Yau MS, Baumgarten M, Hochberg MC. Dynamic effects of depressive symptoms on osteoarthritis knee pain. Arthritis Care Res (Hoboken) 2018;70:80–8.
7. Rathbun AM, Yau MS, Shardell M, Stuart EA, Hochberg MC. Depressive symptoms and structural disease progression in knee osteoarthritis: data from the Osteoarthritis Initiative. Clin Rheumatol 2017;36:155–63.
8. Sugai K, Takeda-Imai F, Michikawa T, Nakamura T, Takebayashi T, Nishiwaki Y. Association between knee pain, impaired function, and development of depressive symptoms. J Am Geriatr Soc 2018;66:570–6.
9. Rosemann T, Gensichen J, Sauer N, Laux G, Szecsenyi J. The impact of concomitant depression on quality of life and health service utilisation in patients with osteoarthritis. Rheumatol Int 2007;27:859–63.
10. Detweiler-Bedell JB, Friedman MA, Leventhal H, Miller IW, Leventhal EA. Integrating co-morbid depression and chronic physical disease management: identifying and resolving failures in self-regulation. Clin Psychol Rev 2008;28:1426–46.
11. Carstensen J, Andersson D, Andre M, Engstrom S, Magnusson H, Borgquist LA. How does comorbidity influence healthcare costs? A population-based cross-sectional study of depression, back pain and osteoarthritis. BMJ Open 2012;2:e000809.
12. Goldberg D. The heterogeneity of "major depression." World Psychiatry 2011;10:226–8.
13. Rathbun AM, Harrold LR, Reed GW. A description of patient- and rheumatologist-reported depression symptoms in an American rheumatoid arthritis registry population. Clin Exp Rheumatol 2014;32:523–32.
14. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA 2003;289:3095–105.
15. Marcus SC, Olfson M. National trends in the treatment for depression from 1998 to 2007. Arch Gen Psychiatry 2010;67:1265–73.
16. Stubbs B, Aluko Y, Myint PK, Smith TO. Prevalence of depressive symptoms and anxiety in osteoarthritis: a systematic review and meta-analysis. Age Ageing 2016;45:228–35.
17. Wang PS, Insel TR. NIMH-funded pragmatic trials: moving on. Neuropsychopharmacology 2010;35:2489.
18. Fried E. Moving forward: how depression heterogeneity hinders progress in treatment and research. Expert Rev Neurother 2017;17:423–5.
19. Fried EI, van Borkulo CD, Cramer AO, Boschloo L, Schoevers RA, Borsboom D. Mental disorders as networks of problems: a review of recent insights. Soc Psychiatry Psychiatr Epidemiol 2017;52:1–10.
20. Nevitt MC, Felson DT, Lester G. The Osteoarthritis Initiative: a knee health study. 2006. URL: https://oai.epi-ucsf.org/datarelease/docs/StudyDesignProtocol.pdf.
21. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. Osteoarthritis Cartilage 2007;15:A1–56.
22. Furukawa TA, Kitamura T, Takahashi K. Time to recovery of an inception cohort with hitherto untreated unipolar major depressive episodes. Br J Psychiatry 2000;177:331–5.
23. Radloff LS. The CES-D scale a self-report depression scale for research in the general population. Appl Psychol Meas 1977;1:385–401.
24. Ulbricht CM, Rothschild AJ, Lapane KL. The association between latent depression subtypes and remission after treatment with citalopram: a latent class analysis with distal outcome. J Affect Disord 2015;188:270–7.
25. Bellamy N. Validation study of WOMAC: a health status instrument for measuring clinically-important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. J Orthop Rheumatol 1988;1:95–108.
26. Ehrich EW, Davies GM, Watson DJ, Bolognese JA, Seidenberg BC, Bellamy N. Minimal perceptible clinical improvement with the Western Ontario and McMaster Universities osteoarthritis index questionnaire and global assessments in patients with osteoarthritis. J Rheumatol 2000;27:2635–41.
27. Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. J Clin Epidemiol 1994;47:1245–51.
28. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957;16:494–502.
29. Lanza ST, Rhoades BL. Latent class analysis: an alternative perspective on subgroup analysis in prevention and treatment. Prev Sci 2013;14:157–68.
30. Schwarz G. Estimating the dimension of a model. Ann Stat 1978;6:461–4.
31. Akaike H. Factor analysis and AIC. Psychometrika 1987;52:317–32.

32. McCaffrey DF, Griffin BA, Almirall D, Slaughter ME, Ramchand R, Burgette LF. A tutorial on propensity score estimation for multiple treatments using generalized boosted models. Stat Med 2013;32:3388–414.
33. Van Loo HM, De Jonge P, Romeijn JW, Kessler RC, Schoevers RA. Data-driven subtypes of major depressive disorder: a systematic review. BMC Med 2012;10:156.
34. Zhu L, Ranchor AV, van der Lee M, Garssen B, Sanderman R, Schroevers MJ. Subtypes of depression in cancer patients: an empirically driven approach. Support Care Cancer 2016;24:1387–96.
35. Gallo JJ, Rabins PV, Lyketsos CG, Tien AY, Anthony JC. Depression without sadness: functional outcomes of nondysphoric depression in later life. J Am Geriatr Soc 1997;45:570–8.
36. Gallo JJ, Rabins P, Anthony J. Sadness in older persons: 13-year follow-up of a community sample in Baltimore, Maryland. Psychol Med 1999;29:341–50.
37. Gallo JJ, Rabins PV. Depression without sadness: alternative presentations of depression in late life. Am Fam Physician 1999;60: 820–6.
38. Penninx BW, Milaneschi Y, Lamers F, Vogelzangs N. Understanding the somatic consequences of depression: biological mechanisms and the role of depression symptom profile. BMC Med 2013;11:129.
39. Chen LS, Eaton WW, Gallo JJ, Nestadt G. Understanding the heterogeneity of depression through the triad of symptoms, course and risk factors: a longitudinal, population-based study. J Affect Disord 2000;59:1–11.
40. Loas G. Vulnerability to depression: a model centered on anhedonia. J Affect Disord 1996;41:39–53.
41. Lamers F, Beekman A, Van Hemert A, Schoevers R, Penninx B. Six-year longitudinal course and outcomes of subtypes of depression. Br J Psychiatry 2016;208:62–8.
42. Veltman E, Lamers F, Comijs H, De Waal M, Stek M, Van der Mast R, et al. Depressive subtypes in an elderly cohort identified using latent class analysis. J Affect Disord 2017;218:123–30.
43. Bühler J, Seemüller F, Läge D. The predictive power of subgroups: an empirical approach to identify depressive symptom patterns that predict response to treatment. J Affect Disord 2014;163:81–7.
44. Gleicher Y, Croxford R, Hochman J, Hawker G. A prospective study of mental health care for comorbid depressed mood in older adults with painful osteoarthritis. BMC Psychiatry 2011;11:147.
45. Addington A, Gallo J, Ford D, Eaton W. Epidemiology of unexplained fatigue and major depression in the community: the Baltimore ECA follow-up, 1981–1994. Psychol Med 2001;31:1037–44.
46. Pincus T, Hassett AL, Callahan LF. Criterion contamination of depression scales in patients with rheumatoid arthritis: the need for interpretation of patient questionnaires (as all clinical measures) in the context of all information about the patient. Rheum Dis Clin North Am 2009;35:861–4.
47. Collins JE, Katz JN, Dervan EE, Losina E. Trajectories and risk profiles of pain in persons with radiographic, symptomatic knee osteoarthritis: data from the osteoarthritis initiative. Osteoarthritis Cartilage 2014;22:622–30.
48. Rathbun AM, Reed GW, Harrold LR. The temporal relationship between depression and rheumatoid arthritis disease activity, treatment persistence and response: a systematic review. Rheumatology (Oxford) 2013;52:1785–94.
49. Rathbun AM, Shardell MD, Stuart EA, Yau MS, Gallo JJ, Schuler MS, et al. Pain severity as a mediator of the association between depressive symptoms and physical performance in knee osteoarthritis. Osteoarthritis Cartilage 2018;26:1453–60.
50. Cg N. Osteoarthritis care and management in adults. London: National Institute for Health and Care Excellence; 2014.

Arthritis Care & Research
Vol. 72, No. 5, May 2020, pp 679–684
DOI 10.1002/acr.23883

American College of Rheumatology
Empowering Rheumatology Professionals

Comparative Responsiveness of Outcome Measures for the Assessment of Pain and Function in Osteoarthritis of the First Metatarsophalangeal Joint

Hylton B. Menz,[1] Maria Auhl,[2] Jade M. Tan,[2] Pazit Levinger,[3] Edward Roddy,[4] and Shannon E. Munteanu[2]

Objective. The present study was undertaken to assess the comparative responsiveness of outcome measures used for the assessment of pain and function in individuals with osteoarthritis (OA) of the first metatarsophalangeal (MTP) joint.

Methods. Eighty-eight patients (mean ± SD age 57.2 ± 10.2 years) with OA of the first MTP joint who participated in a randomized trial completed the Foot Health Status Questionnaire (FHSQ), the Foot Function Index Revised Short Form (FFI-RS), and 100-mm visual analog scales (VAS) of pain and stiffness at baseline and 12 weeks. Responsiveness of the subscales for each outcome measure was determined using paired *t*-tests, Cohen's *d* coefficient, the standardized response mean (SRM), and the Guyatt index (GI). Sample size estimations were calculated based on minimal important differences (MIDs).

Results. All outcome measures were sensitive to change and demonstrated at least medium effect sizes. Three outcome measures exhibited large or very large effect sizes for Cohen's *d* coefficient, the SRM, and the GI: the FHSQ pain subscale (*d* = 1.03; SRM 1.10, GI score 1.30), the FFI-RS pain subscale (*d* = 1.09; SRM 1.05, GI score 1.73), and the 100-mm VAS of pain severity while walking (*d* = 1.22; SRM 1.07, GI score 1.78). Sample size calculations indicated that between 20 and 33 participants per group would be required to detect MIDs using these measures.

Conclusion. The FHSQ pain subscale, FFI-RS pain subscale, and the 100-mm VAS of pain severity while walking are the most responsive outcome measures for the assessment of pain and function in individuals with OA of the first MTP joint. These findings provide useful information to guide researchers in selecting appropriate outcome measures for use in future clinical trials.

INTRODUCTION

Osteoarthritis (OA) of the first metatarsophalangeal (MTP) joint is the most common form of foot OA, affecting 8% of individuals over 50 years of age (1). OA of the first MTP joint is characterized by joint pain and stiffness and impairs both foot-specific and general health-related quality of life (2). Nearly three-fourths of those with the condition report it to be disabling (1). Increasing radiographic severity of OA of the first MTP joint is associated with an increased prevalence of pain, deformity, and decreased joint range of motion, which suggests that it may be a progressive disorder that has an accumulative impact on the load-bearing function of the foot (3).

Despite the high prevalence and burden of OA of the first MTP joint, relatively little research has been undertaken to evaluate commonly used treatments for this condition (4), and there is no consensus as to which outcome measures should be used to assess treatment effectiveness. Of the 4 randomized trials that have been undertaken to assess interventions for OA of the first MTP joint, 2 used generic visual analog scales (VAS) of pain (5,6), and 2 used the pain domain of the Foot Health Status Questionnaire (FHSQ) (7,8) as the primary outcome measure. However, recent reviews of measures of foot function, foot health, and foot pain indicate that while several questionnaires have been developed, very few of these outcome

ACTRN: 12613001245785.

Supported by the National Health and Medical Research Council of Australia (grant 1049085). Dr. Menz is recipient of a National Health and Medical Research Council of Australia Senior Research Fellowship (award 1135995).

[1]Hylton B. Menz, PhD, DSc: La Trobe University, Melbourne, Victoria, Australia, and Keele University, Staffordshire, UK; [2]Maria Auhl, BSc, Jade M. Tan, BPod (Hons), Shannon E. Munteanu, PhD: La Trobe University, Melbourne, Victoria, Australia; [3]Pazit Levinger, PhD: La Trobe University and National Ageing Research Institute, Melbourne, Victoria, Australia; [4]Edward Roddy, MD: Keele University, Staffordshire, UK, and Academic Rheumatology Centre, Midlands Partnership NHS Trust, Stoke-on-Trent, UK.

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

Address correspondence to Hylton B. Menz, PhD, DSc, School of Allied Health, Human Services and Sport, La Trobe University, Melbourne 3086, Victoria, Australia. E-mail: h.menz@latrobe.edu.au.

Submitted for publication October 30, 2018; accepted in revised form March 19, 2019.

SIGNIFICANCE & INNOVATIONS

- This is the first study to evaluate the comparative responsiveness of outcome measures for osteoarthritis of the first metatarsophalangeal joint.
- The most responsive measures were the Foot Health Status Questionnaire pain subscale, the Foot Function Index Revised Short Form pain subscale, and the 100-mm visual analog scale of pain severity while walking.
- These findings will assist in the design of future clinical trials investigating this common and disabling condition.

measures have undergone adequate psychometric evaluation. In particular, the ability of these instruments to detect changes in foot health status, commonly referred to as responsiveness, has not been examined in detail, thereby limiting their use in clinical trials (9–14).

Of the available foot-specific questionnaires, the FHSQ and Foot Function Index (FFI) have undergone the most extensive psychometric development and would appear to be the most appropriate measures to use in clinical trials (12). However, VAS scores of pain have also been widely used in both the foot-specific (13) and general rheumatology (15) literature and are also worthy of consideration. Therefore, as part of a randomized trial comparing the effectiveness of prefabricated foot orthoses and rocker-sole footwear for the treatment of OA of the first MTP joint (8,16), we compared the responsiveness of 3 outcome measures: the FHSQ (17), the FFI Revised Short Form (FFI-RS) (18), and 100-mm VAS for pain and stiffness specifically focused on the first MTP joint. We also estimated the sample size requirements using these outcome measures for future clinical trials.

MATERIALS AND METHODS

Study design. This study was undertaken as part of a larger randomized trial, the details of which have been published previously (8,16). Briefly, the trial was a parallel-group randomized trial with a 12-week follow-up period, with participants randomly allocated to receive either prefabricated foot orthoses or rocker-sole footwear. Key inclusion criteria for the trial were pain in the first MTP joint rated at least 20 mm on a 100-mm VAS on most days for at least 12 weeks, and <64° of dorsiflexion range of motion of the first MTP joint. Key exclusion criteria included previous surgery on the first MTP joint, significant deformity of the first MTP joint including hallux valgus, cognitive impairment, or a history of recurrent falls (8,16). Radiographic OA of the first MTP joint was documented using a standardized atlas (19). The trial was registered (Australian New Zealand Clinical Trials Registry ID: ACTRN12613001245785), the La Trobe University Human Ethics Committee provided ethics approval (number 13-003), and all participants provided written informed consent.

Outcome measures. To assess responsiveness, 3 sets of outcome measures were evaluated: the FHSQ (17), the FFI-RS (18), and 4 100-mm VAS of pain and stiffness specifically focused on the first MTP joint.

The FHSQ consists of 13 questions reflecting 4 foot health–related subscales with a recall period of the past week: pain (4 questions), function (4 questions), footwear (3 questions), and general foot health (2 questions) (17). The pain and function domains are the most commonly used in clinical trials, and the pain domain was the primary outcome measure in this study. Each question is scored on a 5-point scale, and individual scores are then recoded, tabulated, and finally transformed to a scale ranging from 0 (indicating poorest foot health) to 100 (indicating best foot health). The FHSQ demonstrates a high degree of content, criterion, and construct validity and high retest reliability (17) and has been shown to be responsive to change in clinical trials of foot orthoses for plantar heel pain (20) and extra-depth footwear for older individuals with foot pain (21). A recent review recommended the use of the FHSQ in clinical trials of rheumatologic foot disorders (12). The FHSQ pain and function subscales were measured at baseline and at 4, 8, and 12 weeks. For the purpose of the analysis in this study, only the baseline and 12-week scores were used.

The original FFI consisted of 23 questions divided into 3 subscales: pain (9 questions), disability (9 questions), and activity limitation (5 questions), with a recall period of the past week (22). The FFI was subsequently revised to incorporate additional questions relating to the psychosocial burden of foot pain (23), and 2 versions were proposed: a long-form consisting of 68 questions, and a short form version (FFI-RS) containing 34 questions divided into 5 subscales: pain (7 questions), stiffness (7 questions), difficulty (11 questions), activity limitation (3 questions), and social aspects (6 questions) (18). Each question is scored on a 4-point scale, with higher scores representing worse foot health. To obtain a subscale score, the item scores are added and divided by the maximum total possible and then multiplied by 100. The FFI-RS demonstrates similar psychometric properties to the long-form version, which has demonstrated content validity, construct validity, and reliability (23).

In this study, the FFI-RS pain, stiffness, and difficulty subscales were measured at baseline and 12 weeks. It was necessary to modify the wording of the pain and stiffness subscales because the original questionnaire required participants to report pain and stiffness both when wearing shoes and when wearing custom shoe inserts. Because this would have created confusion in our trial (as participants were allocated to either foot orthoses or rocker-sole footwear), we merged these 2 questions and used the generic phrase “the intervention provided” rather than “shoes” and “custom inserts.” This resulted in the pain subscale having 5 questions and the stiffness subscale having 6 questions rather than the original 7 questions.

In addition to the FHSQ and FFI-RS outcome measures, we also asked participants to report, in relation to their big toe joint:

Table 1. Participant characteristics (n = 88)*

Age, mean ± SD years	57.2 ± 10.2
Sex, female	51 (58.0)
Height, mean ± SD cm	165.7 ± 8.7
Weight, mean ± SD kg	79.1 ± 14.1
Body mass index, mean ± SD kg/m^2	28.7 ± 4.6
General health, mean ± SD	
Short Form 12 physical score	45.0 ± 10.4
Short Form 12 mental score	53.9 ± 8.6
Clinical features†	
Pain duration, median (range) months	33 (4–420)
First MTP joint range of motion, mean ± SD degrees	40.1 ± 12.8
Pain on palpation	88 (100)
Palpable dorsal exostosis	84 (96)
Joint effusion	29 (33)
Pain on motion of first MTP joint	81 (92)
Hard-end feel when dorsiflexed	78 (89)
Crepitus	57 (66)
Radiographic features‡	
Dorsal osteophytes	81 (92)
Dorsal joint space narrowing	76 (86)
Lateral osteophytes	73 (83)
Lateral joint space narrowing	76 (86)
Radiographic OA of the first MTP joint§	61 (70)

* Values are the number (%) unless indicated otherwise. MTP = metatarsophalangeal; OA = osteoarthritis.
† For full descriptions of tests, see Zammit et al (41).
‡ Score >0 using Menz et al atlas (19).
§ At least 1 score of 2 for osteophytes or joint space narrowing from either view, using case definition from Menz et al atlas (19).

(a) the amount of pain they experienced in the past week while walking over a flat surface, (b) the amount of pain they experienced in the past week at rest, (c) the severity of stiffness they experienced in the past week after first awakening in the morning, and (d) the severity of stiffness they experienced in the past week while sitting, lying, or resting later in the day. Each of these 4 questions was accompanied by a 100-mm VAS, with the pain questions anchored by the statements "no pain" and "worst pain possible" and the stiffness questions anchored by the statements "not stiff at all" to "most stiff possible." To avoid clustering of scores, no verbal descriptors at intermediate points were used (15). We chose VAS rather than numerical rating scales because the VAS have been used in 2 of the 4 clinical trials for this condition (5,6).

Statistical analysis. Statistical analyses were undertaken using Statistics, version 25 (IBM) and Excel (Microsoft Corporation). To evaluate the responsiveness of the outcome measures, 4 different effect size statistics were used (24–28): (a) a paired *t*-test to test the null hypothesis that there was no change in the mean scores from baseline to the 12-week follow-up; (b) Cohen's *d* coefficient, calculated as the mean change scores between baseline and 12-week follow-up divided by the SD of the baseline scores (29), (c) the standardized response mean (SRM), calculated as the mean change scores between baseline and 12-week follow-up divided by the SD of the differences between the baseline and 12-week follow-up scores (30), and (d) the Guyatt index (GI), which represents the magnitude and variability in change scores for an outcome measure relative to the minimal important difference (MID) of the measure (24). The MID for each measure was calculated as the mean change score in participants who improved, minus the mean change score in participants who did not improve or whose symptoms worsened (24). The formulae used to calculate these responsiveness measures are provided in Supplementary Appendix A, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23883/abstract. To aid interpretation, the following effect size benchmarks were used: negligible effect size (<0.15), small effect size (≥0.15 and <0.40), medium effect size (≥0.40 and <0.75), large effect size (≥0.76 and <1.10), and very large effect size (≥1.10) (31). Sample size estimations for each outcome measure were calculated using the Statistics Sample Power 3.0 Plug-In (IBM), using the calculated MIDs and SDs of the 12-week follow-up scores at 80% and 90% power, 5% alpha, and assuming no dropouts.

RESULTS

Participant characteristics. For the clinical trial from which these data were derived, 102 participants (45 men and 57 women) were randomly allocated to receive prefabricated foot orthoses (n = 52) or rocker-sole footwear (n = 50) (8). By the 12-week follow-up, there were 5 withdrawals in the orthoses group and 5 withdrawals in the footwear group. An additional 4 participants were missing 12-week follow-up data, leaving a total of 88 participants with complete data for this analysis (37 men and 51 women). Characteristics of these participants are shown in Table 1.

Responsiveness. Means and SDs of baseline and 12-week follow-up scores for each outcome measure are shown in Table 2, and the 4 responsiveness statistics are shown in Table 3. All outcome measures were sensitive to change (paired *t*-tests significant at $P < 0.001$) and exhibited at least medium effect sizes for Cohen's

Table 2. Mean ± SD scores for each outcome measure at baseline and 12-week follow-up*

	Baseline	12-week follow-up
FHSQ pain†	53.3 ± 19.5	73.4 ± 15.8
FHSQ function†	68.8 ± 24.0	81.9 ± 17.8
FFI-RS pain‡	42.2 ± 17.4	23.3 ± 16.2
FFI-RS stiffness‡	34.8 ± 20.8	21.6 ± 17.3
FFI-RS difficulty‡	39.2 ± 24.9	26.5 ± 19.7
VAS pain severity at rest‡	34.0 ± 25.4	16.6 ± 19.5
VAS pain severity while walking‡	47.8 ± 21.8	21.2 ± 19.4
VAS stiffness severity in the morning‡	35.6 ± 26.2	19.6 ± 21.9
VAS stiffness severity later in the day‡	35.7 ± 26.7	16.3 ± 18.6

* FHSQ = Foot Health Status Questionnaire; FFI-RS = Foot Function Index Revised Short Form; VAS = visual analog scale.
† Higher scores represent improved health status.
‡ Lower scores represent improved health status.

Table 3. Responsiveness of the outcome measure subscales*

	Paired *t*-test		Cohen's *d* coefficient		SRM		GI		
	t	*P*	Value	Interpretation	Value	Interpretation	Value	Interpretation	MID
FHSQ pain	–10.25	<0.001	1.03	Large	1.10	Very large	1.30	Very large	11.1
FHSQ function	–7.90	<0.001	0.55	Medium	0.85	Large	1.23	Very large	9.7
FFI-RS pain	9.84	<0.001	1.09	Large	1.05	Large	1.73	Very large	14.7
FFI-RS stiffness	6.37	<0.001	0.63	Medium	0.68	Medium	1.17	Very large	10.3
FFI-RS difficulty	6.57	<0.001	0.51	Medium	0.70	Medium	1.42	Very large	12.1
VAS pain severity at rest	6.65	<0.001	0.69	Medium	0.72	Medium	0.68	Medium	6.7
VAS pain severity while walking	10.00	<0.001	1.22	Very large	1.07	Large	1.78	Very large	17.7
VAS stiffness severity in the morning	6.09	<0.001	0.61	Medium	0.65	Medium	1.31	Very large	13.0
VAS stiffness severity later in the day	6.40	<0.001	0.73	Medium	0.69	Medium	1.12	Very large	11.9

* SRM = standardized response mean; GI = Guyatt index; MID = minimal important difference; FHSQ = Foot Health Status Questionnaire; FFI-RS = Foot Function Index Revised Short Form; VAS = visual analog scale.

d coefficient, SRM, and GI score. Three outcome measures exhibited large or very large effect sizes: the FHSQ pain subscale (*d* = 1.03; SRM 1.10, GI score 1.30), the FFI-RS pain subscale (*d* = 1.09; SRM 1.05, GI score 1.73), and the 100-mm VAS of pain severity while walking (*d* = 1.22; SRM 1.07, GI score 1.78).

Sample size estimations. Based on a secondary analysis of the changes observed in our trial, sample size estimations for the 9 outcome measures ranged from 20 to 135 participants per group at 80% power, and 27 to 180 participants per group at 90% power (Table 4). The smallest sample size requirement was for VAS pain severity while walking (n = 20 at 80% power and n = 27 at 90% power), and the largest was for VAS pain severity at rest (n = 135 at 80% power and n = 180 at 90% power).

DISCUSSION

This is the first study to evaluate the responsiveness of outcome measures of foot pain and disability in individuals with OA of the first MTP joint. Using data obtained from a clinical trial of foot orthoses and rocker-sole footwear (8), we evaluated 3 outcome measures: the FHSQ (17), the FFI-RS (23), and VAS of pain and stiffness. We applied the following 4 responsiveness statistics: paired *t*-tests, Cohen's *d* coefficient, the SRM, and the Guyatt index. We found that all outcome measures were sensitive to change and demonstrated at least medium effect sizes across all responsiveness measures. Of these measures, the FHSQ pain subscale, the FFI-RS pain subscale, and the 100-mm VAS of pain severity while walking demonstrated the highest responsiveness and would therefore appear to be appropriate outcome measures to use in future clinical trials of interventions for OA of the first MTP joint.

Table 4. Sample size estimates for each outcome measure at 80% and 90% power, assuming 2-group comparison and no dropouts*

	80%	90%
FHSQ pain	33	44
FHSQ function	54	72
FFI-RS pain	21	27
FFI-RS stiffness	46	61
FFI-RS difficulty	43	57
VAS pain severity at rest	135	180
VAS pain severity while walking	20	27
VAS stiffness severity in the morning	46	61
VAS stiffness severity later in the day	40	53

* Values are the number of participants. FHSQ = Foot Health Status Questionnaire; FFI-RS = Foot Function Index Revised Short Form; VAS = visual analog scale.

The responsiveness of the FHSQ pain and function subscales reported here is consistent with previous studies demonstrating responsiveness of this outcome measure in clinical trials of foot orthoses for plantar heel pain (20) and extra-depth footwear in older individuals with generalized foot pain (21). The MIDs (11.1 for pain and 9.7 for function) are also similar to those reported for plantar heel pain (13 and 7, respectively) (32). However, it is not possible to compare our findings relating to the FFI-RS to previous literature because although the original version of the FFI has also been demonstrated to be responsive to change following foot surgery (33,34), there are no responsiveness data available for the FFI-RS (18). Pain assessment using the VAS has also been shown to be responsive in clinical trials of plantar heel pain (32), with the MID for first step pain (19 mm) being similar to pain while walking reported here (17.7 mm).

Recent Osteoarthritis Research Society International guidelines for clinical trials of rehabilitation interventions did not recommend a specific outcome measure for trials investigating OA. However, as a general principle, the guidelines suggest that pain measures should specify the timeframe of pain recall and the type of pain (e.g., at rest or during movement) (35). All 3 outcome measures we used asked participants to report symptoms in the

past week, which has been shown to be the most reliable recall period for chronic musculoskeletal pain (36). Similarly, these outcome measures evaluate pain during different activities such as standing, walking, or negotiating stairs. Interestingly, the VAS of pain while walking was more responsive than pain at rest, which is consistent with the view that the burden associated with OA of the first MTP joint relates primarily to difficulty walking (37). It would therefore appear that these outcome measures reliably capture changes in domains that are of importance to individuals with this condition.

Reduced range of motion is a cardinal feature of OA of the first MTP joint, which alters load distribution through the foot (38) and has a dose-response relationship with radiographic severity (3). In this context, perceived stiffness of the first MTP joint would seem to be an important outcome measure to include in clinical trials. However, although both the FFI-RS stiffness and VAS stiffness measures were responsive to change, the associated effect sizes were generally smaller than measures related to pain. A possible explanation for this observation is that the interventions used in the trial (orthoses and rocker-sole footwear) do not change the available range of motion within the first MTP joint but achieve some reduction in perceived stiffness by reducing pain and inflammation. In contrast, cheilectomy, a surgical procedure in which the dorsal exostosis is excised, results in significant increases in first MTP joint range of motion (39). As such, the stiffness outcome measures may demonstrate greater responsiveness following joint-preserving surgery and should therefore be considered for inclusion in future surgical trials.

The findings of this study need to be interpreted in the context of several key limitations. First, there is no widely accepted, gold standard approach for assessing responsiveness of outcome measures, and each statistical approach has inherent limitations (25,28). However, the 4 statistics we used resulted in a consistent pattern of responsiveness across the outcome measure subscales. Second, there is no consensus regarding the most appropriate question or number of response levels in determining the anchor used to define the MID (28), and the MID for symptom deterioration may be different from the MID for symptomatic improvement and require further investigation. Third, due to the nature of the trial upon which this analysis was based, it was necessary to alter the wording of some FFI-RS questions, which may have altered its psychometric properties, reliability, and validity. As such, our findings may differ from applications of the FFI-RS in its original form. Fourth, it is possible that the responsiveness of these outcome measures may be different for other types of interventions. Fifth, we have only tested the responsiveness of the measures over a 1-week recall period, so their responsiveness over a longer recall period is unknown. Sixth, because this study was embedded in a randomized trial, the findings are only generalizable to the patient population who met the inclusion criteria for the trial. Finally, the FHSQ has not yet undergone Rasch analysis, a statistical technique that evaluates whether overall scores summed from ordinal items can be considered to be linear, interval-level variables (40).

In conclusion, this study has shown that the FHSQ pain subscale, the FFI-RS pain subscale, and the 100-mm VAS of pain severity while walking are the most responsive outcome measures in individuals with OA of the first MTP joint receiving an orthotic or rocker-sole footwear intervention. These findings provide useful information to guide researchers in selecting appropriate outcome measures for use in future clinical trials and assist in determining sample size requirements. Further investigation is required to evaluate the broader psychometric properties of these measures in individuals with OA of the first MTP joint.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. Menz 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. Menz, Levinger, Roddy, Munteanu.
Acquisition of data. Menz, Auhl, Tan, Munteanu.
Analysis and interpretation of data. Menz, Auhl, Tan, Levinger, Roddy, Munteanu.

REFERENCES

1. Roddy E, Thomas MJ, Marshall M, Rathod T, Myers H, Menz HB, et al. The population prevalence of symptomatic radiographic foot osteoarthritis in community-dwelling older adults: the Clinical Assessment Study of the Foot. Ann Rheum Dis 2015;74:156–63.
2. Bergin SM, Munteanu SE, Zammit GV, Nikolopoulos N, Menz HB. Impact of first metatarsophalangeal joint osteoarthritis on health-related quality of life. Arthritis Care Res (Hoboken) 2012;64:1691–8.
3. Menz HB, Roddy E, Marshall M, Thomas MJ, Rathod T, Myers H, et al. Demographic and clinical factors associated with radiographic severity of first metatarsophalangeal joint osteoarthritis: cross-sectional findings from the Clinical Assessment Study of the Foot. Osteoarthritis Cartilage 2015;23:77–82.
4. Roddy E, Menz HB. Foot osteoarthritis: latest evidence and developments. Ther Adv Musculoskelet Dis 2018;10:91–103.
5. Shamus J, Shamus E, Gugel RN, Brucker BS, Skaruppa C. The effect of sesamoid mobilization, flexor hallucis strengthening, and gait training on reducing pain and restoring function in individuals with hallux limitus: a clinical trial. J Orthop Sports Phys Ther 2004;34:368–76.
6. Baumhauer JF, Singh D, Glazebrook M, Blundell C, De Vries G, Le IL, et al. Prospective, randomized, multi-centered clinical trial assessing safety and efficacy of a synthetic cartilage implant versus first metatarsophalangeal arthrodesis in advanced hallux rigidus. Foot Ankle Int 2016;37:457–69.
7. Munteanu SE, Zammit GV, Menz HB, Landorf KB, Handley CJ, Elzarka A, et al. Effectiveness of intra-articular hyaluronan (Synvisc, hylan G-F 20) for the treatment of first metatarsophalangeal joint osteoarthritis: a randomised placebo-controlled trial. Ann Rheum Dis 2011;70:1838–41.
8. Menz HB, Auhl M, Tan JM, Levinger P, Roddy E, Munteanu SE. Effectiveness of foot orthoses versus rocker-sole footwear for first metatarsophalangeal joint osteoarthritis: randomized trial. Arthritis Care Res (Hoboken) 2016;68:581–9.

9. Button G, Pinney S. A meta-analysis of outcome rating scales in foot and ankle surgery: is there a valid, reliable, and responsive system? Foot Ankle Int 2004;25:521–5.
10. Martin RL, Irrgang JJ. A survey of self-reported outcome instruments for the foot and ankle. J Orthop Sports Phys Ther 2007;37:72–84.
11. Walmsley S, Williams AE, Ravey M, Graham A. The rheumatoid foot: a systematic literature review of patient-reported outcome measures. J Foot Ankle Res 2010;3:12.
12. Riskowski JL, Hagedorn TJ, Hannan MT. Measures of foot function, foot health, and foot pain: American Academy of Orthopedic Surgeons Lower Limb Outcomes Assessment: Foot and Ankle Module (AAOS-FAM), Bristol Foot Score (BFS), Revised Foot Function Index (FFI-R), Foot Health Status Questionnaire (FHSQ), Manchester Foot Pain and Disability Index (MFPDI), Podiatric Health Questionnaire (PHQ), and Rowan Foot Pain Assessment (ROFPAQ). Arthritis Care Res (Hoboken) 2011;63:S229–39.
13. Hunt KJ, Hurwitt D. Use of patient-reported outcome measures in foot and ankle research. J Bone Joint Surg Am 2013;95:e118.
14. Shultz S, Olszewski A, Ramsey O, Schmitz M, Wyatt V, Cook C. A systematic review of outcome tools used to measure lower leg conditions. Int J Sports Phys Ther 2013;8:838–48.
15. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). Arthritis Care Res (Hoboken) 2011;63 Suppl 11:S240–52.
16. Menz HB, Levinger P, Tan JM, Auhl M, Roddy E, Munteanu SE. Rocker-sole footwear versus prefabricated foot orthoses for the treatment of pain associated with first metatarsophalangeal joint osteoarthritis: study protocol for a randomised trial. BMC Musculoskelet Disord 2014;15:86.
17. Bennett P, Patterson C, Wearing S, Baglioni T. Development and validation of a questionnaire designed to measure foot-health status. J Am Podiatr Med Assoc 1998;88:419–28.
18. Budiman-Mak E, Conrad KJ, Mazza J, Stuck RM. A review of the foot function index and the foot function index: revised. J Foot Ankle Res 2013;6:5.
19. Menz HB, Munteanu SE, Landorf KB, Zammit GV, Cicuttini FM. Radiographic classification of osteoarthritis in commonly affected joints of the foot. Osteoarthritis Cartilage 2007;15:1333–8.
20. Landorf KB, Keenan AM. An evaluation of two foot-specific, health-related quality-of-life measuring instruments. Foot Ankle Int 2002;23:538–46.
21. Menz HB, Auhl M, Ristevski S, Frescos N, Munteanu SE. Comparison of the responsiveness of the Foot Health Status Questionnaire and the Manchester Foot Pain and Disability Index in older people. Health Qual Life Outcomes 2014;12:158.
22. Budiman-Mak E, Conrad K, Roach K. The Foot Function Index: a measure of foot pain and disability. J Clin Epidemiol 1991;44:561–70.
23. Budiman-Mak E, Conrad K, Stuck R, Matters M. Theoretical model and Rasch analysis to develop a revised Foot Function Index. Foot Ankle Int 2006;27:519–27.
24. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chron Dis 1987;40:171–8.
25. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol 2000;53:459–68.
26. Beaton DE. Understanding the relevance of measured change through studies of responsiveness. Spine 2000;25:3192–9.
27. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol 2003;56:395–407.
28. 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.
29. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale (NJ): Erlbaum; 1988.
30. Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. J Clin Epidemiol 1997;50:869–79.
31. Thalheimer W, Cook S. How to calculate effect sizes from published research articles: a simplified methodology. 2002. URL: http://work-learning.com/effect_sizes.htm.
32. Landorf KB, Radford JA, Hudson S. Minimal Important Difference (MID) of two commonly used outcome measures for foot problems. J Foot Ankle Res 2010;3:7.
33. SooHoo NF, Vyas R, Samimi D. Responsiveness of the foot function index, AOFAS clinical rating systems, and SF-36 after foot and ankle surgery. Foot Ankle Int 2006;27:930–4.
34. Madeley NJ, Wing KJ, Topliss C, Penner MJ, Glazebrook MA, Younger AS. Responsiveness and validity of the SF-36, Ankle Osteoarthritis Scale, AOFAS Ankle Hindfoot Score, and Foot Function Index in end stage ankle arthritis. Foot Ankle Int 2012;33:57–63.
35. Fitzgerald GK, Hinman RS, Zeni J Jr, Risberg MA, Snyder-Mackler L, Bennell KL. OARSI Clinical Trials Recommendations: design and conduct of clinical trials of rehabilitation interventions for osteoarthritis. Osteoarthritis Cartilage 2015;23:803–14.
36. Perrot S, Marty M, Legout V, Moyse D, Henrotin Y, Rozenberg S. Ecological or recalled assessments in chronic musculoskeletal pain? A comparative study of prospective and recalled pain assessments in low back pain and lower limb painful osteoarthritis. Pain Med 2011;12:427–36.
37. Beeson P, Phillips C, Corr S, Ribbans WJ. Hallux rigidus: a cross-sectional study to evaluate clinical parameters. Foot 2009;19:80–92.
38. Menz HB, Auhl M, Tan JM, Buldt AK, Munteanu SE. Centre of pressure characteristics during walking in individuals with and without first metatarsophalangeal joint osteoarthritis. Gait Posture 2018;63:91–6.
39. Nawoczenski DA, Ketz J, Baumhauer JF. Dynamic kinematic and plantar pressure changes following cheilectomy for hallux rigidus: a mid-term followup. Foot Ankle Int 2008;29:265–72.
40. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum 2007;57:1358–62.
41. Zammit GV, Munteanu SE, Menz HB. Development of a diagnostic rule for identifying radiographic osteoarthritis in people with first metatarsophalangeal joint pain. Osteoarthritis Cartilage 2011;19:939–45.

Arthritis Care & Research
Vol. 72, No. 5, May 2020, pp 685–691
DOI 10.1002/acr.23902

American College of Rheumatology
Empowering Rheumatology Professionals

Association of Comorbid Interphalangeal Joint Pain and Erosive Osteoarthritis With Worse Hand Function in Individuals With Symptomatic Thumb Base Osteoarthritis

Leticia A. Deveza,[1] Sarah R. Robbins,[1] Vicky Duong,[1] Anne Wajon,[2] Edward A. Riordan,[1] Kai Fu,[1] Ray Jongs,[3] Win Min Oo,[1] and David J. Hunter[1]

Objective. Hand osteoarthritis (OA) trials often target exclusively the thumb base joint, although concomitant widespread interphalangeal (IP) joint involvement is frequent. We aimed to compare hand strength and function between individuals with isolated thumb base OA and those with coexistent IP joint pain and erosive OA.

Methods. Baseline data from a thumb base OA trial were analyzed (n = 204). Participants were age ≥40 years with symptomatic and radiographic thumb base OA. Only the index hand was included. Self-reported IP joint pain (in any proximal, distal, or thumb IP joint), hand function score (Functional Index for Hand Osteoarthritis questionnaire [range 0–30]), and hand grip and tip-pinch strength test results were obtained at baseline. Radiographs were scored for OA severity at each joint (Kellgren/Lawrence grade) and for the presence of erosive OA at the thumb base or IP joints. Multiple linear regression was used adjusting for age, sex, body mass index, and radiographic thumb base OA severity.

Results. Compared to individuals with isolated thumb base OA (62%), those with concomitant IP joint pain (17%) and erosive OA (21%) had significantly worse hand function (β = 1.82 [95% confidence interval (95% CI) 0.36, 3.28] and β = 1.47 [95% CI 0.74, 2.88], respectively). In addition, coexistence of erosive OA was independently associated with lower grip and tip-pinch strength (β = –5.14 [95% CI –7.58, –2.70] and β = –0.61 [95% CI –1.05, –0.17], respectively).

Conclusion. Concomitant IP joint pain and erosive OA are associated with worse hand function in individuals with thumb base OA. Patient stratification based on these characteristics may improve the design of future thumb base OA trials.

INTRODUCTION

Hand osteoarthritis (OA) is a common condition, with higher prevalence among older postmenopausal women (1). The prevalence of symptomatic hand OA was estimated to vary from 5% to 26% in large population-based surveys and was highest among elderly American women (1,2). It has been recently estimated that, by age 85 years, ~40% of the population (47% of women and 24% of men) will experience symptoms of hand OA (3). Although it is often a neglected disease, individuals who are affected by it generally experience difficulties or inability to perform daily activities due to poor hand strength and function in addition to chronic joint pain, which are the hallmarks of this disease (2). As a consequence, overall health-related quality of life is compromised and is similar or only slightly less to that observed in patients with rheumatoid arthritis (4,5). In addition, the erosive hand OA subset has been associated with worse pain and function compared to other inflammatory arthritis affecting the hands (6).

It is generally accepted that hand OA comprises 3 different phenotypes with potentially distinct risk factors and pathogenesis: thumb base OA, erosive OA, and nodal or interphalangeal (IP) OA (1). Nevertheless, overlap between these phenotypes frequently

The Efficacy of Combined Conservative Therapies on Clinical Outcomes in Base of Thumb Osteoarthritis (COMBO) Trial was supported by the National Health and Medical Research Council (program grant APP1091302) and the Lincoln Centre for Bone and Joint Diseases. Dr. Hunter's work was supported by the National Health and Medical Research Council (Practitioner Fellowship).

[1]Leticia A. Deveza, MD, Sarah R. Robbins, BPthy, PhD, Vicky Duong, BSc (Hons), DPT, Edward A. Riordan, MBBS, BSC, Kai Fu, MBBS, Win Min Oo, MBBS, MMedSc, RhMSUS, David J. Hunter, MBBS, PhD: The University of Sydney, Sydney, New South Wales, Australia; [2]Anne Wajon, BAppSc (Phty), MBA, PhD: Macquarie University, New South Wales, Sydney, Australia; [3]Ray Jongs, MPhil: Royal North Shore Hospital, Sydney, New South Wales, Australia.

Dr. Hunter has received consulting fees from Pfizer, Merck Serono, TissueGene, and TLC (less than $10,000 each). No other disclosures relevant to this article were reported.

Address correspondence to Leticia A. Deveza, MD, Rheumatology Department, Royal North Shore Hospital, Reserve Road, St. Leonards 2065, New South Wales, Australia. E-mail: leticia.alle@sydney.edu.au.

Submitted for publication December 4, 2018; accepted in revised form April 9, 2019.

SIGNIFICANCE & INNOVATIONS

- Concomitant interphalangeal (IP) joint pain and erosive osteoarthritis (OA) are common in individuals with thumb base OA and are associated with worse hand function.
- A higher number of painful IP joints, but not the number of IP joints affected by radiographic OA, was independently associated with more severe functional impairment in individuals with thumb base OA.
- Whether these characteristics influence treatment response in individuals receiving treatment targeted to the thumb base joint is still not well established.

occurs in the same individual (7–10). Hand OA trials often target 1 specific hand OA phenotype (e.g., erosive or thumb base OA) due to their presumed distinct pathogenesis and treatment options (9). Furthermore, the recently updated recommendations for hand OA management by the European League Against Rheumatism have also endorsed individualization of treatment, taking into account factors such as location and severity (11). However, the frequent co-occurrence of IP and erosive involvement in individuals with thumb base OA may detrimentally influence overall hand strength and function (8,10), which are core outcome measures assessed in hand OA trials to evaluate treatment effects (12). This, in turn, may have important implications for patient stratification and clinical trial design in thumb base OA because widely used instruments to assess hand function are not specific to the thumb (12). As yet, it is not well established if the overlap between hand OA phenotypes influences clinical outcomes in trials targeting exclusively 1 phenotype (12).

The main aim of this study was to compare hand strength and function between individuals with isolated symptomatic thumb base OA and those with coexistent widespread IP joint pain and erosive OA. Pain at the base of the thumb and patient global assessment (PtGA) of the thumb base OA were compared as secondary analyses. As a secondary aim, we also investigated the effect of the number of painful IP joints and the number of IP joints affected by radiographic OA on hand strength and function in this relatively large population of individuals with symptomatic thumb base OA.

MATERIALS AND METHODS

Study design. We conducted an exploratory analysis of baseline data from the Efficacy of Combined Conservative Therapies on Clinical Outcomes in Base of Thumb Osteoarthritis (COMBO) Trial (13). COMBO was a randomized controlled trial assessing the efficacy of a combination of conservative therapies for thumb base OA, including education about OA and joint protection, hand exercises, a splint for the base of the thumb, and diclofenac sodium gel over 6 weeks, compared to education and joint protection alone. The trial was approved by the local ethics committee (HREC/15/HAWKE/479). The data used in this study were collected at baseline before randomization.

Participants. Participants were recruited from the community and a research volunteer database. Eligible participants were required to fulfill the following criteria: age ≥40 years; pain at the base of the thumb at least one-half of the days in the previous month; average pain score of ≥40 on a 100-mm visual analog scale (VAS) over the past 30 days and in the 48 hours prior to the screening visit; score ≥6 on the Functional Index for Hand Osteoarthritis (FIHOA) (range 0–30) (14), and radiographic evidence of thumb base OA as read by a trained rheumatologist (LAD) (Kellgren/Lawrence [K/L] grade ≥2) (15). Participants with a known diagnosis of crystal-related or autoimmune arthritis, hemochromatosis, or fibromyalgia were excluded. We also excluded participants who had the following: hand surgery in the previous 6 months; intraarticular hyaluronic acid injection in the affected joint in the previous 6 months; intraarticular steroid injection in the affected joint in the previous month; significant injury to the affected joint in the previous 6 months; and any other self-reported hand condition likely contributing to the pain at the base of the thumb (e.g., scaphoid fracture, carpal tunnel syndrome, DeQuervain's tendinopathy, trigger thumb, joint infection, diabetic neuropathy, pain referred from the neck, pain following hand or wrist trauma or surgery). Additional exclusion criteria have been described in the trial's protocol (13).

Baseline clinical and radiographic assessment. Demographic and clinical data including age, height, weight, and self-reported pain in the IP joints were collected at the baseline visit. Self-reported IP joint pain was assessed by asking participants to point on their own index hand to the joints in which they experience pain. Presence of pain in the thumb IP joint and second to fifth proximal and distal IP joints were considered in this analysis. A posteroanterior-view radiograph of both hands was obtained at baseline and scored by a trained rheumatologist (LAD) for OA severity according to K/L grade (15) and presence of central erosions (present versus absent), according to the Osteoarthritis Research Society International (OARSI) atlas (16), in the thumb base and IP joints. Erosive OA was defined as the presence of a characteristic OA erosion in at least 1 of the joints assessed in the index hand. The intra- and interrater reliability assessment of the K/L scores at the first carpometacarpal (CMC) joint showed substantial agreement (17) (weighted $\kappa = 0.76$ and 0.87, respectively). Intrarater reliability was calculated by 1 rater using 20 radiographs with a 1-year interval between readings (LAD), and interrater reliability was calculated using 22 radiographs read by 2 raters (LAD and DJH).

Outcomes. The outcomes used in this study have been promoted as core outcome measures in the most recent recommendations by the OARSI for the design and conduct of hand OA trials (12). The outcomes were collected at baseline from all

included participants. Self-reported hand function was assessed by the FIHOA (14), a questionnaire composed of 10 items scored from 0 to 3 (higher scores indicating poorer hand function), to evaluate the degree of difficulty involved in performing distinct activities involving the hand. This tool has been extensively used in hand OA research and was shown to be valid, reliable, and feasible (18). Hand strength was assessed by the grip and tip-pinch strength tests using a hand dynamometer (Jamar) and a pinch gauge (B&L Engineering), respectively, and measured in kilograms. Participants were sitting with both feet flat on the ground and elbow flexed at 90°. For the tip-pinch strength test, participants were asked to pinch the pinch gauge with their thumb and index finger forming an O shape. The maximum measurement of 3 attempts for each strength outcome was used in the analysis. A VAS (range 0–100) was used to quantify average pain at the base of the thumb in the previous 48 hours. PtGA of the thumb base condition was assessed on a VAS with the question, "Considering all the ways your thumb arthritis affects you, how have you been during the last 48 hours?" (0 indicating "very well," 100 indicating "very poor").

Disease subsets. Only the index hand, defined by average VAS pain in the previous 48 hours, was included in the analysis. If both thumb bases were equally painful, participants nominated the worst hand (i.e., the one that caused more difficulties to the participant), and it was included as the index.

Participants were divided into 1 of 3 groups based on the presence of self-reported IP joint pain defined as pain in ≥1 IP joint (among all proximal and distal IP and thumb IP joints) and presence of erosive OA on radiograph in either thumb base or IP joints: 1) isolated thumb base OA (group 1); 2) thumb base OA and concomitant IP joint pain (nonerosive) (group 2); 3) thumb base OA and erosive OA (group 3). Group 3 was based on a radiographic definition and included participants with and without IP joint pain (n = 22 and n = 21, respectively). Due to the small sizes of these groups, they were not analyzed separately.

Statistical analysis. The distribution of thumb base OA severity and other demographic characteristics between the 3 groups were summarized and compared using the chi-square test for categorical variables and the analysis of variance (ANOVA) test for continuous variables. Clinical outcomes were compared across the groups using the ANOVA test as the data were normally distributed on assessment of histograms. Multiple linear regression was used to compare hand function, strength, thumb base pain, and PtGA across the groups, adjusting for age, sex, body mass index (BMI), and radiographic thumb base OA severity. The adjusted difference between the groups in the outcome (beta coefficient and 95% confidence intervals [95% CIs]) was presented along with the *P* value. *P* values less than 0.05 were considered significant.

Categories for the number of painful IP joints and number of IP joints affected by radiographic OA (K/L grade ≥2) were created based on its distribution: 0, 1–5, and 6–9. To assess the effect of the number of painful IP joints and number of IP joints affected by radiographic OA on our main clinical outcomes, linear regression models were fitted, adjusting for age, sex, BMI, radiographic thumb base OA severity, and disease group.

RESULTS

All randomized participants (n = 204) were included in this analysis. The mean ± SD age was 65 ± 10 years, and 75% of participants were female (Table 1). The majority of participants had isolated thumb base OA (n = 126, 62%), while 17% (n = 35) had concomitant IP joint pain without radiographic erosions, and 21% (n = 43) had erosive OA. Radiographic IP OA defined as K/L grade ≥2 in any IP joint was present in 73% (n = 149) of participants, while the presence of IP joint pain was less common (n = 57, 28%). Women more often had symptomatic involvement of the IP joints and erosive OA. Radiographic thumb base OA was more severe in group 3 (Table 2).

Table 1. Baseline characteristics of the study population*

Characteristic	Value
Age, mean ± SD years	65 ± 10
Female sex	155 (75.6)
BMI, mean ± SD kg/m^2	28.6 ± 6.5
Right hand index	118 (57.6)
Function (FIHOA score), mean ± SD	10.6 ± 3.9
Grip strength, mean ± SD kg	21.2 ± 9.3
Tip-pinch strength, mean ± SD kg	3.2 ± 1.3
Thumb base pain, mean ± SD (range 0–100)	57.7 ± 13.6
Patient global assessment, mean ± SD (range 0–100)	39.4 ± 21.7
K/L grade	
2	88 (43.1)
3	80 (39.2)
4	36 (17.6)
Presence of IP joint pain	57 (27.9)
Number of painful IP joints	
0	147 (72.1)
1–5	39 (19.1)
6–9	18 (8.8)
Presence of radiographic OA in IP joint (any IP joint with K/L grade ≥2)	149 (73)
Number of joints affected by radiographic OA†	
0	55 (27.0)
1–5	108 (52.9)
6–9	41 (20.1)
Hand OA group	
Isolated thumb base OA (group 1)	126 (61.8)
Thumb base and symptomatic IP joint (group 2)	35 (17.2)
Thumb base and erosive OA (group 3)	43 (21.1)

* Values are the number (%) unless indicated otherwise. BMI = body mass index; FIHOA = Functional Index for Hand Osteoarthritis; K/L = Kellgren/Lawrence (grade); IP = interphalangeal; OA = osteoarthritis.
† K/L grade ≥2.

Table 2. Comparison of baseline characteristics and clinical outcomes between the 3 groups*

	Group 1, isolated thumb base OA (n = 126)	Group 2, thumb base OA plus symptomatic IP joint (nonerosive) (n = 35)	Group 3, thumb base OA plus erosive OA (n = 43)	*P*†
Age, years	64 ± 10	64 ± 7	66 ± 13	0.342
Female sex, no. (%)	88 (69.8)	29 (82.9)	38 (88.4)	0.028‡
BMI, kg/m^2	27.9 ± 5.9	30.0 ± 8.3	29.6 ± 6.3	0.135
Thumb base K/L grade, no. (%)				0.010
2	59 (46.8)	18 (51.4)	11 (25.6)	
3	50 (39.7)	13 (37.1)	17 (39.5)	
4	17 (13.5)	4 (11.4)	15 (34.9)	
Function (FIHOA score) (range 0–30)	9.8 ± 3.5	11.9 ± 4.4	11.9 ± 4.1	0.001‡
Grip strength, kg	23.5 ± 9.7	19.6 ± 7.2	15.6 ± 7.0	<0.001‡
Tip-pinch strength, kg	3.4 ± 1.3	3.3 ± 1.3	2.6 ± 1.1	0.002‡
Thumb base pain (range 0–100)	56.4 ± 13.5	60.3 ± 15.1	59.7 ± 12.1	0.182
Patient global assessment (range 0–100)	35.8 ± 22.1	47.7 ± 21.2	43.3 ± 18.7	0.006‡

* Values are the mean ± SD unless indicated otherwise. OA = osteoarthritis; IP = interphalangeal; BMI = body mass index; K/L = Kellgren/Lawrence (grade); FIHOA = Functional Index for Hand Osteoarthritis.
† All 3 groups.
‡ Significant.

There were statistically significant differences between the groups in all clinical outcomes except thumb base pain (Table 2). Overall, hand function, grip and pinch strength, and PtGA scores were less severe in individuals in group 1 compared to those in the other 2 groups. Individuals in group 3 displayed the lowest hand strength scores (mean ± SD grip strength 15.6 ± 7 kg in group 3 versus 23.5 ± 9.7 kg and 19.6 ± 7.2 kg in groups 1 and 2, respectively; mean ± SD pinch strength 2.6 ± 1.1 kg in group 3 versus 3.4 ± 1.3 kg and 3.3 ± 1.3 kg in groups 1 and 2, respectively), while mean ± SD PtGA was poorest in group 2 (47.7 ± 21.2 in group 2 versus 35.8 ± 22.1 and 43.3 ± 18.7 in groups 1 and 3, respectively). Compared to group 1, adjusted analysis for age, sex, BMI, and radiographic thumb base severity (Table 3) showed persistence of significantly worse hand function scores in groups 2 (β = 1.82 [95% CI 0.36, 3.28]) and group 3 (β = 1.47 [95% CI 0.74, 2.88]), worse grip and pinch strength in group 3 (β = –5.14 [95% CI –7.58, –2.70] and β = –0.61 [95% CI –1.05, –0.17], respectively), and worse PtGA in group 2 (β = 11.76 [95% CI 3.72, 19.80]).

A higher number of painful IP joints and IP joints affected by radiographic OA were associated with worse hand strength and function in unadjusted analysis (Table 4) (statistically significant for all associations except for tip-pinch strength and the number of IP joints affected by radiographic OA). However, in adjusted analysis, only the association between increasing number of painful IP joints and worse hand function was statistically significant (1–5 joints β = 1.99 [95% CI 0.41, 3.58] and 6–9 joints β = 2.53 [95% CI 0.36, 4.69], compared to no symptomatic IP joint).

DISCUSSION

In this study, we aimed to investigate potential differences in clinical outcomes, particularly hand strength and function, in individuals with symptomatic thumb base OA with and without coexistent IP joint pain and radiographic evidence of erosive OA. We found that individuals with isolated involvement of the thumb base joint had less severe impairment in hand function and strength

Table 3. Association between hand osteoarthritis (OA) group and clinical outcomes*

	Unadjusted				Adjusted†			
Outcome	Thumb base OA plus symptomatic IP joint (nonerosive)	*P*	Thumb base OA plus erosive OA	*P*	Thumb base OA plus symptomatic IP joint (nonerosive)	*P*	Thumb base OA plus erosive OA	*P*
Function (FIHOA score)	2.13 (0.67, 3.58)	0.004‡	2.12 (0.77, 3.47)	0.002‡	1.82 (0.36, 3.28)	0.015‡	1.47 (0.74, 2.88)	0.039‡
Grip strength	–3.92 (–7.25, –0.60)	0.021‡	–7.90 (–10.98, –4.83)	<0.001‡	–2.13 (-4.67, 0.40)	0.099	–5.14 (–7.58, –2.70)	<0.001‡
Tip-pinch strength	–0.15 (–0.63, 0.32)	0.525	–0.79 (–1.24, –0.34)	0.001‡	–0.05 (–0.50, 0.40)	0.828	–0.61 (–1.05, –0.17)	0.006‡
Thumb base pain	3.90 (–1.21, 9.01)	0.134	3.37 (–1.35, 8.10)	0.160	3.20 (–2.00, 8.40)	0.226	2.14 (–2.85, 7.15)	0.398
PtGA	1.88 (3.84, 19.91)	0.004‡	7.53 (0.11, 14.96)	0.047‡	11.76 (3.72, 19.80)	0.004‡	7.27 (–0.45, 15.00)	0.065

* Values are the beta coefficient (95% confidence interval) unless indicated otherwise. IP = interphalangeal; FIHOA = Functional Index for Hand Osteoarthritis; PtGA = patient global assessment.
† Adjusted for age, sex, body mass index, and radiographic thumb base OA severity.
‡ Significant.

Table 4. Effect of the number of painful interphalangeal (IP) joints and the number of IP joints affected by radiographic osteoarthritis (OA) on hand strength and function in individuals with thumb base OA*

	Unadjusted		Adjusted†	
	β (95% CI)	*P*	β (95% CI)	*P*
Function (FIHOA score)				
Number of IP joints affected by ROA‡				
0 (ref.)				
1–5	0.99 (–0.29, 2.28)	0.130	0.47 (–0.82, 1.77)	0.470
6–9	2.13 (0.53, 3.74)	0.009§	0.84 (–0.98, 2.67)	0.363
Number of symptomatic IP joints				
0 (ref.)				
1–5	2.40 (1.04, 3.77)	0.001§	1.99 (0.41, 3.58)	0.014§
6–9	2.96 (1.07, 4.85)	0.002§	2.53 (0.36, 4.69)	0.022§
Grip strength				
Number of IP joints affected by ROA				
0 (ref.)				
1–5	–2.63 (–5.60, 0.29)	0.081	–0.50 (–2.75, 1.74)	0.658
6–9	–7.19 (–10.88, –3.50)	<0.001§	–1.53 (–4.69, 1.62)	0.339
Number of symptomatic IP joints				
0 (ref.)				
1–5	–4.33 (–7.56, –1.11)	0.009§	–0.13 (–2.78, 2.75)	0.993
6–9	–7.37 (–11.84, –2.91)	0.001§	–3.71 (–7.50, 0.06)	0.054
Tip-pinch strength				
Number of IP joints affected by ROA				
0 (ref.)				
1–5	0.05 (–0.36, 0.48)	0.792	0.22 (–0.17, 0.63)	0.268
6–9	–0.48 (–1.01, 0.42)	0.071	0.03 (–0.53, 0.60)	0.897
Number of symptomatic IP joints				
0 (ref.)				
1–5	–0.12 (–0.58, 0.33)	0.590	0.31 (–0.18, 0.81)	0.211
6–9	–0.92 (–1.55, –0.28)	0.005§	–0.49 (–1.17, 0.18)	0.154

* 95% CI = 95% confidence interval; FIHOA = Functional Index for Hand Osteoarthritis; ref. = reference; ROA = radiographic OA.
† Adjusted for age, sex, body mass index, hand OA group, and Kellgren/Lawrence (K/L) grade at the first carpometacarpal joint.
‡ K/L grade ≥2.
§ Significant.

and considered themselves overall less affected by the thumb base condition compared to their counterparts with concomitant IP joint pain and radiographic erosions. The significant differences in hand function, assessed by the FIHOA questionnaire, remained despite adjustment for confounders. Similarly, the presence of erosive OA was independently associated with decreased hand strength in this population.

Personalized hand OA treatment according to factors such as OA location (e.g., IP versus thumb base) has been promoted due to the heterogeneity in risk factors and clinical presentation between hand OA phenotypes (e.g., joint laxity as a risk factor for thumb base OA [19] and a more prominent inflammatory component and more severe clinical burden in erosive OA [20]). Accordingly, most thumb base OA trials investigate treatments targeting this joint specifically and often utilize, at least in part, available outcome measures that are not thumb base specific. We propose that the co-occurrence of IP joint pain and erosive OA may be important factors for patient stratification in thumb base OA trials due to their significant association with important outcome measures in hand OA trials. In clinical practice, clinicians often modify the treatment in patients with both thumb base and IP joint pain, for example, by designing supports that address both joints and modifying the exercise program accordingly.

Differences in clinical manifestations between hand OA phenotypes have been previously investigated, and no difference in disability was encountered when the thumb base or the IP joints were the most symptomatic joints (21). In addition, Marshall et al (8) have shown that individuals with concomitant radiographic thumb base and IP OA displayed worse hand function compared to those with isolated IP or thumb base involvement on radiograph. However, contrary to our findings, when clinical data were used, no significant difference was found in hand pain and function, as assessed by the Australian Canadian Osteoarthritis Hand Index questionnaire (22), between individuals with isolated symptomatic thumb base OA and those with concomitant IP joint pain (10). An important limitation of the latter study was the small number of individuals with isolated thumb base OA (n = 20), which may have limited the analysis. Other studies have shown that among the 3 phenotypes, the erosive type is associated with the greatest overall clinical burden (1,7).

The first CMC joint (i.e., thumb base joint) has unique structural and shape characteristics to permit wide movement of the thumb in multiple planes. Hand activities, particularly those involving

strong pinching and grasping, cause an increase in the forces across the first CMC joint (23,24), which is substantially greater compared to the forces across IP joints during the same task (24). As yet, widely used tools for the assessment of treatment effects to conservative therapies in thumb base OA are generally not condition specific. Such tools have been developed and tested to assess outcome following hand surgery in order to better detect variations related to treatment while considering the common coexistence of other hand conditions in the same individual (25). Furthermore, efforts are currently underway to develop new classification criteria for hand OA distinguishing thumb base OA from IP OA in order to reduce heterogeneity in clinical trials. We found a significant impact of IP joint pain and erosive OA on clinical outcomes in thumb base OA, suggesting that tools specific for the thumb base condition or improved patient stratification may be needed in future trials.

There are a few potential limitations in our study. First, this is an observational study and, therefore, we cannot confirm if there are implications of the different groups for treatment response. Second, we did not collect data on pain in the first metacarpophalangeal joint, which is also commonly involved in individuals with thumb base OA. Finally, we used the FIHOA tool to assess self-reported hand function based on the available evidence and recommendations for hand OA trials when the COMBO trial was designed (12,26). However, the FIHOA tool contains items considered outdated at the present, and other tools may be more appropriate for use in future hand OA studies.

In conclusion, concomitant involvement of the IP joints is common in individuals with thumb base OA and is independently associated with worse clinical outcomes except for thumb base pain. These findings suggest that patient stratification based on these characteristics may improve the design of future thumb base OA trials. Future research should examine whether these characteristics influence treatment response in individuals receiving treatment targeted to the thumb base joint.

ACKNOWLEDGMENTS

The authors thank all participants in the COMBO trial for their generous contribution to this study. They also thank the statisticians Dr. Rachel L. O'Connell and Rebecca Asher from the National Health and Medical Research Council Clinical Trials Centre, University of Sydney, Camperdown, New South Wales, Australia, for their valuable advice on the statistical analysis of 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. Deveza 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. Deveza, Robbins, Wajon, Riordan, Jongs, Hunter.

Acquisition of data. Robbins, Duong, Riordan, Fu, Hunter.

Analysis and interpretation of data. Deveza, Oo.

REFERENCES

1. Kloppenburg M, Kwok WY. Hand osteoarthritis: a heterogeneous disorder. Nat Rev Rheumatol 2011;8:22–31.
2. Zhang Y, Niu J, Kelly-Hayes M, Chaisson CE, Aliabadi P, Felson DT. Prevalence of symptomatic hand osteoarthritis and its impact on functional status among the elderly: the Framingham Study. Am J Epidemiol 2002;156:1021–7.
3. Qin J, Barbour KE, Murphy LB, Nelson AE, Schwartz TA, Helmick CG, et al. Lifetime risk of symptomatic hand osteoarthritis: the Johnston County Osteoarthritis Project. Arthritis Rheumatol 2017;69:1204–12.
4. Slatkowsky-Christensen B, Mowinckel P, Loge JH, Kvien TK. Health-related quality of life in women with symptomatic hand osteoarthritis: a comparison with rheumatoid arthritis patients, healthy controls, and normative data. Arthritis Rheum 2007;57:1404–9.
5. Michon M, Maheu E, Berenbaum F. Assessing health-related quality of life in hand osteoarthritis: a literature review. Ann Rheum Dis 2011;70:921–8.
6. Wittoek R, Cruyssen BV, Verbruggen G. Predictors of functional impairment and pain in erosive osteoarthritis of the interphalangeal joints: comparison with controlled inflammatory arthritis. Arthritis Rheum 2012;64:1430–6.
7. Marshall M, Peat G, Nicholls E, van der Windt D, Myers H, Dziedzic K. Subsets of symptomatic hand osteoarthritis in community-dwelling older adults in the United Kingdom: prevalence, inter-relationships, risk factor profiles and clinical characteristics at baseline and 3-years. Osteoarthritis Cartilage 2013;21:1674–84.
8. Marshall M, van der Windt D, Nicholls E, Myers H, Hay E, Dziedzic K. Radiographic hand osteoarthritis: patterns and associations with hand pain and function in a community-dwelling sample. Osteoarthritis Cartilage 2009;17:1440–7.
9. Kloppenburg M. Hand osteoarthritis-nonpharmacological and pharmacological treatments. Nat Rev Rheumatol 2014;10:242–51.
10. Bijsterbosch J, Visser W, Kroon HM, Stamm T, Meulenbelt I, Huizinga TW, et al. Thumb base involvement in symptomatic hand osteoarthritis is associated with more pain and functional disability. Ann Rheum Dis 2010;69:585–7.
11. Kloppenburg M, Kroon FP, Blanco FJ, Doherty M, Dziedzic KS, Greibrokk E, et al. 2018 update of the EULAR recommendations for the management of hand osteoarthritis. Ann Rheum Dis 2019;78:16–24.
12. Kloppenburg M, Boyesen P, Visser AW, Haugen IK, Boers M, Boonen A, et al. Report from the OMERACT Hand Osteoarthritis Working Group: set of core domains and preliminary set of instruments for use in clinical trials and observational studies. J Rheumatol 2015;42:2190–7.
13. Deveza LA, Hunter DJ, Wajon A, Bennell KL, Vicenzino B, Hodges P, et al. Efficacy of combined conservative therapies on clinical outcomes in patients with thumb base osteoarthritis: protocol for a randomised, controlled trial (COMBO). BMJ Open 2017;7:e014498.
14. Dreiser RL, Maheu E, Guillou GB, Caspard H, Grouin JM. Validation of an algofunctional index for osteoarthritis of the hand. Rev Rhum Engl Ed 1995;62 Suppl 1:43S–53S.
15. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957;16:494–502.
16. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. Osteoarthritis Cartilage 2007;15 Suppl A:A1–56.
17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.

18. Visser AW, Boyesen P, Haugen IK, Schoones JW, van der Heijde DM, Rosendaal FR, et al. Instruments measuring pain, physical function, or patient's global assessment in hand osteoarthritis: a systematic literature search. J Rheumatol 2015;42:2118–34.
19. Jonsson H, Eliasson GJ, Jonsson A, Eiriksdottir G, Sigurdsson S, Aspelund T, et al. High hand joint mobility is associated with radiological CMC1 osteoarthritis: the AGES-Reykjavik study. Osteoarthritis Cartilage 2009;17:592–5.
20. Kortekaas MC, Kwok WY, Reijnierse M, Huizinga TW, Kloppenburg M. In erosive hand osteoarthritis more inflammatory signs on ultrasound are found than in the rest of hand osteoarthritis. Ann Rheum Dis 2013;72:930–4.
21. Spacek E, Poiraudeau S, Fayad F, Lefevre-Colau MM, Beaudreuil J, Rannou F, et al. Disability induced by hand osteoarthritis: are patients with more symptoms at digits 2–5 interphalangeal joints different from those with more symptoms at the base of the thumb? Osteoarthritis Cartilage 2004;12:366–73.
22. Bellamy N, Campbell J, Haraoui B, Buchbinder R, Hobby K, Roth JH, et al. Dimensionality and clinical importance of pain and disability in hand osteoarthritis: Development of the Australian/Canadian (AUSCAN) Osteoarthritis Hand Index. Osteoarthritis Cartilage 2002;10:855–62.
23. Anakwe RE, Middleton SD. Osteoarthritis at the base of the thumb. BMJ 2011;343:d7122.
24. Goislard de Monsabert B, Vigouroux L, Bendahan D, Berton E. Quantification of finger joint loadings using musculoskeletal modelling clarifies mechanical risk factors of hand osteoarthritis. Med Eng Phys 2014;36:177–84.
25. Citron N, Hulme CE, Wardle N. A self-administered questionnaire for basal osteoarthritis of the thumb. J Hand Surg Eur Vol 2007;32:524–8.
26. Kloppenburg M, Maheu E, Kraus VB, Cicuttini F, Doherty M, Dreiser RL, et al. OARSI Clinical Trials recommendations: design and conduct of clinical trials for hand osteoarthritis. Osteoarthritis Cartilage 2015;23:772–86.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 72, No. 5, May 2020, pp 692–698
DOI 10.1002/acr.23903

Potential Role of Cost and Quality of Life in Treatment Decisions for Arthritis-Related Knee Pain in African American and Latina Women

Darrell J. Gaskin,[1] Taruja D. Karmarkar,[1] Anne Maurer,[2] Linda Bucay-Harari,[1] Giselle Casillas,[1] Allyson Gittens,[1] Lynne C. Jones,[1] Roland J. Thorpe Jr.,[1] Elliott Tolbert,[1] and James E. Wood[3]

Objective. The present study was undertaken to investigate whether Latina and African American women with arthritis-related knee pain and primary care providers who treat them believe their treatment decisions would benefit from having more information about the impact of treatment on their quality of life, medical care costs, and work productivity.

Methods. We conducted 4 focus groups of Latina and African American women over age 45 years who had knee pain. We also conducted 2 focus groups with primary care providers who treated Latina and African American women for knee pain. The participants were recruited from the community. They were asked their opinions about a decision tool that presented information on a range of treatment options and their impacts on quality of life, medical care costs, and work productivity. They were asked whether providing this information would help them make better treatment decisions. We analyzed the focus group transcripts using ATLAS.ti.

Results. We found that minority women and primary care providers endorsed the use of a decision-making tool that provided information of the impact of treatment on quality of life, medical care costs, and work productivity. Providers felt that patients would benefit from having the additional information but were concerned about its complexity and some patients' ability to comprehend the information.

Conclusion. Latina and African American women could make more informed treatment decisions for their knee pain using a decision-making tool that provides them with significant information about how various treatment options may impact their quality of life, medical care costs, and workforce productivity.

INTRODUCTION

According to the National Health Interview Survey, an estimated 14 million individuals in the US have symptomatic knee osteoarthritis (OA), including >3 million racial/ethnic minorities (1,2). Knee OA presents as a range of clinical manifestations, from mild knee pain causing minimal impairment to severe pain and disability (3). Furthermore, women who are obese are at an increased risk of experiencing knee OA, and this particular comorbid condition can lead to other burdensome or complicating comorbidities such as heart disease or diabetes mellitus (4). Individuals who develop knee OA often find themselves in this cycle: limited mobility leads to weight gain, which leads to increased pain in the joints, which further limits patients' mobility (5–10). In addition to physiologic risk factors, prior research has found that minority women, specifically Latina and African American women, are disproportionately more likely to lose quality-adjusted life years due to knee pain and/or OA (4).

As obesity rates continue to rise in the US (11), it is imperative to understand the potential costs and health outcomes that result from knee OA and associated comorbidities. In 2010, total costs related to pain ranged from $560 to $635 billion, which includes $261 and $300 billion for health care and $299 to $335 billion for lost worker productivity (12).

A substantial portion of these costs is due to arthritis and joint pain (13). Direct health care costs of commercially insured patients

Supported by Zimmer Biomet. Dr. Karmarkar's work was supported by the NIH (T32 National Research Service Award training grant from the Agency for Healthcare Research and Quality).

[1]Darrell J. Gaskin, PhD, MS, Taruja D. Karmarkar, PhD, MHS, Linda Bucay-Harari, MPH, Giselle Casillas, MPP, Allyson Gittens, MHS, Lynne C. Jones, PhD, Roland J. Thorpe Jr., PhD, Elliott Tolbert, PhD: Johns Hopkins University, Baltimore, Maryland; [2]Anne Maurer, MS: Zimmer Biomet, Warsaw, Indiana; [3]James E. Wood, MD: MedStar Harborview Hospital, Baltimore, MD.

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

Address correspondence to Darrell J. Gaskin, PhD, MS, Johns Hopkins University Bloomberg School of Public Health, Health Policy & Management, 624 North Broadway, Suite 441, Baltimore, MD 21205. E-mail: dgaskin1@jhu.edu.

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

SIGNIFICANCE & INNOVATIONS

- Latina and African American women with knee pain want their health providers to give them more information about treatment options.
- Health providers who treat Latina and African American women suggested that a class or tutorial about various treatment options for knee pain might be beneficial for them.
- Providing Latina and African American women information about quality of life and nonmedical cost implications of conservative to aggressive treatment options can help them make better treatment decisions.

diagnosed with OA were more than double the costs of matched controls without OA (13). Annual out-of-pocket expenditures have been shown to increase significantly once OA is diagnosed (14). Direct noninpatient costs the year immediately prior to surgery might be saved by more effective treatment choices (15,16). Clinical guidelines for knee OA recommend nonpharmacologic first-line management, which should include weight loss, healthy eating habits, physical activity, self-management of pain, information/education, and orthosis (17).

By equipping patients with the right tools to make informed decisions, we can motivate the use of high-value, cost-effective treatment options. A shared decision-making tool that can facilitate a discussion between the patient and provider can help limit unnecessary expenditures, lead to better outcomes, improve communication, and help address disparities. We previously used a Markov model to compare lifetime costs and quantify disparities in treatment across different subpopulations of knee OA patients and found that all stakeholders should consider the costs associated with delaying or forgoing therapy (16). We developed a decision-making tool to provide patients with information that they may not otherwise receive during their physician visit. Several components of our tool set it apart from those discussed in the current literature (3,18–21): it incorporates and compares several treatment options; it adjusts for patient demographics and comorbidities; and, most importantly, it includes a cost component that highlights the effects of knee OA in a patient's daily life. Cost information is rarely discussed during a patient–physician visit.

Given that delaying or forgoing therapy was found to be especially costly for minority populations (16), we aimed to understand the potential benefits of a decision-making tool that translates health economic data into patient-friendly information. Our goal was to capture provider and minority patient perceptions of the value of these key pieces of information during the patient–physician interaction and identify potential tool improvements to facilitate its role in the decision-making process.

PATIENTS AND METHODS

Six focus groups were conducted between September 2015 and May 2017. Women were eligible for the study if they had knee pain for >3 months and were age 45 years or older. The Latina focus groups were conducted in Spanish and the African American focus groups were conducted in English. Two focus groups were conducted in English with Maryland health care providers who treat minority women with knee pain. Focus group sessions were tape recorded and transcribed by a professional transcription company. All focus group participants completed a demographic questionnaire prior to the focus group discussion. Patient participants received lunch and a $25 gift card, and providers received dinner from a restaurant in Baltimore. The Latina focus groups were held in meeting rooms at a Latino community-based organization and a community health clinic. The African American focus groups were conducted at a meeting room in a Baltimore hotel. The provider focus groups were held in a closed meeting room in a restaurant. This study was approved by the Johns Hopkins Bloomberg School of Public Health Institutional Review Board.

Study sample. *Patient participants*. African American women were recruited by placing an ad in the *Baltimore Sun* and the *Baltimore Afro-American*, a local newspaper read by a predominantly African American audience. We attempted to recruit Latina women through the *Baltimore Sun* and *The Latin Opinion*, a local newspaper read by a predominantly Latino audience; however, no one replied. Also, efforts in Baltimore to recruit participants through ads in Spanish-language newspapers, distributing fliers in local Spanish-speaking congregations, and distributing fliers in the downtown section of the Latino community were unsuccessful. Therefore, we recruited Latina women for the first Latina focus group by referral from a primary care provider in Montgomery County, Maryland. Participants in the second Latina focus group were recruited by placing fliers in a community health center in Northern Virginia.

Most of the African American women had some college education, were retired, and had insurance, while most of the Latina women in the study sample had a high school education or less, were working or unemployed, and one-half of them had insurance. Most of the African American women had been told by a doctor that they had some form of arthritis, compared to less than one-half of the Latina participants. Demographic and health characteristics for both groups can be found in Table 1.

Provider participants. Two different sampling methods were used to recruit providers in Maryland. They were recruited independently from the patient participants. Convenience sampling was conducted by sending fliers to Johns Hopkins Community Physicians and placing advertisements in the newsletter for the Nurse Practitioner Association of Maryland. Snowball sampling was also used to reach health care providers in Baltimore, Maryland. Participants in the provider focus group included 3

Table 1. Demographic and health characteristics for African American and Latina focus group participants*

Characteristic	African American women (n = 9)	Latina women (n = 12)
Last time you had pain, how much pain did you have?		
A lot	3 (33.3)	4 (33.3)
Moderate	4 (44.4)	4 (33.3)
A little		4 (33.3)
Missing (did not answer question)	2 (22.2)	
Education level		
Completed 8th grade or less		5 (41.7)
Completed 9th–11th grade		2 (16.7)
Completed 12th grade	2 (22.2)	2 (16.7)
Completed some college	3 (33.3)	1 (8.3)
Received a bachelor's degree	3 (22.2)	2 (16.7)
Received a master's degree	1 (11.1)	
Doctoral or professional degree	1 (11.1)	
Work status		
Employed full time		1 (8.3)
Employed part time		4 (33.3)
Illness or sick leave		1 (8.3)
Unemployed		3 (25.0)
Permanently disabled	1(11.1)	1 (8.3)
Retired	8 (88.9)	1 (8.3)
Covered by any health insurance	9 (100.0)	6 (50.0)
Have you ever been diagnosed with some form of arthritis, RA, gout, lupus, or fibromyalgia?	8 (88.9)	5 (41.7)
Have you ever been diagnosed with hypertension?	7 (77.8)	6 (50.0)
Have you ever been diagnosed with diabetes mellitus?	5 (55.6)	11 (91.7)
Have you ever been diagnosed with obesity?	7 (77.8)	6 (50.0)

* Values are the number (%). RA = rheumatoid arthritis.

internists, 2 family practitioners, 1 physician assistant, 1 orthopedic surgeon, and 7 nurse practitioners. All but 1 reported that they serve Latina or African American women in their practice. The providers had been practicing medicine from 18 months to 30 years.

Focus group guides. *Patients*. The participants were asked a series of questions about their knee pain, the types of health care professionals they had seen for their pain, prescription medication usage, and what they had done as their knee pain worsened. We asked about the impact of knee pain on their daily lives, i.e., their ability to work, do household chores, engage in self-care, and participate in recreational activities. We presented the decision tool inputs and outputs to the participants (see Supplementary Appendix A, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23903/abstract). The decision tool shows a menu of 11 individual treatments and potential treatment pathways. The tool provides information about estimated quality of life improvement, treatment costs, and productivity loss associated with 3 different treatment pathways. The patient focus group guide and tool were developed in English and translated into Spanish and back translated to ensure the consistency across the 2 groups of women.

Providers. Health care providers were queried about the patients they see who experience knee pain, how common severe knee pain was among their patients, how they treated these patients, and at what point they referred their patients to a specialist. We explained each section of the decision tool and asked if and how they would use this information to discuss treatment options with their patients.

The decision tool. We highlighted the tool inputs, which include demographic details, health and health insurance information, and treatment options, and reviewed tool outputs, which include total cost of treatment, out-of-pocket expenses, pain level, quality of life, and lost income.

Data analysis. Interview transcriptions were deidentified and verified for accuracy. Using a deductive approach, each question from the focus group guide was translated into a series of predetermined codes designed to compare and contrast responses from participants. General questions (e.g., seeing a professional about their knee pain or not) were categorized as "yes" or "no." Probing questions that were meant to obtain more detail (e.g., a professional's treatment for pain) were initially coded as responses that addressed their respective topics, which was used to generate a thematic report. Using this report, each probing question was analyzed individually for commonalities that supported

final categorization. An initial coder coded each transcript, which was subsequently reviewed by a second coder. If there were any discrepancies, a third coder would provide a final review. An experienced qualitative researcher (ET) participated in codebook development, piloting, and analysis. Before the codebook was applied, it was piloted twice. The transcripts were analyzed using ATLAS.ti qualitative software to apply the constant comparative method (22). Specifically, each question was analyzed individually for commonalities that allowed for categorization based on the codebook. These categories were examined for common themes.

RESULTS

Patient experiences of pain. For most African American participants, pain began slowly and gradually worsened over the course of years; for a few, however, it happened quickly due to an activity or accident. Several Latina participants described having moderate-to-severe knee pain that lasted for years and worsened depending on their activities, while 4 Latina participants described having mild pain. Women from both groups found themselves unable to do everyday activities such as walking or climbing up and down flights of stairs. One African American woman stated, "I find that I don't use the stairs nearly as much as I used to. Even if it's just going a flight up I'll hop on the elevator, obviously, if there's an elevator there." Standing for long periods of time proved to be difficult for participants whose jobs required them to do so.

Family played a significant role in the women's concern about knee pain. Participants in the Latina focus group described how important it was to be able to spend time with their families and not have to constantly worry about treating their knee pain. One participant reported, "For me, being with my granddaughters. Being with my family, my husband, my children, and my grandchildren. For me that is important."

For some African American participants, knee pain had an effect on their family members, as they described how they were unable to keep up with family members and often relied on them for different types of support, including financial assistance. One participant did not want to strain her family: "Well, I have insurance that I know will pay, and then my children if I got sick they'll pay, but I don't want them to pay because they all have their own families. And I've suffered with it this long so I don't want to take anything from them."

Seeking treatment for knee pain. All participants discussed their pain with a health care provider and were typically referred to an orthopedic specialist. Providers of African American participants did diagnostic procedures, administered cortisone shots, provided prescriptions, and often recommended surgery. In order to alleviate pain outside the clinical setting, African American participants discussed using heating pads, knee braces, medication, and ice. These techniques alleviated enough pain for some participants, while others either decided to have surgery or dealt with their pain. Participants who opposed surgery were vehemently against any type of what they called "cutting."

In order to alleviate pain, Latina participants described taking medications, using sprays, and elevating their legs or sleeping with a pillow between them. Latina participants were typically referred to an orthopedist who obtained radiographs, provided cortisone shots, physical therapy, and referred them to surgery, if necessary. Most Latina participants mentioned a lack of information from providers. If pain continued to worsen, participants discussed additional diagnostic testing and surgical options.

Patient considerations of cost of care. When asked to discuss cost of care information, 2 African American participants said it would be very helpful for someone who did not have insurance or had limited income. While pain levels and conversations with their providers primarily drove their decisions, participants did see value in having information about cost and quality of life and described the information as beneficial. One participant noted that while she chose to treat her pain aggressively, other patients may want a more measured approach: "You know, when you see this, if this had been presented to me as a patient early on, I probably would have still followed the same path of treatment there because I was very aggressive in trying to make sure. But I think if I had been a person who had been a little scared of surgery or something like that, seeing something like this and charts like this would have helped me with my decision-making to realize that in the long run, I'm going to pay with days missed from work, dollars to dollars."

Latina participants found the information to be helpful and described it as an opportunity to learn: "I like it because you can learn. It makes you think of what can be done because with this condition… Having someone explain it to you, what you can do… Having this condition, arthritis, the doctor says that it won't go away, that it is for life, there is no improvement."

Some participants found the cost information to be confusing because the estimates seemed too low, noting that costs in the real world would likely be much higher. The tool was still seen as useful, however, as participants understood that the information was an example. One participant commented: "I agree that this does not reflect the numbers that we are paying for medical, especially when we don't have assistance to health insurance. We pay a lot for medicines and visits to the doctor and all that. But I think that what you have tried to do is to give us an idea of what more or less they are trying to say. I think that it would be important to increase the costs a little or the numbers that are on this table."

The impact of knee pain on work. Participants acknowledged that knee pain affects their ability to work. One participant indicated that her knee pain caused her to retire early: "I retired from the state of Maryland. I was a nurse, LPN. And I worked in rehab because of my knees. And I said, I came out on early retirement because I didn't want to come out using a cane being

unable to move or function." Another worried about how knee pain impacted her ability to work and earn a living: "But the problem I have is, if I do not work, who will pay my bills? I pay $1,300 for my apartment, together with my brother. There is 2 of us. My part is $645. I pay my car, which is $500, and insurance and all that. And if I do not work… even if I have to drag my little leg, I'm going to work! Where am I going to [get] income for my household? And my medicines that are expensive!"

Treatment of knee pain by providers. There was variation in the frequency with which providers treated patients with knee pain, ranging from 10% to "very common." Providers discussed using scales to assess pain and activity levels, although descriptive assessments were also used, as patients may be able to relate more to an example than a number. In order to treat knee pain, providers described over-the-counter and prescription painkillers, physical therapy, and nonsteroidal antiinflammatory drugs. Referring a patient to an orthopedic surgeon was described as the last step.

Utility of the decision tool for providers. Some providers needed clarification as to whether higher or lower numbers represented better or worse pain and activity levels. Regarding usability of the tool with patients, providers discussed how it could provide valuable information to help patients make short- and long-term treatment decisions. Lifetime costs of different treatments, potential lost income, and quality of life were seen as particularly useful information.

Providers indicated that the tool might cause confusion between clinicians and patients because patients may not understand the complex information during a brief provider visit. Physicians explained that they ignore cost when deciding what the next steps for a patient might be: "But just to say that we have excellent access, so I wouldn't not order [magnetic resonance imaging] because of cost, and I wouldn't not refer for [physical therapy] because of cost, just because of the way our clinic is set up, which I know is unusual for patients who don't have insurance."

Provider suggestions for the decision tool. Providers suggested that the information be simplified and offered to patients during a class before a clinical visit. Providers agreed that patients would benefit from information about quality-adjusted additional years of life and lifetime cost but suggested simplifying the terms.

Influence of comorbidities. All but 1 of the African American participants and all the Latina participants had at least 1 of the following 3 comorbidities: hypertension, diabetes mellitus, or obesity/overweight (Table 1). One primary care provider explained that his priority would be to treat hypertension and treating knee pain might be an "afterthought."

DISCUSSION

Knee pain is an early harbinger of OA (23). Although the pathogenesis of knee OA is not clearly established, different interventions have been shown to be effective in interrupting the natural history of this disease (23–25). Best practices recommend nonsurgical management of knee OA, including biomechanical interventions, weight loss, healthy eating habits, physical activity, strength training, intraarticular corticosteroids, self-management of pain, education, and orthoses (17). The effectiveness of these treatment modalities hinges on the engagement and compliance of the patient with their own treatment. Shared decision-making is a first step in this engagement (26–31). This communication can be enhanced with the use of a shared decision-making tool (18–21). Our goal is to develop a shared decision-making tool for knee OA that provides a list of treatment alternatives, is interactive by allowing the information to be personalized to the specific patient, and includes comparative outcome measures of pain, function, and financial impact of lost productivity.

Latina and African American women who have been treated for knee pain and health care providers who serve them indicate that information regarding cost of treatment, productivity loss associated with different treatment options, and health outcomes would inform their treatment decisions. The women involved in our focus groups expressed a variety of factors that they would take into account when making decisions about treatment, such as pain, the ability to participate in daily activities, the burden on family members, and cost of care. While many of these women sought help for their varying levels of knee pain, lack of knowledge and information likely played a large role in their treatment decisions. Notably, Latina women reported a lack of information provided to them. Patients, providers, and health purchasers recognize that there is a role for decision and communication aides in orthopedic practice (27,28). Shared decision-making can improve patients' satisfaction with the care they receive (32).

Many women expressed support for the decision tool in that it provided them with greater knowledge, noting that it was nice to see all the treatment options available. Both African American and Latina women felt that the information was beneficial by providing more transparency about economic feasibility. Providers indicated that the decision-making tool would be useful for their patients in making both short- and long-term decisions, particularly the information about lifetime costs, potential lost income, and quality of life. While some patients and providers discussed the complexity of the cost information, the different groups indicated overall that it would be valuable information.

An important strength of this study includes the active engagement of patients in the development of a decision-making tool. Specifically, by seeking the input of the intended audience, it informed translation of traditional health economic output to patient-friendly information. Additionally, we learned that patients would find the financial implications of their treatment decisions

helpful. Further, the providers offered insight into how they care for these particular patient groups given unique clinical challenges. The information provided to patients included results from a previously developed Markov model based on economic and epidemiologic studies in the literature. While the tool does not provide an exact prediction of any single patient experience, it shows the most likely statistical outcome across similar populations.

Some limitations to our study should be considered when interpreting the results. The women who participated in our focus groups were primarily from the Baltimore–Washington metropolitan area and may not be representative of the views of Latina and African American and women across the US. Given geographic variation in access to and delivery of health care, these differences may be enhanced. Further, the African American participants were older than the Latina women that participated in our focus groups. While we recognize this and other differences between the 2 patient groups and the providers that care for them, our intent was not to conduct a comparison of these groups but to understand patient perspective for both underserved patient populations. Certain health economic outputs, when presented in patient-friendly terms, are helpful and should be provided for patient-centered, shared decision-making. Despite these limitations, our qualitative approach to this study allowed us to capture the patient experiences and preferences in a way that other quantitative methods would not provide.

In conclusion, the increased complexity of chronic conditions combined with a debilitating disease such as knee OA necessitates increased patient engagement and, at a minimum, that they understand that the health care choices they make today, or delay making, will impact their health and productivity. Further, since the disease presents more frequently in minority women, it is imperative that the clinical community provide a tool that supports a robust conversation between provider and patient to optimize treatment. The model-based decision-making tool we presented during our focus group sessions has the potential to provide useful information about the effects on quality of life and the value of various knee pain treatments. The key to effective communication will be translating the terminology and information into clear, relatable terms and illustrations. Our findings demonstrate that there are multiple factors that contribute to a woman's decision to pursue a specific type of treatment for her knee pain, including the ability to do daily activities. In terms of costs, patients were intrigued by the concept of lost productivity and lost income. The patients from these focus groups had insurance coverage and referred to that frequently, so direct medical costs were not an evident concern in the discussion. Although cost is just 1 of many factors to consider when determining a treatment plan, our discussions with both patients and physicians demonstrate that patients can make more informed decisions with information about the estimated costs of care and potential effects on productivity. With the innovative Markov-based decision-making tool, older African American and Latina women can be empowered with integral information to make important decisions about their knee pain 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. Gaskin 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. Gaskin, Karmarkar, Maurer, Jones, Thorpe, Wood.

Acquisition of data. Gaskin, Karmarkar, Maurer, Bucay-Harari, Casillas, Gittens, Jones, Wood.

Analysis and interpretation of data. Gaskin, Karmarkar, Maurer, Gittens, Jones, Tolbert.

ROLE OF THE STUDY SPONSOR

Zimmer Biomet had no role in the study design or in the collection, analysis, or interpretation of the data, the writing of the manuscript, or the decision to submit the manuscript for publication. Publication of this article was not contingent upon approval by Zimmer Biomet.

REFERENCES

1. Deshpande BR, Katz JN, Solomon DH, Yelin EH, Hunter DJ, Messier SP, et al. Number of persons with symptomatic knee osteoarthritis in the US: impact of race and ethnicity, age, sex, and obesity. Arthritis Care Res (Hoboken) 2016;68:1743–50.
2. Vina ER, Kwoh CK. Epidemiology of osteoarthritis: literature update. Curr Opin Rheumatol 2018;30:160–7.
3. Bozic KJ, Belkora J, Chan V, Youm J, Zhou T, Dupaix J, et al. Shared decision making in patients with osteoarthritis of the hip and knee: results of a randomized controlled trial. J Bone Joint Surg Am 2013;95:1633–9.
4. Losina E, Walensky RP, Reichmann WM, Holt HL, Gerlovin H, Solomon DH, et al. Impact of obesity and knee osteoarthritis on morbidity and mortality in older Americans. Ann Intern Med 2011;154:217–26.
5. MacKay C, Jaglal SB, Sale J, Badley EM, Davis AM. A qualitative study of the consequences of knee symptoms: "It's like you're an athlete and you go to a couch potato." BMJ Open 2014;4:e006006.
6. Neogi T. The epidemiology and impact of pain in osteoarthritis. Osteoarthritis Cartilage 2013;21:1145–53.
7. Fitzgerald GK, White DK, Piva SR. Associations for change in physical and psychological factors and treatment response following exercise in knee osteoarthritis: an exploratory study. Arthritis Care Res (Hoboken) 2012;64:1673–80.
8. Johnson VL, Hunter DJ. The epidemiology of osteoarthritis. Best Pract Res Clin Rheumatol 2014;28:5–15.
9. Batsis JA, Zbehlik AJ, Barre LK, Bynum J, Pidgeon D, Bartels SJ. Impact of obesity on disability, function, and physical activity: data from the osteoarthritis initiative. Scand J Rheumatol 2015;1–8.
10. Centers for Disease Control and Prevention. Obesity trends in adults with arthritis. URL: http://www.cdc.gov/arthritis/resources/spotlights/obesity-trends.htm.
11. National Institute of Diabetes and Digestive and Kidney Diseases. Overweight and obesity statistics. URL: https://www.niddk.nih.gov/health-information/health-statistics/overweight-obesity.
12. Gaskin DJ, Richard P. The economic costs of pain in the United States. J Pain 2012;13:715–24.
13. Le TK, Montejano LB, Cao Z, Zhao Y, Ang D. Health care costs in US patients with and without a diagnosis of osteoarthritis. J Pain Res 2012;5:23–30.

14. Kotlarz H, Gunnarsson CL, Fang H, Rizzo JA. Insurer and out-of-pocket costs of osteoarthritis in the US: evidence from national survey data. Arthritis Rheum 2009;60:3546–53.
15. Bedard NA, Dowdle SB, Anthony CA, DeMik DE, McHugh MA, Bozic KJ, et al. The AAHKS Clinical Research Award: what are the costs of knee osteoarthritis in the year prior to total knee arthroplasty? J Arthroplasty 2017;32:S8–S10.
16. Karmarkar T, Maurer A, Parks ML, Mason T, Bejinez-Eastman A, Harrington M, et al. A fresh perspective on a familiar problem: examining disparities in knee osteoarthritis using a Markov model. Med Care 2017;55:993–1000.
17. McAlindon TE, Bannuru RR, Sullivan MC, Arden NK, Berenbaum F, Bierma-Zeinstra SM, et al. OARSI guidelines for the non-surgical management of knee osteoarthritis. Osteoarthritis Cartilage 2014;22:363–88.
18. Marrin K, Wood F, Firth J, Kinsey K, Edwards A, Brain KE, et al. Option grids to facilitate shared decision making for patients with osteoarthritis of the knee: protocol for a single site, efficacy trial. BMC Health Serv Res 2014;14:160.
19. Laba T, Brien J, Fransen M, Jan S. Patient preference for adherence to treatment for osteoarthritis: the Medication Decisions in Osteoarthritis Study (MEDOS). BMC Muskuloskelet Disord 2013;14:160.
20. Shue J, Karia RJ, Cardone D, Samuels J, Shah M, Slover JD. A randomized controlled trial of two distinct shared decision-making aids for hip and knee osteoarthritis in an ethnically diverse patient population. Value Health 2016;19:487–93.
21. Bozic KJ, Chenok KE, Schindel J, Chan V, Huddleston JI III, Braddock C III, et al. Patient, surgeon, and healthcare purchaser views on the use of decision and communication aids in orthopaedic surgery: a mixed methods study. BMC Health Serv Res 2014;14:366.
22. Glaser B, Strauss A. The discovery of grounded theory. London: Weidenfeld and Nicholson; 1967. p. 24, 288–304.
23. Owens C, Conaghan PG. Improving joint pain and function in osteoarthritis. Practitioner 2016;260:17–20.
24. Marley J, Tully MA, Porter-Armstrong A, Bunting B, O'Hanlon J, Atkins L, et al. The effectiveness of interventions aimed at increasing physical activity in adults with persistent musculoskeletal pain: a systematic review and meta-analysis. BMC Musculoskelet Disord 2017;18:482.
25. Manheim LM, Dunlop D, Song J, Semanik P, Lee J, Chang RW. Relationship between physical activity and health-related utility among knee osteoarthritis patients. Arthritis Care Res (Hoboken) 2012;64:1094–8.
26. McClellan FM, Wood JE Jr, Fahmy SM, Jones LC. Musculoskeletal health disparities: health literacy, cultural competency, informed consent, and shared decision making. J Long Term Eff Med Implants 2014;24:195–204.
27. Youm J, Chenok KE, Belkora J, Chiu V, Bozic KJ. The emerging case for shared decision making in orthopaedics. Instr Course Lect 2013;62:587–94.
28. Fraenkel L, Rabidou N, Wittink D, Fried T. Improving informed decision-making for patients with knee pain. J Rheumatol 2007;34:1894–8.
29. Page AE. Safety in surgery: the role of shared decision-making. Patient Saf Surg 2015;9:24.
30. Suarez-Almazor ME, Richardson M, Kroll TL, Sharf BF. A qualitative analysis of decision-making for total knee replacement in patients with osteoarthritis. J Clin Rheumatol 2010;16:158.
31. Carmona-Teres V, Moix-Queralto J, Pujol-Ribera E, Lumillo-Gutierrez I, Mas X, Batlle-Gualda E, et al. Understanding knee osteoarthritis from the patients' perspective: a qualitative study. BMC Musculoskelet Disord 2017;18:225.
32. Altin SV, Stock S. The impact of health literacy, patient-centered communication and shared decision-making on patients' satisfaction with care received in German primary care practices. BMC Health Serv Res 2016;16:450.

Arthritis Care & Research
Vol. 72, No. 5, May 2020, pp 699–704
DOI 10.1002/acr.23897

American College of Rheumatology
Empowering Rheumatology Professionals

Determinants of Positive Temporal Artery Biopsies in the Veterans Health Administration National Database Cohort

Sarah H. Chung,[1] Meredith B. Morcos,[1] and Bernard Ng[2]

Objective. This study sought to determine the effect of temporal artery biopsy (TAB) postfixation length, laterality, age, and prior prednisone exposure on TAB positivity utilizing the Veterans Health Administration national database.

Methods. Subjects with procedure code for TAB between 1999 and 2017 were queried, and pathology reports were reviewed manually. Demographic, laboratory, and prescription data were extracted. Multivariate analyses and logistic regression were run using Stata, version 13.0.

Results. A total of 3,057 pathology reports were reviewed; 306 biopsies (10%) were designated positive. The likelihood of a positive TAB significantly correlated with TAB postfixation length of >3.0 cm (odds ratio [OR] 1.58 [95% confidence interval (95% CI) 1.06, 2.36], $P < 0.05$) as well as with bilateral biopsy in 1 sitting (OR 1.83 [95% CI 1.29, 2.59], $P < 0.01$). Positive TAB also significantly correlated with age >71 years. Prednisone administration up to and beyond 42 days prior to TAB did not influence TAB result.

Conclusion. This retrospective study examined predictors of TAB positivity and utilized national data collected on US veterans over the span of 18 years. The results suggest consideration of pursuing initial bilateral TAB or achieving a TAB postfixation length of at least 3 cm to improve yield. The results also agree with prior studies showing that pre-TAB steroid exposure does not appear to affect yield even up to and beyond 42 days prior to biopsy.

INTRODUCTION

The diagnosis of giant cell arteritis (GCA), a granulomatous large-vessel vasculitis, is often still elusive even with improved understanding of its pathophysiology and advancements in treatment approaches. Temporal artery biopsy (TAB) remains the gold-standard diagnostic tool despite its poor sensitivity due to skip inflammatory lesions inherent to the disease (1,21). Regardless of the longstanding practice of obtaining a TAB to achieve this diagnosis, there is still no definitive consensus regarding the specific issue of optimal TAB length that should be obtained by surgeons. A 1994 study by Achkar et al described the standardized approach at the Mayo Clinic, where the practice was to pursue a specimen of 3–4 cm. If the specimen was negative on frozen sections at multiple levels, then surgeons pursued biopsy of the contralateral vessel (2). Using this method, positive TAB results were found in 33% of referred patients, a higher yield than that of subsequent reported studies (see Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23897/abstract). In real-life practice, however, this level of yield is rarely achieved, likely due to the heterogeneity among institutions regarding which specialty performs these procedures. Length of specimen determined by the surgeon is a decision influenced by multiple factors, including patient preference, surgeon experience, intraoperative issues, guidance from the referring provider, and more (3–6). Furthermore, since the Mayo Clinic publication (2), there have been numerous subsequent retrospective studies that negate the "bigger is better" concept, but these studies have been limited by small sample sizes and narrow practice settings (see Supplementary Table 1).

Another area of historic concern regarding the diagnostic yield of TAB has been pre-TAB exposure to glucocorticoids. High-dose glucocorticoids are often promptly initiated when GCA is clinically suspected, before a TAB sample is obtained, to avoid ischemic complications, which are the largest source of morbidity and mortality in this condition. The majority of studies have found that pre-TAB glucocorticoid exposure does not decrease yield; however, there is still disagreement with regard to the length of treatment, which may affect the results of biopsy (7–9). This is of particular importance because the use of diagnostic ultrasound has become

[1]Sarah H. Chung, MD, Meredith B. Morcos, MD: University of Washington Medical Center, Seattle; [2]Bernard Ng, MBBS, MSc, MMed, FRCP, FRCP, FACP, FACR, FAAAAI: University of Washington Medical Center and Veterans Health Administration Puget Sound, Seattle, Washington.

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

Address correspondence to Sarah H. Chung, MD, University of Washington Medical Center, 1959 NE Pacific Street, Box 356428, Seattle, WA 98195. E-mail: sarahhc@uw.edu.

Submitted for publication December 11, 2018; accepted in revised form March 26, 2019.

SIGNIFICANCE & INNOVATIONS

- This study utilizes a national database containing data on US veterans and demonstrates that longer-length temporal artery biopsy (TAB) is more likely to yield a positive result.
- A higher rate of positive TAB results was observed in those who underwent initial bilateral TAB compared to unilateral TAB.
- The initiation of prednisone at a dosage of ≥30 mg per day beyond 42 days prior to TAB did not affect the yield of the TAB.

more widespread yet is limited by the brisk disappearance of the halo sign and other characteristic findings soon after initiation of steroids (10).

Currently, there are no published recommendations from the American College of Rheumatology regarding optimal TAB length. The 2009 European League Against Rheumatism (EULAR) and 2010 British Society for Rheumatology/British Health Professionals in Rheumatology guidelines on GCA management suggest pursuing a TAB length of at least 1 cm (11,12). The EULAR guidelines further provide a grade C recommendation against bilateral biopsies, citing that this may not add significantly to diagnostic yield and additionally commenting that the TAB should not be delayed beyond 1–2 weeks of commencing glucocorticoid therapy.

This study aimed to examine objective predictors of TAB positivity among US veterans, an elderly and largely male population, using a national database collecting pathology reports, laboratory data, and prescription information over the span of 18 years. Specifically, this study sought to determine the association of TAB positivity with postfixation TAB length, laterality of biopsy, age, and prior prednisone exposure.

MATERIALS AND METHODS

Subjects and data collection. This study was approved by the Veterans Health Administration (VHA) and the University of Washington institutional review boards. Subjects with a procedure code for TAB between 1999 and 2017 were queried through the VHA national database. TAB pathology reports were reviewed manually and designated positive or negative based on the pathologist's description of findings. Interrater reliability was calculated from a blinded, 2-researcher manual calculation using a random sample of 50 cases and had near perfect agreement, determined using Cohen's kappa statistics (agreement regarding TAB final diagnosis $\kappa = 0.8169$; agreement regarding TAB length $\kappa = 0.934$). The following data were extracted for multivariate analysis: 1) postfixation TAB length, 2) laterality, 3) whether bilateral TAB was performed in 1 sitting, 4) age at TAB, 5) sex, 6) self-reported ethnicity. Postfixation TAB length categories were organized as follows: 1) <10 mm, 2) 10 to <15 mm, 3) 15 to <20 mm, 4) 20 to <25 mm, 5) 25 to <30 mm, 6) ≥30 mm. Additionally, if available, prescription data regarding oral prednisone were collected for each subject, including dispense date, mg dispensed, and prescription directions for patients. Data on other forms of glucocorticoids were not extracted for this study. The total prednisone dosage prior to the TAB date was calculated using a generated algorithm validated against results from a blinded, 2-researcher manual calculation using a random sample of 50 cases. Interrater reliability had near perfect agreement, determined using Cohen's kappa statistics ($\kappa = 0.897$).

Study variables and clinical definitions. The dependent variable for multivariate logistic regression was a positive TAB. Independent variables included age, sex, postfixation TAB length, TAB laterality, and treatment with high-dose prednisone.

Definition of positive biopsy. A positive TAB result was defined as the presence of medial inflammatory infiltrate with mononuclear or granulomatous features in the intima or media layers of the artery and/or positive TAB result designated in the final impression by the pathologist. Specimens that were not artery were excluded; those that involved inflammation of the vasa vasorum were deemed negative. Evidence of a fragmented internal elastic lamina alone was deemed negative. Indeterminate results (i.e., inconclusive, healed arteritis) were also categorized as negative.

Definition of high-dose prednisone treatment. Subjects who had received a total daily dosage of prednisone of ≥30 mg per day were included in the investigation and categorized into 5 groups based on the prednisone dispense date relative to the TAB date: 1) 0–14 days prior to TAB, 2) >14–28 days prior to TAB, 3) >28–42 days prior to TAB, 4) >42 days prior to TAB, and 5) first dispensed after TAB. Subjects who were treated with prednisone at a dosage of <30 mg per day were excluded from the analysis.

Statistical analyses. Univariate analyses were performed between positive TAB result and age, sex, ethnicity, postfixation TAB length, and laterality to determine the independent variables for the multivariate regression. Multivariate logistic regression was subsequently performed using TAB result as the outcome variable, adjusting for age, postfixation TAB length, laterality, and pre-TAB prednisone exposure. A sensitivity analysis was performed using the same multivariate logistic regression equation. This analysis excluded indeterminate results (i.e., inconclusive, healed arteritis, fragmentation of the internal elastic lamina alone, inflammation of the vasa vasorum, etc.) rather than categorizing these reports as negative.

RESULTS

A total of 3,057 TAB pathology reports were reviewed. A total of 306 biopsies (10%) were designated positive per pathology report. Only 93 pathology reports (3%) specified the presence of multinucleated giant cells and/or granulomatous inflammation (see

Supplementary Table 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23897/abstract).

Demographics: sex, ethnicity, and age. Of the 11,984 subjects in the VHA database with a GCA diagnosis code, 94% (11,257) were men and 6% (727) were women (Table 1). Nearly one-half of the TAB cohort either did not report their ethnicity (44%) or did not know their ethnicity (7%) (Table 2). Of those who had data on self-reported ethnicity, the majority were white (n = 1,129; 37%), followed by African American (n = 287; 9%). Asian/Pacific Islanders (16%) and Hispanic/Latinos (13%) demonstrated higher frequency of positive TAB results compared to other self-reported ethnicity groups.

The majority (36%) of subjects in the TAB cohort were 61–70 years of age (Table 3). When compared to a reference category of 61–70 years of age, age >71 years was significantly associated with the odds of TAB positivity (odds ratio [OR] 1.87, $P < 0.05$ if age 71–80 years; OR 2.80, $P < 0.05$ if age >80 years). These results were unchanged in the sensitivity analysis (see Supplementary Table 3, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23897/abstract).

Postfixation length of biopsy. Of the 3,057 biopsies reviewed, 3,043 reported postfixation specimen length, and 2,983 reported laterality (right, left, bilateral). The likelihood of a positive TAB result was significantly associated with postfixation TAB length of ≥30 mm (OR 1.58 [95% confidence interval (95% CI) 1.06, 2.36], $P < 0.05$) (Table 4) when compared to a reference category of <10 mm, as well as with bilateral biopsy performed in 1 sitting (OR 1.83 [95% CI 1.29, 2.59], $P < 0.05$) (Table 5) when compared to unilateral biopsy. An additional postfixation TAB length category (15 to <20 mm group; OR 1.70 [95% CI 1.01, 2.06], $P < 0.05$) also reached significance in the sensitivity analysis (see Supplementary Table 3 at http://onlinelibrary.wiley.com/doi/10.1002/acr.23897/abstract). The association between initial bilateral biopsy and positive TAB result was unchanged in the sensitivity analysis (see Supplementary Table 3 at http://onlinelibrary.wiley.com/doi/10.1002/acr.23897/abstract). The likelihood of TAB positivity increased incrementally with longer biopsies. Neither sex nor age influenced TAB length or the likelihood that a bilateral TAB was performed.

Table 1. Characteristics of the study subjects*

Characteristic	Values
Total GCA diagnosis (ICD-9, ICD-10)	11,984
TAB performed	3,057 (26)
Prednisone ≥30 mg per day, dispensed among total cohort	6,161 (51)
Prednisone ≥30 mg per day, dispensed among TAB cohort	2,012 (66)
Mean age, years	72.5
Median age, years	73.1
Male sex	11,257 (94)

* Values are the number (%) unless indicated otherwise. GCA = giant cell arteritis; ICD-9 = International Classification of Diseases, Ninth Revision; ICD-10 = International Statistical Classification of Diseases and Related Health Problems, Tenth Revision; TAB = temporal artery biopsy.

Table 2. Self-reported ethnicity among TAB cohort*

Self-reported ethnicity	No. TABs performed	Positive TAB result	Negative TAB result
Missing	1,337	145 (11)	1,192 (89)
Unknown	206	13 (6)	193 (94)
White	1,129	116 (10)	1,013 (90)
African American	287	19 (7)	268 (93)
Hispanic/Latino	76	10 (13)	66 (87)
Asian/Pacific Islander	19	3 (16)	16 (84)
American Indian	3	0 (0)	3 (100)

* Values are the number (%) unless indicated otherwise. TAB = temporal artery biopsy.

Pre-TAB prednisone exposure. Of the 3,057 biopsies reviewed, 2,012 subjects were prescribed prednisone at a dosage of ≥30 mg per day at some point during a range of 90 days prior to and 90 days post-TAB. The majority (73%) of these subjects were prescribed prednisone 0–14 days prior to TAB (Table 6). The duration of treatment with prednisone prior to biopsy was not influenced by age, postfixation TAB length, or laterality. There was a negative association on univariate analysis with prednisone initiation and male sex (coefficient –2.97 [95% CI –5.6, 0.36], $P < 0.05$). After adjustment for sex, there was no significant association between time to prednisone initiation up to 42 days prior to TAB and either positive TAB result or presence of giant cells if described in the pathology report (Table 6 and Supplementary Table 2, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.23897/abstract). Positive TAB was significantly associated with the first prednisone treatment after the TAB date. These results were unchanged in the sensitivity analysis (see Supplementary Table 3, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.23897/abstract).

DISCUSSION

This retrospective study utilizing national data from the US veteran population reveals an increased yield of positive TAB results with longer biopsies in a literature landscape of discrepant findings regarding this issue (see Supplementary Table 1, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.23897/abstract). Our results found an OR of 1.8 for obtaining a positive TAB result with an initial bilateral biopsy when compared to unilateral biopsy. The positive biopsy rate was 17% in the bilateral biopsy group compared to 9% in the unilateral biopsy group, and of the 238 bilateral biopsies, there was a 94.5% concordance rate. Durling et al examined a cohort of 250 initial bilateral biopsy patients, achieving a postfixation length of at least 1 cm, and found a discordance rate of 17% among the TAB-positive group and a

Table 3. Multivariate logistic regression of age on temporal artery biopsy (TAB) positivity, adjusting for postfixation TAB length and laterality*

Age at TAB, years	No. (%)†	No. (%) TAB positive‡	OR	95% CI	*P*
<50	23 (1)	1 (4)	0.55	0.73, 4.14	0.561
51–60	307 (10)	10 (3)	0.58	0.33, 1.02	0.061
61–70	1,107 (36)	66 (6)	Ref.	–	–
71–80	894 (29)	110 (12)	1.87	1.41, 2.50	0.000§
>80	726 (24)	119 (16)	2.80	2.11, 3.72	0.000§

* Age as a categorical variable produces the reported odds ratio (OR) when compared to the reference category (61–70 years). 95% CI = 95% confidence interval; ref. = reference.
† Percentages use a denominator of 3,057 (TAB performed).
‡ Percentages use the denominator of the total number in the age category.
§ Significant.

discordance rate of 27% between localization of symptoms and laterality of positive TAB (13). These findings are not insignificant when considering the consequences of the diagnosis and the justification required for long-term immunosuppressive therapy, particularly in intermediate probability cases. Logistically, there may be benefits to simultaneous initial biopsy compared to sequential biopsy, including the use of the same instruments and procedure equipment and less travel time for the patient (13). There are no studies published on the rate of morbidities related to the TAB procedure, but according to various publications, the complications involved in this procedure appear to be limited to local irritation or factors related to not obtaining the artery (i.e., neuralgia, venous specimen, etc.) (3,13–15).

This study reveals the heterogeneous approach to obtaining TAB length, presumably determined by surgeon discretion, as evidenced by the roughly equal distribution among postfixation biopsy-length categories. This raises the compelling question of whether to provide a standardized approach for surgeons who perform these procedures. The consideration of specimen shrinkage with both frozen section and formalin fixation must be underscored. A 2012 study examining 62 biopsies found a mean shrinkage length of 4.61 mm with a 2.97-mm SD (16). Thus, achieving a postfixation TAB length of 3.0 cm may mean obtaining a specimen of 3.5–4.0 cm intraoperatively in an effort to increase yield, particularly in cases of indeterminate probability.

The initiation of moderate-to-high–dose prednisone did not appear to affect the yield of TAB even when started beyond 42 days prior to the TAB. This observation aligns well with a recent prospective study by Maleszewski et al that examined serial TABs during the first year of therapy for biopsy-proven GCA (17). Granulomatous infiltrates decreased slowly in a time-dependent fashion, with 71% of cases demonstrating persistent granulomatous inflammation at 3 months and still 25% of cases at 12 months despite brisk clinical improvement of symptoms. Our results affirm that clinicians should not be deterred in initiating glucocorticoid therapy prior to TAB and suggest that biopsy may still be valuable despite significant time receiving glucocorticoid therapy, in contrast to magnetic resonance imaging or ultrasound, modalities in which the sensitivity of GCA is reported to decline rapidly within the first week of glucocorticoid exposure (10).

Of those subjects who underwent TAB and were treated with prednisone at some point, a positive TAB result was only significantly associated with first prednisone treatment after the TAB date. While this observation could be due to the possibility that treatment with prednisone prior to TAB does reduce yield, the lack of a corresponding relationship between pre-TAB duration of treatment and TAB positivity rate renders this explanation equivocal. We suspect that this observation is more aptly explained by the possibility that a positive TAB result prompted the provider to initiate treatment.

Table 4. Multivariate logistic regression of postfixation temporal artery biopsy (TAB) length on TAB positivity, adjusting for age and laterality*

Postfixation TAB length	No. (%)†	No. (%) TAB positive‡	OR	95% CI	*P*
<10 mm	606 (20)	46 (8)	Ref.	–	–
10 to <15 mm	680 (22)	59 (9)	1.15	0.80, 1.65	0.465
15 to <20 mm	640 (21)	69 (11)	1.42	0.99, 2.03	0.053
20 to <25 mm	501 (17)	55 (11)	1.30	0.89, 1.90	0.179
25 to <30 mm	265 (9)	31 (9)	1.44	0.92, 2.24	0.106
≥30 mm	351 (12)	42 (12)	1.58	1.06, 2.36	0.026§

* Postfixation TAB length as a categorical variable produces the reported odds ratio (OR) when compared to the reference category (<10 mm). 95% CI = 95% confidence interval; ref. = reference.
† Percentages use a denominator of 3,043 (TAB with length reported).
‡ Percentages use the denominator of the total number in the postfixation TAB length category.
§ Significant.

Table 5. Multivariate logistic regression of laterality on temporal artery biopsy (TAB) positivity, adjusting for age and postfixation TAB length*

TAB laterality	No. (%)†	No. (%) TAB positive‡	OR	95% CI	*P*
Unilateral biopsy	2,746 (92)	254 (9)	Ref.	–	–
Bilateral biopsy	237 (8)	41 (17)	1.83	1.29, 2.59	0.000§

* Bilateral biopsy produces an odds ratio (OR) of 1.83 when compared to unilateral biopsy. 95% CI = 95% confidence interval; ref. = reference.
† Percentages use a denominator of 2,983 (TAB with laterality reported).
‡ Percentages use the denominator of the total number in the laterality category.
§ Significant.

Associations between GCA and ethnicity were unable to be performed, largely due to the fact that the majority of subjects in the cohort did not have this data available. Of those who did self-report ethnicity and who underwent TAB, the positivity rate was lower in the African American group (7%) compared to white participants (10%). Interestingly, the rates of positive TAB results among the Hispanic/Latino group (13%) and the Asian/Pacific Islander group (16%) were higher than among white participants. While the ethnicity distribution in this VHA cohort does not reflect that of the general population, these observations illuminate the noteworthy prevalence of GCA in other nonwhite groups.

Associations between TAB positivity and elevated erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP) level were also not included in this study because the analysis was deemed problematic. In the database, the majority of subjects who underwent TAB did not have an abnormal ESR or CRP level within 180 days pre- and post-TAB. Inflammatory markers are not only widely used as the initial screening test for GCA, but the sensitivity of abnormal markers is high; thus, the discrepancy was thought to be due to errors in the data set, i.e., subjects who received the initial evaluation at an outside non-VHA facility (18).

It is important to note several limitations to this study, the most significant of which are the poor sensitivity of TAB itself, with a false negative rate cited up to 25%, and the unknown prevalence of biopsy-negative GCA in this cohort (1,2,4,19). We also acknowledge the lack of clinical data to correlate the diagnosis of GCA, a condition that is characterized by symptomatology (i.e., headache, jaw claudication, visual symptoms) and response to immunosuppressive therapy. Due to the lack of access to clinical data, we were unable to examine TAB results in the context of other related diagnoses, namely large vessel vasculitis, Takayasu arteritis, and polymyalgia rheumatica, which could account for treatment with prednisone. Of 23 subjects under age 50 years who underwent TAB in this data set, 1 subject (age 31 years) had a positive TAB result, which was manually reviewed by all authors independently with agreement that the histopathologic description was compatible with GCA. It has been reported that GCA can indeed occur in patients younger than 50 years of age; however, we must acknowledge the possibility that this data set did not represent a pure GCA cohort (20). Case reports and small case series have described temporal artery involvement in non-GCA vasculitides, such as antineutrophil cytoplasmic antibody–associated vasculitis, polyarteritis nodosa, and cryoglobulinemia. And since we did not have access to chart data, we were unable to confirm the ultimate diagnoses after TAB (21). Owing to the large retrospective nature of this study, confirmation and agreement of the pathology report with a fixed specimen could not be achieved, and we were unable to confirm outside glucocorticoid sources beyond prescription services through the VHA system. Additionally, pathology reports were heterogeneous in the degree of description regarding biopsy specimens, which makes this study susceptible to information bias. Last, despite the large cohort size, the overall rate of positive TAB results in this study was low (10%). The low TAB positivity rate may be due to the largely male cohort (GCA has a higher prevalence in women) or might reflect the referral base that influences pretest probability of biopsy yield.

Table 6. Multivariate logistic regression of prednisone treatment duration on temporal artery biopsy (TAB) positivity, adjusting for age, sex, postfixation TAB length, and laterality*

Days relative to TAB date	No. (%)†	No. (%) TAB positive‡	OR	95% CI	*P*
Post-TAB	224 (11)	43 (19)	1.79	1.24, 2.58	0.002§
0–14 prior to TAB	1,474 (73)	173 (12)	Ref.	–	–
>14–28 prior to TAB	213 (11)	22 (10)	0.85	0.53, 1.37	0.513
>28–42 prior to TAB	51 (3)	4 (8)	0.65	0.23, 1.82	0.411
>42 prior to TAB	50 (3)	6 (12)	1.01	0.42, 2.41	0.983

* Duration (days) that prednisone was first dispensed for a diagnosis of giant cell arteritis is the categorical variable producing an odds ratio (OR) of positive TAB result when compared to the reference category (0–14 days). 95% CI = 95% confidence interval; ref. = reference.
† Percentages use a denominator of 2,012 (prednisone dispensed among TAB cohort).
‡ Percentages use the denominator of the total number in the prednisone treatment duration category.
§ Significant.

The current study, however, contributes to the literature by examining objective predictors of TAB positivity utilizing a large national database rather than case series, and our results correlate with the current understanding regarding the histopathology of this disease. Clinically, these results are relevant because they not only support the early initiation of prednisone in cases suspicious for GCA but also make a case for the persistent value of TAB, despite its limited sensitivity, in assisting in the diagnosis of GCA for patients who have already had significant exposure to glucocorticoids. Our results also suggest consideration of pursuing initial bilateral TAB or achieving a postfixation TAB length of at least 3 cm to improve yield.

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. Chung 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. Chung, Morcos, Ng.
Acquisition of data. Chung, Morcos, Ng.
Analysis and interpretation of data. Chung, Ng.

REFERENCES

1. Buttgereit F, Dejaco C, Matteson EL, Dasgupta B. Polymyalgia rheumatica and giant cell arteritis: a systematic review. JAMA 2016; 315:2442–58.
2. Achkar AA, Lie JT, Hunder GG, O'Fallon WM, Gabriel SE. How does previous corticosteroid treatment affect the biopsy findings in giant cell (temporal) arteritis? Ann Intern Med 1994;120:987–92.
3. Au CP, Sharma NS, McCluskey P, Ghabrial R. Increase in the length of superficial temporal artery biopsy over 14 years. Clin Exp Ophthalmol 2016;44:550–4.
4. Banz Y, Stone JH. Why do temporal arteries go wrong? Principles and pearls from a clinician and a pathologist. Rheumatology (Oxford) 2018;57 Suppl 2:ii3–10.
5. Grossman C, Ben-Zvi I, Barshack I, Bornstein G. Association between specimen length and diagnostic yield of temporal artery biopsy. Scand J Rheumatol 2017;46:222–5.
6. Mahr A, Saba M, Kambouchner M, Polivka M, Baudrimont M, Brochériou I, et al. Temporal artery biopsy for diagnosing giant cell arteritis: the longer, the better? Ann Rheum Dis 2006;65:826–8.
7. Cavazza A, Muratore F, Boiardi L, Restuccia G, Pipitone N, Pazzola G, et al. Inflamed temporal artery: histologic findings in 354 biopsies, with clinical correlation. Am J Surg Pathol 2014;38:1360–70.
8. Hocevar A, Rotar Z, Jese R, Semrl SS, Pizem J, Hawlina M, et al. Do early diagnosis and glucocorticoid treatment decrease the risk of permanent visual loss and early relapse in giant cell arteritis: a prospective longitudinal study. Medicine (Baltimore) 2016;95: e3210.
9. Narváez J, Bernad B, Roig-Vilaseca D, García-Gómez C, Gómez-Vaquero C, Juanola X, et al. Influence of previous corticosteroid therapy on temporal artery biopsy yield in giant cell arteritis. Semin Arthritis Rheum 2007;37:13–9.
10. Hauenstein C, Reinhard M, Geiger J, Markl M, Hetzel A, Treszl A, et al. Effects of early corticosteroid treatment of magnetic resonance imaging and ultrasound findings in giant cell arteritis. Rheumatology (Oxford) 2012;51:1999–2003.
11. Dasgupta B, Borg FA, Hassan N, Alexander L, Barraclough K, Bourke B, et al. BSR and BHPR guidelines for the management of giant cell arteritis. Rheumatology (Oxford) 2010;49:1594–7.
12. Mukhtyar C, Guillevin L, Cid MC, Dasgupta B, de Groot K, Gross W, et al. EULAR recommendations for the management of large vessel vasculitis. Ann Rheum Dis 2009;68:318–23.
13. Durling B, Toren A, Patel V, Gilberg S, Weis E, Jordan D. Incident of discordant temporal artery biopsy in the diagnosis of giant cell arteritis. Can J Ophthalmol 2014;49:157–61.
14. Boyev LR, Miller NR, Green WR. Efficacy of unilateral versus bilateral temporal artery biopsies for the diagnosis of giant cell arteritis. Am J Opthalmol 1999;128:211–5.
15. Danesh-Meyer HV, Savino PJ, Eagle RC Jr, Kubis KC, Sergott RC. Low diagnostic yield with second biopsies in suspected giant cell arteritis. J Neuroopthalmol 2000;20:213–5.
16. Murchison AP, Bilyk JR, Eagle RC Jr, Savino PJ. Shrinkage revisited: how long is long enough? Ophthalmic Plast Reconstr Surg 2012;28:261–3.
17. Maleszewski JJ, Younge BR, Fritzlen JT, Hunder GG, Goronzy JJ, Warrington KJ, et al. Clinical and pathological evolution of giant cell arteritis: a prospective study of follow-up temporal artery biopsies in 40 treated patients. Mod Pathol 2017;30:788–96.
18. Kermani TA, Schmidt J, Crowson CS, Ytterberg SR, Hunder GG, Matteson EL, et al. Utility of erythrocyte sedimentation rate and C-reactive protein for the diagnosis of giant cell arteritis. Semin Arthritis Rheum 2012;41:866–71.
19. Dejaco C, Brouwer E, Mason JC, Buttgereit F, Matteson EL, Dasgupta B. Giant cell arteritis and polymyalgia rheumatica: current challenges and opportunities. Nat Rev Rheumatol 2017;13: 578–92.
20. Albreiki D, Al Belushi F, Patel V, Farmer J. When a temporal artery biopsy reveals a diagnosis other than temporal arteritis: eosinophilic granulomatosis with polyangiitis. Can J Ophthalmol 2016;51: e108–9.
21. Hall S, Persellin S, Lie JT, O'Brien PC, Kurland LT, Hunder GG. The therapeutic impact of temporal artery biopsy. Lancet 1983;2: 1217–20.

Arthritis Care & Research
Vol. 72, No. 5, May 2020, pp 705–710
DOI 10.1002/acr.23334

American College of Rheumatology
Empowering Rheumatology Professionals

BRIEF REPORT

Joint Distribution and Two-Year Outcome in 347 Patients With Monoarthritis of Less Than Sixteen Weeks' Duration

Ellen Sauar Norli,[1] Gina Hetland Brinkmann,[2] Tore Kristian Kvien,[3] Olav Bjørneboe,[4] Anne Julsrud Haugen,[5] Halvor Nygaard,[6] Cathrine Thunem,[7] Elisabeth Lie,[3] and Maria Dahl Mjaavatten[3]

Objective. The present study was undertaken to investigate the joint distribution and 2-year outcome of patients with recent-onset monoarthritis.

Methods. Adult patients with clinically apparent monoarthritis of ≤16 weeks' duration were included in a multicenter 2-year longitudinal study. Clinical characteristics, joint distribution, development of chronic inflammatory rheumatic disease (CIRD), as well as classification criteria according to the American College of Rheumatology (ACR)/European League Against Rheumatism (EULAR) 2010 criteria for RA were studied. Predictors for development of CIRD were analyzed by multivariable logistic regression analyses.

Results. The knee (49.3%), ankle (16.7%), and wrist (14.1%) were the most frequently affected joints among the 347 included patients. A total of 91 patients (26.2%) developed CIRD during follow-up; 21 (6.1%) were diagnosed with RA, and 16 (4.6%) with psoriatic arthritis. Longer duration of joint swelling, joint localization, and anti–citrullinated protein antibody (ACPA) and rheumatoid factor (RF) positivity were independent predictors of CIRD. Six of 58 patients (10.3%) with ankle monoarthritis and 21 of 49 patients (42.9%) with wrist monoarthritis developed CIRD during follow-up. The 2010 ACR/EULAR Criteria for RA identified all patients diagnosed with seropositive RA at an early stage, mostly within 3 months.

Conclusion. Approximately one-fourth of patients with recent-onset monoarthritis developed CIRD over 2 years. Patients presenting with ankle arthritis rarely developed CIRD, whereas patients presenting with wrist arthritis more frequently did so. Longer duration of joint swelling and ACPA and RF positivity were also predictive of CIRD. Our findings facilitate the early identification of patients with monoarthritis who have an unfavorable prognosis.

INTRODUCTION

Monoarthritis is inflammatory swelling of a single joint or, in the case of the wrist and ankle, a joint unit. The major causes are crystal-induced arthritis, osteoarthritis, infection, trauma, mechanical derangement, and chronic inflammatory rheumatic diseases (CIRD) such as rheumatoid arthritis (RA). Early recognition of CIRD is essential to maintain function and minimize joint damage. Most studies of early arthritis have focused on polyarthritis because patients with polyarthritis tend to have the worst prognosis (1). Nevertheless, previous studies have indicated that ~30–60% of patients presenting with monoarthritis also develop CIRD (2–6). The studies have, however, generally included patients with longstanding arthritis and have often been small and retrospective (2–4,6–8). Knowledge is limited regarding recent-onset monoarthritis (2).

The aims of this study were to investigate the joint distribution of monoarthritis and to identify predictors of CIRD development in a large, prospective, multicenter, very early monoarthritis cohort. We further studied the applicability of the American College of Rheumatology (ACR)/European League Against Rheumatism (EULAR) 2010 criteria for RA in early monoarthritis (9).

Supported by The Norwegian Extra Foundation for Health and Rehabilitation (grants through EXTRA and the Norwegian Rheumatism Association) and the South-Eastern Norway Regional Health Authority.

[1]Ellen Sauar Norli, MD: Martina Hansens Hospital, Sandvika, Norway, Diakonhjemmet Hospital, and University of Oslo, Oslo, Norway; [2]Gina Hetland Brinkmann, MD: Østfold Hospital Trust, Grålum, Norway, and Diakonhjemmet Hospital, Oslo, Norway; [3]Tore Kristian Kvien, MD, PhD, Elisabeth Lie, MD, PhD, Maria Dahl Mjaavatten, MD, PhD: Diakonhjemmet Hospital, Oslo, Norway; [4]Olav Bjørneboe, MD: Martina Hansens Hospital, Sandvika, Norway; [5]Anne Julsrud Haugen, MD, PhD: Østfold Hospital Trust, Grålum, Norway; [6]Halvor Nygaard, MD: Lillehammer Hospital For Rheumatic Diseases, Lillehammer, Norway; [7]Cathrine Thunem, MD: Betanien Hospital, Skien, Norway.

Dr. Kvien has received speaking fees and/or consulting fees from AbbVie, Biogen, Bristol-Myers Squibb, Boehringer Ingelheim, Celltrion, Eli Lilly and Company, Epirus, Janssen, Merck-Serono, MSD, Mundipharma, Novartis, Oktal, Orion Pharma, Hospira/Pfizer, Roche, Sandoz, and UCB (less than $10,000 each). No other disclosures relevant to this article were reported.

Address correspondence to Ellen Sauar Norli, MD, Department of Rheumatology, Martina Hansens Hospital, Pb 823, 1306 Sandvika, Norway. E-mail: ellensnorli@yahoo.no.

Submitted for publication March 9, 2017; accepted in revised form August 1, 2017.

SIGNIFICANCE & INNOVATIONS

- Few patients with recent-onset monoarthritis seemed to develop chronic inflammatory rheumatic disease (CIRD) over 2 years.
- Patients with ankle monoarthritis had an especially low likelihood of developing CIRD, and none of these patients were diagnosed with rheumatoid arthritis (RA) or psoriatic arthritis.
- The American College of Rheumatology/European League Against Rheumatism 2010 criteria for RA performed well for patients with early seropositive monoarthritis but did not identify the seronegative RA patients.

MATERIALS AND METHODS

Study population and data collection. We included patients from the Norwegian Very Early Arthritis Clinic (NOR-VEAC) who presented with monoarthritis. NOR-VEAC is a 2-year prospective, multicenter, observational cohort in which 1,118 patients (ages 18–75 years) with clinically apparent arthritis in ≥1 joint of ≤16 weeks' duration were included between 2004 and 2010. The rheumatology departments involved established a dedicated track for receiving patients with early arthritis within 2 weeks. Primary care physicians were trained by their local rheumatology department to recognize arthritis early and were requested to refer all patients directly in order to minimize doctor's delay and ensure inclusion of a large proportion of patients with recent-onset arthritis. Exclusion criteria were arthritis due to crystal deposits, trauma, osteoarthritis, mechanical joint lesion, and septic arthritis. Data were collected at baseline and after 3, 6, 12, and 24 months by rheumatologists and study nurses, and the patients reported health status by questionnaires. A study nurse contacted patients not attending prescheduled study visits by telephone. Specific diagnostic tests such as HLA–B27, joint fluid analysis, and ultrasound were not required per protocol but were performed at the discretion of the treating rheumatologist. Anti–citrullinated protein antibodies (ACPAs), as well as immunoglobulin M and A rheumatoid factor (RF) were analyzed post hoc for research purposes for all patients. Details regarding design and data collection have been described elsewhere (10).

Patients with any follow-up data were included in the current study. For patients lost to follow-up before 2 years, the last registered data were used in a last observation carried forward approach. The study was approved by the Regional Ethics Committee, and patients gave informed consent.

Outcome. The final clinical diagnoses were made by the treating rheumatologist and coded according to the World Health Organization International Classification of Diseases, 10th revision. Patients with RA, psoriatic arthritis (PsA), other chronic spondyloarthritides (SpA) (ankylosing spondylitis, axial SpA, and inflammatory bowel disease–related arthritis), and connective tissue disease (CTD) were included in the CIRD group. Patients with undifferentiated arthritis (UA) or reactive arthritis (ReA) were also assigned to this outcome group if they received disease modifying antirheumatic drugs (DMARDs) during follow-up, had persistent joint swelling, or received glucocorticoids during the last 3 months of observation. The remainder of the patients, referred to as non-CIRD patients, thus included those in whom the arthritis had resolved without DMARDs, as well as patients diagnosed with gout or degenerative disorders during follow-up. Further, we

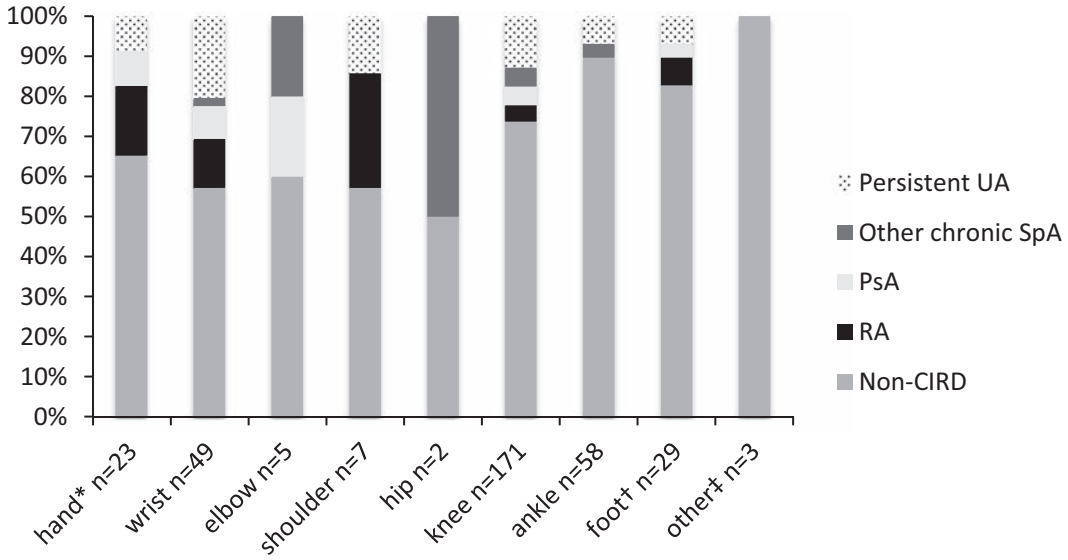

Figure 1. Overview of outcome according to joint distribution at baseline. Other chronic SpA are ankylosing spondylitis, axial SpA, and inflammatory bowel disease–associated arthritis. * = small joints of the hand; † = small joints of the foot; ‡ = 1st carpometacarpal joint (n = 2) and acromioclavicular joint (n = 1); UA = undifferentiated arthritis; SpA = spondyloarthritides; PsA = psoriatic arthritis; RA = rheumatoid arthritis; CIRD = chronic inflammatory rheumatic disease.

retrospectively applied the 2010 ACR/EULAR Criteria for RA at baseline and cumulatively at each visit, excluding patients with a clinical diagnosis other than RA.

Statistical analysis. Baseline characteristics were compared between CIRD and non-CIRD patients using the chi-square test, independent samples *t*-test, or the Wilcoxon-Mann-Whitney test, as appropriate. Multivariable logistic regression analyses were performed, with CIRD as the dependent variable. Clinically relevant variables (body mass index, smoking status, coffee consumption, education, joint localization, duration of joint swelling, tender joint count in 28 joints [TJC28], erythrocyte sedimentation rate [ESR], C-reactive protein level, Health Assessment Questionnaire Disability Index score, ACPA and RF positivity) with a univariable $P < 0.20$, as well as sex and age, were included in the full model. *P* values less than 0.05 were considered significant. Furthermore, explorative analyses of the development of different CIRDs and fulfillment of the 2010 ACR/EULAR Criteria for RA were performed. Statistical analyses were performed using SPSS, version 23.

RESULTS

Patient characteristics and outcome. Overall, 32.6% (364 of 1,118) of the patients in NOR-VEAC presented with monoarthritis, of whom 347 patients had follow-up data and were included in the current study (see Supplementary Figure 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23334/abstract). The most commonly affected joints were the knee (49.3%), ankle (16.7%), and wrist (14.1%) (Figure 1).

Table 1. Comparisons of baseline demographics and disease characteristics of patients with and without chronic inflammatory rheumatic disease (CIRD)*

	CIRD n = 91	non-CIRD n = 256	*P*
Demographics			
Age, mean ± SD years	46.0 ± 15.4	45.5 ± 13.7	0.78
Female, %	61.5	48.8	0.04
BMI, mean ± SD kg/m^2	26.4 ± 5.1	25.8 ± 3.9	0.71
Current daily smoker, %	38.5	27.0	0.04
Ever smoker, %	63.7	57.1	0.27
Coffee consumption ≥5 cups per day, %	25.6	26.3	0.89
Education (college/university), %	39.6	50.0	0.09
Disease characteristics			
Joint localization			
Small joints of the hand	8 (8.8)	15 (5.9)	0.33
Wrist	21 (23.1)	28 (10.9)	0.004
Elbow	2 (2.2)	3 (1.2)	NA
Shoulder	3 (3.3)	4 (1.6)	NA
Hip	1 (1.1)	1 (0.4)	NA
Knee	45 (49.5)	126 (49.2)	0.97
Ankle	6 (6.6)	52 (20.3)	0.003
Small joints of the foot	5 (5.5)	24 (9.4)	0.25
First carpometacarpal joint	0 (0.0)	2 (0.8)	NA
Acromioclavicular joint	0 (0.0)	1 (0.4)	NA
Other disease characteristics			
Duration of joint swelling, median (25th, 75th percentiles) days	38.0 (16.0–66.0)	19.5 (6.0–40.0)	<0.001
ACPA positive, %	18.9	2.4	<0.001
RF positive, %	21.1	3.6	<0.001
ACPA and/or RF positive, %	27.8	4.8	<0.001
ESR, median (25th, 75th percentiles) mm	22.0 (12.0–37.0)	16.0 (8.0–31.0)	0.03
CRP, median (25th, 75th percentiles) mg/liter	12.0 (4.0–29.0)	9.0 (3.0–26.0)	0.29
Joint pain VAS, mean ± SD mm	52.9 ± 28.6	46.7 ± 25.2	0.04
Fatigue VAS, mean ± SD mm	36.1 ± 31.0	31.5 ± 29.0	0.25
Patient global VAS, mean ± SD mm	50.6 ± 25.1	45.9 ± 25.3	0.12
Assessor global VAS, mean ± SD mm	29.0 ± 15.9	24.8 ± 14.4	0.04
SF-36 PCS score, mean ± SD	34.3 ± 10.6	36.9 ± 10.8	0.06
SF-36 MCS score, mean ± SD	49.9 ± 9.1	51.0 ± 11.1	0.13
HAQ DI score, mean ± SD	0.66 ± 0.57	0.59 ± 0.55	0.29
TJC28, median (25th, 75th percentiles)	1.0 (1.0–1.0)	1.0 (0.0–1.0)	0.001

* Values are the number (%) unless indicated otherwise. Differences between the groups were examined with the Wilcoxon-Mann-Whitney test, independent samples *t*-test, and chi-square test, as appropriate. BMI = body mass index; NA = not appropriate; ACPA = anti–citrullinated protein antibody; RF = rheumatoid factor (immunoglobulin A and/or M); ESR = erythrocyte sedimentation rate; CRP = C-reactive protein level; VAS = visual analog scale; SF-36 PCS = Short Form-36 Health Survey physical component summary; SF-36 MCS = SF-36 mental component summary; HAQ DI = Health Assessment Questionnaire Disability Index; TJC28 = tender joint count in 28 joints.

The majority of the patients (73.8%) did not develop CIRD. A total of 21 patients were diagnosed with RA (6.1%), 16 were diagnosed with PsA (4.6%), 12 were diagnosed with other chronic SpA (3.5%), and 42 were diagnosed with persistent UA (12.1%), which together constitute the CIRD group (n = 91). No patients were diagnosed with CTD. Further details about final clinical diagnoses in the CIRD and non-CIRD groups are given in Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23334/abstract. Baseline characteristics of CIRD and non-CIRD patients are compared in Table 1. Female sex, smoking, wrist involvement, long duration of joint swelling, positive ACPA, positive RF, high ESR, and high TJC28 were associated with CIRD development, while ankle involvement was significantly less frequent in the CIRD group. DMARDs were administered to 58.2% of the CIRD patients. A total of 50.5% in the CIRD group versus 13.3% in the non-CIRD group received systemic corticosteroids during follow-up. The corresponding proportions for intraarticular corticosteroids and treatment with nonsteroidal antiinflammatory drugs were 58.2% versus 52.3% and 63.7% versus 55.9%, respectively.

Two-year follow-up data were available for 219 patients (63.1%). Compared to the overall group of patients (n = 347), the completers were more often women (58.0% versus 52.2%) and had a longer median duration of joint swelling (27 days versus 23 days) (see Supplementary Table 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23334/abstract). Details regarding follow-up data are available in Supplementary Tables 3 and 4.

Predictors for CIRDs. Since both ankle and wrist monoarthritis were univariably associated with the outcome, we used a categorical variable comparing ankle, wrist, and other joints for the multivariable logistic regression analysis (Table 2). The odds ratio for CIRD was 2.0 (95% confidence interval [95% CI] 1.0–4.2) for monoarthritis of the wrist and 0.5 (95% CI 0.2–1.2) for monoarthritis of the ankle compared to all other joints. Longer duration of joint swelling and ACPA and RF positivity were also significantly associated with CIRD, whereas smoking, education, ESR, and TJC28 were not retained in the multivariable model.

Patients with monoarthritis of the wrist, shoulder, or the small joints of the hand more often developed RA (12.2%, 28.6%, and 17.4%, respectively, compared to 6.1% for the whole group) (Figure 1). For PsA, no clear trend was observed. Patients with ankle monoarthritis (n = 58) had the lowest rate of CIRD (10.3%), and no patients presenting with monoarthritis of the ankle developed RA or PsA.

Classification according to the 2010 ACR/EULAR Criteria for RA. A total of 17 patients (4.9%) fulfilled the 2010 ACR/EULAR Criteria for RA at baseline; the majority had arthritis of the wrist (n = 7) or other small joints (n = 7). Three patients with large joint monoarthritis fulfilled the criteria due to additional tender joints. During follow-up, 9 additional patients fulfilled the criteria, of whom 6 had large joint arthritis at baseline.

We also compared clinical diagnosis to the 2010 ACR/EULAR Criteria for RA fulfillment. Of the 16 patients with a final clinical diagnosis of seropositive RA, 56.3% were diagnosed within 3 months, as opposed to 93.8% if the 2010 ACR/EULAR Criteria for RA had been used for diagnostic purposes. The 5 patients with a final clinical diagnosis of seronegative RA were all diagnosed after 6 to 24 months and did not fulfill the 2010 ACR/EULAR Criteria for RA during follow-up. Conversely, 10 patients who never received a clinical diagnosis of RA fulfilled the 2010 ACR/EULAR Criteria for RA during follow-up, and their respective final diagnoses were persistent UA (5 patients), PsA (1 patient), and non-CIRD (4 patients).

Table 2. Comparisons of baseline demographics and disease characteristics of patients with and without chronic inflammatory rheumatic disease (CIRD)*

	Univariable analyses			Multivariable analysis†		
	OR	95% CI	*P*	OR	95% CI	*P*
Age, years	1.00	0.99–1.02	0.78	0.99	0.97–1.01	0.42
Male (ref.: female)	0.60	0.37–0.97	0.04	0.72	0.42–1.23	0.23
Current smoker (ref.: former or never)	1.69	1.02–2.80	0.04	–	–	–
Education (college/university) (ref.: lower education)	0.66	0.40–1.07	0.09	–	–	–
Joint localization (wrist vs. ankle vs. other)			0.001			0.03
Swollen wrist vs. other	2.06	1.09–3.89	0.03	2.03	0.97–4.24	0.06
Swollen ankle vs. other	0.32	0.13–0.77	0.01	0.46	0.18–1.16	0.10
Duration of joint swelling, weeks	1.11	1.05–1.17	<0.001	1.09	1.03–1.16	0.002
ACPA positive (ref.: negative)	9.51	3.62–25.00	<0.001	3.83	1.25–11.74	0.02
RF positive (ref.: negative)	7.20	3.12–16.60	<0.001	3.47	1.30–9.23	0.01
ESR, mm/hour	1.01	1.00–1.02	0.17	–	–	–
TJC28	1.31	1.07–1.62	0.01	–	–	–

* OR = odds ratio; 95% CI = 95% confidence interval; ref. = reference; ACPA = anti–citrullinated protein antibody; RF = rheumatoid factor (immunoglobulin A and/or M); ESR = erythrocyte sedimentation rate; TJC28 = tender joint count in 28 joints.
† Hosmer-Lemeshow goodness-of-fit test, $P = 0.208$.

DISCUSSION

This is the first study to prospectively investigate outcomes in a large, unselected cohort with recent-onset monoarthritis (2,4). Approximately one-third of the patients in NOR-VEAC presented as having monoarthritis, of whom 26.2% developed CIRD. RA was diagnosed in 6.1% during follow-up. In a follow-up study, Binard et al found that among 27 patients with monoarthritis of <1 year duration, 11 patients had persistent arthritis, and no patients were diagnosed with RA (4). Among 23 other patients presenting with monoarthritis at baseline but with a history of involvement of other joints during the year before baseline, they reported 7 patients with persistent disease and 14 with RA (4). Another follow-up study including 32 patients with undifferentiated monoarthritis of 1–6 months disease duration identified 2 patients with RA (8). However, in most previous studies of monoarthritis, the proportions with CIRD and RA development are higher, with approximate frequencies of 43.3–65.8% and 7.9–18.1%, respectively (2,3,6). We believe longer duration of joint swelling at baseline is the major reason for this difference compared to our study. However, direct comparison of results between other studies and our study is challenged by the retrospective design of most former studies, as well as differences in patient selection, joints studied, and outcome definitions. DMARDs were prescribed to 58.2% of the patients with CIRD in our study. We believe this rather low percentage may be explained by the expected beneficial prognosis of monoarthritis, as well as the inclusion of persistent ReA and UA in the CIRD group.

Type of joint affected, duration of joint swelling, and ACPA and RF positivity were found to be independent predictors of CIRD development. Monoarthritis of the wrist or small joints of the hand involved the highest risk of patients developing CIRD, especially RA, which is in line with the early arthritis population in general and also results from retrospective studies assessing chronic monoarthritis (2,6,11). Ankle involvement has previously been shown to predict diagnostic outcomes other than RA in a very early arthritis cohort (not only monoarthritis patients) and to predict peripheral SpA in a retrospective monoarthritis study (2,12). These findings are partially in accordance with the current study, which demonstrated that very early ankle monoarthritis was infrequently associated with CIRD.

Among the 26 patients who fulfilled the 2010 ACR/EULAR Criteria for RA during follow-up, 84.6% developed CIRD. This finding is in line with data from the entire NOR-VEAC cohort, as well as with validation studies of the criteria in early arthritis populations with more joints affected (13,14). Use of the criteria facilitated earlier recognition of RA among seropositive patients. Notably, the seronegative RA patients never fulfilled the 2010 ACR/EULAR Criteria for RA, supporting the findings of Nordberg et al, who showed that the 2010 criteria only capture seronegative patients with extensive joint involvement (15).

In this study, the diagnosis of arthritis and the decision to start DMARDs were both based on clinical judgment, leaving the use of diagnostic tools, classification criteria, and treatment guidelines to the discretion of the rheumatologist. This might be a limitation when it comes to comparing the results with those from future studies of early monoarthritis. Also, distinction between true arthritis and inflammation of juxtaarticular structures might be difficult. Another limitation of our study is the loss to follow-up. We used a last observation carried forward method to handle this. This method carries the risk of bias because it is an unrealistic assumption that the last measured outcome is frozen in time. Nevertheless, in our study, we believe this method resulted in less bias than only to study patients with 2-year follow-up or to impute other outcomes. Recurrence of arthritis in some patients classified as non-CIRD and resolution of arthritis in some patients classified as CIRD can therefore not be excluded. Moreover, some subgroups were small. This was an exploratory analysis in a cohort study, and the findings should ideally be replicated in an independent cohort.

The prospective, multicenter design, the short duration of joint swelling examined, as well as the large size of the cohort are the main strengths of our study. Early identification of arthritis and prompt referral of all patients by general practitioners contribute to the external validity.

In conclusion, monoarthritis is common in patients with very early arthritis. Our study further demonstrated that approximately one-fourth of recent-onset monoarthritis patients developed CIRD over 2 years. The likelihood of developing CIRD was highest in patients with monoarthritis of the wrist and lowest in patients with monoarthritis of the ankle, while no patients with ankle monoarthritis developed RA or PsA. The results facilitate early identification of monoarthritis patients at risk of developing CIRD.

ACKNOWLEDGMENTS

The authors thank all the participating patients, nurses, laboratory workers, and clinicians in the NOR-VEAC study group who made the study 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. Norli 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. Norli, Brinkmann, Kvien, Lie, Mjaavatten.
Acquisition of data. Norli, Brinkmann, Kvien, Bjørneboe, Haugen, Nygaard, Thunem, Mjaavatten.
Analysis and interpretation of data. Norli, Brinkmann, Kvien, Lie, Mjaavatten.

REFERENCES

1. Hazes JM, Luime JJ. The epidemiology of early inflammatory arthritis. Nat Rev Rheumatol 2011;7:381–90.
2. Jeong H, Kim AY, Yoon HJ, Park EJ, Hwang J, Kim H, et al. Clinical courses and predictors of outcomes in patients with mono-

arthritis: a retrospective study of 171 cases. Int J Rheum Dis 2014; 17:502–10.
3. Blocka KL, Sibley JT. Undiagnosed chronic monarthritis: clinical and evolutionary profile. Arthritis Rheum 1987;30:1357–61.
4. Binard A, Alassane S, Devauchelle-Pensec V, Berthelot JM, Jousse-Joulin S, Chales G, et al. Outcome of early monoarthritis: a followup study. J Rheumatol 2007;34:2351–7.
5. Mjaavatten MD, van der Heijde D, Uhlig T, Haugen AJ, Nygaard H, Sidenvall G, et al. The likelihood of persistent arthritis increases with the level of anti-citrullinated peptide antibody and immunoglobulin M rheumatoid factor: a longitudinal study of 376 patients with very early undifferentiated arthritis. Arthritis Res Ther 2010;12:R76.
6. Inaoui R, Bertin P, Preux PM, Trèves R. Outcome of patients with undifferentiated chronic monoarthritis: retrospective study of 46 cases. Lancet 2004;71:209–13.
7. Fletcher MR, Scott JT. Chronic monarticular synovitis: diagnostic and prognostic features. Ann Rheum Dis 1975;34:171–6.
8. Kaarela K, Tiitinen S, Luukkainen R. Long-term prognosis of monoarthritis: a follow-up study. Scand J Rheumatol 1983;12: 374–6.
9. 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.
10. Mjaavatten MD, Haugen AJ, Helgetveit K, Nygaard H, Sidenvall G, Uhlig T, et al. Pattern of joint involvement and other disease characteristics in 634 patients with arthritis of less than 16 weeks' duration. J Rheumatology 2009;36:1401–6.
11. Funovits J, Aletaha D, Bykerk V, Combe B, Dougados M, Emery P, et al. The 2010 American College of Rheumatology/ European League Against Rheumatism classification criteria for rheumatoid arthritis: methodological report phase I. Ann Rheum Dis 2010;69:1589–95.
12. Abhishek A, de Pablo P, Cader MZ, Buckley CD, Raza K, Filer A. Diagnostic outcomes associated with ankle synovitis in early inflammatory arthritis: a cohort study. Clin Exp Rheumatol 2014;32:533–8.
13. Norli ES, Brinkmann GH, Kvien TK, Bjorneboe O, Haugen AJ, Nygaard H, et al. Self-limiting arthritis among patients fulfilling the 2010 ACR/EULAR classification criteria for rheumatoid arthritis in a very early arthritis cohort. Semin Arthritis Rheum 2016;46:272–8.
14. Radner H, Neogi T, Smolen JS, Aletaha D. Performance of the 2010 ACR/EULAR classification criteria for rheumatoid arthritis: a systematic literature review. Ann Rheum Dis 2013;73:114–23.
15. Nordberg LB, Lillegraven S, Lie E, Aga AB, Olsen IC, Hammer HB, et al. Patients with seronegative RA have more inflammatory activity compared with patients with seropositive RA in an inception cohort of DMARD-naive patients classified according to the 2010 ACR/ EULAR criteria. Ann Rheum Dis 2017;76:341–5.

Arthritis Care & Research
Vol. 72, No. 5, May 2020, pp 711–722
DOI 10.1002/acr.23896

American College of Rheumatology
Empowering Rheumatology Professionals

Patients' Perspectives and Experience of Psoriasis and Psoriatic Arthritis: A Systematic Review and Thematic Synthesis of Qualitative Studies

Daniel Sumpton,[1] Ayano Kelly,[2] David J. Tunnicliffe,[3] Jonathan C. Craig,[4] Geraldine Hassett,[5] Diana Chessman,[5] and Allison Tong[3]

Objective. To describe the range and depth of perspectives and experiences of patients with psoriasis and psoriatic arthritis to inform gaps in patient-centered care.

Methods. We searched MEDLINE, Embase, PsycINFO, and CINAHL to April 2018. Thematic synthesis was used to analyze the findings.

Results. We included 56 studies involving 1,484 adult patients with psoriasis (n = 1,147) and psoriatic arthritis (n = 337). Six themes (and subthemes) were identified: suffering uncontrollable and ongoing upheaval (dictating life choices and course, disrupting family and social roles, limited by debilitating symptoms, unstoppable and far-reaching fatigue), weighed down by mental load (anxiety provoked by the volatility of symptoms, dreading deterioration, struggling with unrecognized distress, helpless and nihilistic), harboring shame and judgement (marked as unhygienic and contagious, rejected and isolated, hiding away and resenting own appearance, pain and embarrassment in intimacy), demoralized by inadequacies and burden of therapy (disappointed by unmet expectations of treatment benefit, daily drudgery, deterred by unpalatable or inconvenient treatments, disempowered by lack of personalized care), gaining control (making sense of the condition, accepting a new health status, regaining independence and normality, attuning to the body), and making confident treatment decisions (trading off perceptible benefits against safety and convenience, relying on family input, seeking empowering and reassuring relationships).

Conclusion. Patients with psoriasis and psoriatic arthritis contend with disruption in their functioning, roles, and life course and have unmet expectations about treatment. Enhanced therapeutic relationships, addressing treatment expectations and supporting psychosocial needs may improve satisfaction and outcomes.

INTRODUCTION

Psoriasis and psoriatic arthritis (PsA) are coexisting autoimmune inflammatory conditions with shared pathophysiology and overlapping therapies (1). Psoriasis affects ~3% of the population, and ~30% of patients with psoriasis will develop PsA (2). Given the high rates of undiagnosed and undertreated PsA among individuals with psoriasis, early screening is used to improve access to early treatment and prevent joint damage (3). Patients with PsA and psoriasis have a higher risk of comorbidities, including depression, metabolic disease, and cardiovascular mortality compared with the general population (4).

There is high treatment burden and dissatisfaction among patients with psoriasis and PsA (5) due to a perceived lack of medication efficacy, side effects, and the need for monitoring of treatments (6,7). Underdiagnosis of PsA and the variability in how rheumatologists and dermatologists prescribe disease-modifying treatments may lead to undertreatment of PsA (8). While tighter disease control has been shown to improve disease activity (9), there remains a mismatch between patient and physician perception of the severity of the condition, potentially leading to differences in the

[1]Daniel Sumpton, MBBS, FRACP: Concord Repatriation General Hospital, Sydney School of Public Health, The University of Sydney, and The Children's Hospital Westmead, Sydney, New South Wales, Australia; [2]Ayano Kelly, MBBS, FRACP: The Children's Hospital Westmead, Sydney, New South Wales, and Australian National University, Canberra, Australia; [3]David J. Tunnicliffe, PhD, Allison Tong, PhD: Sydney School of Public Health, The University of Sydney, and The Children's Hospital Westmead, Sydney, New South Wales, Australia; [4]Jonathan C. Craig, PhD, FRACP: Flinders University, Adelaide, South Australia, Australia; [5]Geraldine Hassett, PhD, FRACP, Diana Chessman, PhD, FRACP: Liverpool Hospital, Sydney, New South Wales, Australia.

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

Address correspondence to Daniel Sumpton, MBBS, FRACP, Rheumatology Department, Level 6 West, Concord Repatriation General Hospital, Concord, New South Wales 2139, Australia. E-mail: daniel.sumpton@health.nsw.gov.au.

Submitted for publication December 10, 2018; accepted in revised form March 26, 2019.

SIGNIFICANCE & INNOVATIONS

- Patients with psoriasis and psoriatic arthritis contend with severe psychosocial impacts of their disease and perceive that their mental health burden and stigmatization due to skin disease is underrecognized.
- Experiences of inadequate therapeutic relationships, a perception of treatment ineffectiveness, and fear of medications were common and have the potential to reduce medical engagement and treatment adherence.
- Patients yearn for trusting and personalized relationships with clinicians to gain confidence in management and overcome fears of taking medication.

perception of need for treatment escalation (10,11). International guidelines (12,13) emphasize the need for a shared decision-making approach between clinicians and patients, which requires an understanding of the goals and values of patients.

Systematic review and synthesis of multiple qualitative studies can provide an in-depth understanding of the experience of patients with both psoriasis and PsA across health care contexts and patient populations. The aim of this study was to describe the range and depth of experiences and perspectives of patients with psoriasis and PsA, which may inform strategies to improve satisfaction with treatment and overall health and quality of life outcomes.

MATERIALS AND METHODS

Selection criteria. We used the Enhancing Transparency of Reporting the Synthesis of Qualitative Research framework to report this study (14). Qualitative studies that describe the experiences of adult patients ages ≥18 years and diagnosed with psoriasis or PsA were eligible. Studies involving patients with different conditions were included if at least 1 participant had either psoriasis or PsA. Epidemiologic studies, case reports, basic science, nonprimary research articles (letters, editorials,

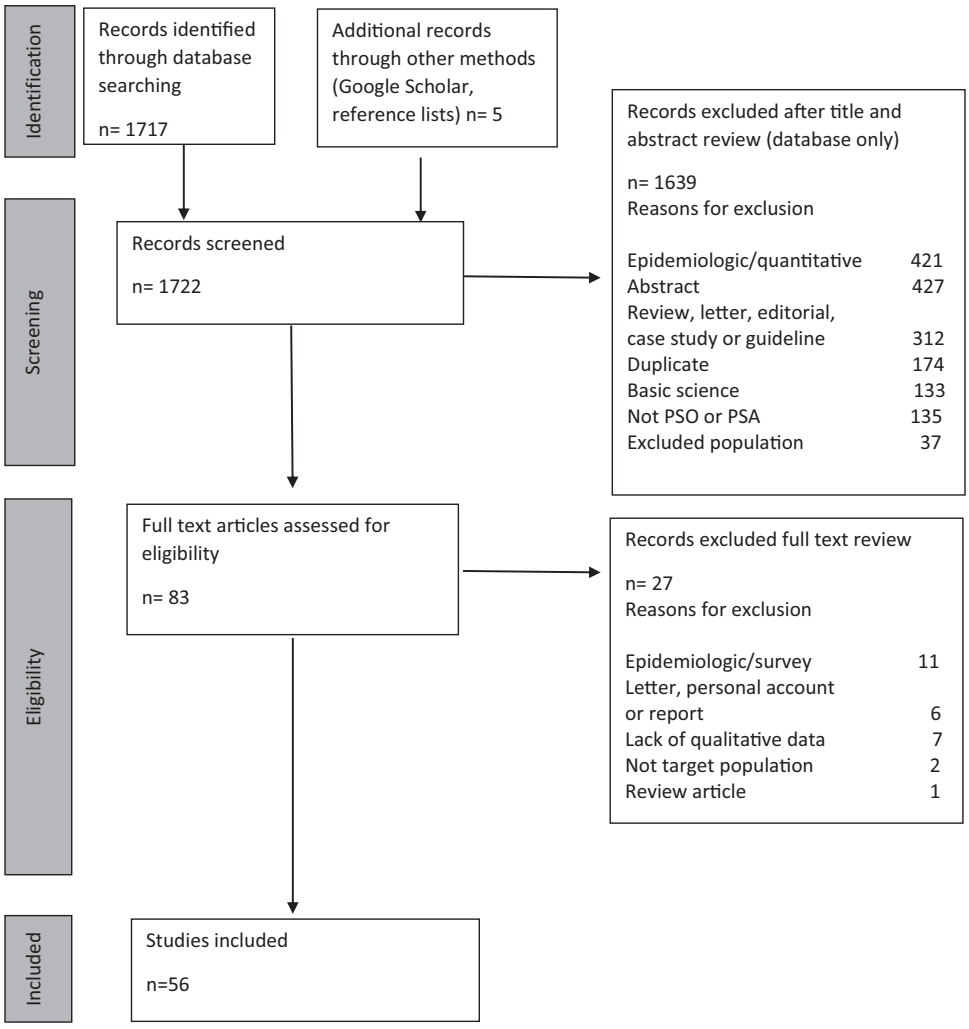

Figure 1. Search results. PSO = psoriasis; PSA = psoriatic arthritis.

reviews), and non-English articles were excluded due to a lack of resources.

Data sources and searches. We searched MEDLINE, Embase, CINAHL and PsycINFO from database inception to April 29, 2018 using comprehensive search strategies (see Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23896/abstract). We also searched reference lists of relevant studies and Google Scholar. DS screened the titles and abstracts and discarded those that did not meet the eligibility criteria. The full-text of the remaining studies were then assessed for eligibility (Figure 1).

Appraisal of transparency of reporting. The transparency of reporting used in interview and focus-group studies was evaluated using the adapted Consolidated Criteria for Reporting Qualitative Health Research (15). This framework includes criteria specific to the research team, study methods, context of the study, analysis, and interpretations. Independent assessment was undertaken by 3 reviewers (DS, AK, and DJT). Any discrepancies were discussed until consensus was reached or resolved in further discussion with AT.

Data analysis. Thematic synthesis was used for analysis (16). For each article, DS inductively identified preliminary concepts. An initial coding structure with the preliminary concepts was discussed among all authors to revise the preliminary themes and subthemes to ensure that the breadth and depth of data were reflected in the analysis. All text in the Results (including quotations and the author's themes/description of the themes and their interpretations of the data) and Discussion sections were then imported into HyperRESEARCH software, version 3.7.3 (ResearchWare). DS performed line-by-line coding of each article.

RESULTS

Literature search. We identified 56 studies involving 1,484 patients with either psoriasis (n = 1,147, range 1–104) or PsA (n = 337, range 1–89) (Table 1). Study characteristics are shown in Table 1 and in Supplementary Table 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23896/abstract. Studies were conducted in 19 countries, with the majority of studies from Europe, the UK, the US, or Australia. The majority of studies used interviews (61%) or focus groups (21%). Patients' ages ranged from 18 to 86 years.

Completeness of reporting. There was variability in the comprehensiveness of reporting of the 26 items in the modified Consolidated Criteria for Reporting Qualitative Research framework (see Supplementary Table 3, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23896/abstract). The median number of items reported was 13 (interquartile range 9–15). Twenty-five studies (46%) reported theoretical saturation, 35 (65%) reported multiple researcher involvement in data collection and analysis, and 45 studies (83%) provided participant quotations.

Table 1. Characteristics of the included studies*

Study characteristics	Values
Year of publication	
1985–2010	22 (39)
2011–2018	34 (61)
Patient population	
Psoriasis	33 (59)
Psoriatic arthritis (PsA)	6 (11)
Mixed skin with psoriasis	9 (16)
Mixed rheumatic disease with PsA	7 (12)
Psoriasis and PsA	1 (2)
Region	
Europe (excluding UK)	16 (29)
UK	15 (27)
US	9 (16)
Australia/New Zealand	8 (14)
Multinational	5 (9)
Iran	1 (2)
South Africa	1 (2)
Canada	1 (2)
Sample size	
1–10	16 (29)
11–20	11 (20)
21–30	13 (23)
31–40	6 (11)
41–50	5 (9)
51–100	3 (5)
100–110	2 (4)
Method of data collection	
Interviews	37 (61)
Focus groups	12 (21)
Interviews and focus groups	3 (5)
Video recording and interviews	1 (2)
Observation and interviews	1 (2)
Written postcards	1 (2)
Written free text	1 (2)

* Values are the number (%).

Synthesis. We identified 6 themes: suffering uncontrollable and ongoing upheaval, weighed down by mental load, harboring shame and judgement, demoralized by inadequacies and burden of therapy, gaining control, and making confident treatment decisions. The themes and subthemes are described in the following section, with illustrative text provided in Table 2. The themes below reflect the perspectives of patients with psoriasis and PsA unless otherwise specified. The conceptual links among themes are shown in Figure 2.

Suffering uncontrollable and ongoing upheaval. "I can't control it, it controls me. That's the frustrating sense that I can't beat it" (17).

Table 2. Illustrative quotations (references cited are in Supplementary Appendix A)*

Themes, illustrative quotations, and text	Contributing studies
Suffering uncontrollable and ongoing upheaval	
Dictating life choices and course	1,3,5–9,11,12,19,21,26,32–35,37,39, 42,51,53–55
"That's how I used to look [points at picture on mantelpiece], it's like I miss myself. Running things, being in charge, being in the office, being up in management and now I'm just like this now." (37)	
"I couldn't function. I couldn't function at all, and I'm used to being active and doing lots of stuff, and then I couldn't do anything. I hardly left the house for months." (35)	
"I had to give up nursing because of the symptoms on my feet and my hands. I have always been a nurse, so when I had to stop I almost [felt like I] died." (37)	
Some respondents expressed fatalistic beliefs and an external locus of control. (5)	
Disrupting family and social roles	5,7–9,11,12,19,21,22,26,32, 33,35,37,41,45,46,51,53,55,56
"I've got a 21-year-old son, and I'm thinking, when he has babies, am I going to be a proper grandma." (8)	
"Sometimes it has been hopeless. It isn't just me, you know. It is just as much my wife and the rest of the family who are affected." (55)	
"I just worry about the burden to my family and all that, that they have to look after me." (8)	
Many patients stated that the disease engenders conflict situations, especially in relation to their spouse or domestic partner. (55)	
Limited by debilitating symptoms	1,2,7,8,9,11,12,13,17,19,21,30–32, 35,36,38,39,41,42,44,45,46,50
"Psoriasis is very, very sore. The itchiness drives you mad. You can't sleep and you scratch your arms and legs to pieces." (56)	
"The polite British thing is to say, oh yeah I'm fine thanks. But actually I'm thinking, every bone in my body hurts and I'm so tired I want to cry." (8)	
Patients reported consequences of living with PsA that affected all areas of their social, work, and family lives but most prominently described how physically restricting PsA was. (8)	
Unstoppable and far-reaching fatigue	9,35,41,42
"And [I] keep myself going but it's really difficult because the tiredness just takes you over, takes over…it is hard. It's hard because you can't keep your eyelids open, you're fighting it." (35)	
"It's as if your brain's fatigued, you know, it's as if it's something else, it's not just tired, it's beyond that." (9)	
Participants' accounts of the ineffectiveness of some treatments to ameliorate fatigue highlight the potential for nonpharmacological approaches. (9)	
Weighed down by mental load	
Anxiety provoked by the volatility of symptoms	2,5–9,12,13,19,22,26,35,37,39,45,51,55
"I'll be thinking about it way too much, and then I'll start getting…affecting my skin, because the stress will make it outbreak, and then the outbreak, I'll want to itch, and just scratch." (12)	
"Patients said they were unable to visualize a future for themselves and often expressed suicidal ideation." (8)	
The psychological impact of psoriasis was characterized by constant worry, a struggle for control and a fear of stress triggering symptoms. (2)	
Dreading deterioration	5,8,9,22,38,39,41,44,53
"How fast it's going to degenerate…I don't want to be some person who's sitting in a chair somewhere unable to move, and that sometimes makes me a bit anxious." (8)	
"It's depressing…you can't see it getting any better…just worse and worse and worse… might be a slight improvement, but the general trend…it is very depressing." (38)	
Patients feared their condition would rapidly and unexpectedly deteriorate. (8)	
Struggling with unrecognized distress	1,2,4,5,8,9,11,19,21,22,26, 27,37,39,41,42,45, 46,50,51,56
"People don't realize how uncomfortable it is. If it looks bad, they sympathize, people don't realize unless it looks bad. It can look okay but be really painful." (19)	
"She's got MS, but she's less disabled than I am, and that's what's annoying." (8)	
The failure of practitioners explicitly to acknowledge in consultations the feelings of stigma and lack of control often associated with psoriasis was commonly reported in both primary and secondary care. (38)	
Helpless and nihilistic	26,37–39,42,51,56
"Bleak. I'm going to have this forever, I'm never going to get away from it, it's going to ruin my life, I'm going to be an old maid." (26)	
"There's nothing you can do. There's not a hope in hell that [creams] control it at all. There's no point." (39)	
A feature of depression in this study was the feelings of helplessness and nihilism engendered by the 'incurability' of respondents' psoriasis. Suicidal ideation was prominent in some participants. (26)	

(Continued)

Table 2. *(Cont'd)*

Themes, illustrative quotations, and text	Contributing studies
Harboring shame and judgement	
Marked as unhygienic and contagious	2,5,7,11,19–22,25,26,37,39,42,44,51,53,55
"My relatives do not come to our house…They say your disease is contagious and we might get infected." (11)	
"People take a step back, it can be horrifying, they give you a funny look, as if you have 3 heads." (19)	
Patients fear being stigmatized and fear that others will think the disease is contagious. (37)	
Rejected and isolated	2,3,5,7,8,11,19,21,25,26,29,37,39,41,42,44,45, 50,51,53,55,56
"We almost never talk about it, so I feel sort of alone with my disease and all that it entails." (55)	
"I want to be a positive person and talk to other people. But the psoriasis stops me from seeking the contact I want with others. I'm afraid of rejection." (55)	
Hiding away and resenting own appearance	5,6,7,9,16,19–22,26,28,29, 32,37,38,41,42,48, 50, 51,53,55,56
"The ability to disfigure has spoiled the best years of my life, turning what was beautiful into something ugly and undesirable." (5)	
For many, experiencing their own bodies as offensive, unattractive, and ugly was a source of feelings of sadness, vulnerability, irritation, and despair. (55)	
"Media stuff and the advertisements for the soft skin on television makes me feel embarrassed…it also makes me feel less attractive sexually to people, because I'm not the ideal, I'm not the perfect image." (26)	
Pain and embarrassment in intimacy	6,7,16,21,28,32,42,55
"There's no sex life. First of all, you don't want to expose yourself…I'm single and I don't want to think about dating because I'm not ready to share with a stranger." (7)	
"It is frustrating. Skin should be smooth. Skin is contact. It is my closest contact with my surroundings. Expressions of love revolve around skin. Through skin. It is frustrating to have a damper placed on one's contact with other people." (55)	
Negative effects on sexual experience encompassed physical effects such as mechanical friction, cracking, and pain and psychosocial effects such as embarrassment and feeling stigmatized. (7)	
Demoralized by inadequacies and burden of therapy	
Disappointed by unmet expectations of treatment benefit	2,5-10,17,21,24,27,37–40,42,44,46,51,54–56
"It was a waste of time. I shouldn't have gone because, I will be honest with you, I have probably got a cupboard full of medication that there is nothing I can do with." (46)	
"That's all I've ever felt since I got psoriasis, was frustrated. Frustrated that there wasn't an answer, there wasn't a solution." (42)	
Patients also expressed unrealistic expectations of their treatments and stated that their doctor had not communicated what to expect. (2)	
Daily drudgery	3–5,45,50,51,53–55
"You know, I grease whenever I want to. I can't follow any kind of regimen, I have no time for that." (45)	
"Treating your symptoms takes 2 and a half hours out of each and every day whilst the embarrassment of cleaning up continual masses of fallen 'scales' exhausts me." (5)	
Deterred by unpalatable or inconvenient treatments	2,4,5,9,13,19,22,32,37–39,41–45,49,51,54,55
"Just the thought of taking more medication and taking that long term, that bit worries me." (9)	
"It is just not natural" and "I don't want this in my body." (43)	
"And in the end when I'm up to fairly long times I feel like a dishrag afterwards…and then you ask yourself, is it worth it?" (44)	
The burden of adhering to monitoring requirements was too much for some participants, who found accessing services at the appropriate time incompatible with other commitments. (9)	
Disempowered by lack of personalized care	2,3,5,8–10,20–22,24,27, 38,39,43,45,46, 49,52,54,56
"I was given nothing when I was first diagnosed, they never told me how I would feel, they never told me what my body would feel like. If I'd have had that at least I could have prepared myself." (8)	
"You need to talk to us more as individuals because clearly…one size doesn't fit all." (10)	
"I have never experienced that health care personnel have asked me how I feel about my body." (21)	
"After I got handed a bunch of cream with no details on how to use them, I totally lost the trust in doctors." (49)	
Participants suggested that their treatment tended to be prescribed on an arbitrary basis with little thought being given to how it may fit with their lifestyle. They wanted an agreed, personalized care plan that would fit into their daily lives. (10)	

(Continued)

Table 2. *(Cont'd)*

Themes, illustrative quotations, and text	Contributing studies
Gaining control	
Making sense of the condition	3,10,14,19,21,22,35,38–40,42,45,46,52,54,56
"I try to understand the disease, and I try to help myself as much as I possibly can… whether it's doing the [physical therapy] exercises every day or reading up on things or trying to maintain a reasonable diet." (35)	
"It's not my fault is one of the biggest things. By knowing all these things affect [psoriasis], it's not something I've done…that's made this happen…by having the leaflets…I felt more confident to go and sit and have breakfast in a coffee shop." (40)	
It was only under the conditions of becoming more informed about their psoriatic condition that participants felt sufficiently empowered to assume some control over the symptoms. (56)	
Accepting a new health status	6,7,10,13,17,33,37,42,43,46,53,55,56
"It has become a part of yourself; you learn to live with it." (53)	
"I now understand the lifestyle implications of the condition and have adjusted my lifestyle appropriately…wearing nonabrasive clothing, eating more healthily." (56)	
Adaptation and acceptance: the need to adapt day-to-day routines and to acknowledge that such changes are necessary. (33)	
Regaining independence and normality	9,40,41,49
"If I could have anything it would be independence, it would be to be able to be as fast as everybody else, it will be able to drive my own car, go out when I wanted to go out, come in and lock my own front door, and not have somebody to come in to help with the shower." (9)	
"I live alone and I want to keep my independence." (9)	
Engagement with the materials…encouraged an increased sense of personal control and/or prompted new ways of thinking about self management. (40)	
Attuning to the body	8,13,15,35,37,41,49,55,56
"Right, I'm having a lazy day today…I get the burning sensation, and then I think, right, just calm down now." (35)	
"I try to exercise and eat well because, for me, a healthy body equals a healthy mind and a healthy mind will help to reduce my symptoms, as I will be less stressed." (37)	
"Listen to the body" was a state where the informants learned to understand their individual body signals and how the arthritis fluctuations affected them. (13)	
Making confident treatment decisions	
Trading off perceptible benefits against safety and convenience	2,4,9,18,22,23,24,37,42,43,45,46,49,54,55
"Control of the disease is vital for us, that's for sure, I definitely need control of the disease, but without the doctor I cannot have it, because there are no remedies I can use by myself." (4)	
"It's a balance of the quality of life and what you're risking." (23)	
The need for efficient treatment exceeded the perceived risk and participants coped with these emotions by reconceptualizing the risk of experiencing side effects and the dream of being cured of psoriasis. (22)	
Relying on family input	14,21,41,42,43,46,49,54,56
"If my wife hadn't supported me, I wouldn't have coped with my psoriasis as well as I do." (21)	
Mothers were also portrayed as supporting and contributing to the cognitive aspects of decision-making. (14)	
Seeking empowering and reassuring relationships	8,10,18,21,22,27,39,42,43,45,46,52,54,56
Being interested in the patient's needs, doubts, and fears, and a thoughtful response to these items were mentioned as important, as was the way of providing information about the medication and its side effects by the rheumatologist. (43)	
"You need someone who understands your situation there and then, why you have become who you are, no one knows, not even yourself. But to have a receiver, someone to sit down and talk with. That, I think is very valuable." (45)	
"If there had been a better dialogue between me and the physicians, we would probably have found an efficient treatment for me earlier…I think I have an overview of what works for me and what doesn't. I'm able to make decisions in collaboration with the physicians and the nurses." (22)	

* Direct quotes from study participants are in quotation marks. Text from the Results and Discussion sections of the articles is in italics. References cited on this table are available in Supplementary Appendix A, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23896/abstract.

Dictating life choices and course. Participants viewed their disease as an overpowering force, described as a "monster" (18) or "the enemy within" (19) that "damaged" (20) their life. The disease changed their life trajectory and led to loss of opportunity in career, study, or relationships: "I was invited to join, go to the college, but I didn't go because I couldn't move or do anything, so I just stayed at home like a prisoner" (20).

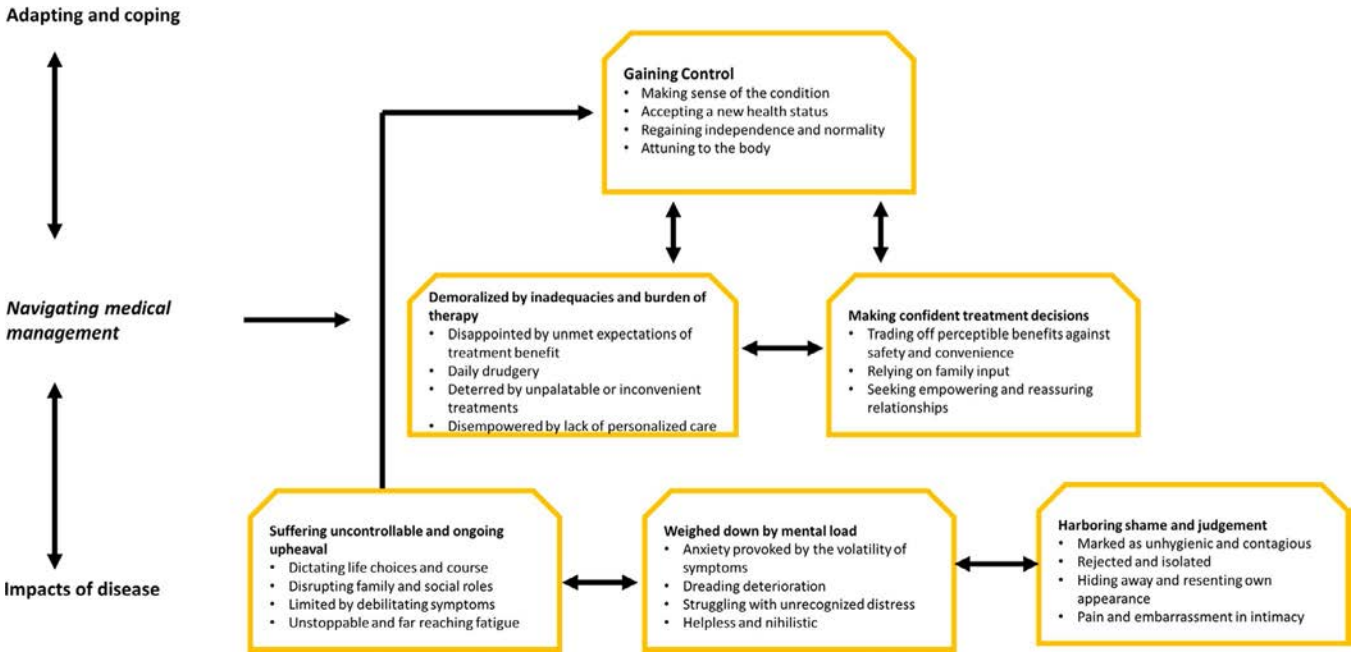

Figure 2. Thematic schema. Patients with psoriasis and psoriatic arthritis contend with continual upheaval in their lives due to debilitating symptoms, fatigue, and disruption to normal roles. Shame associated with skin disease, social isolation, unrecognized mental health burden, and a perception of inadequate management further compounds their psychosocial burden. Patients attempt to cope by making sense of their condition, regaining independence, or adjusting to their new health status. Empowering therapeutic relationships that help patients gain confidence in management through perceptible benefit may overcome fears, meet previously unmet needs, and improve health-related quality of life.

Disrupting family and social roles. Pain and fatigue associated with psoriasis and arthritis took away independence and removed individuals from normal roles in the family: "It has a knock-on effect…on my work, my marriage, my poor husband" (21). Participants were worried about the effects of the disease on their family, describing themselves as a "burden" (20) or questioning their ability to support family.

Limited by debilitating symptoms. Skin symptoms were "burning" (22) and "painful" (22) to the extent that patients wanted to self-harm or "rip (their) skin off" (23). Pain and joint dysfunction were disabling, such that patients could not leave their house or even "brush their hair" (21). For patients who developed arthritis in addition to psoriasis, the burden of symptoms was viewed as cumulative and required a trade-off in improving skin and joint symptoms: "If I adjust my clothing to my rashes, my joints become cold which leads to more pain. If I dress warm because of my joint symptoms, my skin gets worse" (24).

Unstoppable and far-reaching fatigue. For participants with PsA, fatigue was beyond the typical tiredness. Instead, it was conveyed as an all-encompassing sensation of bodily shutdown: "like I needed plugging in…the battery had gone" (21). Fatigue was inextricably linked with a lack of motivation, loss of appetite, and bodily pain.

Weighed down by mental load. "The worry is always there that this is going to get worse and worse" (25).

Anxiety provoked by the volatility of symptoms. Participants were perpetually concerned by unpredictable symptoms. They were always on guard, awaiting an attack of skin or joint flare. For patients with psoriasis, their hypervigilance exacerbated emotional stress, paradoxically leading to a skin flare.

Dreading deterioration. Participants expected that their overall condition would "deteriorate exponentially" (20). Those with PsA believed they had lost normal function permanently and thus were unable to "visualize a future for themselves" (20). The speed of disease progression was also a concern.

Struggling with unrecognized distress. Patients felt unsupported and sometimes angry due to the perceived lack of medical attention or understanding among their family and community about psychological symptoms: "I was about to break down mentally…I wrote it on the questionnaire…but none of the physicians or nurses asked me about it" (26).

Helpless and nihilistic. Patients believed they were "incurable" (27) and thought that treatment and medical care were futile. They struggled with having a lack of control over their disease, leading to hopelessness and a "bleak" (17) future. Some expressed suicidality and considered that death offered a better alternative than the torment of their condition.

Harboring shame and judgement. "You feel ostracized as if you were a monster. The only thing missing is pitch forks and torches" (18).

Marked as unhygienic and contagious. Patients with psoriasis reported being labeled as "dirty" (28) or "contagious" (29) and felt others avoided contact with them for fear of catching psoriasis: "My son was saying how disgusting it was. He didn't like me touching him" (18). This generated embarrassment and shame in patients with psoriasis, prompting a guarded and distrustful approach in social interactions.

Rejected and isolated. Patients with psoriasis held a notion that they were "set apart" (17): "I feel as if I belong to a different race" (29). Being stigmatized by family members, having failed relationships or being bullied because of their condition caused them to feel alienated and segregated from others.

Hiding away and resenting own appearance. Patients with psoriasis internalized feelings of being "disgusting" because they observed that their skin repulsed others. For some females, regarding their skin as flawed led to a sense of reduced femininity: "I battled with feeling unfeminine, ugly…all the things that a young woman should not feel" (19). Patients were embarrassed in public by their skin or swollen joints, covered up and avoided situations where their skin was exposed, including sporting activities.

Pain and embarrassment in intimacy. Patients with psoriasis were self-conscious about their body in sexual encounters, even with established partners. The skin was regarded as intricate to the sensory aspects of intimacy, leading to a sense of being deprived of normal sexual activity. Particularly for men, physical pain due to genital psoriasis or peripheral and axial PsA also meant that participants refrained from being intimate with partners.

Demoralized by inadequacies and burden of therapy. "You need to talk to us more as individuals because clearly…one size doesn't fit all. We know, unfortunately, that it is so different for each of us" (30).

Disappointed by unmet expectations of treatment benefit. Patients with psoriasis were disillusioned by the lack of response to treatment because they had expected that treatment would change their lives. This disappointment led some to feel trapped in a cycle of hopelessness in which a new medication provided initial hope but subsequently failed to live up to the expectation of cure. Some felt resigned to dissatisfaction when clinicians reinforced the hopelessness of treatment: "there's nothing to be done" (29).

Daily drudgery. The routine of managing wet dressings, constant cleaning of skin flakes, and applying topical creams in psoriasis was an arduous daily task that consumed energy and time: "You sit there…and flake off…this can take hours and hours" (31). On the other hand, some patients describe the liberation from daily applications and cleaning afforded by effective and more convenient treatments such as biologic therapies.

Deterred by unpalatable or inconvenient treatments. Topical treatments for psoriasis were "smelly" (32) or "greasy" (32), while disease-modifying antirheumatic drugs (DMARDs) were viewed by some as "poison" (33) with the potential for organ damage or to shorten their lives (26). For some, the need for constant monitoring of treatment with regular blood tests and clinical review exacerbated the fear of long-term medication use. Some discontinued medications because of the cost, having limited access to blood test monitoring, and to avoid side effects, including fatigue and loss of appetite. For some, competing priorities meant they could not access treatments such as daily ultraviolet therapy.

Disempowered by lack of personalized care. At diagnosis, participants felt that information given by clinicians did not prepare them for facing the impacts of disease on their lifestyle. Management was perceived to be automated, mechanistic, or a series of trial and error (30): "If I go back to see my dermatologist for 'thirty seconds,' he just fills out the prescription" (34). Participants felt disregarded if concerns about body image or mental health were not broached or addressed properly in clinical consultations. Some who reported having short consultations with their clinician felt "robbed" (32) by being excluded from treatment decision-making.

Gaining control. "The most important thing is to do my job, and I'm managing that. When I come home, I mostly sleep. I don't have enough energy to be social" (35).

Making sense of the condition. Receiving information and education empowered participants to gain a sense of control over their psoriasis and made them feel valued by their doctors. Of particular importance was understanding the pathophysiology of their condition, comprehending the link between psoriasis and arthritis, and gaining broader knowledge of the treatment available: "I didn't even know there were things for psoriasis other than the ointment" (36). Using written information as a tool to engage patients in care was just as important as the information conveyed: "It would open up a conversation" (36).

Accepting a new health status. Learning to adjust to a chronic disease and accept a different lifestyle involved changing daily hygiene routines and choosing appropriate clothing. Some strived to "learn to live with it" (24) or "not let it get (them) down" (37).

Regaining independence and normality. Regaining independence and a sense of normality was a triumph over illness: "If I could have anything it would be independence…to be able to be as fast as everybody else…able to drive my own car, go out when I wanted to go out, come in and lock my own front door and not have somebody to come in to help with the shower" (25).

Attuning to the body. Patients with PsA learned to recognize bodily pain and fatigue earlier, such that flares of pain could be avoided. Mindfulness techniques and relaxation methods were used to slow pain down and reduce stress associated with skin flares. Proactive changes to diet, exercising, and adopting a healthy lifestyle were perceived to help control disease.

Making confident treatment decisions. "I have always said I want a good quality even though it is short; I don't want to live until I am 90 and be curled up in a ball somewhere. I would rather keep taking the injections and keep going" (25).

Trading off perceptible benefits against safety and convenience. Patients with psoriasis and PsA wanted to see and feel tangible change after taking the medications, which would mitigate fears of medication use: "Patients spoke about pain and not about the preventive effects of taking DMARDs for potential joint damage" (33). Willingness to accept a biologic therapy was higher among those who had already used and experienced the benefits of a biologic. For these patients, fear of needles and side effects was outweighed by perceptible experience of benefit (18).

Relying on family input. Participants depended on family members and partners to provide support in the management of psoriasis and arthritis. Encouragement from wives to seek medical care was described as a stabilizing influence for males otherwise disengaged from medical care. Family members or close friends referenced their own experiences of illness and medication taking and influenced patients' decisions to accept or decline therapy. This influence could either negatively or positively affect patient attitudes toward willingness to accept treatment. For young patients with inflammatory arthritis, including PsA, mothers reinforced positive attitudes by supporting treatment decisions and monitoring adherence.

Seeking empowering and reassuring relationships. Patients wanted continuity of care with clinicians who were experts in treating psoriasis and PsA. Within a consult, patients valued care that used active listening and was individualized, with consideration of issues beyond physical symptoms, such as mental health and stigmatization. Trust and an empathetic communication style were highlighted as reasons to accept and initiate disease-modifying drugs. Participants who felt that clinicians acknowledged their fears and doubts about medications, such as the risk of long-term organ damage or side effects, were more willing to accept treatment. Patients expected clinicians to balance competing messages of reassurance in their management, informing patients of the potential limitations of treatment efficacy, but also conveying confidence and optimism in their management.

DISCUSSION

Patients with psoriasis and PsA felt that the disease disrupted the course of their life, work, and family roles and could not be controlled. They feared deterioration of their clinical condition and felt that their concerns and distresses were trivialized by others. Patients felt burdened by treatment inadequacy, and immunosuppressive medications were regarded as toxic and causing long-term organ damage. They felt empowered when they developed an understanding of the pathophysiology and link between psoriasis and arthritis and gained a broader insight into treatment options to advocate for their use.

Most of these concerns were consistent across patient populations. However, some differences were apparent based on patient demographic, the type of disease, and experience with immunosuppressive medications. Some females were concerned with bodily shame impacting on their self-esteem and intimacy (38). Men with genital psoriasis had greater concerns about impacts on sexual health due to physical pain (37). Younger patients with PsA relied heavily on family and their mothers in making treatment decisions (39). Some males needed the support from their female partners to engage in therapy. Fatigue was severe and pervasive for patients with PsA in particular. Feeling contagious, having bodily shame, and the drudgery of daily treatment routines were specific to patients with psoriasis. Compared to patients who had never used biologic medication, patients who had experienced biologic medications were less likely to fear side effects due to the perceptible benefits seen and felt.

Patients with other rheumatic and skin conditions have expressed similar challenges and beliefs compared with our findings. Stigmatization attributed to visible changes in the skin have also been reported among patients with systemic sclerosis (40). Similar to a recent study on the goals of patients with rheumatoid arthritis (41), patients with PsA and psoriasis valued treatments due to perceptible improvement (symptom alleviation and achieving independence) over preventing damage.

In this review, we used established methodology and software to systematically assess and code all relevant data (14,16). Triangulation among researches in analyzing the findings ensured that the full range and depth of data were reflected in the findings. However, there are some potential limitations to the study. We recognize that the majority of studies were performed in predominantly English-speaking, high-income countries, and thus the transferability of the findings to other populations is uncertain. Due to ethical and feasibility reasons, we were unable to analyze the primary qualitative data of included studies. There were relatively few studies in PsA, highlighting the need for further research in this population.

There are a number of implications for the care of patients with psoriasis and PsA arising from this review. The experience of dealing with 2 overlapping conditions created a cumulative burden compared to patients with inflammatory arthritis or psoriasis alone. Several new biologic agents are now available for treatment of moderate-to-severe psoriasis and PsA. Treatment regimens for psoriasis patients should be tailored to meet the specific needs of the disease severity, the impact on quality of life, the response to previous therapies, and the presence of comorbidities such as PsA (42,43). While the prevalence of depression and anxiety in these conditions is higher than in the general population (44), specific factors that contribute to this problem are less well understood and emphasized in the clinical setting. Our review shows that the psychosocial impacts of disease are severe, with a sense of feeling controlled by disease, shame about skin, and lack of recognition of emotional concerns compounding the patient's

mental health burden. Mood disturbance and skin stigmatization should be actively addressed in medical consultation in both diseases. Using validated mental health screening tools, such as the Hospital Anxiety and Depression Scale, has been shown to improve the rate of depression and anxiety diagnosis in PsA (44). Our review also highlights the need for patients to feel that their mental health is not only uncovered but adequately managed. Health services must be able to provide timely and appropriate mental health referral for patients when illness is discovered.

As with other types of inflammatory arthritis (45), patients are concerned about long-term use of systemic immunosuppressive therapies because of potential organ damage. Blood test monitoring used to ensure medication safety does not necessarily reassure patients because they may interpret the need for monitoring to be a warning sign of the inherent toxicity of medications. These perceptions are at odds with cohort studies, showing that systemic therapy reduces the risk of cardiovascular events in patients with psoriasis (46) and PsA (47), potentially improving mortality (48). Addressing misconceptions of treatment harm, reinforcing messages of reducing cardiovascular risk with adequate disease control, and reassurance about monitoring may help reduce anxiety and improve confidence in taking disease-modifying therapies.

Patients with PsA have less immunosuppressive treatment and more persistent disease activity compared to patients with rheumatoid arthritis (7), and there are low levels of reported use of systemic therapies (49). This analysis shows that initial treatments (particularly topical) may be perceived as inefficacious and coupled with management viewed as automated and may have lasting harmful effects on patient engagement with effective therapies. Addressing negative perceptions of prior treatment is important because current guidelines mandate a therapeutic pyramid, whereby the therapies with the most evidence and efficacy are prescribed after failure to the least effective medications. Efforts are needed to address these initial expectations and dissatisfaction with treatment. Our analysis showed that patients wanted empathetic medical encounters that built trust and actively addressed their concerns of medication taking. In psoriasis, improving patient-provider relationships through effective communication, which builds therapeutic trust and empowerment of patients through education about their disease, improves patient adherence and treatment efficacy (4). While patients wanted greater satisfaction in treatment, they also expected realistic advice on potential treatment efficacy. Anticipating a realistic expectation of disease management, making patients aware of treatment options, signposting the plan for therapeutic escalation, and reassurance in the therapeutic journey may mitigate against treatment disengagement.

Guidelines (12,13) highlight the need for shared decision-making between patients and clinicians. This review highlights the perception that management is not viewed as a shared decision, echoing a recent review on shared decision-making in psoriasis (50). Compared to usual care, patient decision aids have been shown to improve participation in care and align care with patient preferences (51). Patient decision aids have been shown to be useful in rheumatoid predominant inflammatory arthritis (52), and they have been minimally studied in psoriasis and PsA, with no clarity on their effectiveness (50). There is a need for further research in psoriasis and PsA to optimize and evaluate shared decision-making models in these conditions.

Patients with PsA and psoriasis contend with psychosocial challenges due to a sense of life disruption, fear of deterioration, and the perception of their mental health burden and unmet treatment needs. Validation of this burden, aspiring to a holistic management approach, and addressing low treatment expectations and fear of medications may build trusting therapeutic relationships and improve overall quality of life and engagement with medical therapy.

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. Sumpton 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. Sumpton, Kelly, Tunnicliffe, Craig, Hassett, Chessman, Tong.

Acquisition of data. Sumpton, Kelly, Tunnicliffe, Tong.

Analysis and interpretation of data. Sumpton, Kelly, Tunnicliffe, Craig, Hassett, Chessman, Tong.

REFERENCES

1. Potenza MC, Peris K, Berardesca E, Bianchi L, Richetta A, Bernardini N, et al. Use of biological drugs in patients with psoriasis and psoriatic arthritis in Italy: results from the PSONG survey. Dermatol Ther 2018;31:1.
2. Gladman DD, Antoni C, Mease P, Clegg DO, Nash P. Psoriatic arthritis: epidemiology, clinical features, course, and outcome. Ann Rheum Dis 2005;64 Suppl 2:ii14–7.
3. Mishra S, Kancharla H, Dogra S, Sharma A. Comparison of four validated psoriatic arthritis screening tools in diagnosing psoriatic arthritis in patients with psoriasis (COMPAQ Study). Br J Dermatol 2017;176:765–70.
4. Soleymani T, Reddy SM, Cohen JM, Neimann AL. Early recognition and treatment heralds optimal outcomes: the benefits of combined rheumatology-dermatology clinics and integrative care of psoriasis and psoriatic arthritis patients. Curr Rheumatol Rep 2017;20:1.
5. Mercy KM, Gordon KB, Paller AS. Patient satisfaction and quality of life in psoriasis and psoriatic arthritis. JAMA 2014;312:2676–7.
6. Kavanaugh A, Helliwell P, Ritchlin CT. Psoriatic arthritis and burden of disease: patient perspectives from the Population-Based Multinational Assessment of Psoriasis and Psoriatic Arthritis (MAPP) survey. Rheumatol Ther 2016;3:91–102.
7. Helliwell P, Coates L, Chandran V, Gladman D, de Wit M, FitzGerald O, et al. Qualifying unmet needs and improving standards of care in psoriatic arthritis. Arthritis Care Res (Hoboken) 2014;66:1759–66.
8. Van de Kerkhof PC, Reich K, Kavanaugh A, Bachelez H, Barker J, Girolomoni G, et al. Physician perspectives in the management of psoriasis and psoriatic arthritis: results from the population-based Multinational Assessment of Psoriasis and Psoriatic Arthritis survey. J Eur Acad Dermatol Venereol 2015;29:2002–10.

9. Coates LC, Moverley AR, McParland L, Brown S, Navarro-Coy N, O'Dwyer JL, et al. Effect of tight control of inflammation in early psoriatic arthritis (TICOPA): a UK multicentre, open-label, randomised controlled trial. Lancet 2015;386:2489–98.
10. Furst DE, Tran M, Sullivan E, Pike J, Piercy J, Herrera V, et al. Misalignment between physicians and patient satisfaction with psoriatic arthritis disease control. Clin Rheumatol 2017;36:2045–54.
11. Okubo Y, Tsuruta D, Tang AC, Inoue S, Torisu-Itakura H, Hanada T, et al. Analysis of treatment goal alignment between Japanese psoriasis patients and their paired treating physicians. J Eur Acad Dermatol Venereol 2018;32:606–14.
12. Coates LC, Kavanaugh A, Mease PJ, Soriano ER, Acosta-Felquer ML, Armstrong AW, et al. Group for Research and Assessment of Psoriasis and Psoriatic Arthritis 2015 treatment recommendations for psoriatic arthritis. Arthritis Care Res (Hoboken) 2016;68:1060–71.
13. Gossec L, Smolen JS, Ramiro S, de Wit M, Cutolo M, Dougados M, et al. European League Against Rheumatism (EULAR) recommendations for the management of psoriatic arthritis with pharmacological therapies: 2015 update. Ann Rheum Dis 2016;75:499–510.
14. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. BMC Med Res Methodol 2012;12:181.
15. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19:349–57.
16. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008;8:45.
17. Magin P, Adams J, Heading G, Pond D, Smith W. The psychological sequelae of psoriasis: results of a qualitative study. Psychol Health Med 2009;14:150–61.
18. Narayanan S, Guyatt V, Franceschetti A, Hautamaki EL. Disease burden and patient reported outcomes among patients with moderate to severe psoriasis: an ethnography study. Psoriasis (Auckl) 2015;5:25–33.
19. Bundy C, Borthwick M, McAteer H, Cordingley L, Howells L, Bristow P, et al. Psoriasis: snapshots of the unspoken: using novel methods to explore patients' personal models of psoriasis and the impact on well-being. Br J Dermatol 2014;171:825–31.
20. Chisholm A, Pearce CJ, Chinoy H, Warren RB, Bundy C. Distress, misperceptions, poor coping and suicidal ideation in psoriatic arthritis: a qualitative study. Rheumatology (Oxford) 2016;55:1047–52.
21. Moverley AR, Vinall-Collier KA, Helliwell PS. It's not just the joints, it's the whole thing: qualitative analysis of patients' experience of flare in psoriatic arthritis. Rheumatology (Oxford) 2015;54:1448–53.
22. Martin ML, Gordon K, Pinto L, Bushnell DM, Chau D, Viswanathan HN. The experience of pain and redness in patients with moderate to severe plaque psoriasis. J Dermatolog Treat 2015;26:401–5.
23. Pariser D, Schenkel B, Carter C, Farahi K, Brown TM, Ellis CN. A multicenter, non-interventional study to evaluate patient-reported experiences of living with psoriasis. J Dermatolog Treat 2016;27:19–26.
24. Uttjek M, Nygren L, Stenberg B, Dufaker M. Marked by visibility of psoriasis in everyday life. Qual Health Res 2007;17:364–72.
25. Dures E, Hewlett S, Lord J, Bowen C, McHugh N, Tillett W. Important treatment outcomes for patients with psoriatic arthritis: a multisite qualitative study. Patient 2017;10:455–62.
26. Khoury LR, Skov L, Moller T. Facing the dilemma of patient-centred psoriasis care: a qualitative study identifying patient needs in dermatological outpatient clinics. Br J Dermatol 2017;177:436–44.
27. Uhlenhake EE, Kurkowski D, Feldman SR. Conversations on psoriasis: what patients want and what physicians can provide. A qualitative look at patient and physician expectations. J Dermatolog Treat 2010;21:6–12.
28. Bewley A, Burrage DM, Ersser SJ, Hansen M, Ward C. Identifying individual psychosocial and adherence support needs in patients with psoriasis: a multinational two-stage qualitative and quantitative study. J Eur Acad Dermatol Venereol 2014;28:763–70.
29. Khoury LR, Danielsen PL, Skiveren J. Body image altered by psoriasis: a study based on individual interviews and a model for body image. J Dermatolog Treat 2014;25:2–7.
30. Ersser SJ, Cowdell FC, Latter SM, Healy E. Self-management experiences in adults with mild-moderate psoriasis: an exploratory study and implications for improved support. Br J Dermatol 2010;163:1044–9.
31. Blome C, von Usslar K, Augustin M. Feasibility of using qualitative interviews to explore patients' treatment goals: experience from dermatology. Patient 2016;9:261–9.
32. Watson T, de Bruin GP. Impact of cutaneous disease on the self-concept: an existential-phenomenological study of men and women with psoriasis. Dermatol Nurs 2007;19:351–6.
33. Pasma A, Van't Spijker A, Luime JJ, Walter MJ, Busschbach JJ, Hazes JM. Facilitators and barriers to adherence in the initiation phase of disease-modifying antirheumatic drug (DMARD) use in patients with arthritis who recently started their first DMARD treatment. J Rheumatol 2015;42:379–85.
34. Kennedy C. Psoriasis narratives: how qualitative research is of value in dermatology research. Int J Dermatol 2006;45:1107–9.
35. Gronning K, Lomundal B, Koksvik HS, Steinsbekk A. Coping with arthritis is experienced as a dynamic balancing process: a qualitative study. Clin Rheumatol 2011;30:1425–32.
36. Nelson PA, Kane K, Pearce CJ, Bundy C, Chisholm A, Hilton R, et al. 'New to me': changing patient understanding of psoriasis and identifying mechanisms of change. The Pso Well patient materials mixed-methods feasibility study. Br J Dermatol 2017;177:758–70.
37. Cather JC, Ryan C, Meeuwis K, Potts Bleakman AJ, Naegeli AN, Edson-Heredia E, et al. Patients' perspectives on the impact of genital psoriasis: a qualitative study. Dermatol Ther (Heidelb) 2017;7:447–61.
38. Magin P, Heading G, Adams J, Pond D. Sex and the skin: a qualitative study of patients with acne, psoriasis and atopic eczema. Psychol Health Med 2010;15:454–62.
39. Hart RI, Foster HE, McDonagh JE, Thompson B, Kay L, Myers A, et al. Young people's decisions about biologic therapies: who influences them and how? Rheumatology (Oxford) 2015;54:1294–301.
40. Sumpton D, Thakkar V, O'Neill S, Singh-Grewal D, Craig JC, Tong A. "It's not me, it's not really me". Insights from patients on living with systemic sclerosis: an interview study. Arthritis Care Res (Hoboken) 2017;69:1733–42.
41. Hulen E, Ervin A, Schue A, Evans-Young G, Saha S, Yelin EH, et al. Patient goals in rheumatoid arthritis care: a systematic review and qualitative synthesis. Musculoskeletal Care 2017;15:295–303.
42. Kaushik SB, Lebwohl MG. Psoriasis: which therapy for which patient. Psoriasis comorbidities and preferred systemic agents. J Am Acad Dermatol 2019;80:27–40.
43. Kaushik SB, Lebwohl MG. Psoriasis: Which therapy for which patient: focus on special populations and chronic infections. J Am Acad Dermatol 2019;80:43–53.
44. McDonough E, Ayearst R, Eder L, Chandran V, Rosen CF, Thavaneswaran A, et al. Depression and anxiety in psoriatic disease: prevalence and associated factors. J Rheumatol 2014;41:887–96.
45. Kelly A, Tymms K, Tunnicliffe DJ, Sumpton D, Perera C, Fallon K, et al. Patients' attitudes and experiences of disease-modifying antirheumatic drugs in rheumatoid arthritis and spondyloarthritis: a qualitative synthesis. Arthritis Care Res (Hoboken) 2018;70:525–32.

46. Ahlehoff O, Skov L, Gislason G, Gniadecki R, Iversen L, Bryld LE, et al. Cardiovascular outcomes and systemic anti-inflammatory drugs in patients with severe psoriasis: 5-year follow-up of a Danish nationwide cohort. J Eur Acad Dermatol Venereol 2015;29: 1128–34.
47. Lee JL, Sinnathurai P, Buchbinder R, Hill C, Lassere M, March L. Biologics and cardiovascular events in inflammatory arthritis: a prospective national cohort study. Arthritis Res Ther 2018;20:171.
48. Gladman DD. Recent advances in understanding and managing psoriatic arthritis. F1000Res 2016;5:2670.
49. Tveit KS, Duvetorp A, Ostergaard M, Skov L, Danielsen K, Iversen L, et al. Treatment use and satisfaction among patients with psoriasis and psoriatic arthritis: results from the NORdic PAtient survey of Psoriasis and Psoriatic arthritis (NORPAPP). J Eur Acad Dermatol Venereol 2019;33:340–54.
50. Larsen MH, Hagen KB, Krogstad AL, Wahl AK. Shared decision making in psoriasis: a systematic review of quantitative and qualitative studies. Am J Clin Dermatol 2019;20:13–29.
51. Stacey D, Legare F, Col NF, Bennett CL, Barry MJ, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2017;4:Cd001431.
52. Nota I, Drossaert CH, Taal E, Vonkeman HE, Haagsma CJ, van de Laar MA. Evaluation of a patient decision aid for initiating disease modifying anti-rheumatic drugs. Arthritis Res Ther 2016;18:252.

Arthritis Care & Research
Vol. 72, No. 5, May 2020, pp 723–733
DOI 10.1002/acr.24127

American College of Rheumatology
Empowering Rheumatology Professionals

Multicenter Qualitative Study Exploring the Patient Experience of Digital Ulcers in Systemic Sclerosis

Michael Hughes,[1] John D. Pauling,[2] Jennifer Jones,[3] Christopher P. Denton,[4] Robyn T. Domsic,[5] Tracy M. Frech,[6] Ariane L. Herrick,[7] Dinesh Khanna,[8] Marco Matucci-Cerinic,[9] Lorraine McKenzie,[10] Lesley Ann Saketkoo,[11] Rachael Gooberman-Hill,[12] and Andrew Moore[3]

Objective. Digital ulcers (DUs) are a major cause of disease-related morbidity and are a difficult-to-treat vascular complication of systemic sclerosis (SSc). Demonstrating treatment efficacy has traditionally focused on clinician assessment of DUs alone. No existing patient-reported outcome (PRO) instrument captures the multifaceted impact of SSc-DU. We report the findings of a multicenter qualitative research study exploring the patient experience of SSc-DU.

Methods. Patient focus groups were conducted across 3 scleroderma units, following a topic guide devised by SSc patients, experts, and experienced qualitative researchers. A purposive sampling framework ensured that the experiences of a diverse group of patients were captured. Focus groups were audio recorded, and information was transcribed, anonymized, and analyzed using inductive thematic analysis. We continued focus groups until thematic saturation was achieved.

Results. Twenty-nine SSc patients with a history of DU disease participated in 4 focus groups across the UK (Bath, Manchester, and London). Five major interrelated themes (and subthemes) were identified that encompass the patient experience of SSc-DU: disabling pain and hypersensitivity; deep and broad-ranging emotional impact; impairment of physical and social activity; factors aggravating occurrence, duration, and impact; and mitigating, managing, and adapting.

Conclusion. The patient experience of SSc-DU is multifaceted and comprises a complex interplay of experiences associated with significant pain and morbidity. Patient experiences of SSc-DU are not captured using existing SSc-DU outcomes. Our findings will inform the development of a novel PRO instrument to assess the severity and impact of SSc-DU for use in future SSc-DU clinical trials.

INTRODUCTION

Digital ulcers (DUs) are a major cause of pain and disability in patients living with systemic sclerosis (SSc) (1). DUs are common, with approximately half of patients reporting a history of ulceration, and 5–10% of patients with SSc at any time have a current ulcer (2,3). DUs have a major impact on the quality of life and hand function, including occupation (4). Although we have a number of treatments available to both prevent and heal SSc-DU (5–8), a third of patients are affected by refractory DU disease (9).

In general, demonstrating treatment efficacy in previous clinical trials has been based on clinician assessment of ulcer healing and/or new ulcer occurrence alone (1). However, the agreement among SSc experts to classify SSc-DU is poor to moderate at

Supported by the Scleroderma Clinical Trials Consortium and the NIHR Manchester Biomedical Research Centre.

[1]Michael Hughes, BSc (Hons), MSc, MBBS, MRCP (UK) (Rheumatology), PhD: University of Manchester, Manchester, Royal Hallamshire Hospital, and Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK; [2]John D. Pauling, BMedSci, BMBS, PhD, FRCP: Royal National Hospital for Rheumatic Diseases and University of Bath, Bath, UK; [3]Jennifer Jones, PhD, Andrew Moore, BSc (Hons), PhD: Bristol Medical School, Bristol, UK; [4]Christopher P. Denton, PhD, FRCP: Royal Free Hospital, University College London, London, UK; [5]Robyn T. Domsic, MD, MPH: University of Pittsburgh Medical Center, Pittsburgh, PA; [6]Tracy M. Frech, MD, MS: University of Utah and Salt Lake Veterans Affairs Medical Center, Salt Lake City; [7]Ariane L. Herrick, MD, FRCP: University of Manchester, Salford Royal NHS Foundation Trust, and Manchester Academic Health Science Centre, Manchester, UK; [8]Dinesh Khanna, MD, MS: University of Michigan, Ann Arbor; [9]Marco Matucci-Cerinic, FRCP, FBSRhon: University of Florence, Florence, Italy; [10]Lorraine McKenzie: Patient representative (contact via Dr. Herrick, University of Manchester, Manchester, UK); [11]Lesley Ann Saketkoo, MD, MPH: Tulane University School of Medicine, New Orleans Scleroderma and Sarcoidosis Patient Care and Research Center, and University Medical Center Comprehensive Pulmonary Hypertension Center, New Orleans, Louisiana; [12]Rachael Gooberman-Hill, PhD: Bristol Medical School, NIHR Bristol Biomedical Research Centre, and University Hospitals Bristol NHS Foundation Trust, Bristol, UK.

Dr. Hughes has received speaking fees from Actelion (less than $10,000). Dr. Pauling has received consulting fees from Actelion and Boehringer Ingelheim (less than $10,000 each) and has received speaking fees and research grants from Actelion (more than $10,000). Dr. Denton has received consulting fees from Roche, Actelion, GlaxoSmithKline, Sanofi, Inventiva, Boehringer Ingelheim, Leadiant, Galapagos, and Bayer (less than $10,000 each) and has received research grants from

SIGNIFICANCE & INNOVATIONS

- Existing systemic sclerosis (SSc)–digital ulcer (DU) outcome measures do not capture the complete patient experience of SSc-DU.
- The patient experience of SSc-DU comprises interrelated factors that contribute to the significant morbidity of SSc-DU.
- Five major interrelated themes were identified: disabling pain and hypersensitivity; deep and broad-ranging emotional impact; impairment of physical and social activity; factors aggravating occurrence, duration, and impact; and mitigating, managing, and adapting.
- The interplay between the themes suggests that the presence of SSc-DU can have a considerable impact on patients' physical and psychological wellbeing, impairing physical and social activities, and that patients expend great effort in remaining vigilant and managing their condition, often in innovative ways.
- Our findings can be used to inform the development of a novel patient-reported outcome instrument to assess the severity and impact of SSc-DU.

best (10–12). Interrater agreement is not improved with the provision of clinical (real-world) contextual information (e.g., the severity of pain and duration of the lesion) (11). Recent negative clinical trials of promising therapies for SSc-DU (13,14) have led to calls for a fresh approach to establishing treatment efficacy in SSc-DU (15–17).

No studies have specifically explored the patient experience of SSc-DU, although studies examining broader symptom burden in scleroderma have identified the major impact that SSc-DU can have for patients, as the following quotation attests: "The pain that you felt in your fingers as they were dying was so excruciating that you almost begged to say please cut it off" (reproduced from [18]).

Previous attempts to quantify the impact of SSc-DU have used legacy patient-reported outcome (PRO) instruments to assess broader aspects of SSc disease severity and function (19,20). There was limited or no SSc patient participation in the development of many of these instruments (21). The patient perspective captured by PRO instruments provides insight into the patient experience of disease that can not be assessed using clinician-reported instruments (22). Regulatory bodies, such as the Federal Drug Administration, seek target patient population involvement in PRO instrument development to ensure that instruments fully capture the way patients feel and function (23). A thorough understanding of the patient experience of SSc-DU is necessary to ensure that a future PRO instrument captures the multifaceted impact of DUs. Against this background, the aim of the current study was to comprehensively explore the experiences, attitudes, and perspectives of patients with SSc-DU. A further aim was to inform the development of a future SSc-DU PRO instrument.

PATIENTS AND METHODS

Study management. The development and conduct of the study were overseen by a dedicated steering committee that comprised SSc experts (MH, JDP, CPD, RTD, TMF, ALH, DK, MM-C, LAS), 2 patient research partners, and a team of experienced qualitative methodologists. The study was approved by the East Midlands–Nottingham 1 Research Ethics Committee (18/EM/0018), and all participants provided written informed consent.

Study design. A multicenter qualitative research study comprising patient focus groups was undertaken at scleroderma centers across the UK (Bath, Manchester, and London). Patient focus groups create an open environment in which a broad range of experiences can be expressed and explored and can often enable some (but not necessarily all) sensitive issues to be discussed more freely than in a one-to-one interview setting (24).

Participants. Adult SSc patients (ages >18 years) with a history of SSc-DU, fulfilling the 2013 American College of Rheumatology/European League Against Rheumatism classification criteria for SSc (25), and with sufficient language skills to participate in a focus group discussion were enrolled. A purposive sampling framework ensured the enrollment of a diverse cohort comprising a 60:40 split between limited and diffuse cutaneous SSc (26), early and established disease (≤3 and >3 years since first non-Raynaud's phenomenon symptom, respectively), a spectrum of historical DU disease, sex (aiming for 5:1 female predominance), and ethnicity (e.g., with Caucasian and Black British). The focus groups sought to include 6–10 participants to enable

GlaxoSmithKline, CSL Behring, and Inventiva. Dr. Domsic has received consulting fees from Eicos Sciences (less than $10,000). Dr. Herrick has received consulting fees from Boehringer-Ingelheim, Camurus, and Gesynta (less than $10,000 each), has received speaking fees from Actelion (less than $10,000), and has received research funding from Gesynta. Dr. Khanna has received consulting fees from Actelion, Acceleron, Arena, Bayer, Bristol-Myer Squibb, Chemomab, GlaxoSmithKline, Genentech/Roche, and Galapa Mitsubishi Tanabi (less than $10,000 each) and from CSL Behring, Corbus, and Boehringer Ingelheim (more than $10,000 each), has received grants from Bayer, Bristol-Myers Squibb, and Pfizer, and has stock options in Sciences, Inc. Dr. Matucci-Cerinic has received speaking fees from Actelion and Biogen (less than $10,000 each). Dr. Saketkoo consulting fees from Boehringer Ingelheim and Eicos Sciences (less than $10,000 each) and from CSL Behring and Mallinckrodt/United Therapeutics (more than $10,000 each) and has received research grants from Boehringer Ingelheim and Eicos Sciences. No other disclosures relevant to this article were reported.

Address correspondence to Michael Hughes BSc (Hons), MSc, MBBS, MRCP (UK) (Rheumatology), PhD, Consultant Rheumatologist, Department of Rheumatology, Royal Hallamshire Hospital, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, S10 2JF, UK. E-mail: Michael.hughes-6@postgrad.manchester.ac.uk.

Submitted for publication July 15, 2019; accepted in revised form December 10, 2019.

open discussion, while ensuring that each participant had the opportunity to express their personal experiences, interact, and offer alternative opinions should they wish. A minimum of 2 to 3 focus groups was expected to be necessary to achieve thematic saturation, but the intention was to continue enrollment until there was consensus that no meaningful new experiences were being shared by participants or that warranted further exploration by the investigators (27).

Data collection. Each focus group lasted approximately 1 hour and all were facilitated by MH, with the first focus group also facilitated by JDP and AM to ensure that there were no issues including a need to revise the topic guide (which was not the case). Focus groups were facilitated by rheumatologists (MH and JDP) with experience in the clinical heterogeneity and management of patients with SSc. Focus groups were held within hospitals but outside of clinical areas, in a quiet ambient environment without external distraction. The focus group lead facilitator (MH) is a rheumatologist with an interest in SSc and was not directly involved in the clinical care of the participants. JDP is a rheumatologist with an interest in SSc, and AM is an experienced qualitative researcher/methodologist. A relaxed environment in which each participant's views were sought, valued, and respected enabled individuals to share experiences of SSc-DU, and that setting allowed others to express similar or opposing views. The focus groups were audio recorded, and information was subsequently transcribed verbatim, with all the context anonymized. A topic guide was developed with input from the study steering committee (see Supplementary Appendix A, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24127/abstract). Each focus group started with broad open questions asking participants to describe their experience of their disease and DU history. Focus groups adopted an adaptive study design enabling incompletely explored or newly emerging themes to be investigated to ensure that thematic saturation was achieved.

Data analysis. Qualitative analysis was conducted by JJ and AM, both experienced qualitative methodologists, with further input from the wider team (MH, JDP, RG-H, and patient partners). NVivo 11 software was used to manage and interrogate the data. Transcribed data were analyzed using thematic analysis (27). First, JJ read and reread transcripts to ensure familiarity with the content. Information relevant to patients' experience and understanding of DUs was then coded using descriptive labels. Codes that occurred repeatedly, or that shared conceptual similarities, were then grouped together to form initial categories. The

Table 1. Demographics and clinical phenotype of enrolled participants according to purposive sampling framework*

Demographics/clinical phenotype	Bath	Manchester	London (1)	London (2)	Overall
Participants	8	7	6	8	29
Age, mean ± SD years	66.1 ± 12.6	61.6 ± 12.2	50.4 ± 12.4	59.5 ± 12.8	59.9 ± 13.3
Sex F:M ratio	7:1	7:0	3:3	3:5	20:9
Disease subtype					
LcSSc	8	6	2	4	20
DcSSc	0	1	4	4	9
RP duration, mean ± SD years	20.7 ± 19.9	17.9 ± 15.9	23.1 ± 22.1	13.6 ± 9.5	18.5 ± 16.6
Disease duration, mean ± SD years†	14.3 ± 11.2	10.9 ± 7.3	13.9 ± 12.6	13.2 ± 12.2	12.8 ± 9.7
Early vs. established disease‡					
Early	0	1	1	0	2
Established	8	6	5	8	27
History of DU					
1 previous	1	1	0	1	3
2–4 previous	3	3	2	1	9
≥5 previous	4	3	4	6	17
Ethnicity					
White/Caucasian	7	6	5	5	23
Black British	0	1	1	2	4
Asian	1	0	0	1	2
Vasodilator medication used§					
None	1	2	1	2	6
Calcium channel blocker	5	2	1	2	10
Phosphodiesterase type-5 inhibitor	5	4	4	5	18
Endothelin receptor antagonist	3	2	2	2	9

* Values are the number unless indicated otherwise. LcSSc = limited cutaneous systemic sclerosis; DcSSc = diffuse cutaneous systemic sclerosis; RP = Raynaud's phenomenon; DU = digital ulcer.
† Since first non-RP symptom.
‡ Early and established disease (≤3 and >3 years since first non-RP symptom, respectively).
§ Indication not specified and includes SSc-RP, SSc-DU, SSc-pulmonary artery hypertension, and/or systemic hypertension/cardiovascular risk.

initial set of codes and categories was then discussed with the wider team (MH, JDP, RG-H, and AM) to ensure they captured all elements from the focus group. The coding framework was then applied to subsequent transcripts, and any newly identified codes were added as appropriate. The focus group facilitators decided when data saturation had been reached (28). Codes were collated and grouped into themes and subthemes. Coded data within each theme were checked to ensure internal coherence (fit within the pattern of the theme) and external representativeness (fit within the whole data set). JJ and AM regularly discussed the conceptual development of the themes and subthemes and an analysis debriefing meeting was convened involving JJ, JDP, RG-H, and AM to discuss the final theme groupings and the conceptual map describing the interrelationship of the respective themes.

Our approach was both deductive, in the sense that the research team examined preconceived considerations on the impact of DUs (derived from an earlier comprehensive literature review (20) and how participants understood and managed them, for the purposes of developing a PRO instrument, and inductive, in the sense that there was no preexisting coding frame and the developing codes were derived from and grounded in the data themselves (29).

RESULTS

Twenty-nine patients with SSc participated in 4 focus groups conducted in Bath (n = 8), Manchester (n = 7), and 2 focus groups in London (n = 6 and n = 8). Our a priori purposive sampling framework ensured that we studied a broad population of patients with SSc and DU disease (Table 1). Thematic saturation was felt to have been achieved after 4 focus groups.

Five major themes emerged that together constitute the patient experience of SSc-DU: 1) disabling pain and hypersensitivity, 2) deep and broad-ranging emotional impact, 3) impairment of physical and social activity, 4) factors aggravating occurrence, duration, and impact, and 5) mitigating, managing, and adapting to ulcers. The 5 constituent themes (and subthemes) can be arranged within a conceptual map of the patient experience of SSc-DU (Figure 1).

Theme 1: disabling pain and hypersensitivity. Our study found that pain is a cardinal symptom of SSc-DU and is often very severe (question [Q]1–4) (Table 2). Participants used a wide range of words and phrases to describe the severity of pain such as: "excruciating," "pain that could reduce you to tears," "agonizing," and "unbearable." Participants often described the pain as pulsatile or throbbing in nature (Q5, Q6), including a pressure-like effect (Q7). Not all participants used the word "pain" to describe the physical discomfort of SSc-DU; other expressions included "soreness," "tenderness," or "discomfort." The level of reported pain was often considered as being disproportionate to the size of the DU (Q8). DU pain can radiate to the other digits and proximally (Q9, Q10). Coexistent infection of the ulcer increases DU pain (Q11), and some participants reported that changes in temperature can worsen DU pain (Q12, Q13). Many participants described pain in the areas where previous ulcers had occurred, whereas others said the area was tender, sore, or sensitive and could be aggravated by touch or exposure to cold (Q13–15). Other sensations in areas of previous ulcers included tingling nerve-like sensations and partial or complete numbness (Q16, Q17). One participant said, "It's never the same again" (participant 6, Manchester group) when talking about the area where previous ulceration had occurred. Due to the severity of DU pain, some participants suggested that invasive procedures (including digital amputation) may be both necessary and appropriate to relieve symptoms (Q1, Q4, Q5, Q7, Q18, Q24). Across all the focus groups, participants

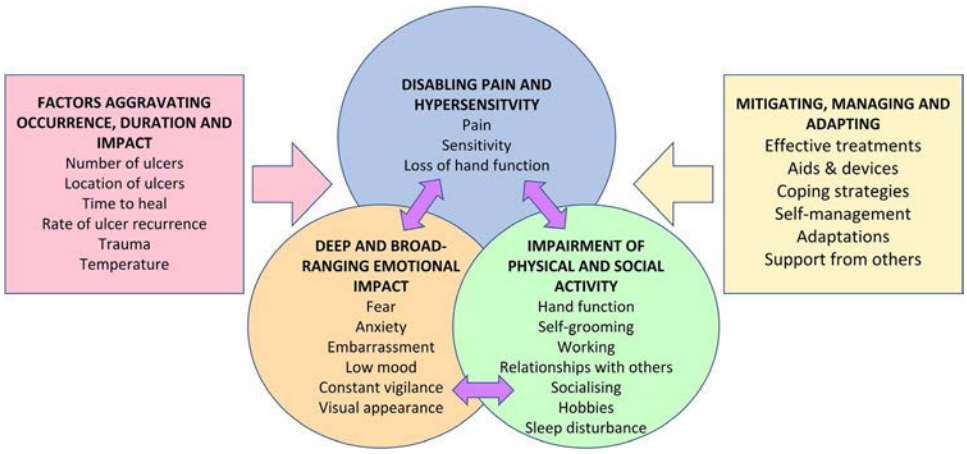

Figure 1. A conceptual map comprising the 5 major interrelated themes that constitute the patient experience of digital ulcers and systemic sclerosis.. The manifestation of pain that is often unbearable affects both the day-to-day functioning of the individual and their psychological well-being. For example, an inability to physically manipulate the world through their hands can lead to avoidance of activities or social interaction and subsequently cause low mood. This impairment can be supported through the use of aids and devices, such as gloves, or with help from other people.

Table 2. Quotes supporting the "disabling pain and hypersensitivity" theme of the patient experience of SSc-DU*

Subtheme: Q (subject and group)	Quotation
Pain:	
1 (P1 M1)	The pain is just unbearable, in fact you just want to chop your finger off, don't you? You think, well, I'd rather have my finger chopped off than have that pain. I've got to the point where I think just take it off. I can't stand it.
2 (P8 B1)	When the pain is really bad you, you just rock back and forward like this.
3 (P1 M1)	The pain, I just wanted to sit on the floor and cry…the pain is the worst thing I've had.
4 (P6 M1)	I just want it off. It needs to go, it gets that bad. You think, sorry, you feel like you want to bang your head to refer the pain somewhere else, just to relieve it.
Pulsatile/throbbing:	
5 (P7 L2)	You just want to take your finger off, that's how bad it is. The pulsating pain…
6 (P7 L2)	Like someone's getting a nail and hammering a nail right through the tip…And keep going and going, because it just keeps going through the finger.
Pressure: 7 (P2 B1)	If I could have taken my nail off just to release the pressure I would have done.
Pain disproportionate to ulcer size: 8 (P5 M1)	It's quite incongruous the amount of pain from the minimal amount of disruption to your thumb.
Radiation:	
9 (P2 B1)	The pain started actually in the finger bed, and I could feel it tracking along the finger and it dipped down into the first joint, so I could actually feel the pain in between the 2 joints.
10 (P7 B1)	So the ulcer is in the middle but I'll still get pain in the index and ring finger, which is equivalent to the ulcer pain but there's nothing there.
Infection: 11 (P5 L1)	I try not to get mine infected because then the pain level goes up.
Temperature:	
12 (P2 L1)	It's almost impossible to go in, in the summer when they've got the air conditioning on, it's not just the frozen aisles, it's the whole supermarket…if you've got an ulcer, the change in temperature will make the ulcer sensitive like a nerve, you can really feel it.
13 (P1 B1)	I don't go near the freezer for that reason, but even a cold bottle of milk in the winter, if you take it out of the fridge that's enough to set things off…where I've had the ulcers, particularly that one it, it becomes painful.
14 (P3 M1)	It's really tender if I just catch me finger now, but luckily I haven't had any more since then, it just left a lot of tenderness on, on the tips of me fingers… it's just the pain where I had the ulcer that's where it's straight away, the cold, as soon as I go out.
Pain/sensitivity at sites of past ulcers:	
15 (P1 L1)	It's the very end of the fingers, it's extremely sensitive, and it doesn't matter if it looks like an ulcer sort of wound, or it might be completely healed up, it can still be extremely sensitive to touch.
16 (P3 B1)	Just a slight tingly nerve sensation now, no pain.
Numbness at sites of past ulcers: 17 (P1 B1)	It's a bit numb.
Considered need for invasive procedures: 18 (P7 M1)	It's just so painful that the idea of cutting my finger open to take it out seems better than having that pain all the time.
Need to validate pain:	
19 (P4 B1)	Going back to what you said just now about people seeing it, sometimes you almost want to show, because you can't explain the pain you get with them, you almost want to show people this is what it's causing. My family's seen them obviously but I couldn't get it across.
20 (P1 B1)	Seems a bit feeble ringing and saying I can't come to work 'cause my finger's hurting, doesn't it?
21 (P2 L1)	Some people don't understand the pain we're going through.
22 (P1 M1)	But you could cry with them, it is, you could sit down and cry, and you can't explain to anyone in your family how bad the pain is.
Description of severity:	
23 (P5 L1)	It was very difficult to be an electrician. I think the difficulty is the severity of the winter, as you get the ulcer appear during the winter and then it's the amount of time after the winter they take to heal up.
24 (P5 B1)	I've had the 2 digital ulcers, touch wood that's healed up. I thought I was going to lose this finger at 1 stage.

* Q refers to the numbered quote cited in the text. SSc = systemic sclerosis; DU = digital ulcer; P = participant; B1 = Bath group; M1 = Manchester group; L1, L2 = London groups.

talked about the need to validate the pain they experienced with friends, family, and colleagues (Q19–22). Participants described the severity of their ulcers in different ways. These included the need for hospitalization, the time to heal, changes in their life (e.g., giving up work or hobbies) due to ulcers, and previous/risk of amputation (Q23, Q24). There was a wide variety in the reported location (fingertips, over the small joints, under the nails, and on the sides of the fingers) of DUs among participants. Some experienced ulcers in different locations on the hands, whereas others tended to only get ulcers in 1 area.

Theme 2: deep and broad-ranging emotional impact. Related to the severity of pain, most participants shared a constant fear of the development of new DUs (Q25), and many considered the development of further lesions inevitable (Q26, Q27) (Table 3). Participants experienced anxiety/uncertainty regarding how severe

Table 3. Quotes supporting the "deep and broad-ranging emotional impact" theme of the patient experience of SSc-DU*

Subtheme: Q (subject and group)	Quotation
Fear: 25 (P1 M1)	I don't particularly want to go out when I've got one because I'm so frightened of getting in the car and banging or, you know, picking my keys up and banging it.
Unavoidable recurrence of DUs:	
26 (P5 L1)	I don't know if it's a good or bad thing but I've got used to having them, so it becomes a way of life…when I used to maybe have 1 a year, I used to think it was quite a big deal but then since getting 5 or 6 a year, it doesn't become a big deal any more, you just get used to it.
27 (P4 L2)	Because you can probably guarantee you are going to get another one sometimes…I don't see how you can prevent it, if it's going to happen, it's going to happen. I don't see how it could be.
Anxiety/uncertainty: 28 (P7 M1)	I'm still learning about the whole thing so it changes every day. I call it the Hunger Games, when something starts to get better something else happens and you don't know what is happening, so the answer is I don't know what brings them, I don't know what I do wrong or not wrong…it's one of the worst things about the disease because it makes you scared and it makes you nervous, irritable.
Depression/anger/uncertainty about the future:	
29 (P2 L2)	It affected me quite a bit, yes…it really depressed me at that time.
30 (P1 M1)	It's like a black cloud, isn't it? It doesn't tend to go away, does it? Some days you just think, well, I'm not thinking about it and then other days it gets you down a bit, don't it?"
31 (P7 M1)	Just angry all the time, because you have to be conscious and you can't relax.
32 (P6 L2)	It really, it ruins the day, it changes your life.
Constant vigilance:	
33 (P3 M1)	You do feel very cautious, if you, if you do have a bang then you're more aware that you're not to do things for the next few days in case it goes really bad.
34 (P1 M1)	Well I've got to be particularly careful now if me nails grow, especially at the side I've got to try and cut them…and then of course you're worried when you cut them that you're not going to do any damage as well, so it's a bit difficult really.
Anger: 35 (P7 M1)	Just angry all the time because you have to be conscious and you can't relax… And it affects you, yes, it affects you and it affects the kids, it affects everything around you. You have to tell yourself all the time, you've got this, you have to, you have to remember your hand all the time.
Embarrassment/hiding/protecting ulcers:	
36 (P8 B1)	I used to hide mine under the table cloth at a function…Embarrassment, probably.
37 (P4 B1)	You don't want other people to be distressed at seeing them, also it's protection against infection.
38 (P2 B1)	So I kept them covered up and I've got photographs in my bag that I took for my own record really, you know, and my son said last time "don't you let me see those, I don't want to see them," but even the doctors never looked at my fingers when I had the ulcers.
39 (P7 L2)	Sometimes it looks awful, all the skin peeled back and it's all exposed, yeah, you just hide it…I just don't want people to look at it as well, I feel conscious sometimes.
40 (P4 B1)	If I was going out to a social function or meeting friends or something I would put plasters on, because it's better for someone to see plasters than, you know, and your friends get used to the fact of, how's your hands, you know.

* Q refers to the numbered quote cited in the text. SSc = systemic sclerosis; DU = digital ulcer; P = participant; B1 = Bath group; M1 = Manchester group; L1, L2 = London groups.

each ulcer would be, whether they were treating the ulcer correctly, and how long it would take the ulcer to heal (Q28). Although most participants did not explicitly say that ulcers caused them depression (Q29), they mentioned many associated emotions (in addition to anxiety and embarrassment), including uncertainty/fear for the future and anger (Q29–32). Participants described the need for a constant level of vigilance to prevent the development of new DUs and infection of intercurrent ulcers (Q33, Q34). Participants described many different emotions associated with the ulcers, from panic, anxiety, fear, and irritability to anger (Q35). Participants did not forget about the past impact of the ulcers, and some described frightening times (with current ulcers) when they were perhaps unsure whether they would need to have part of their finger amputated (Q24). Patients also experienced embarrassment and distress due to the physical appearance of SSc-DU and took a range of actions to hide DUs from others (Q36–40).

Theme 3: impairment of physical and social activity. The physical and psychological impact of SSc-DU was closely related to impact on physical and social functioning (Table 4). Patients' interactions with the world and other people were characterized by an avoidance of pain and a constant vigilance during physical and social interaction. Participants reported how DUs impacted their ability to use their hands during activities of daily living (Q41–46), including self-care/grooming (Q38, Q47, Q48), hobbies (Q49), and domestic activities (e.g., cooking and household chores) (Q13, Q50, Q51). Activities of daily living that were taken for granted became foregrounded, such as the ability to reach their hands into pockets, a bag, or a purse (Q45, Q52, Q53), difficulty driving (Q30, Q54), sleeping (Q55), and challenges when shopping (Q12, Q56, Q57). The impact of DUs on work varied among the participants. For some participants, ulcers had not severely impacted their work, whereas others had to change

Table 4. Quotes supporting the "impairment of physical and social activity" theme of the patient experience of SSc-DU*

Subtheme: Q (subject and group)	Quotation
Hand function and activities of daily living:	
41 (P1 L1)	Just trying to handle things with your fingers, you just have to careful you don't drop a tea cup, your dexterity goes.
42 (P5 L1)	Where the ulcers were, sort of like stop the movement in your hands so I wasn't able to do these things that I needed to do.
43 (P4 L2)	To actually bend the fingers where your ulcers are actually on top of the knuckles is practically impossible.
44 (P4 B1)	That's the thing, that's what I say, I can get things done, but I cannot do it at the speed that I used to before.
45 (P5 L1)	Putting things in bags, lifting stuff, you can't actually grip stuff so I just feel really clumsy.
46 (P6 M1)	It's like opening a bag of crisps if you're out for a drink, I can't open the crisps.
Self-care/grooming:	
47 (P1 M1)	I'm frightened of catching it. You don't want to get dressed in case you've got to zip something up and you catch it.
48 (P5 L1)	Even just going and brushing our teeth…it's painful when our hands are sore and ulcerated.
Hobbies: 49 (P2 B1)	I've had to stop doing things like knitting… Because they flare up straight away and open and it doesn't matter whether I use natural fibers, it's just the irritation of my skin so I had to give up knitting… I have to be very careful gardening.
Domestic activities:	
50 (P6 L1)	Two years ago I can do nothing really, so I needed help, my daughter, husband, everyone doing something at home. I could do nothing, cooking.
51 (P8 L2)	It's impossible to make the bed, I can't put my hand, I can't put the sheet under.
Putting hands in pockets/bags/purse:	
52 (P4 M1)	When it starts to crust over that, that's when I can't go in me bag, you know, and you just tip everything out to find what you want and then scoop everything back up.
53 (P1 L1)	Putting your hand in your pocket can be horrendous if you hit a key or something like that.
Difficulty driving: 54 (P4 B1)	Things to try and protect it, 'cause you're guaranteed knocks on every single day, you carefully put the ignition key in the car, you still knock this one on the steering wheel and things like that.
Sleep disturbance: 55 (P1 M1)	It's like somebody's sticking a needle in your finger when you're trying to go to sleep, you could hold your hand in the air.
Shopping:	
56 (P4 L1)	Going to supermarkets I can't go up and down the fridge aisle. I have to stand there and wait and think about do I need anything down there, but even just going into a supermarket, it's just too cold… Because you have to balance your bags so that you can carry them, if they're rushing you, you're just dropping everything in and it's all falling out and it just becomes a disaster.
57 (P3 L1)	There's always that doubt in the checkouts, you know, not only are they not hassling me, but I'm sort of thinking I'm holding the queue up and I suddenly hear this voice behind me saying, "you don't have to rush you know." People are nice I find.
Change in working/occupation:	
58 (P3 L1)	Well I was a programmer, so it wasn't a difficult job to carry on doing.
59 (P7 B1)	Obviously it's affected a lot of the work that I do as well. There's only 50% of the work that I used to do that I can continue to do now, with the digital ulcers, but it's just knowing what you can and can't get away with anymore.
Financial concerns: 60 (P7 B1)	Most people have said you need to change your job, but once you're set up and you're established and you've got a wife, kids, a mortgage and bills to pay, it's impossible to go back and start as tea boy again somewhere else, so you carry on but you've got to try and adjust what you do to maintain your income, that's the biggest difficulty I've had so far.
Concealing ulcers:	
61 (P4 B1)	You don't want other people to be distressed at seeing them, also it's protection against infection and also, you know, if you're going out to any social function I will bandage… I did go to my daughter's wedding, which was in all of this, and so I did wear my black gloves all through the wedding.
62 (P5 L1)	I don't know if it's them or myself thinking, oh are they thinking I'm contagious or that kind of thing, because they look horrible when they're at their worst, but now I'll try to keep them, I'll keep them covered if they're… I wouldn't go out anywhere without them being covered, but still when you're covered in a million plasters, that doesn't look nice either.
Change in caring roles within the family: 63 (P7 B1)	It changes the way you have to think of it, everything that you do. I mean the wife says to me, do you want to take the kids down the fair, and the first thing I have to do is check the temperature outside, you know. If it's 20° or less, I'll bail out, I wouldn't bother going, but it's not nice because you miss out on a lot of life experiences with your family.

* Q refers to the numbered quote cited in the text. SSc = systemic sclerosis; DU = digital ulcer; P = participant; B1 = Bath group; M1 = Manchester group; L1, L2 = London groups.

Table 5. Quotes supporting the "factors aggravating occurrence, duration, and impact" and "mitigating, managing, and adapting" themes of the patient experience of SSc-DU*

Subtheme: Q (subject and group)	Quotation
Factors aggravating occurrence, duration, and impact	
Number of ulcers:	
64 (P3 M1)	I have only had 1 ulcer, which was really quite bad. And it, I was put on a drip in hospital with, is it epoprostenol, twice to see if that would help, but it didn't and they ended up going to theatre to have it cleaned out, and that's the only ulcer I've ever had.
65 (P4 B1)	I was diagnosed with limited scleroderma approximately 28 years ago, which started with an ulcer in 1 finger and just gradually got worse over the years, with anything up to 4 or 5 ulcers every winter, which sometimes cleared up in the summer, yeah, so on-going.
66 (P2 L1)	When I was first diagnosed, ulcers weren't really a problem. I might have 1 a year, but as the scleroderma has progressed, I have had up to 10 ulcers at a time on my hands, in different degrees of severity.
Ulcers heal slower in the winter:	
67 (P4 M1)	They get easier in the summer, they heal better.
68 (P4 L1)	They just would erupt through the whole winter, and then I've got to wait till the middle to the end of the summer, then I get a short respite.
Mitigating, managing, and adapting	
Indication that treatment is effective:	
69 (P1 B1)	Being able to sleep during the night with the bearable pain would be an absolutely added bonus.
70 (P7 B1)	Well, within 5 days that finger healed up more than it did in 3 months, so the minute I came in on the iloprost…certainly the 5 days I spent here last week, I wouldn't be as healed up as I am now, and able to work again.
71 (P7 L2)	If the pain stops.
72 (P4 L2)	It helps it, calm it down, to stop being hurting.
73 (P5 L1)	I just find it keeps them at bay. I worry that if I was to lengthen it again it would just be worse, so yeah, it sort of helped the aggression that you say, the inflammation and things.
Burden of treatments:	
74 (P4 B1)	I think I would definitely say it's helped a lot and it's kept me out of hospital. I've managed, the ulcers are still taking several weeks, if not months to heal, but they do heal without the need to intervene with iloprost on top and a stay in hospital, presumably that's an extra cost to the NHS and it's better for me 'cause I'm not in hospital.
75 (P3 L1)	It takes forever to get them on and get them off and then you realize that the reason they're hurting more than usual is you made a complete mess of putting on last time and you've got to start again.
Coping strategies/aids and devices:	
76 (P8 B1)	I keep a pair of gloves up on top of the fridge freezer to do just that, you know, to take anything out from the freezer.
77 (P4 L1)	I've got things that help me grip jars.
78 (P1 L2)	I also wear gloves, 'cause every time you hit it on something it flares more, that is a big problem I've got no matter, if you touch it, or anything you touch, once you hit it, it flares up again.
79 (P8 L2)	The other thing that I've done for the last 18 months, I never, ever, wet them, as least as possible to get them wet, so in the shower I've got rubber gloves.
Support from others:	
80 (P4 L1)	They've adapted, my children have, I mean they're grown up now, but they know I'll just call, they walk in, open a bottle, if I'm cooking and if I look, they know which one, which saucepan to get out, they just know, like in and out of cars, and they just know now, and so do my friends. They just know.
81 (P7 M1)	I have 3 children, and I live alone and it's not easy because you have to do everything, so you have to cook, you have to touch water and that is something that terrifies you… It is very difficult but the way to cope about it, I think it is just to explain to them…and they will know that they have to step up to do something of the things so they understand that part, but the other part that you have to live, you have to do it, you have to bath them, you have to do everything else, and you know that you'll be in pain for that time, all the time. You know it's going to happen whether you like it or not.
Adaptations/self-management:	
82 (P1 L2)	It's really good, the pumice stone really helps peel it down.
83 (P7 L2)	I think the hardest thing is trying to treat it, and put bandages on it because it's such awkward positions, you can't keep the bandage on there and do other things.
84 (P2 L1)	It's just a lot of care that I have to take, and just move very, very slowly, be very aware of your space around you, with my ulcers.
85 (P3 M1)	You do feel very cautious, if you do have a bang then you're more aware that you're not to do things for the next few days in case it, it goes really bad.
86 (P1 B1)	When you put it in hot water or cold water, moving from one room to another it would just set the pain off again.

* Q refers to the numbered quote cited in the text. SSc = systemic sclerosis; DU = digital ulcer; P = participant; B1 = Bath group; M1 = Manchester group; L1, L2 = London groups.

roles in the organization or even change jobs completely (Q58, Q59). Some participants described financial concerns from the impact of DUs on their work (Q60). DUs have an impact on social participation, and participants reported taking measures to conceal ulcers with bandages or gloves, both to avoid others seeing them and to reduce the risk of infection (Q61, Q62). A number of participants described difficulties undertaking caring roles within the family, for example, avoiding taking their children outside to play due to the cold weather (Q63), because the cold both exacerbated the pain and aggravated the healing of ulcers or provoked their onset.

Theme 4: factors aggravating occurrence, duration, and impact. There were a number of factors that aggravated the occurrence and duration and impact of ulcers (Table 5). There was variation in the number of ulcers experienced by participants, ranging from experiences of solitary DUs to recurrent episodes of refractory digital ulceration (Q24, Q26, Q64–66). There was variation among participants on the time to DU healing (weeks, months, or even years). The length of time to heal was often related to the season and treatment. Most participants reported that over the winter ulcers took longer to heal (Q67), or they did not heal at all until the summer (Q68). Most participants seemed to be able to identify where previous ulcers had occurred, either based on how they looked or how they felt or both (Q13–16).

Theme 5: mitigating, managing, and adapting to ulcers. Participants used a variety of ways to describe whether a treatment had been effective or not (Table 5). This variety included whether and how quickly the ulcer had healed, whether there had been a reduced rate of recurrence of the ulcer, how the appearance of the ulcer had changed, whether the level of pain was reduced, the positive impact on other activities such as sleeping, whether the participant thought circulation had improved, whether the wound dressing had been effective in protecting the ulcer, and whether the risk of amputation was reduced (Q69–73). As well as the effectiveness of treatment, participants also alluded to the burden of treatment. This burden could mean the need for hospitalization or the burden of medication, the duration (time) of receiving treatments, or the severity of associated side effects, and the time and ease of putting on bandages (Q74, Q75). Participants discussed a range of coping strategies to manage DUs, including different ways in which they had adapted or used support to cope with their ulcers. This adaptation included using a device or aid to help manage ulcers (Q76, Q77), strategies to avoid causing pain or prevent a new ulcer developing (Q78, Q79), and getting help or support (paid or unpaid) from others (Q80). Several participants talked about how their children have adapted to the condition and help the patient cope with limited function (Q80). However, some participants noted that avoiding all activities that may aggravate the ulcer was not possible, especially if they have young children (Q81). Participants described a variety of techniques they used to manage their ulcers, from the earliest stages of development to when the ulcer is visible and active. These techniques included using "home remedies" and alternative treatments (Q82), wound care (Q83), the vigilance associated with self-management (Q84, Q85), and avoiding behaviors (e.g., cold exposure) that the patients consider can cause ulcers (Q86).

DISCUSSION

The current study is the first, to the best of our knowledge, to specifically explore the multifaceted patient experience of SSc-DU. We have identified 5 major interrelated themes (and subthemes) that constitute the patient experience of SSc-DU that we have organized within a conceptual map of SSc-DU. The major themes comprised disabling pain and hypersensitivity; deep and broad-ranging emotional impact; impairment of physical and social activity; factors aggravating occurrence, duration, and impact; and mitigating, managing, and adapting to SSc-DU.

The multicenter study design and purposive sampling framework ensured that we captured the experiences from a broad cohort of SSc patients and the spectrum of SSc-DU disease (from solitary DUs to recurrent refractory disease). Thematic analysis of the focus group transcripts was conducted by experienced qualitative researchers without direct experience in the management of SSc-DU, avoiding the potential bias that preconceptions held by scleroderma clinicians might have introduced. The study benefited from a broad international steering committee of SSc experts, qualitative researchers, and patient research partners.

Painful physical symptoms and signs were the most important experiences of SSc-DU. Pain is the cardinal symptom of SSc-DU and is often very severe. Patients often consider the severity of pain disproportionate to the physical size of DUs. Infection and changes in temperature can worsen DU pain. The physical symptoms of DUs result in considerable psychological distress, and impaired hand function impacts all the activities of daily living, including occupation and social interactions. Many patients describe a constant state of vigilance both during and between episodes of ulceration. There are a number of aggravating factors, including the number and severity of DUs. Of interest, participants reported that the ulcers took longer to heal during the winter and residual symptoms at sites of previous DUs. In particular, dysesthesias and paresthesias could suggest persistent nerve damage. Patients with SSc make considerable efforts to both prevent and manage DUs (e.g., avoiding trauma and preventing infection) and describe a wide range of coping strategies and adaptations. The patient experience of DUs mirrors that of Raynaud's phenomenon, in which patients report the need for constant vigilance and self-management (30). Overall, our themes show similarities to those reported by Nakayama et al (30), who conducted a systematic review and thematic analysis of 26 studies with 463 patients to explore patients' perspectives and experiences living with SSc. The 6 key themes were distressing appearance transformation,

palpable physical limitations, social impairment, navigating uncertainty, alone and understood, and gradual acceptance and relative optimism (30). Furthermore, DUs (along with Raynaud's phenomenon and calcinosis) were described as "being intensely painful by some patients," were "emotionally distressing," and "limited patients' ability to work, go outdoors," "or even walk" (30).

As previously described, in earlier clinical trials of SSc-DU, primary assessment of treatment efficacy has focused on clinician assessment of DU presence alone (occurrence and persistence) and has largely overlooked the patient experience of SSc-DU. Legacy PRO instruments assessing function and interference capture patient experiences relevant to SSc-DU, but those instruments are limited by the inclusion of redundant items that are less relevant to SSc-DU (e.g., the inclusion of nonhand domains of the Health Assessment Questionnaire disability index). The recent development of an SSc-specific PRO instrument, the Hand Disability in SSc DUs, was developed through modification of the Cochin Hand Function Scale, including qualitative patient interviews to assess the impact of DUs on hand function in patients with SSc (31). However, to date, other important experiences of SSc-DU (e.g., psychological impacts and social participation) have been comparatively overlooked. The development of a novel PRO instrument that captures the broader patient experience of SSc-DU (e.g., relationships and body image dissatisfaction) would be valuable for assessing interventions in clinical trials but also in clinical practice, where there is a dearth of practice-based evidence examining the comparative efficacy of different pharmacologic, surgical, and wound-care protocols. Furthermore, even after ulcer healing, patients can still suffer from significant residual pain and anxiety of future DUs. Therefore, effective ulcer treatments (and PRO instruments) should also modify future patient (negative) experiences of DU disease even after ulcer healing.

Our analysis has not addressed potential differences in experiences relating to DUs occurring at different locations on the hands (e.g., fingertip versus extensor). The etiopathogenesis (and patient experience) of different types of DU may differ, although all types of DU are generally accepted to have an ischemic contribution (32,33). Therefore, future efforts to develop a dedicated PRO instrument for assessing SSc-DU should explore different experiences (including treatment effects) at different ulcer locations. We also highlight the fact that we only recruited a relatively small number of patients with early disease. This situation is likely due to the need in our study to include a large majority of patients with a significant burden (history) of digital vascular disease, which usually takes time (years) to accrue. There were differences observed in the clinical and demographic characteristics (e.g., age and sex) of participants who participated in the 4 focus groups. For example, the majority of patients in Bath and Manchester had the limited subset of the disease, whereas approximately equal numbers of patients had diffuse disease in the 2 London focus groups.

We did not entirely achieve our intended purposive sampling framework, but we were satisfied that we had captured the experiences of a broad spectrum of patients and did not feel this gap was a barrier to achieving the study's aims. Due to the rarity and heterogeneity of the disease, identifying and enrolling patients with specific phenotypes to studies of this nature is not always possible. We also excluded participants who could not speak English. Although our focus groups were conducted only in the UK, previous studies (including multinational recruiting clinical trials) have demonstrated no important differences in DU disease between countries. In our study, we captured limited information on the impact of SSc-DU on intimate relationships (19). If the data had been collected during one-to-one interviews, then comments on the impact of DUs on intimate relationships likely would have arisen and should be considered in the design of future research. We will explore such themes in a 1:1 setting during future cognitive debriefing of a provisional item-bank for the proposed DU PRO instrument.

It should be highlighted that treating clinicians (MH and JDP) facilitated the focus groups, which could have impacted the reflexivity of the research and introduced potential bias, for example, by shaping the discussion and/or limiting patients' willingness to discuss certain aspects of their experience. However, mitigating factors include the study topic guide that was developed with support from patient insight partners and was used to inform the structure of the focus groups. Patients were only known to 1 individual clinician at 1 geographic location. Furthermore, while background clinical knowledge of SSc was essential to successfully facilitate the focus groups, the analysis of data was led by 2 independent researchers (JJ and AM), to mitigate this potential source of bias.

In conclusion, ours is the first study to examine the multifaceted patient experience of SSc-DU. Traditional clinical trial end points are not currently designed to capture the patient experience of SSc-DU, which should be a key priority for demonstrating meaningful treatment benefit. The resultant themes and subthemes from our study provide a unique insight into the patient experience of SSc-DU. This work could form the basis of a novel PRO instrument to assess the impact and severity of SSc-DU to support much needed new treatment approaches for SSc-DU.

ACKNOWLEDGMENT

The authors thank Ms Jane Withey for her contribution to the project steering committee as a patient research partner.

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. Hughes 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. Hughes, Pauling, Jones, Denton, Domsic, Frech, Herrick, Khanna, Matucci-Cerinic, McKenzie, Saketkoo, Gooberman-Hill, Moore.

Acquisition of data. Hughes, Pauling, Jones, Moore.

Analysis and interpretation of data. Hughes, Pauling, Jones, Denton, Domsic, Frech, Herrick, Khanna, Matucci-Cerinic, McKenzie, Saketkoo, Gooberman-Hill, Moore.

REFERENCES

1. Hughes M, Herrick AL. Digital ulcers in systemic sclerosis. Rheumatology (Oxford) 2017;56:14–25.
2. Khimdas S, Harding S, Bonner A, Zummer B, Baron M, Pope J. Associations with digital ulcers in a large cohort of systemic sclerosis: results from the Canadian Scleroderma Research Group registry. Arthritis Care Res (Hoboken) 2011;63:142–9.
3. Ennis H, Vail A, Wragg E, Taylor A, Moore T, Murray A, et al. A prospective study of systemic sclerosis-related digital ulcers: prevalence, location, and functional impact. Scand J Rheumatol 2013;42:483–6.
4. Mouthon L, Carpentier PH, Lok C, Clerson P, Gressin V, Hachulla E, et al. Ischemic digital ulcers affect hand disability and pain in systemic sclerosis. J Rheumatol 2014;41:1317–23.
5. Hughes M, Ong VH, Anderson ME, Hall F, Moinzadeh P, Griffiths B, et al. Consensus best practice pathway of the UK Scleroderma Study Group: digital vasculopathy in systemic sclerosis. Rheumatology (Oxford) 2015;54:2015–24.
6. Wigley FM, Seibold JR, Wise RA, McCloskey DA, Dole WP. Intravenous iloprost treatment of Raynaud's phenomenon and ischemic ulcers secondary to systemic sclerosis. J Rheumatol 1992;19: 1407–14.
7. Matucci-Cerinic M, Denton CP, Furst DE, Mayes MD, Hsu VM, Carpentier P, et al. Bosentan treatment of digital ulcers related to systemic sclerosis: results from the RAPIDS-2 randomised, double-blind, placebo-controlled trial. Ann Rheum Dis 2011;70:32–8.
8. Hachulla E, Hatron PY, Carpentier P, Agard C, Chatelus E, Jego P, et al. Efficacy of sildenafil on ischaemic digital ulcer healing in systemic sclerosis: the placebo-controlled SEDUCE study. Ann Rheum Dis 2016;75:1009–15.
9. Matucci-Cerinic M, Krieg T, Guillevin L, Schwierin B, Rosenberg D, Cornelisse P, et al. Elucidating the burden of recurrent and chronic digital ulcers in systemic sclerosis: long-term results from the DUO registry. Ann Rheum Dis 2016;75:1770–6.
10. Baron M, Chung L, Gyger G, Hummers L, Khanna D. Consensus opinion of a North American working group regarding the classification of digital ulcers in systemic sclerosis. Clin Rheumatol 2014;33:207–14.
11. Hughes M, Roberts C, Tracey A, Dinsdale G, Murray A, Herrick AL. Does the clinical context improve the reliability of rheumatologists grading digital ulcers in systemic sclerosis? Arthritis Care Res (Hoboken) 2016;68:1340–5.
12. Hughes M, Tracey A, Bhushan M, Chakravarty K, Denton CP, Dubey S, et al. Reliability of digital ulcer definitions as proposed by the UK Scleroderma Study Group: a challenge for clinical trial design. J Scleroderma Relat Disord 2018;3:170–4.
13. Khanna D, Denton CP, Merkel PA, Krieg T, Le Brun FO, Marr A, et al. Effect of macitentan on the development of new ischemic digital ulcers in patients with systemic sclerosis: DUAL-1 and DUAL-2 randomized clinical trials. JAMA 2016;315:1975–88.
14. Seibold JR, Wigley FM, Schiopu E, Denton CP, Silver RM, Steen VD, et al. Digital ulcers in SSc treated with oral treprostinil: a randomized, double-blind, placebo-controlled study with open-label follow-up. J Scleroderma Relat Disord 2017;2:42–9.
15. Suliman YA, Bruni C, Johnson SR, Praino E, Alemam M, Borazan N, et al. Defining skin ulcers in systemic sclerosis: systematic literature review and proposed World Scleroderma Foundation (WSF) definition. J Scleroderma Relat Disord 2017;2:115–20.
16. Pauling JD, Nagaraja V, Khanna D. Insight into the contrasting findings of therapeutic trials of digital ischaemic manifestations of systemic sclerosis. Curr Treatm Opt Rheumatol 2019;5:85.
17. Li W, Frech TM. The critical need for accurately defining digital ulcers in scleroderma. J Scleroderma Relat Disord 2017;2:69–71.
18. Suarez-Almazor ME, Kallen MA, Roundtree AK, Mayes M. Disease and symptom burden in systemic sclerosis: a patient perspective. J Rheumatol 2007;34:1718–26.
19. Hughes M, Pauling JD. Exploring the patient experience of digital ulcers in systemic sclerosis. Semin Arthritis Rheum 2019;48:888–94.
20. Pauling J, Frech TM, Domsic RT, Hudson M. Patient participation in patient-reported outcome instrument development in systemic sclerosis. Clin Exp Rheumatol 2017;106:184–92.
21. Kirwan JR, Bartlett SJ, Beaton DE, Boers M, Bosworth A, Brooks PM, et al. Updating the OMERACT filter: implications for patient-reported outcomes. J Rheumatol 2014;41:1011–5.
22. Administration U.S.D.o.H.a.H.S.F.a.D. Guidance for industry: patient-reported outcome measures. Use in medical product development to support labeling claims. 2009.
23. Krueger R, Casey M. Focus groups: a practical guide for applied research. 3rd ed. Thousand Oaks (CA): Sage; 2000.
24. 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.
25. LeRoy EC, Black C, Fleischmajer R, Jablonska S, Krieg T, Medsger TA, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J Rheumatol 1988;15:202–5.
26. Guest G, Bunce A, Johnson L. How many interviews are enough?: an experiment with data saturation and variability. Field Meth 2006;18:59–82.
27. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3:77–101.
28. Glaser B, Strauss A. Discovery of grounded theory: strategies for qualitative research. Chicago: Aldine; 1967.
29. Pauling JD, Domsic RT, Saketkoo LA, Almeida C, Withey J, Jay H, et al. A multinational qualitative research study exploring the patient experience of Raynaud's phenomenon in systemic sclerosis. Arthritis Care Res (Hoboken) 2018;70:1373–84.
30. Nakayama A, Tunnicliffe DJ, Thakkar V, Singh-Grewal D, O'Neill S, Craig JC, et al. Patients' perspectives and experiences living with systemic sclerosis: a systematic review and thematic synthesis of qualitative studies. J Rheumatol 2016;43:1363–75.
31. Khanna DK, Poiraudeau S, Gelhorn H, Hunsche E, Papadakis K, Perchenet L. Development and content validity of the Hand Disability in Systemic Sclerosis—Digital Ulcers (HDISS-DU) scale [abstract]. Arthritis Rheum 2011;63 Suppl 10:1863.
32. Hughes M, Moore T, Manning J, Wilkinson J, Dinsdale G, Roberts C, et al. Reduced perfusion in systemic sclerosis digital ulcers (both fingertip and extensor) can be increased by topical application of glyceryl trinitrate. Microvasc Res 2016;111:32–6.
33. Hughes M, Murray A, Denton C, Herrick A. Should all digital ulcers be included in future clinical trials of systemic sclerosis related digital vasculopathy? Med Hypotheses 2018;116:101–4.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 72, No. 5, May 2020, pp 734–737

LETTERS

DOI 10.1002/acr.24158

Not positive about antinuclear antibody–negative lupus: comment on the article by Choi et al

To the Editor:

The value of a large systemic lupus erythematosus (SLE) inception cohort cannot be overstated, but the findings of the study by Choi et al, recently published in *Arthritis Care & Research*, regarding the prevalence of "ANA [antinuclear antibody]–negative lupus" is debatable (1). The topline message from the article is that within the 1,132 patients enrolled in the group, 6.2% were anticellular antibody negative. We need to emphasize, however, that this subset of patients had been receiving treatment for months prior to enrollment. Antibody status at the time of enrollment is therefore not necessarily reflective of what their antibody status had been when their illness was first identified and therapy begun. Forty-six percent of ANA-negative patients had taken high-dose glucocorticoids prior to enrollment. High-dose steroids have been shown to possibly effect a change in ANA status (2). Beneficial information might be gained from looking at the pre-enrollment laboratory data (i.e., obtained before treatment was begun) on the 71 ANA-negative patients to see whether some of them had previously been antibody positive.

A smaller concern is the chosen cutoff to define negative ANA as a titer of <1:160. While this limit is a generally agreed-upon cutoff value, it conflicts with the updated SLE classification criteria, which uses 1:80 as a threshold (the cutoff of 1:160 was undoubtedly chosen because the study was designed long before the new American College of Rheumatology [ACR] classification criteria came into being) (3). Possibly some of the 71 anticellular antibody negative individuals may not have been negative according to the new ACR criteria. If samples from the time of enrollment were banked, investigating how many would test positive employing the ACR's less stringent definition would be of interest.

ANA testing is, of course, by no means infallible. We and others have shown that ANA test results from the same sample can differ between testing kits (4,5). Samples that are shown to have high-titer ANA from patients with true disease via one kit may have undetectable antibodies by another. Crosschecking across different kits, while not feasible in the real world, might prove useful in a study like this, to be employed on a limited basis (i.e., in antibody-negative individuals).

Lastly, overdiagnosis of SLE is an issue that has not been quantified, but prior studies have demonstrated that it occurs (6). The authors note that past longitudinal studies show that many ANA-negative SLE patients tend to have a benign course, only experiencing arthritic and photosensitive manifestations. Could it be that some of these ANA-negative patients do not have SLE, and if so, could the same account for some of the ANA-negative SLE individuals in longitudinal lupus cohorts?

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

Aryeh M. Abeles, MD
NYU Langone Medical Center
New York, NY

1. Choi MY, Clarke AE, St Pierre Y, Hanly JG, Urowitz MB, Romero-Diaz J, et al. Antinuclear antibody-negative systemic lupus erythematosus in an international inception cohort. Arthritis Care Res (Hoboken) 2019;71:893–902.
2. Weisbart RH, Colburn K. Effect of corticosteroids on serum antinuclear antibodies in man. Immunopharmacology 1984;8:97–101.
3. 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.
4. Abeles AM, Gomez-Ramirez M, Abeles M, Honiden S. Antinuclear antibody testing: discordance between commercial laboratories. Clin Rheumatol 2016;35:1713–8.
5. Pisetsky DS, Spencer DM, Lipsky PE, Rovin BH. Assay variation in the detection of antinuclear antibodies in the sera of patients with established SLE. Ann Rheum Dis 2018;77:911–3.
6. Narain S, Richards HB, Satoh M, Sarmiento M, Davidson R, Shuster J, et al. Diagnostic accuracy for lupus and other systemic autoimmune diseases in the community setting. Arch Intern Med 2004;164:2435–41.

DOI 10.1002/acr.24152

Reply

To the Editor:

We appreciate the comments by Dr. Abeles about our recent publication on ANA (referred to as anticellular antibody [ACA] in our publication and in this response)–negative sera in SLE that used biobanked serum samples from the Systemic Lupus International Collaborating Clinics (SLICC) inception cohort. We concur with many of the comments and concerns, although we prefer to think that the issues raised are not so much "debatable" as requiring further study. Certainly, the challenges of sensitivity, specificity, commutability, and standardization of the ACA test performed by the HEp-2 immunofluorescence assay (IFA) are not new and date to the adoption of the ACA IFA test for the diagnosis of SLE (for reviews, see refs. 1–4).

We agree that the ACA status at the time of enrollment in our study may not be reflective of the ACA status when SLE was first identified nor of the ACA status at any time during the preclinical and postclinical diagnosis. A significant proportion (46.5%) of the ACA-negative patients in our study were prescribed high-dose glucocorticoids (GC) at or prior to enrollment, and multivariable analysis showed that GC treatment was associated with a greater likelihood of being ACA negative. We hypothesized that this finding might be attributed to the GC treatment, but we also indicated that the impact of GC on our results was speculative, because we did not know the ACA status prior to initiation of GC treatment. Dr. Abeles' citation of a study of 8 anti-DNA positive SLE patients using an in-house anti-DNA enzyme-linked immunoassay (5) is intriguing but does not provide definitive evidence for the influence of GC on anti–double-stranded DNA, or by implication ACA status. Not only was that study performed on a small selected SLE patient group, it only evaluated the anti-DNA titers for up to 30 days afterward, it did not evaluate the dynamics of anti-DNA status over time (including anti-DNA status prior to GC administration), and it did not have a control group with untreated SLE or other controls. Clearly, well-designed studies are needed to clarify the influence of GC on ACA or any autoantibody titers. Unfortunately, biobanked serum samples prior to enrollment in the SLICC inception cohort and prior to initiation of GC were not available, so this question is not directly answerable for our cohort.

Another approach would have been to compile the ACA results performed at each SLICC center prior to enrollment. However, although we considered this approach, we concluded that such a process would not be helpful because: 1) data and information about the ACA kits, techniques, and ACA IFA assay parameters as performed locally were not available, and obtaining them retroactively would have been an unwieldy task; 2) some centers changed their assays and some test parameters (i.e., screening dilutions changed) during the study; and 3) determining the criteria used by technologists at each center to define a positive ACA was not possible (i.e., did the definition of a positive result include cytoplasmic and mitotic staining?). After much deliberation, we concluded that inclusion of such data lacked commutability and would be open to misinterpretation. Accordingly, we stated: "The results obtained at a single center were used for the ACA analysis in this study because the ACA analyses performed at each regional site had a wide variation in testing parameters (date of test performance, serum screening dilutions, test kits and protocols, microscopes, readers, etc.) and thus were not comparable across sites."

Because the initiation of the SLICC cohort in 1999 antedated the SLICC criteria (6) or the more recent European League Against Rheumatism (EULAR)/ACR criteria (7,8), all enrolled patients in our cohort fulfilled the ACR criteria (6). In our study, we used an ACA titer cutoff of <1:160 to increase the specificity of the ACA for SLE at the expense of sensitivity (9). When the HEp-2 IFA test was repeated at a serum dilution of 1:80 on 67 of 71 of the available ACA-negative samples, 50 of 67 (74.6%) of the ACA-negative sera and 50 of 1,133 (4.4%) of the total cohort remained negative at the lower serum dilution (data not published). Therefore, 4.4% of patients from our study would still not have been classified as having SLE according to the new EULAR/ACR criteria. As to the question of whether some of the ACA-negative patients in our study did not have SLE, which could account for some of the ACA-negative individuals in our study, notably all patients fulfilled the ACR criteria, including those with a negative ACA. Furthermore, classification criteria are not diagnostic criteria (10) and are intended to identify SLE patients for research studies, such as ours, while the diagnosis of SLE remains in the jurisdiction of an appropriately trained physician.

In summary, Dr. Abeles raises some important questions that underscore persisting gaps in our knowledge of ACA status in SLE. These include the impact of GC on both antibody titers and specificity, the harmonization and commutability of ACA testing, and the constantly changing approach to classification and diagnosis of SLE. These issues are all addressable by well-designed studies and a concerted approach to ACA standardization.

Marvin J. Fritzler, PhD, MD
Ann E. Clarke, MD, MSc
May Y. Choi, MD
University of Calgary
Calgary, Alberta, Canada
Yvan St. Pierre, MSc
McGill University Health Centre
Montreal, Quebec, Canada

1. Fritzler MJ. The antinuclear antibody test: last or lasting gasp? [editorial]. Arthritis Rheum 2011;63:19–22.
2. Choi MY, Fritzler MJ. Autoantibodies in SLE: prediction and the p value matrix. Lupus 2019;28:1285–91.
3. Bossuyt X, Claessens J, Belmondo T, De Langhe E, Westhovens R, Poesen K, et al. Harmonization of clinical interpretation of antinuclear antibody test results by solid phase assay and by indirect immunofluorescence through likelihood ratios. Autoimmun Rev 2019;18:102386.
4. Pisetsky DS, Bossuyt X, Meroni PL. ANA as an entry criterion for the classification of SLE. Autoimmun Rev 2019;18:102400.
5. Weisbart RH, Colburn K. Effect of corticosteroids on serum antinuclear antibodies in man. Immunopharmacology 1984;8:97–101.
6. 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.
7. 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. Ann Rheum Dis 2019;78:1151–9.
8. 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.
9. Tan EM, Feltkamp TE, Smolen JS, Butcher B, Dawkins R, Fritzler MJ, et al. Range of antinuclear antibodies in "healthy" individuals. Arthritis Rheum 1997;40:1601–11.
10. 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.

DOI 10.1002/acr.24153

2019 update of the American College of Rheumatology–recommended rheumatoid arthritis disease activity measures: comment on the article by England et al

To the Editor:

I read the article by England et al (1), recently published in *Arthritis Care & Research*, with great interest; it discusses the appropriate metrics to be used by practicing rheumatologists employing a treat-to-target approach. The article is comprehensive and well written. I would like to make 2 comments and clarify 1 point, however.

Table 1 contains a clinically significant error that should be clarified. The table states that the cut points for disease activity states such as remission, low, moderate, and high disease activity are exactly the same for the Disease Activity Score in 28 joints (DAS28) using the erythrocyte sedimentation rate (ESR) and the DAS28 with the C-reactive protein (CRP) level. This idea is a popular misconception among rheumatologists and pharmaceutical companies, but multiple times these cut points have been shown to overestimate response for DAS28-CRP and suggest that more patients are in remission and have low disease activity and that fewer have high disease activity (Figure 1). This error has major implications in treating to target, because the use of the DAS28-CRP will suggest that more patients are at target than is accurate. The table should be amended to show the correct cut points for DAS28-CRP for disease activity states that are lower than for DAS28-ESR.

The 5 disease activity metrics that were recommended by the committee are appropriate and have face validity. However, I would have thought the committee might suggest a combination of metrics, such as the Clinical Disease Activity Index (CDAI) or Simplified Disease Activity Index and the Routine Assessment of Patient Index Data 3 (RAPID3). The American College of Rheumatology (ACR) endorses the treat-to-target recommendations, which state that the metrics used must include 1 that assesses joint counts, such as the CDAI, but patient-reported outcomes, such as the RAPID3, should be assessed as well (2–7).

Finally, I would like to commend the committee for stating that the Multi-Biomarker Disease Activity is not a preferred metric for following disease activity, and their analysis of this test shows that its usefulness is inconclusive.

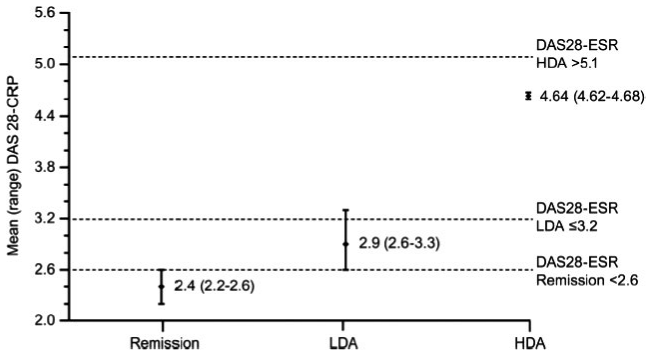

Figure 1. Relationship between Disease Activity Score in 28 joints (DAS28) using the erythrocyte sedimentation rate (ESR) and the DAS28 using the C-reactive protein (CRP) level (ref. 3). HDA = high disease activity; LDA = low disease activity.

Roy Fleischmann, MD
University of Texas
Dallas, TX

1. England BR, Tiong BK, Bergman MJ, Curtis JR, Kazi S, Mikuls TR, et al. 2019 update of the American College of Rheumatology recommended rheumatoid arthritis disease activity measures. Arthritis Care Res (Hoboken) 2019;71:1540–55.
2. Fleischmann R, van der Heijde D, Koenig A, Pedersen R, Szumski A, Marshall L, et al. How much does Disease Activity Score in 28 joints ESR and CRP calculations underestimate disease activity compared with the Simplified Disease Activity Index? Ann Rheum Dis 2015;74:1132–7.
3. Fleischmann R, van der Heijde D, Gardiner P, Szumski A, Marshall L, Bananis E. DAS28-CRP and DAS28-ESR cut-offs for high disease activity in rheumatoid arthritis are not interchangeable. RMD Open 2017;3:e00038.
4. Favalli E, Becciolini A, Biggioggero M, Marchesoni A, Meroni P. Is there a need for new thresholds to define remission and low disease activity by Disease Activity Score 28 calculated with C-reactive protein? Real-life data from a local registry. Ann Rheum Dis 2015;74:e5.
5. Matsui T, Kuga Y, Kaneko A, Nishino J, Eto Y, Chiba N, et al. Disease Activity Score 28 (DAS28) using C-reactive protein underestimates disease activity and overestimates EULAR response criteria compared with DAS28 using erythrocyte sedimentation rate in a large observational cohort of rheumatoid arthritis patients in Japan. Ann Rheum Dis 2007;66:1221–6.
6. 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.
7. Smolen J, Breedveld F, Burmester G, Bykerk V, Dougados M, Emery P, et al. Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. Ann Rheum Dis 2016;75:3–17.

DOI 10.1002/acr.24157

Reply

To the Editor:

We thank Dr. Fleischmann for his comments on the working group's recommendations on rheumatoid arthritis (RA) disease activity measures. While several studies have shown that DAS28-CRP may underestimate disease activity compared to DAS28-ESR (1–4), there also exists conflicting evidence that the 2 scores have good agreement (5,6). More stringent cutoffs for DAS28-CRP have been proposed by both Inoue et al (7) and Fleischmann et al (8) to improve agreement with DAS28-ESR disease activity categorization, although the thresholds

Table 1. Proposed cutoffs for DAS28-CRP by disease activity categories and DAS28-ESR thresholds*

Disease activity category and DAS28-ESR cutoffs	Remission <2.6	Low 2.6 to <3.2	Moderate 3.2 to ≤5.1	High >5.1
Inoue et al (7)	<2.3	2.3 to <2.7	2.7 to ≤4.1	>4.1
Fleischmann et al (1)	<2.4	2.4 to <2.9	≥2.9†	–
Fleischmann et al (8)†	–	–	–	>4.6

* DAS28 = Disease Activity Score in 28 joints; CRP = C-reactive protein: ESR = erythrocyte sedimentation rate.
† Low and high cutoffs for moderate disease activity not established in the same study. From independent studies, moderate disease activity range could be considered 2.9 to ≤4.6.

proposed by the 2 authors differ somewhat (Table 1). Providers, or those designing clinical trials, may choose to use these more stringent DAS28-CRP cutoffs. However, substituting alternative thresholds is premature at this time, because only limited validation of these alternate cutoffs has been completed to date (9). There is a need for validation in varied RA patient populations and clinical settings, as well as to determine whether individuals discordant in their categorization of disease activity by DAS28-CRP versus DAS28-ESR have a distinct clinical course. Conservative adoption of the more stringent cutoffs that Inoue et al and Fleischmann et al suggest is particularly important, with increasing interest among health care payers to score quality of care using these metrics. Certainly, selection of appropriate cutoffs will be revisited with future updates of the recommended RA disease activity measures. Regarding the use of multiple metrics, the committee did not recommend against complementary use of multiple RA disease activity measures in routine clinical care. Indeed, 15% of the >50,000 RA patients with any RA disease activity data represented in the ACR Rheumatology Informatics System for Effectiveness registry had providers who measured >1 disease activity metric as part of routine clinical care (10). Providers may choose to use multiple RA disease activity measures regularly or in special clinical circumstances. In addition, there is no requirement of joint count assessment in ACR RA treatment guidelines, only the use of recommended RA disease activity measures, of which 3 of the 5 contain joint counts (11).

Bryant R. England, MD
Ted R. Mikuls, MD, MSPH
James R. O'Dell, MD
Veterans Administration Nebraska–Western Iowa Health Care System and University of Nebraska Medical Center
Omaha, NE
Benedict K. Tiong, MD
Veena K. Ranganath, MD, MS, RhMSUS
University of California Los Angeles, David Geffen School of Medicine
Los Angeles, CA
Martin J. Bergman, MD
Drexel University College of Medicine
Philadelphia, PA
Jeffrey R. Curtis, MD, MS, MPH
University of Alabama at Birmingham
Birmingham, AL
Salahuddin Kazi, MD
University of Texas Southwestern
Dallas, TX
Alex Limanni, MD
Arthritis Centers of Texas
Dallas, TX
Lisa G. Suter, MD
Yale University
New Haven, CT
and Veterans Affairs Medical Center
West Haven, CT
Kaleb Michaud, PhD
University of Nebraska Medical Center
Omaha, NE
and FORWARD, The National Databank for Rheumatic Diseases
Wichita, KS

1. Fleischmann R, van der Heijde D, Koenig AS, Pedersen R, Szumski A, Marshall L, et al. How much does Disease Activity Score in 28 joints ESR and CRP calculations underestimate disease activity compared with the Simplified Disease Activity Index? Ann Rheum Dis 2015;74:1132–7.
2. Matsui T, Kuga Y, Kaneko A, Nishino J, Eto Y, Chiba N, et al. Disease Activity Score 28 (DAS28) using C-reactive protein underestimates disease activity and overestimates EULAR response criteria compared with DAS28 using erythrocyte sedimentation rate in a large observational cohort of rheumatoid arthritis patients in Japan. Ann Rheum Dis 2007;66:1221–6.
3. Wells G, Becker JC, Teng J, Dougados M, Schiff M, Smolen J, et al. Validation of the 28-joint Disease Activity Score (DAS28) and European League Against Rheumatism response criteria based on C-reactive protein against disease progression in patients with rheumatoid arthritis, and comparison with the DAS28 based on erythrocyte sedimentation rate. Ann Rheum Dis 2009;68:954–60.
4. Matsui T, Kuga Y, Nishino J, Kaneko A, Eto Y, Tohma S. Comparison of composite disease activity indices for rheumatoid arthritis. Mod Rheumatol 2011;21:134–43.
5. Crowson CS, Rahman MU, Matteson EL. Which measure of inflammation to use? A comparison of erythrocyte sedimentation rate and C-reactive protein measurements from randomized clinical trials of golimumab in rheumatoid arthritis. J Rheumatol 2009;36:1606–10.
6. Nielung L, Christensen R, Danneskiold-Samsoe B, Bliddal H, Holm CC, Ellegaard K, et al. Validity and agreement between the 28-joint Disease Activity Score based on C-reactive protein and erythrocyte sedimentation rate in patients with rheumatoid arthritis. Arthritis 2015;2015:401690.
7. Inoue E, Yamanaka H, Hara M, Tomatsu T, Kamatani N. Comparison of Disease Activity Score (DAS)28- erythrocyte sedimentation rate and DAS28- C-reactive protein threshold values. Ann Rheum Dis 2007;66:407–9.
8. Fleischmann RM, van der Heijde D, Gardiner PV, Szumski A, Marshall L, Bananis E. DAS28-CRP and DAS28-ESR cut-offs for high disease activity in rheumatoid arthritis are not interchangeable. RMD Open 2017;3:e000382.
9. Favalli EG, Becciolini A, Biggioggero M, Marchesoni A, Meroni PL. Is there a need for new thresholds to define remission and low disease activity by Disease Activity Score 28 calculated with C reactive protein? Real life data from a local registry. Ann Rheum Dis 2015;74:e5.

10. Yun H, Chen L, Xie F, Patel H, Boytsov N, Zhang X, et al. Do patients with moderate or high disease activity escalate rheumatoid arthritis therapy according to treat-to-target principles? Results from the ACR's RISE registry. Arthritis Care Res (Hoboken) 2020;72:166–75.
11. 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.

DOI 10.1002/acr.24162

Did antiphospholipid antibodies limit intervention efficacy for postoperative total knee arthroplasty–related thrombotic event prevention? Comment on the article by Smith et al

To the Editor:

Reduction of thrombotic events among individuals subsequent to total knee arthroplasty was the subject of an intriguing article by Smith et al, recently published in *Arthritis Care & Research* (1). The authors reported that prophylactic aspirin, rivaroxaban, low molecular heparin, fondaparinux, and warfarin reduced the deep vein thrombosis rate by only 15%, at best. A possible explanation may be modulation of platelet and vascular function by antiphospholipid antibodies (2). Antiphospholipid antibodies are not rare (3).

If antiphospholipid antibodies are present, use of aspirin might be considered, but monitoring platelet function to assure efficacy is essential (4). The effect of aspirin on platelet function is highly variable, with many patients requiring more than the "standard" 81-mg dose, noting that even 975 mg (3 tablets) may not be sufficient (5). The very convenient fractionated heparins and factor X antagonists are ineffective in preventing thrombotic events in individuals with antiphospholipid antibodies (6,7).

Limited efficacy of warfarin used in the study by Smith et al (1) may similarly be a dose effect. The standard prothrombin time international normalized ratios (INRs) range used in monitoring warfarin is inadequate in the presence of antiphospholipid antibodies (6). Failure to pursue the 3.0–3.5 INR-recommended range may explain reduced intervention efficacy reported by Smith et al (1).

Thus, assessment for the presence of antiphospholipid antibodies in patients, especially in those with a prior history of thrombotic events, seems worthwhile, with prophylactic action when identified. Identification of antiphospholipid antibodies would be especially important for individuals undergoing surgery, as the presence of antiphospholipid antibodies requires modification of standard intervention (8). Which antibodies should be assessed? IgG, IgM, and IgA antibodies to anticardiolipin, β_2-glycoprotein I, and antiphosphatidylserine/prothrombin are also pertinent.

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

Bruce Rothschild, MD
IU Health
Muncie, IN

1. Smith SR, Katz JN, Losina E. Cost-effectiveness of alternative anticoagulation strategies for postoperative management of total knee arthroplasty patients. Arthritis Care Res (Hoboken) 2019;71:1621–9.
2. Cohen D, Berger SP, Steup-Beekman GM, Bloemenkamp KW, Bajema IM. Diagnosis and management of the antiphospholipid syndrome [editorial]. BMJ 2010;340:2541.
3. Garcia D, Erkan D. Diagnosis and management of the antiphospholipid syndrome. N Eng J Med 2018;378:2010–21.
4. Rothschild BM. Comparative anti-platelet activity of COX-1 NSAIDS versus aspirin, encompassing regimen simplification and gastroprotection: a call for a controlled study. Reumatismo 2004;56:89–93.
5. Perneby C, Wallén NH, Rooney C, Fitzgerald D, Hjemdahl P. Dose- and time-dependent antiplatelet effects of aspirin. Thromb Haemost 2006;95:652–8.
6. Ziakas PD, Pavlou M, Voulgarelis M. Heparin treatment in antiphospholipid syndrome with recurrent pregnancy loss: a systematic review and meta-analysis. Obstet Gynecol 2010;115:1256–62.
7. Pengo V, Denas G, Zoppellaro G, Jose SP, Hoxha A, Ruffatti A, et al. Rivaroxaban vs warfarin in high-risk patients with antiphospholipid syndrome. Blood 2018;132:1365–71.
8. Khamashta MA, Cuadrado MJ, Mujic F, Taub NA, Hunt BJ, Hughes GR. The management of thrombosis in the antiphospholipid-antibody syndrome. N Engl J Med 1995;332:993–7.

DOI 10.1002/acr.24161

Reply

To the Editor:

Dr. Rothschild provides important insights regarding the study of anticoagulation in individuals with antiphospholipid syndrome (APS). The presence of any of several antiphospholipid antibodies can result in a hypercoagulable state leading to increased incidence of arterial and/or venous thromboses (1). APS occurs in 30–50 individuals per 100,000 in the general population and is classically considered in young patients with multiple pregnancy losses. However, patients with APS are at risk of deep vein thrombosis and pulmonary embolism as well. Indeed, these thrombotic events are more frequent disease manifestations in patients with APS than with obstetric complications (2). Further, because antiphospholipid antibodies result in platelet dysfunction, anticoagulation strategies in patients with APS generally call for either higher doses or alternative medication regimens. For these reasons, testing for antiphospholipid antibodies is recommended in patients presenting with thrombotic events (3).

Our analyses focused on venous thrombotic events following total knee arthroplasty (TKA) in patients with osteoarthritis. The likelihood that any of our subjects had APS is low, given the low prevalence of APS in the population. Nonetheless, we appreciate the clinical insights Dr. Rothschild provides. We were not able to specifically assess the cost-effectiveness of anticoag-

ulation strategies following TKA in patients with hypercoagulable conditions, because the published data on efficacy of anticoagulants are not stratified by preexisting prothrombotic states, and the proportion of individuals with thrombosis associated with antiphospholipid antibodies following TKA is unknown. Future analyses should evaluate the appropriate postoperative anticoagulation strategies for patients with underlying disease processes that contribute to higher incidences of thrombotic events.

Savannah R. Smith, BA
George Washington University School of Medicine and Health Sciences
Washington, DC
and Brigham and Women's Hospital
Boston, MA

Jeffrey N. Katz, MD, MSc
Brigham and Women's Hospital and Harvard Medical School
Boston, MA
Elena Losina, PhD
Brigham and Women's Hospital, Harvard Medical School
and Boston University School of Public Health
Boston, MA

1. Corban MT, Duarte-Garcia A, McBane RD, Matteson EL, Lerman LO, Lerman A. Antiphospholipid syndrome: role of vascular endothelial cells and implications for risk stratification and targeted therapeutics. J Am Coll Cardiol 2017;69:2317–30.
2. Duarte-Garcia A, Pham MM, Crowson CS, Amin S, Moder KG, Pruthi RK, et al. The epidemiology of antiphospholipid syndrome: a population-based study. Arthritis Rheumatol 2019;71:1545–52.
3. Giannakopoulos B, Krilia SA. The pathogenesis of the antiphospholipid syndrome. N Engl J Med 2013;368:1033–44.