Arthritis Care & Research

Aims and Scope

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

Volume 71, No. 3, March 2019

Arthritis Care & Research

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

Cover design: Sandra Pulmano

This journal is printed on acid-free paper.

Arthritis Care & Research

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

VOLUME 71 • March 2019 • NO. 3

Gout

Systemic Sclerosis

Cover image: The figure on the cover (from Harris et al, page 392) is a directed acyclic graph (based on creation and evaluation of directed acrylic graphs by Textor et al) modeling the relationship between potential confounding key variables for pregnancy outcomes in systemic lupus erythematosus and rheumatoid arthritis.

Arthritis Care & Research
Vol. 71, No. 3, March 2019, pp 331–333
DOI 10.1002/acr.23822

AMERICAN COLLEGE OF RHEUMATOLOGY
EDUCATION • TREATMENT • RESEARCH

EDITORIAL

We All Agree: Nobody Likes Maintenance of Certification

David R. Karp

Rheumatologists love a debate. A "Great Debate" has become one of the highlights of the American College of Rheumatology's Annual Scientific Meeting. Our consultations often include lengthy differential diagnoses so that even the rarest of rheumatological conditions is not missed. Our treatment guidelines discuss the pros and cons of multiple therapies so that we can choose the best option for the individual patient. However, there seems to be one issue upon which rheumatologists are in agreement — their unhappiness with the current process of Maintenance of Certification (MOC) required by the American Board of Internal Medicine (ABIM). In this issue of *Arthritis Care & Research*, Sawalha and Coit present the data from their survey of US rheumatologists on the topic of MOC (1). Not surprisingly, the overwhelming opinion is negative.

The ABIM began certifying physicians in the specialty of rheumatology in 1971. Certification was voluntary, life-long, and based solely on training and achieving a passing score on a secure examination. For diplomates certified in 1990 and beyond, the initial certification became limited to 10 years. To maintain certification, an additional high-stakes, secure examination was required. Additional requirements were added by the ABIM so that by 2014, maintaining certification required the physician to continuously earn MOC "points" by completing medical knowledge assessments that were separate from the certifying examination, as well as earning points from quality improvement projects incorporating a Practice Assessment, Patient Voice, and Patient Safety (2). Backlash against these new MOC requirements was extensive, and came from both the medical community (3–5) and lay press (6). Concerns were raised about the amount of time it took to meet all the MOC requirements, requiring time away from patients and academic activities, as well as family. The costs of obtaining MOC points and preparing for the examination, as well as the costs of the examination itself were felt to be excessive (7), and potentially financially enriching the ABIM. Perhaps most important was the criticism that maintenance of certification, in any form, has not been shown to result in measurable improvement in physician performance in terms of patient outcome, patient satisfaction, or cost of care (8,9). In response to these criticisms, the ABIM issued a public apology, and temporarily suspended the quality improvement portion of MOC (10).

Sawalha and Coit have undertaken an analysis of rheumatologists' perceptions of MOC in an effort to provide a quantitative and qualitative assessment of how MOC affects the day-to-day work of these physicians. An electronic survey was sent by e-mail invitation to 3,107 US rheumatologists. The survey consisted of 19 single/multiple choice or ranking questions and one open-ended comment question. A total of 515 (16.6%) of the invitees responded. Nearly equal percentages of respondents were in academic practice work settings and community practice, 40% were female, and 88.3% of the respondents were between the ages of 40 and 69 years. Overall, respondents had negative impressions of MOC. Three of 4 believed that MOC added no significant value over continuing medical education (CME) and nearly two-thirds of the respondents expressed the opinion that rheumatology certification should be life-long. More than 60% of respondents had negative impressions about the value of MOC in improving patient care. The opinions of rheumatologists were most unified concerning their opinions on the negative effects of the current MOC program. Approximately 70% of the rheumatologists agreed with the idea that MOC resulted in time away from patient care, time away from family, and personal psychological stress. Nearly 90% agreed that MOC imposed an unfair financial burden without proven benefit, and more than 60% of respondents also agreed with the idea that MOC contributes to physician burnout, early retirement from practice, and ultimately, a reduction in the number of practicing US rheumatologists.

There are several limitations in the study by Sawalha and Coit. The authors invited ~3,100 rheumatologists to participate. This is slightly more than half of the practicing adult and pediatric rheumatologists in the US. The survey response rate, while low, is in keeping with the rate of most external surveys; however, there is no discussion of whether the characteristics of the respondents

The opinions and assertions contained herein are those of the author and do not necessarily represent those of the American College of Rheumatology or its Board of Directors.

Dr. Karp is a former member of the American College of Rheumatology Board of Directors, and the American College of Rheumatology Executive Committee.

David R. Karp, MD, PhD: University of Texas Southwestern Medical Center, Dallas.

Address correspondence to David R. Karp, MD, PhD University of Texas Southwestern Medical Center, Harry Hines Boulevard, Room Y8.302, Dallas, TX 75390. E-mail: David.Karp@UTSouthwestern.edu.

Submitted for publication November 19, 2018; accepted December 4, 2018.

reflect the demographics, geographic distribution, MOC participation, or practice setting of rheumatologists overall. Importantly, the survey did not appear to include rheumatology fellows-in-training. This group of physicians will participate in MOC for the longest time and their opinions should be solicited. The survey asked questions about current/past certification and MOC participation, but those responses were not reported and we do not know if they correlated with answers to other questions. Although negative opinions about MOC were expressed by an overwhelming majority of respondents, there are some minority opinions worth mentioning. Approximately one-third of respondents felt MOC was "somewhat valuable" or "very valuable" for improving patient care and more than one-fifth of respondents felt that MOC actually achieved that goal. Increased patient confidence and reassurance were also mentioned by approximately 15% of respondents. As we go forward, these positive opinions and goals should be some of the tenets of MOC because they reflect physicians' desire to provide better patient care.

Three-quarters of the respondents thought that the state legislatures should pass laws that prevent the use of MOC participation as a requirement for employment, hospital privileges, or insurance contracts. In fact, legislative bills doing just that have been considered in the majority of states, and have been passed in a handful, including Texas (11). Opinions vary on the impact of this legislation, because the interested parties have other criteria they can use to allow or limit physician participation or judge quality. Dr. David Johnson, former Chair of the ABIM Board (and in the interest of disclosure, Department Chair at my institution) points out the risk of such laws (12). Medicine has always enjoyed the privilege of self-regulation. Physicians set the standards for entry into the profession and continued practice. Once we begin to give authority to others (e.g., state governments) to decide how the profession of medicine is regulated, we lose some of that autonomy. The counter argument put forth by physicians who are against all forms of MOC is that the ABIM has already usurped that autonomy, a feeling echoed by the numerous anti-MOC websites and threats of anti-trust lawsuits against the ABIM (13). A recent entrant into the MOC argument is the Anti-Trust Division of the US Department of Justice. In an opinion letter by this agency written to support an anti-MOC bill in the Maryland legislature (14), it was stated that although MOC laws would diminish the autonomy of medicine's self-regulation (e.g., through hospital privileges), the American Board of Medical Specialists (ABMS) did have an effective anti-competitive monopoly on certification. The agency stated that having more certifying organizations would result in a process that was easier and cheaper, and physicians "may pass along some of that extra time and lower costs in the form of savings or extra care for consumers." I'm not sure that these legal and legislative efforts will have that particular effect, but their existence underscores the passionate dislike of the process across a wide spectrum of physicians.

If rheumatologists are unified in their dislike of the current MOC program, what should they do to improve their knowledge, medical care, and patient outcomes? In the results of the survey by Sawalha and Coit (1), approximately 65% of respondents believed that MOC helped them stay current with new knowledge in rheumatology, the same percentage who believe MOC is redundant with CME activities. I think this says rheumatologists favor lifelong learning, but in a format that they can control and with a focus on education. Many of the concerns about MOC center on the choices (and costs) of obtaining medical knowledge points, and on the format of the secure examination. Nearly every physician, trainee, and student I know carries a device in their pocket that can store and immediately access more medical information than most of us will ever know. The idea that physicians must remember enough of that information to pass an 8-hour (or even 3-hour) examination seems foreign in a time when medical school lectures have been replaced by team-based learning exercises and hospital rounds are less about quizzing learners' memory and more about using information to support diagnoses and plan personalized, high-value care. Pedagogy, or instructor-controlled learning, is "out" and andragogy, or learner-controlled education is "in." Heutagogy, an extension of andragogy is now making its way into life-long learning strategies. In heutagogy, the learner not only directs their education, but also determines what should be learned, the method of learning, and reflects on how that learning has influenced their actions and beliefs (15). In this context, physicians would not only extend their competency to diagnose and treat, but also their capability to apply that knowledge through communication, teamwork, and creative problem solving. This method is well-suited to distance learning, mobile technologies, and even use of social media. In practical terms, each physician would maintain certification through a program of actions and assessments of their own design, choosing those that best reflect their knowledge gaps and practice needs. MOC activities, which could include specific CME, would meet a standard but would not be prescribed as "one size fits all."

Where do we go from here? This study is not the only one to assess physician or even rheumatologist opinion. Last summer, the American College of Rheumatology (ACR) conducted an in-depth survey of its members, including fellows-in-training. More than 1,800 people responded. The key results of the survey were presented at the ACR 2018 Annual Scientific Meeting by Dr. Carol Langford. Those results also demonstrated widespread dissatisfaction with MOC. Based on member concerns, the ACR is investigating other options for rheumatology certification beyond the ABIM. In an effort to address the issue of MOC for all physicians, the ABMS, the overarching certification organization in the US, has convened the Vision Initiative, a commission that has begun to study MOC and will provide recommendations to the ABIM and the other constituent boards in February 2019. The commission has recently published the results of its own stakeholder survey in which 36,392 people responded (including 34,616 physicians, 1,373 non-physician providers, and 403 members of the general public) (16). The findings of the Vision Initiative survey were sim-

ilar to those of Sawalha and Coit: only 12% of physicians value MOC, 41% do not value it at all, and 46% have mixed feelings. At least for now a "wait and see" attitude seems best until the Vision Initiative commission provides its full report. Although possible, it seems unlikely that they will ignore the concerns of the majority of physicians. I would hope that the ABMS support a change in MOC that allows physicians to use high-quality, self-designed professional development to demonstrate ongoing improvement in the capability to provide medical care. Until then, I, like many others, will continue to get my MOC points. They won't be cheap, they won't be easy, but I hope that I'll learn something.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Sawalha A, Coit P. Evaluating the perception among rheumatologists of maintence of board certification in the United States. Arthritis Care Res [Hoboken] 2019;71:337–42.
2. Baron RJ, Johnson D. The American Board of Internal Medicine: evolving professional self-regulation. Ann Intern Med 2014;161: 221–3.
3. Centor RM, Fleming DA, Moyer DV. Maintenance of certification: beauty is in the eyes of the beholder. Ann Intern Med 2014;161:226–7.
4. Fisher WG, Schloss EJ. Medical specialty certification in the United States-a false idol? J Interv Card Electrophysiol 2016;47:37–43.
5. American College of Rheumatology. Position Statement, Maintenance of certification. URL: https://www.rheumatology.org/Portals/0/Files/ACR%20MOC%20Postion%20Statement%202015.pdf.
6. Eichenwald K. To the barricades! The doctor's revolt against ABIM is succeeding! URL: https://www.newsweek.com/abim-american-board-internal-medicine-doctors-revolt-372723.
7. Sandhu AT, Dudley RA, Kazi DS. A Cost Analysis of the American Board of Internal Medicine's Maintenance-of-Certification program. Ann Intern Med 2015;163:401–8.
8. Gray BM, Vandergrift JL, Johnston MM, Reschovsky JD, Lynn LA, Holmboe ES, et al. Association between imposition of a Maintenance of Certification requirement and ambulatory care-sensitive hospitalizations and health care costs. JAMA 2014;312:2348–57.
9. Hayes J, Jackson JL, McNutt GM, Hertz BJ, Ryan JJ, Pawlikowski SA. Association between physician time-unlimited vs time-limited internal medicine board certification and ambulatory patient care quality. JAMA 2014;312:2358–63.
10. American Board of Internal Medicine. ABIM announces immediate changes to MOC program. URL: https://www.abim.org/news/abim-announces-immediate-changes-to-moc-program.aspx.
11. LegiScan. Texas Senate Bill 1148. URL https://legiscan.com/TX/bill/SB1148/2017.
12. Johnson DH. Maintenance of Certification and Texas SB 1148: a threat to professional self-regulation. JAMA 2017;318:697–8.
13. Bloomberg Law. Doctors riled by rec-certification process plan to sue board. URL: http://www.changeboardrecert.com/documents/BloombergLawPlantoSue.pdf.
14. United States Department of Justice. Opinion letter on Maryland House Bill 857. URL: https://www.justice.gov/atr/page/file/1092791/download.
15. Blaschke LM. Heutagogy and lifelong learning: a review of heutagogical practice and self-determined learning. International Rev Research Open Distance Learning 2012;13:16.
16. American Board of Medical Specialties. Stakeholder beliefs about the future of continuing certification survery findings. URL: https://visioninitiative.org/wp-content/uploads/2018/07/Vision_for_the_Future_Stakeholder_Survey_Summary.pdf.

AMERICAN COLLEGE OF RHEUMATOLOGY
EDUCATION • TREATMENT • RESEARCH

Arthritis Care & Research
Vol. 71, No. 3, March 2019, pp 334–336
DOI 10.1002/acr.23727

EDITORIAL

Diagnosing Fibromyalgia as a Disease, an Illness, a State, or a Trait?

Don L. Goldenberg

In this issue of *Arthritis Care & Research*, Wolfe and colleagues report that there were considerable disparities between rheumatologists' clinical diagnosis of fibromyalgia and a criteria-based diagnosis; in their study, Wolfe et al used a modification of the 2010 American College of Rheumatology (ACR) preliminary diagnostic criteria (1). Is this surprising? Based on their findings, do we need to further modify the fibromyalgia classification criteria or better educate our rheumatology trainees and staff? How do we apply diagnostic criteria to clinical practice?

Since fibromyalgia became the generally accepted label for chronic, widespread musculoskeletal pain in 1981, rheumatologists have often expressed doubt regarding its pathophysiology, diagnostic utility, and therapy. In order to provide diagnostic clarity and consensus, the ACR endorsed initial criteria for the classification of fibromyalgia in 1990 (2). After 20 years of utilization of these 1990 criteria in rheumatology practice and research, the criteria were modified in 2010 (3) and again in 2011 (4). It is therefore timely to review the evolution of these sets of fibromyalgia criteria.

Classification criteria for the diagnosis of fibromyalgia, as well as for common illnesses such as depression, tension-type headaches, and irritable bowel syndrome, are established by a panel of experts, then field-tested in previously diagnosed patients and in healthy controls or in patients with illnesses that have similar characteristics. Until biomarkers or genetic testing provides more objective disease markers, these common conditions will continue to be symptom-based diagnoses. As such, they are often considered to be "soft" diagnoses.

Our panel of experts recommended various signs and symptoms for the 1990 fibromyalgia criteria, which were then tested in 300 patients with a previous diagnosis of fibromyalgia and in control patients with rheumatic disorders that could be confused with fibromyalgia (2). The combination of widespread pain and mild or greater tenderness in ≥11 of 18 tender point sites provided good sensitivity and specificity and were selected by consensus as the 1990 classification criteria. Implied, although not formally part of the criteria, fibromyalgia was to be considered only in individuals whose symptoms could not be explained by another rheumatic or medical disorder. It was reassuring that fibromyalgia did not differ in those patients with a concurrent rheumatic disease, such as rheumatoid arthritis (RA), thereby abandoning the term secondary fibromyalgia.

After 20 years of use of these criteria in clinical practice and research, the fibromyalgia criteria were modified in 2010 (3). It was apparent that the tender point examination was difficult to standardize and not widely practiced, especially in primary care. The focus on specific tender point locations implied peripheral tissue pathology, whereas it had become clear that fibromyalgia was a manifestation of central pain sensitization. These 2010 criteria emphasize symptoms, such as fatigue and sleep disturbances, and measure symptom severity over time, recognizing the changing nature of those symptoms. These 2010 criteria were slightly modified in 2011, eliminating the physician's estimate of the extent of somatic symptoms and substituting patient self-reported symptoms, thereby eliminating the need for physician involvement in diagnosis (4).

The so-called "gold standard" in each of these criteria sets, as well as those used for depression, headaches, and other common illness, is expert opinion. In evaluating the sensitivity and specificity of the 2010 criteria, the gold standard was the 1990 criteria. It was explicitly stated that these 2010 criteria were "not meant to replace the ACR classification criteria" (3). We also stated that the 2011 criteria modification were for epidemiologic and clinical studies and "not to be used for self-diagnosis" (4).

However, Wolfe and colleagues chose these 2011 self-report fibromyalgia criteria as their gold standard for a fibromyalgia diagnosis, despite voicing concern regarding the "self-report nature of fibromyalgia" (1). Their finding of a significant discrepancy between these 2011 criteria-based diagnoses and a physician-based diagnosis of fibromyalgia should not be surprising. Jones et al reported that the prevalence of fibromyalgia was significantly higher when using these 2011 criteria compared to the 1990 or 2010 criteria (5). Vincent and colleagues found that 1% of the gen-

Don L. Goldenberg, MD: Oregon Health Sciences University, Portland, and Tufts University School of Medicine, Medford, Massachusetts.

Address correspondence to Don L. Goldenberg, MD, 1125 NW 12th Avenue, Portland, OR 97209. E-mail: dongoldenberg44@gmail.com.

Submitted for publication August 2, 2018; accepted in revised form August 14, 2018.

eral population in Olmsted County, Minnesota, had a diagnosis of fibromyalgia documented in the medical record by a health care provider, but 6% met the 2011 modified fibromyalgia criteria (6).

A number of problems arise when using the 2011 criteria for the diagnosis of fibromyalgia. Patients are more likely to report symptoms, and the physician-completed versus patient-completed symptom severity scales correlated poorly in the study by Jones et al (5). In the 2010 criteria, it was stated explicitly, rather than implied, that "The patient does not have a disorder that would otherwise explain the pain" (3). The 2011 modification requires that the patient have the medical expertise to answer that query.

Could we get better agreement with a physician-based diagnosis of fibromyalgia if we used different classification criteria? Dean et al observed no difference in the association of pain with other core features of fibromyalgia, regardless of whether chronic, widespread pain, or multisite pain definitions were used (7). The American Pain Society has recommended a new taxonomy for chronic pain disorders incorporating 5 dimensions: 1) core diagnostic criteria, 2) common features, 3) common medical co-morbidities, 4) neurobiologic, psychosocial, and functional consequences, and 5) potential illness mechanisms, risk factors, and protective factors, and this classification system has been applied to fibromyalgia (8).

It also should be noted that the 1990, 2010, and 2011 criteria were designated as classification criteria, not diagnostic criteria. Classification criteria are designed to provide uniformity to clinical trials and epidemiologic studies. They are not intended for patient diagnosis, which is the responsibility of the evaluating physician. The sensitivity and specificity of any classification criteria, whether for RA, systemic lupus erythematosus, or fibromyalgia, are measured by their agreement with expert opinion.

Wolfe et al maintained the notion that the diagnosis of fibromyalgia is on a continuum, rather than dichotomous, and coined the term fibromyalgianess to express the trait(s) of fibromyalgia. Indeed, fibromyalgia belongs on the spectrum of chronic widespread pain, which, using the 1990 ACR definition, is present in 10–15% of the general population (Figure 1) (9,10). Fibromyalgia falls on the extreme end of that pain severity and chronicity scale. Depending on classification criteria, its prevalence range is 1–5%.

Every symptom-based medical condition is defined by symptom severity. Chronic widespread pain is no different from headaches or depression. At the extremes of chronicity and severity, symptoms can morph from illness to a diagnosable disease, and the trait become a state, a defined illness with potentially serious consequences (Figure 1). Furthermore, as symptoms wax and wane, the state of fibromyalgia may come and go. Wolfe and his colleagues are concerned with overdiagnosis of fibromyalgia symptoms as a disease, defined as a disorder of structure or function. They worry that, promoted by patient-targeted pharmaceutical advertising, this results in medicalization of common symptoms. Other investigators maintain that symptom-based conditions, such as depression or chronic pain, should be recognized as a disease, which results in better awareness and management.

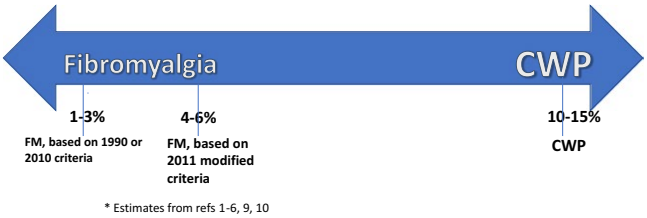


Figure 1. Prevalence of fibromyalgia (FM) and chronic widespread pain (CWP) in general population studies.

Wolfe et al conclude that expert physicians often misdiagnose fibromyalgia (1). This conclusion implies that published criteria are superior to expert clinical judgment for individual patient diagnosis. This notion fails to account for the multiple variables in the clinical encounter, not the least of which is time spent in evaluating and categorizing each patient's multiple symptoms. Only the health care provider can perform the physical examination and interpret psychosocial factors and co-morbid illness that clarify a diagnosis. We may incorporate classification criteria in our electronic health care records, but classification criteria never replace clinical acumen.

It is true that primary care physicians often are not comfortable making the diagnosis of fibromyalgia. It does not necessarily follow that their comfort level would improve if they had better knowledge of the fibromyalgia criteria. It is the responsibility of the rheumatology community to better educate our primary care colleagues regarding diagnosis and differential diagnosis. Rheumatologists also should be available to confirm a fibromyalgia diagnosis and treatment plan, just as psychiatrists do for depression or neurologists for headache disorders. Rheumatologists are best able to strategize subsets of fibromyalgia patients and assist in individualized management, particularly in the most challenging scenarios.

During the past decade, it has become clear that concurrent fibromyalgia is common in every rheumatic disease and complicates diagnoses, disease assessment, and therapy (11,12). Wolfe et al found an even greater discrepancy between physician-based and criteria-based diagnosis in fibromyalgia patients with a concurrent rheumatic disease, including RA or systemic lupus erythematosus (1). This suggests that greater rheumatology education be devoted to the co-morbidity of fibromyalgia with the traditional rheumatic diseases and the adverse impact from that co-occurrence (12). Rheumatologists, partnering with pain specialists, have also demonstrated that fibromyalgia severity scores correlate with poor outcome in patients with chronic spine pain and following joint replacements (13,14). We need to advise primary care physicians, orthopedists, and other specialists on the impact of fibromyalgia/chronic widespread pain in a vast array of systemic and regional musculoskeletal pain disorders.

The diagnostic gold standard for fibromyalgia will continue to be the rheumatologist's expert opinion, not classification cri-

teria, no matter how well-refined and intentioned. This is the only way to capture the variability and severity of interrelated symptoms as they play out over time. There is no short-cut. We are the go-to experts for patients and our colleagues for the diagnosis of fibromyalgia, whether or not we readily accept that role.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Wolfe F, Schmukler J, Jamal S, Castrejon I, Gibson KA, Srinivasan S, et al. Diagnosis of fibromyalgia: disagreement between fibromyalgia criteria and clinician based fibromyalgia diagnosis in a university clinic. Arthritis Care Res 2019;71:343–51.
2. Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL, et al. The American College of Rheumatology 1990 criteria for the classification of fibromyalgia. Arthritis Rheum 1990;33:160–72.
3. Wolfe F, Clauw DJ, Fitzcharles M, Goldenberg DL, Katz RS, Mease P, et al. The American College of Rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity. Arthritis Care Res (Hoboken) 2010;62:600–10.
4. Wolfe F, Clauw DJ, Fitzcharles MA, Goldenberg DL, Häuser W, Katz RS, et al. Fibromyalgia criteria and severity scales for clinical and epidemiological studies: a modification of the ACR Preliminary Diagnostic Criteria for Fibromyalgia. J Rheumatol 2011;38: 1113–22.
5. Jones GT, Atzeni F, Beasley M, Flüß E, Sarzi-Puttini P, Macfarlane GJ. The prevalence of fibromyalgia in the general population: a comparison of the American College of Rheumatology 1990, 2010, and modified 2010 classification criteria. Arthritis Rheumatol 2015;67:568–75.
6. Vincent A, Lahr BD, Wolfe F, Clauw DJ, Whipple MO, Oh TH, et al. Prevalence of fibromyalgia: a population-based study in Olmsted County, Minnesota, utilizing the Rochester Epidemiology Project. Arthritis Care Res (Hoboken) 2013;65:786–92.
7. Dean LE, Arnold L, Crofford L, Bennett R, Goldenberg D, Fitzcharles MA, et al. Impact of moving from a widespread to multisite pain definition on other fibromyalgia symptoms. Arthritis Care Res (Hoboken) 2017;69:1878–86.
8. Arnold LM, Bennett RM, Crofford LJ, Dean LE, Clauw DJ, Goldenberg DL, et al. AAPT diagnostic criteria for fibromyalgia. J Pain 2018. E-pub ahead of print.
9. Papageorgiou AC, Silman AJ, Macfarlane GJ. Chronic widespread pain in the population: a seven year follow up study. Ann Rheum Dis 2002;61:1071–4.
10. Mansfield KE, Sim J, Jordan JL, Jordan KP. A systematic review and meta-analysis of the prevalence of chronic widespread pain in the general population. Pain 2016;157:55–64.
11. Basu N, Kaplan CM, Ichesco E, Larkin T, Harris RE, Murray A, et al. Neurobiologic features of fibromyalgia are also present among rheumatoid arthritis patients. Arthritis Rheumatol 2018;70:1000–7.
12. Salaffi F, Gerardi MC, Atzeni F, Batticciotto A, Talotta R, Draghessi A, et al. The influence of fibromyalgia on achieving remission in patients with long-standing rheumatoid arthritis. Rheumatol Int 2017;37:2035–42.
13. Brummett CM, Goesling J, Tsodikov A, Meraj TS, Wasserman RA, Clauw DJ, et al. Prevalence of the fibromyalgia phenotype in patients with spine pain presenting to a tertiary care pain clinic and the potential treatment implications. Arthritis Rheumatol 2013;65:3285–92.
14. Brummett CM, Urquhart AG, Hassett AL, Tsodikov A, Hallstrom BR, Wood NI, et al. Characteristics of fibromyalgia independently predict poorer long-term analgesic outcomes following total knee and hip arthroplasty. Arthritis Rheumatol 2015;67:1386–94.

Arthritis Care & Research
Vol. 71, No. 3, March 2019, pp 337–342
DOI 10.1002/acr.23823

AMERICAN COLLEGE OF RHEUMATOLOGY
EDUCATION • TREATMENT • RESEARCH

Evaluating the Perception Among Rheumatologists of Maintenance of Board Certification in the US

Amr H. Sawalha and Patrick Coit

Objective. There continues to be a debate about the value and purpose of maintenance of certification (MOC) programs in the US. The goal of this study is to assess the impact, value, and purpose of MOC programs in rheumatology.

Methods. A survey was sent to 3,107 rheumatologists in the US. The survey addressed how rheumatologists perceive the value and impact of MOC programs on rheumatology practice and patient care.

Results. A total of 515 rheumatologists completed this survey. The majority (74.8%) believed there was no significant value in MOC, beyond what is already achieved from continuing medical education. Most rheumatologists did not believe MOC was valuable in improving patient care (63.5%), and the majority felt that the primary reason for creating MOC was either the financial well-being of board-certifying organizations (43.4%) or to satisfy administrative requirements in health systems (30%). Although 65.6% perceived that staying current with new medical knowledge was a positive impact of MOC programs, the MOC was perceived to result in time away from providing patient care (74.6%) and time away from family (74%). When asked about potential effects of requiring MOC, 77.7% reported physician burnout, 67.4% early physician retirement, and 63.9% anticipated an effect on reducing the overall number of practicing rheumatologists.

Conclusion. The majority of rheumatologists do not believe there is significant value for MOC programs. There is evidence for lack of trust in board-certifying organizations, and rheumatologists believe MOC programs contribute to physician burnout, early retirement, and loss in the rheumatology workforce.

INTRODUCTION

Board certification in medical specialties in the US was initially introduced at the turn of the twentieth century to provide a mechanism to demonstrate the clinical competence of practicing physicians. Achieving board certification for a specific specialty or subspecialty used to be a landmark life-long accomplishment for physicians after completing their initial medical training. Board certification in the US is overseen by the American Board of Medical Specialties (ABMS), an umbrella self-governing organization that includes specialty board members such as the American Board of Internal Medicine (ABIM) and the American Board of Pediatrics. Over time, organizations entrusted by physicians to conduct and issue board certifications ended life-long certification, and began issuing time-limited board certificates that required regular renewals. Although ensuring the continuous medical education of physicians was suggested to be the motive for this change, this move remains controversial because a benefit to patients by this process has not been supported by credible scientific, non-conflicted data, and because it excluded or "grandfathered" physicians who had received their initial board certification at an arbitrary earlier time point (1).

More recent changes in board recertification requirements and the expansion of a controversial maintenance of certification (MOC) program finally caught the public attention and resulted in significant discussions both at the national level and within the medical community concerning the value, motive, and consequences behind this program and board recertification (2,3). It has been suspected that this program financially supports board-certifying organizations that have been growing in size, and although self-declared as non-profit organizations, have actively participated in lobbying activities (4,5). Interestingly, MOC programs operated by a specialty body of the ABMS were included as a measure for physician quality and payment incentives within

Amr H. Sawalha, MD, Patrick Coit, MPH: University of Michigan, Ann Arbor.

Address correspondence to Amr H. Sawalha, MD, Division of Rheumatology, University of Michigan, 1150 W. Medical Center Drive 5520 MSRB1, SPC 5680, Ann Arbor, MI 48109. E-mail: asawalha@umich.edu.

Submitted for publication August 22, 2018; accepted in revised form December 4, 2018.

SIGNIFICANCE & INNOVATIONS

- The survey assessed the value of a maintenance of board certification (MOC) program from the perspective of rheumatologists in the US.
- Overall, there appears to be no perceived significant value for MOC among rheumatologists.
- MOC is perceived to contribute to a shortage in the rheumatology workforce in the US, without proven benefit to patients.
- There is evidence for lack of trust among rheumatologists in board-certifying organizations. A financial conflict of interest in creating MOC is perceived to be a likely reason for creating MOC among rheumatologists.

the Affordable Care Act (6,7). Non-transparent and at times questionable financial practices of board-certifying organizations have surfaced in the public domain, eroding trust in these organizations of physicians and medical societies (4). In response, apologies and restructuring attempts have been made (8), but the core issue of whether recertification and MOC programs positively impact health care has remained unclear (9,10). It is alarming that while no clear noncontroversial benefit to patients from MOC has been demonstrated, the mean estimated cost of the ABIM MOC program for all internal medicine physicians is estimated to be ~$5.7 billion over a 10-year period, of which $561 million are fees payable to the ABIM (11).

Board certification, which started as a voluntary achievement and remains so in theory, has become involuntary in practice, making participation in MOC programs mandatory for many if not most physicians in order to maintain employment and clinical privileges, or receive reimbursement. The controversy surrounding the motive behind creating MOC, the value of MOC, and the fear for this requirement to interfere with the ability to practice, has prompted legislatures in several states to prohibit using MOC as a condition for employment, licensure, securing clinical privileges, and reimbursement. Oklahoma was the first state to pass such legislation. The National Board of Physicians and Surgeons (NBPAS) was established as a grass-roots organization to provide an alternative recertification process that would allow physicians to recertify to maintain their practices at a fraction of the cost required by the ABMS and ABIM. To date, NBPAS, which uses participation in continuing medical education (CME) activities as a basis for recertification, has certified over 7,000 physicians and is accepted by 90 hospitals and health systems within the US (12).

Rheumatologists have been very closely following the controversy regarding MOC and board recertification. Indeed, the American College of Rheumatology (ACR) has questioned the value of the ABIM MOC program, and issued a statement raising concerns regarding the cost of MOC, lack of evidence to support beneficial impact of MOC on clinical care, and concerns about the financial stewardship of the ABIM (13). The goal of this survey study was to assess the current perception of practicing US rheumatologists of the value of board recertification and MOC programs, and assess the impact of these programs on rheumatologists and the rheumatology workforce.

MATERIALS AND METHODS

A survey was designed to assess the impact and perceived value of MOC programs in rheumatology. The survey consisted of 20 questions, including 19 closed-ended questions and 1 open-ended question (Supplementary Appendix 1, on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23823/abstract). The questions were designed to address issues relevant to concerns discussed in recent editorials, opinion pieces, and medical society statements concerning MOC programs (2–5,13). The 2 questions pertaining to the perceived positive and negative impact of board recertification and MOC were randomized such that half of the respondents receive each of the 2 questions first. The survey included 1 ranking question, with the order of possible answers for this question also being randomized.

Ten practicing rheumatologists at the University of Michigan, (including a mixture of predominantly clinical faculty and physician scientists who are ABIM board certified and participating in

Table 1. Demographics, primary specialty, and primary work setting of respondents*

	No.	Frequency, %
Primary specialty		
Adult rheumatology	469	91.2
Pediatric rheumatology	33	6.4
Both adult and pediatric	12	2.3
Primary work setting		
Academia/university	223	43.4
Private practice	231	44.9
Government	19	3.7
Industry/pharmaceutical	3	0.6
Other	38	7.4
Age, years		
30–39	29	5.6
40–49	160	31.1
50–59	142	27.6
60–69	152	29.6
≥70	31	6.0
Sex		
Female	202	39.3
Male	312	60.7

* Demographic information was provided by 514 of 515 survey respondents.

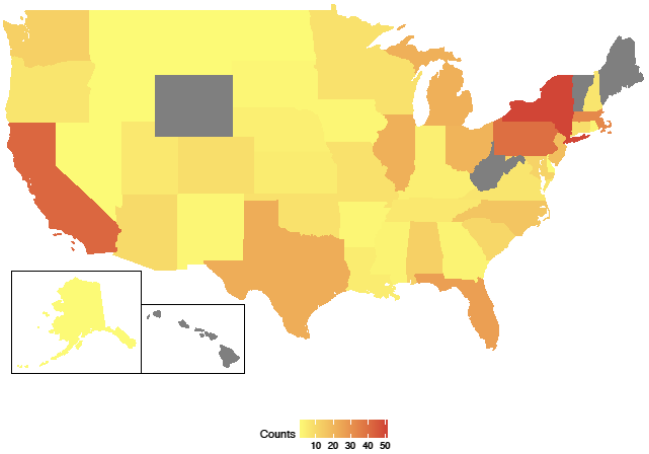


Figure 1. The geographic distribution of rheumatologists who responded to the survey.

MOC) were asked to provide input on the survey questions before the questions were finalized. These 10 rheumatologists and the authors were excluded from taking the survey for the purpose of this study. The survey design and all survey questions were reviewed by staff at the Consulting for Statistics, Computing & Analytic Research at the University of Michigan. The survey was constructed using SurveyMonkey (San Mateo, CA) and sent via email to 3,107 rheumatologists within the US. Invitations to participate in this survey were initially sent out to all 3,107 rheumatologists on March 6, 2018 with responses collected until March 26, 2018 when participation in this survey was closed. Analysis of survey results was performed in the computing environment R.

RESULTS

A total of 515 rheumatologists completed the survey. With an estimated number of ~5,000 practicing rheumatologists in the US, this sample size provides 95% confidence that the expected responses of practicing rheumatologists in the US are within a margin of error of less than 5% of the responses obtained from the rheumatologists who completed this survey.

The majority of respondents were adult rheumatologists (91.2%), while 6.4% were pediatric rheumatologists, and 2.3% were both adult and pediatric rheumatologists. The majority of respondents identified their primary work setting as either private practice (44.9%) or a university or academic setting (43.4%). Of the total respondents, 60.7% were male and 39.3% were female rheumatologists. In terms of age, the majority of respondents

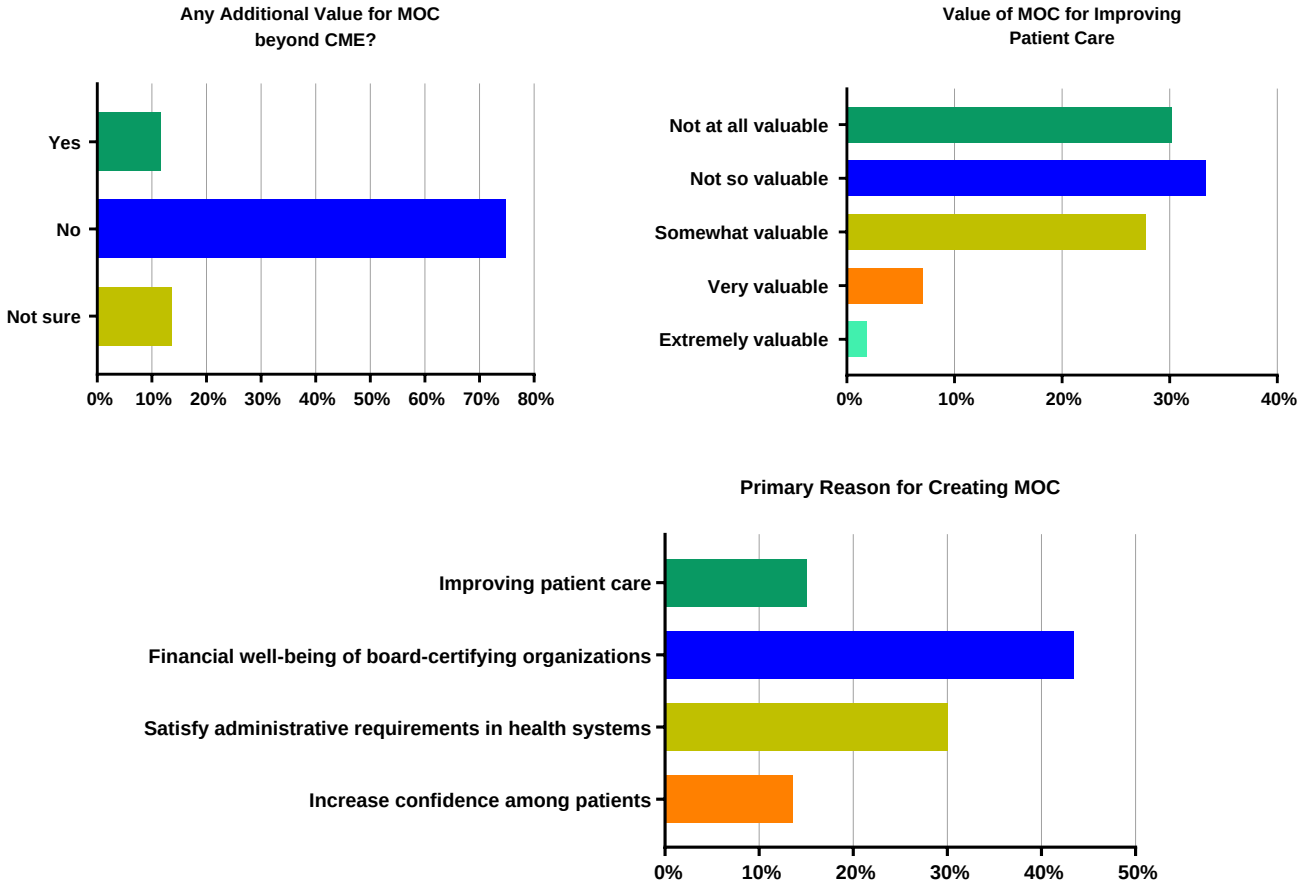


Figure 2. Summary of responses evaluating the value of maintenance of certification (MOC) programs among rheumatologists beyond continuing medical education (CME) and as it pertains to improving patient care, and the perceived reasons for creating MOC programs.

were either age range 40–49 years (31.1%) or 60–69 years (29.6%) (Table 1). Figure 1 shows the geographic distribution for practice locations for the rheumatologists responding to this survey.

The majority of surveyed rheumatologists in the US (74.8%) did not think there was significant additional value in MOC, beyond what is already achieved from CME. Indeed, 63.5% of rheumatologists did not believe board recertification and MOC were valuable in terms of improving patient care (Figure 2). When asked about the primary reason for creating MOC, the majority of rheumatologists felt it was for the financial well-being of board-certifying organizations (43.4%) or to satisfy administrative requirements in health systems (30%). Only 15.1% believed improving patient care was the primary reason for MOC (Figure 2). The majority of rheumatologists believed board certification should be a life-long credential (63.7%).

Notably, when asked about possible negative impacts of MOC, the majority reported that MOC resulted in time away from providing patient care (74.6%), time away from family (74%), and psychological stress (69.7%). In addition, 88.5% of rheumatologists believed that MOC imposes a financial burden on rheumatologists without proven benefits to patients. When asked about possible positive impacts of MOC, 65.6% perceived staying current with new knowledge as a positive impact. Most rheumatologists did not identify patient reassurance, improved quality of patient care, or increased patient satisfaction as possible positive impacts of MOC. When asked about anticipated effects of requiring MOC, 77.7% of the respondents reported physician burnout, 67.4% early physician retirement, and 63.9% anticipated a reduction in the overall number of practicing rheumatologists. Only 14.2% of the respondents believed requiring MOC will improve the overall quality of practicing rheumatologists in the US, and 75.2% favored legislation in their state to remove MOC as a requirement for employment, insurance reimbursement, or securing clinical privileges (Figure 3).

Of interest, 58.9% of the respondents believed board certification in rheumatology should be administered or overseen by other organizations such as the ACR, and 53.7% of the surveyed rheumatologists reported participation in basic, translational, or clinical research in rheumatology. Of the respondents who reported participating in research activities, 39.6% believed MOC was adversely affecting their ability to perform research or research-related activities.

The survey included 1 open-ended question, in which rheumatologists were asked to provide any other relevant thoughts or comments. Comments were received from 186 survey respondents. To summarize these responses, individual ideas or thoughts were grouped into categories that summarized the themes discussed. These groups were ranked based on the number of times a theme was mentioned among all responses. The top 5 themes are listed in Table 2. Overall, these comments echoed the results derived from the closed-ended questions in the survey and stressed issues such as questioning the value of MOC over CME activities, the relevance of material covered by recertification examinations to daily rheumatology clinical practice, the high expense associated with MOC, and the motive behind developing MOC programs.

DISCUSSION

This study was conducted to assess the perception of the value and impact of MOC programs from the perspective of rheumatologists, and how these programs might affect rheumatology practice and the rheumatology workforce in the US. The data derived from this survey study suggest an overall lack of value for these programs, as perceived by prac-

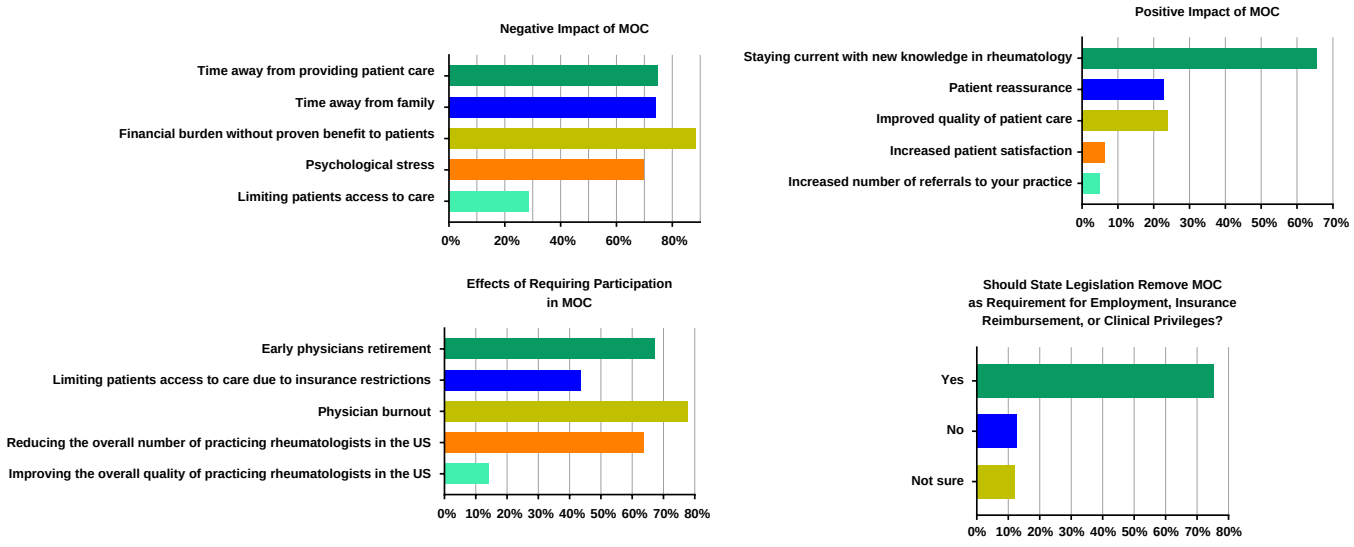


Figure 3. Summary of responses evaluating the impact of maintenance of certification (MOC) programs among rheumatologists.

Table 2. Top ranked themes described by respondents in the open-ended survey question asking for additional thoughts or comments*

Rank	Theme
1	Critical of monetary and time cost of MOC and/or impact on patient care
2	Critical of the MOC idea and process
3	CME is sufficient and more relevant than MOC to stay up to date
4	Critical of ABIM and its motives
5	MOC examination is not relevant to clinical practice

* MOC = maintenance of certification; CME = continuing medical education; ABIM = American Board of Internal Medicine.

ticing rheumatologists, which is consistent with other recent studies across different specialties (10). Physicians, including rheumatologists, are committed to life-long learning and appreciate the importance of keeping up to date with recent knowledge and developments in the field to provide the best possible patient care. It seems that the majority of rheumatologists do not believe that MOC programs are the best means to achieve this goal and ensure physician competence. The majority believe that MOC programs do not add significant value to participation in CME activities, which are already required to maintain and renew state medical licensures. Indeed, unlike MOC programs, CME activities can be more flexible and allow individual rheumatologists to participate in educational activities that are most relevant to their individual practices or the patient populations they manage. In addition, CME activities can be achieved at a fraction of the cost required to participate in MOC programs. A recent cost analysis suggests that the ABIM's MOC is associated with significant testing and time costs for rheumatologists participating in that program, with a mean cost of the program over 10 years of $21,606 per rheumatologist, and a mean cost aggregated for rheumatology as a specialty of $89 million, of which $11 million are fees payable to the ABIM (11).

Importantly, the overwhelming perception of rheumatologists in the US is that enforcing MOC participation results in physician burnout and a reduction in the rheumatology workforce. When asked about whether rheumatologists are required to participate in MOC for employment or insurance reimbursement, 36.2% of rheumatologist reported that they are required to participate in MOC and another 23.7% reported they are not sure if they are required to participate. These data suggest that while the concept of board certification was initially introduced as a voluntary achievement, time-limited certification and now MOC are becoming required for many rheumatologists to sustain employment or practice. The 2015 ACR Workforce Study suggested a current shortage in the rheumatology workforce, and predicted a decline in the number of practicing rheumatologists by 2030 with a supply being outnumbered ~2-fold by the demand for rheumatologists in the US (14,15). If participation in MOC does not have a significant value on improving patient care in rheumatology, as indicated by the perception of practicing rheumatologists, and if it remains a requirement to practice for at least one-third of rheumatologists in the US, then an argument can be made that elimination of MOC might be a way to sustain and improve the rheumatology workforce without compromising quality. To our knowledge, there have not been any studies to show that rheumatologists participating in MOC activities provide better care.

A striking finding from our study is the indication that there seems to be a lack of trust by practicing rheumatologists in the US in board-certifying organizations and their motives. When asked to rank in order what is thought to be the reason for creating MOC programs, financial well-being of board-certifying organizations was the highest ranked answer. Improving patient care, which is the motive claimed by board-certifying organizations, was the lowest ranked reason. Regardless of what the motive might be, these results suggest that practicing physicians, and in this case rheumatologists, do not trust board-certifying organizations. Therefore, we suggest that these organizations revisit their relationships with practicing physicians, and facilitate true collaboration with those physicians to determine the best way to assess and ensure physician competence and knowledge. Of interest, ~60% of rheumatologists believe that alternative organizations, such as the ACR, should be involved in administering or overseeing board certification of rheumatologists.

The NIH has warned that imposing time-consuming MOC programs seems to discourage physician-scientists from maintaining clinical practice (16). This is a potentially serious problem in rheumatology because physician-scientists are the drivers of new discoveries that result in better treatment options and improved care of patients with rheumatologic conditions. Researchers who maintain their own clinical practice are able to stay connected with research questions that are of immediate interest to the diseases being studied and to the patients. In addition, eliminating or reducing the number of physician-scientists who participate in patient care will reduce the number of some of the most qualified and talented physicians who can take care of the most complicated patients in rheumatology and understand the specific disease(s) related to their expertise. Along these lines, our survey revealed that about 40% of rheumatologists who participate in research report a negative impact of MOC on their research careers, as reflected by adversely affecting their ability to perform research or research-related activities.

Limitations of our study include the fact that the survey focused on evaluating perception and attitudes of rheumatologists, and the results cannot be interpreted to provide actual consequences of MOC impact. Nonetheless, the rheumatologist's perceptions and opinions captured in this survey should, at the very least, invite a serious independent nonconflicted evaluation of these programs. The survey did not

address alternative ways for structuring MOC. However, our results indicate that most rheumatologists believe that participating in CME activities is sufficient and they perceive no significant additional value for the MOC program. This belief is consistent with the ACR statement on MOC suggesting that earning MOC points (mandated MOC educational activities) is redundant (13).

In summary, our results suggest that the majority of rheumatologists in the US are concerned about recertification examinations and MOC programs. It appears that these programs are not perceived to be of significant value and do seem to have the potential to contribute to the shortage in the rheumatology workforce in the US. Importantly, there is evidence for eroding credibility and lack of trust in board-certifying organizations among rheumatologists, and a notion that imposing a time limit on board certification and mandating participation in expensive MOC programs is largely driven by financial interests of these organizations rather than improving patient care. The medical community in general and the rheumatology community in particular need to address the gradual transformation of board certification and MOC from a voluntary activity to practically a requirement in order for many physicians to be able to practice medicine and get reimbursed for services provided. It is important to caution against lobbying activities driven by financial interests in setting health care policies, especially in mandating expensive programs such as MOC, in the absence of convincing data that demonstrate improved patient care, which could result in serious consequences in a field such as rheumatology that is threatened by a large shortage in the workforce.

AUTHOR CONTRIBUTIONS

Both authors were involved in drafting the article or revising it critically for important intellectual content, and both authors approved the final version to be published. Dr. Sawalha 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. Sawalha.
Acquisition of data. Sawalha.
Analysis and interpretation of data. Sawalha, Coit.

REFERENCES

1. Hayes J, Jacksn JL, McNutt GM, Hertz BJ, Ryan JJ, Pawlikowski SA. Association between physician time-unlimited vs time-limited internal medicine board certification and ambulatory patient care quality. JAMA 2014;312:2358–63.
2. Centor RM, Fleming DA, Moyer DV. Maintenance of certification: beauty is in the eyes of the beholder. Ann Intern Med 2014;161:226–7.
3. Iglehart JK, Baron RB. Ensuring physicians' competence–is maintenance of certification the answer? N Engl J Med 2012;367:2543–9.
4. Fisher WG, Schloss EJ. Medical specialty certification in the United States-a false idol? J Interv Card Electrophysiol 2016;47:37–43.
5. Kempen PM. Maintenance of Certification - important and to whom? J Community Hosp Intern Med Perspect 2013;April 17:3.
6. Kinney ED. The Affordable Care Act and the Medicare Program: linking Medicare payment to quality performance. NYU Annual Survey of American Law. 2014;Research Paper No. 2014-22
7. The Patient Protection and Affordable Care Act; 124 STAT. 119, Public Law 111–148. 2010; URL:https://www.congress.gov/111/plaws/publ148/PLAW-111publ148.pdf.
8. Baron RJ. ABIM announces immediate changes to MOC Program. URL: https://www.abim.org/news/abim-announces-immediate-changes-to-moc-program.aspx
9. Cook DA, Holmboe ES, Sorensen KJ, Berger RA, Wilkinson JM. Getting maintenance of certification to work: a grounded theory study of physicians' perceptions. JAMA Intern Med 2015;175:35–42.
10. Cook DA, Blachman MJ, West CP, Wittich CM. Physician attitudes about maintenance of certification: a cross-specialty national survey. Mayo Clin Proc 2016;91:1336–45.
11. Sandhu AT, Dudley RA, Kazi DS. A cost analysis of the American Board of Internal Medicine's Maintenance-of-Certification Program. Ann Intern Med 2015;163:401–8.
12. National Board of Physicians and Surgeons (NBPAS). Home page. URL: https://nbpas.org/.
13. American College of Rheumatology. Position Statement maintenance of certification. URL: https://www.rheumatology.org/Portals/0/Files/ACR%20MOC%20Postion%20Statement%202015.pdf.
14. Bolster MB, Bass AR, Hausmann JS, Deal C, Ditmyer M, Greene KL, et al. 2015 American College of Rheumatology Workforce Study: the role of graduate medical education in adult rheumatology. Arthritis Rheumatol 2018;70:817–25.
15. Battafarano DF, Ditmyer M, Bolster MB, Fitzgerald JD, Deal C, Bass AR, et al. 2015 American College of Rheumatology Workforce Study: supply and demand projections of adult rheumatology workforce, 2015-2030. Arthritis Care Res (Hoboken) 2018;70:617–26.
16. US Department of Health & Human Services. Physician-scientist workforce (PSW) report 2014. URL: https://report.nih.gov/Workforce/PSW/challenges.aspx.

Arthritis Care & Research
Vol. 71, No. 3, March 2019, pp 343–351
DOI 10.1002/acr.23731

AMERICAN COLLEGE OF RHEUMATOLOGY
EDUCATION • TREATMENT • RESEARCH

Diagnosis of Fibromyalgia: Disagreement Between Fibromyalgia Criteria and Clinician-Based Fibromyalgia Diagnosis in a University Clinic

Frederick Wolfe,[1] Juan Schmukler,[2] Shakeel Jamal,[2] Isabel Castrejon,[2] Kathryn A. Gibson,[3] Sachin Srinivasan,[4] Winfried Häuser,[5] and Theodore Pincus[2]

Objective. Recent studies have suggested that fibromyalgia is inaccurately diagnosed in the community, and that ~75% of persons reporting a physician diagnosis of fibromyalgia would not satisfy published criteria. To investigate possible diagnostic misclassification, we compared expert physician diagnosis with published criteria.

Methods. In a university rheumatology clinic, 497 patients completed the Multidimensional Health Assessment Questionnaire (MD-HAQ) and the 2010 American College of Rheumatology preliminary diagnostic criteria modified for self-administration during their ordinary medical visits. Patients were evaluated and diagnosed by university rheumatology staff.

Results. Of the 497 patients, 121 (24.3%) satisfied the fibromyalgia criteria, while 104 (20.9%) received a clinician International Classification of Diseases (ICD) diagnosis of fibromyalgia. The agreement between clinicians and criteria was 79.2%. However, agreement beyond chance was only fair (κ = 0.41). Physicians failed to identify 60 criteria-positive patients (49.6%) and incorrectly identified 43 criteria-negative patients (11.4%). In a subset of 88 patients with rheumatoid arthritis (RA), the kappa value was 0.32, indicating slight to fair agreement. Universally, higher polysymptomatic distress scores and criteria-based diagnosis were associated with more abnormal MD-HAQ clinical scores. Women and patients with more symptoms but fewer pain areas were more likely to receive a clinician's diagnosis than to satisfy fibromyalgia criteria.

Conclusion. There is considerable disagreement between ICD clinical diagnosis and criteria-based diagnosis of fibromyalgia, calling into question ICD-based studies. Fibromyalgia criteria were easy to use, but problems regarding clinician bias, meaning of a fibromyalgia diagnosis, and the validity of physician diagnosis were substantial.

INTRODUCTION

Although fibromyalgia is a common diagnosis, it remains a contested disorder to many (1–4). It is an arbitrarily and broadly defined disorder of widespread pain and multiple symptoms that is strongly influenced by culture, context, and social forces (5). Several widely accepted criteria sets have provided definitions and methods of diagnosis since fibromyalgia was first named by Hench in 1976 (6–11). However, a number of studies have suggested that fibromyalgia is overdiagnosed, citing the influence of pharmaceutical company advertising (1,12). Advertising mechanisms that advance the idea of symptoms as disease can lead to illness expansion and overdiagnosis (13,14) and to the indirect suggestion to patients and physicians that common symptoms should be interpreted and diagnosed as fibromyalgia (1,12,15,16). The self-report nature of fibromyalgia symptoms facilitates the uncertainty that surrounds fibromyalgia diagnosis (15). In the current study, we consider underdiagnosis as well as the form of overdiagnosis that lead to diagnosis when fibromyalgia criteria are not satisfied and symptoms are insufficient for diagnosis. Recent evidence has accumulated from population studies that support the idea of common and frequent misdiagnosis (17,18), as well as reports of limited knowledge of fibromyalgia diagnosis and criteria by physicians in the community (19,20). Some have asked whether a "gestalt" diagnosis of

[1]Frederick Wolfe, MD: Arthritis Diseases Center, National Data Bank for Rheumatic Diseases, and University of Wichita School of Medicine, Wichita, Kansas; [2]Juan Schmukler, MD, Shakeel Jamal, MD, Isabel Castrejon, MD, PhD, Theodore Pincus, MD: Rush University Medical Center, Chicago, Illinois; [3]Kathryn A. Gibson, MD: Liverpool Hospital, Sydney, New South Wales, Australia; [4]Sachin Srinivasan, MD: Kansas University Medical Center, Wichita; [5]Winfried Häuser, MD: Klinikum Saarbrücken gGmbH, Saarbrücken, Germany.

Address correspondence to Frederick Wolfe, MD, National Data Bank for Rheumatic Diseases, 1035 North Emporia, Wichita, KS 67214. E-mail: fwolfe@arthritis-research.org.

Submitted for publication April 2, 2018; accepted in revised form August 14, 2018.

SIGNIFICANCE & INNOVATIONS

- Overall agreement between clinicians' diagnosis of fibromyalgia and diagnosis by fibromyalgia criteria is only fair.
- Clinicians fail to identify almost 50% of criteria-positive cases.
- In rheumatoid arthritis patients, diagnostic agreement is slight to fair.
- International Classification of Diseases diagnoses of fibromyalgia do not accurately capture fibromyalgia status.

fibromyalgia has not become the de facto diagnosis in the community (21).

Despite these concerns, there has been no study comparing simultaneous clinical diagnosis with standard research questionnaire-based criteria (9–11). Such a study might address what is the actual definition of fibromyalgia in the clinic. In this study, we used patients receiving care at the Rush University rheumatology clinic to investigate differences in diagnosis and definition of clinical versus criteria-based diagnosis and to understand the clinical meaning and consequences of diagnostic decisions. Our study should also provide data on the validity and reliability of International Classification of Diseases (ICD) diagnoses of fibromyalgia derived from clinical care. As the gold standard for diagnosis, we used the 2011 self-report modification of the American College of Rheumatology (ACR) preliminary diagnostic criteria for fibromyalgia for clinical research (10) of the ACR 2010 preliminary diagnostic criteria for fibromyalgia (ACR criteria) (9). For clinical diagnosis, we used diagnoses as coded using ICD, Tenth Revision (ICD-10). In studies of fibromyalgia, ICD diagnosis has played a major role, including use for case finding (22), symptom description (23), practice and administrative prevalence estimates (24,25), and heath care costs (25,26).

PATIENTS AND METHODS

Patients. The patients were 497 consecutive unselected rheumatology clinic attendees with complete questionnaire data. In a 3-month period (mid April through mid July, 2017), as part of their usual medical care, each patient completed a Multidimensional Health Assessment Questionnaire (MD-HAQ) as well as a questionnaire that assessed fibromyalgia diagnostic variables used in the ACR 2010 preliminary diagnostic criteria for fibromyalgia and its 2011 modification for self-report (9,10,27). The fibromyalgia questionnaire and MD-HAQ were used together to estimate the usefulness of the MD-HAQ in the diagnosis of fibromyalgia. We used these data and ICD diagnosis to measure agreement between a physician diagnosis and published fibromyalgia criteria. The diagnostic usefulness of the MD-HAQ will be reported elsewhere.

Diagnoses. We used clinical diagnoses coded using ICD-10 in Rush University electronic medical records. Up to 6 rheumatic disease diagnoses were chosen and recorded for each patient by the staff rheumatologist at the time of the patient visit. The mean number of ICD diagnoses was 3.6 (range 1–6). Among major rheumatic diagnoses were systemic lupus erythematosus (SLE) (n = 86), osteoarthritis (OA) (n = 152), RA (n = 88), and fibromyalgia (n = 104). The ICD code used to identify fibromyalgia cases was M79.7. Clinical diagnoses of fibromyalgia were made using physician clinical evaluations, the nature of which was determined by each clinician. The study authors made no suggestion to clinicians about how diseases were to be diagnosed, and with respect to fibromyalgia, 2011 criteria instructions were not provided.

All patients completed the questionnaires immediately prior to being seen for their clinic visit. Completed questionnaires were available to clinicians at the time of the clinical examination, but review of the questionnaires by clinicians was voluntary. Physicians were not told that their ICD records would be searched for evidence of fibromyalgia but were aware of the interest of some of the study authors (IC and TP) in fibromyalgia diagnosis. Two of the authors were clinicians who participated in the study. Dr. Pincus is listed as the physician in 1 case and Dr. Schmuker in 5 cases. Patients were a mix of new and returning clinic patients. Of the 104 clinically diagnosed fibromyalgia patients, only 2 had no other medical diagnosis. We did not attempt to separate fibromyalgia into primary and secondary categories, because 98% of the patients had >1 diagnosis. Clinical examinations and diagnoses were provided by 17 physicians (5 fellows [specialist registrars] and 12 staff rheumatologists). After the study was competed, at a physician clinic conference we determined by interview that physician participants had not used written fibromyalgia criteria or checklists for fibromyalgia diagnosis. In addition, no physician acknowledged specific use of published fibromyalgia criteria in reaching a diagnosis.

Fibromyalgia variables. The fibromyalgia criteria questionnaire is composed of the following variables (9–11).

Widespread pain index (WPI) (0–19 scale). The WPI score is a summary count of the number of 19 painful regions from the Regional Pain Scale, a self-reported list of painful regions (28).

Symptom severity scale (SSS) (0–12 scale). The SSS score is the sum of the severity scores of 3 symptoms (fatigue, waking unrefreshed, and cognitive symptoms) (range 0–9) plus the sum (range 0–3) of the number of the following symptoms the patient has been bothered by that occurred during the previous 6 months: 1) headaches (range 0–1), 2) pain or cramps in lower abdomen (range 0–1), and 3) depression (range 0–1).

Polysymptomatic distress (PSD) (0–31 scale). The PSD score (also known as the fibromyalgia symptom score) is the sum of the WPI score and the SSS score. The PSD scale measures the magnitude and severity of fibromyalgia symptoms in those satisfying and those not satisfying the criteria.

The 2011 criteria, modified for self-report from the 2010 ACR preliminary diagnostic criteria. A patient satisfies the modified 2011 fibromyalgia criteria (10) if the following conditions are met: WPI score ≥7 and SSS score ≥5 *or* WPI score 3–6 and SSS score ≥9. The published requirement for symptoms being present for ≥3 months is usually not considered in nonclinical studies when symptoms are being compared and was not used in this study. However, 96% of fibromyalgia criteria–positive patients had symptoms for ≥3 months. Therefore, the exclusion could not have influenced results.

The 2016 modifications of the 2010 and 2011 criteria were not considered in this study, because the study was started before the general promulgation of the criteria (11). The 2016 criteria state specifically that "a diagnosis of fibromyalgia is valid irrespective of other diagnoses. A diagnosis of fibromyalgia does not exclude the presence of other clinically important illnesses." This comment was inserted into the 2016 criteria to make clear that fibromyalgia criteria were valid in the presence of other clinically important illnesses, including those diagnosed using the 2010 and 2011 criteria.

Clinical variables (MD-HAQ). To determine outcome and clinical status, we used the 2-page, single-sheet MD-HAQ (27). The MD-HAQ contains scales that measure physical function, pain, patient's global assessment, fatigue, and anxiety and depression. The MD-HAQ also includes a Rheumatoid Arthritis Disease Activity Index (RADAI) self-report joint count (29), a 60-symptom checklist, and recent medical history. Although designed for use in RA, the RADAI is a general assessment that can be useful in any musculoskeletal disorder (30). Anxiety and depression were assessed by a 0–3 scale; scores ≥2 were considered to indicate substantial problems.

Statistical analysis. No patient refused to complete the questionnaires. However, administrative procedures in the clinic made it impossible to check and correct for missing data. Therefore, research staff determined that only patients with complete data would be studied, and patients with missing data were not entered into the research database. Data were analyzed using Stata version 15.0 (31). We used nonparametric trend analysis for grouped data Table 1, and used the Stata roctab procedure to perform nonparametric receiver operating characteristic curve analysis and to estimate the area under the curve for the analysis of prediction of fibromyalgia 2011 criteria status by PSD score. Agreement

Table 1. Clinical and severity measures according to fibromyalgia classification status*

	Crit−/Clin−	Crit−/Clin+	Crit+/Crit−	Crit+/Clin+
No. (%) of total patients	333 (67.0)	43 (8.7)	60 (12.1)	61 (12.3)
Age, mean ± SD years	54.0 ± 17.0	51.8 ± 12.7	51.5 ± 16.8	48.6 ± 13.0
Female sex, %	77.7	97.7	83.3	96.7
College graduation, %	47.4	30.8	39.7	43.1
BMI, mean ± SD kg/m^2	30.2 ± 8.6	31.2 ± 7.3	30.3 ± 8.2	35.3 ± 8.6
Disabled	6.9	18.6	33.3	32.8
Widespread pain index	3.0 ± 2.4	4.0 ± 2.6	9.2 ± 3.1	11.0 ± 3.5
Symptom scale	3.1 ± 2.9	4.9 ± 2.1	7.6 ± 2.1	9.0 ± 1.8
PSD scale	6.0 ± 3.8	8.9 3.0	16.9 ± 3.4	20.0 ± 4.2
Generalized pain, %	18.6	18.6	78.3	88.5
Symptom count	7.87 ± 5.8	13.2 ± 6.3	19.3 ± 8.0	25.3 ± 9.8
RAPID-3 index	9.5 ± 6.7	14.4 ± 5.5	17.3 ± 4.8	20.0 ± 4.1
Function (MD-HAQ)	1.8 ± 1.9	2.4 ± 1.6	3.5 ± 1.8	4.6 ± 1.8
Pain (MD-HAQ)	4.0 ± 2.9	6.4 ± 2.5	7.0 ± 2.1	8.2 ± 1.4
Global (MD-HAQ)	3.8 ± 2.7	5.6 ± 2.4	6.7 ± 1.9	7.3 ± 2.0
Fatigue (MD-HAQ)	3.4 ± 3.0	6.0 ± 2.0	7.5 ± 1.9	8.4 ± 1.7
RADAI (0–48)	8.3 ± 8.5	14.0 ± 9.6	21.9 ± 11.2	29.3 ± 11.2
Anxiety (substantial), %	6.0	20.9	25.0	42.6
Depression (substantial), %	4.2	11.6	20.0	36.0

* Except for age and college graduation, values for all other variables increased statistically significantly across groups. Crit− = patients who do not satisfy the 2011 criteria for a fibromyalgia (FM) diagnosis; Crit+ = patients who satisfy the 2011 criteria for FM diagnosis; Clin− = patients diagnosed by clinician as not having FM; Clin+ = patients diagnosed by clinician as having FM; BMI = body mass index; PSD = polysymptomatic distress; RAPID-3 = Routine Assessment of Patient Index Data 3; MD-HAQ = Multidimensional Health Assessment Questionnaire; RADAI = Rheumatoid Arthritis Disease Activity Index.

was assessed by kappa analysis (32) and included probabilistic grouping or interval (33,34). Benchmark groupings include: poor = <0.00, slight = 0.00–0.20, fair = 0.21–0.40, moderate = 0.41–0.60, substantial = 0.61–0.80, and almost perfect = 0.81–1.00 (32). Probabilistic benchmarks represent the probability for each coefficient of falling into the selected benchmark interval along with the cumulative probability of exceeding the predetermined threshold associated with the interval.

Ethics and institutional review board (IRB) approval. This study was conducted in accordance with the ethics standards of the responsible committee on human experimentation and with the Helsinki Declaration of 1975, as revised in 1983. The study was approved by the Rush University IRB. The study received an IRB waiver of consent for retrospective use of data.

RESULTS

According to the 2011 criteria, 121 patients (24.3%) could be classified as having fibromyalgia (Table 2). By clinician diagnosis, the fibromyalgia prevalence was 20.9%. Although agreement between methods was 79.2%, agreement beyond chance agreement was only fair, with a kappa score of 0.41 and a probabilistic benchmark interval of 0.20–0.40. Physicians failed to identify 60 criteria-positive patients (49.6%) and incorrectly identified 43 criteria-negative patients (11.4%). Among the 104 patients clinically diagnosed with fibromyalgia, only 61 (58.7%) actually satisfied criteria, and among the 393 not diagnosed with fibromyalgia by clinicians, 60 (15.3%) satisfied the 2011 criteria. Because the number of individuals without fibromyalgia in the community is much greater than the number of those with fibromyalgia, a misclassification rate of 15.3% is clinically and epidemiologically meaningful.

We also examined agreement after restricting the analysis to RA, SLE, and OA. Among the 88 patients with RA, agreement was 84.1%, and the kappa score was 0.32 with a probabilistic benchmark interval of 0.00–0.20. Among the 13 RA patients with criteria-positive fibromyalgia, 5 (38.5%) were identified by clinicians; among those who were criteria-negative, 6 (8.0%) were called positive by clinicians. Even worse results were obtained in SLE. The kappa value was 0.08. Among the 19 SLE patients with criteria-positive fibromyalgia, 4 (21.1%) were identified by clinicians; among those who were criteria-negative, 9 (13.4%) were called positive by clinicians. Finally, among those with OA, the kappa score was 0.51 with a probabilistic benchmark interval of 0.20–0.40. Among the 44 OA patients with criteria-positive fibromyalgia, 28 (63.6%) were identified by clinicians; among those

Table 2. Prevalence and agreement of criteria and clinician assessment in 497 university rheumatology clinic attendees*

	2011 FM criteria–positive	2011 FM criteria–negative	Total
All patients			
Clinician-positive FM diagnosis	61 (50.4)	43 (11.4)	104 (20.9)
Clinician-negative FM diagnosis	60 (49.6)	333 (88.6)	393 (79.1)
Total	121 (24.3)	376 (75.7)	497 (100)
Correct 79.2% (κ = 0.41 [0.20–0.40])†			
RA patients			
Clinician-positive FM diagnosis	5 (38.5)	6 (8.0)	11 (12.6)
Clinician-negative FM diagnosis	8 (61.5)	69 (92.0)	77 (85.5)
Total	13 (14.8)	75 (85.2)	88 (100)
Correct 84.1% (κ = 0.32 [0.00–0.20])†			
SLE patients			
Clinician-positive FM diagnosis	4 (21.1)	9 (13.4)	13 (15.2)
Clinician-negative FM diagnosis	15 (78.9)	58 (86.6)	73 (84.9)
Total	19 (22.1)	67 (77.9)	86 (100)
Correct 72.9% (κ = 0.08 [0.20–0.40])†			
OA patients			
Clinician-positive FM diagnosis	28 (63.6)	14 (13.0)	42 (27.6)
Clinician-negative FM diagnosis	16 (36.4)	94 (87.0)	110 (72.4)
Total	44 (28.9)	108 (71.1)	152 (100)
Correct 80.3% (κ = 0.51 [0.20–0.40])†			

* Except where indicated otherwise, values are the number (%). FM = fibromyalgia; RA = rheumatoid arthritis; SLE = systemic lupus erythematosus; OA = osteoarthritis.
† Values in brackets are probabilistic benchmark intervals, where 0.20–0.40 = fair agreement and 0.00–0.20 = slight agreement.

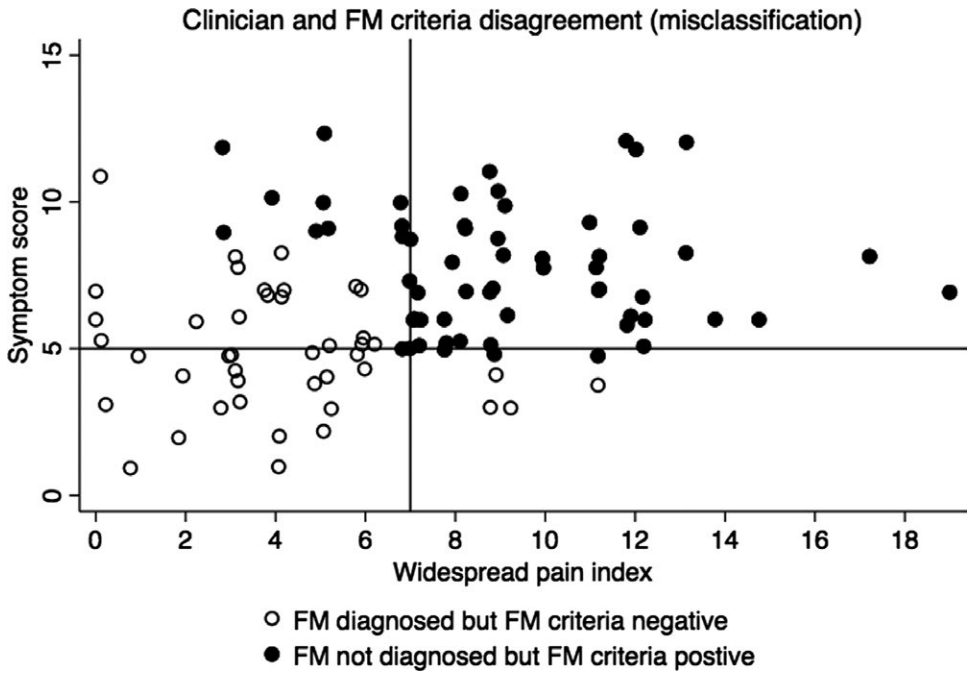


Figure 1. Disagreement between clinician-based diagnosis of fibromyalgia (FM) and diagnosis by FM criteria at the individual patient level. A positive criteria-based diagnosis requires a widespread pain index (WPI) ≥7 and symptom severity scale (SSS) score ≥5 or a WPI of 3–6 and SSS score ≥9.

who were criteria-negative, 14 (13.0%) were called positive by clinicians. In general, the results of the kappa analyses indicated agreement ranging from poor to fair among clinicians' diagnosis and fibromyalgia criteria.

Figure 1 offers insight into how this misclassification occurred. In general, to have fibromyalgia according to the 2010/2011 criteria, a patient must have a WPI score ≥7 and an SSS score ≥5 *or* a WPI score 3–6 and an SSS score ≥9. The black circles in the figure show patients who clearly satisfied criteria but did not receive the diagnosis. The black circles in the left upper quadrant are those patients who satisfied criteria by meeting the WPI of 3–6 and SSS score of ≥9 criteria. The clear circles indicate patients who did not meet criteria but received a fibromyalgia diagnosis from clinic clinicians. Some of these patients had very limited pain and fibromyalgia symptom activity.

Figure 2 provides 3 overlapping distribution curves. The vertical dividing line at PSD = 12 is key, because a patient cannot satisfy fibromyalgia criteria with a PSD score <12. The dashed line on the right half of the graph represents the distribution of scores for criteria-positive patients. The dotted line on the left half is the distribution of mean PSD scores for patients not receiving a clinician diagnosis of fibromyalgia and not satisfying fibromyalgia criteria. The solid line shows the effect of clinician diagnoses that include misclassification. It can be seen that many patients diagnosed with fibromyalgia do not have sufficiently high PSD scores for criteria diagnosis. The reduced height of the clinician curve (solid line) to the right of the dividing line, compared with the criteria-positive line, indicates that many patients who are criteria-positive are not being diagnosed with fibromyalgia, and this observation is supported by the actual data shown in Figure 1.

In these data, we identified 4 groups according to fibromyalgia diagnostic status: 1) criteria-negative/clinical-negative, 2) criteria-negative/clinical-positive, 3) criteria-positive/clinical-negative, and 4) criteria-positive/clinical-positive. In the order specified, these groups contained patients with increasingly more severe symptoms (Table 1). From the perspective of fibromyalgia symptoms, the PSD scores increased as follows: 6.0, 8.9, 16.9, and 20.0 in groups 1, 2, 3, and 4, respectively. The groups are identified by individual density curves in Figure 3, and their mean scores are shown on the x-axis. Importantly, the criteria-negative/clinical-positive overdiagnosis group (second curve from the left) indicates the extent of misdiagnosis in patients with less severe symptoms.

The results shown in Table 1 suggested that factors external to criteria were involved in diagnosis. Therefore, we compared all patients who were diagnosed with fibromyalgia by clinicians with those who met criteria for fibromyalgia. Among the clinically diagnosed patients, 97.7% were women compared with 83.3% of criteria-positive patients ($P = 0.007$), while overall, 82.5% of patients were women. We also found that clinically diagnosed patients had greater SSS scores (7.8 versus 4.9; $P < 0.001$), lower WPI scores (4.0 versus 9.2; $P = 0.000$), and lower PSD scores (8.9 versus 16.9; $P = 0.000$). Thus, clinicians gave greater weight in making a diagnosis to being a women and having increased symptoms and were willing to diagnose patients with

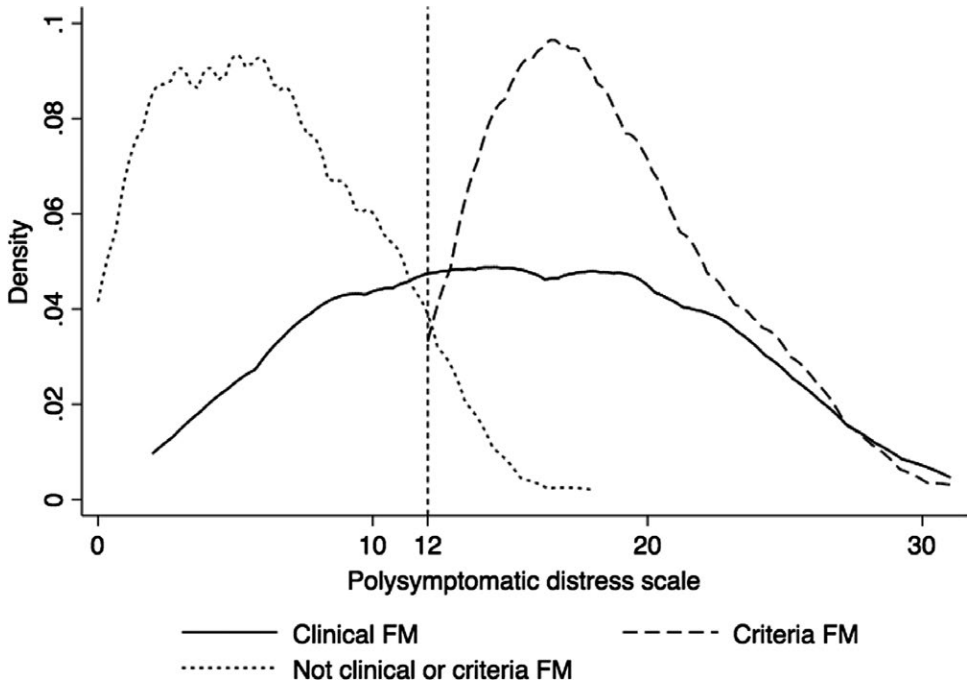


Figure 2. Kernel density estimation for polysymptomatic distress (PSD) in patients with clinician-diagnosed fibromyalgia (FM), those with criteria-diagnosed FM, and patients who did not have a diagnosis of FM.

lower WPI and PSD scores. In addition, a number of significant contextual factors were not formally assessed in this study but may have been important to clinicians and influenced diagnosis, including a history or family history of fibromyalgia, and comorbid conditions such as irritable bowel syndrome and restless leg syndrome.

DISCUSSION

The data of this study provide information on the characteristics of patients diagnosed with fibromyalgia. In a 2016 study of the US population using the National Health Interview Survey, Walitt et al showed that the self-reported diagnosis of fibromyalgia is likely

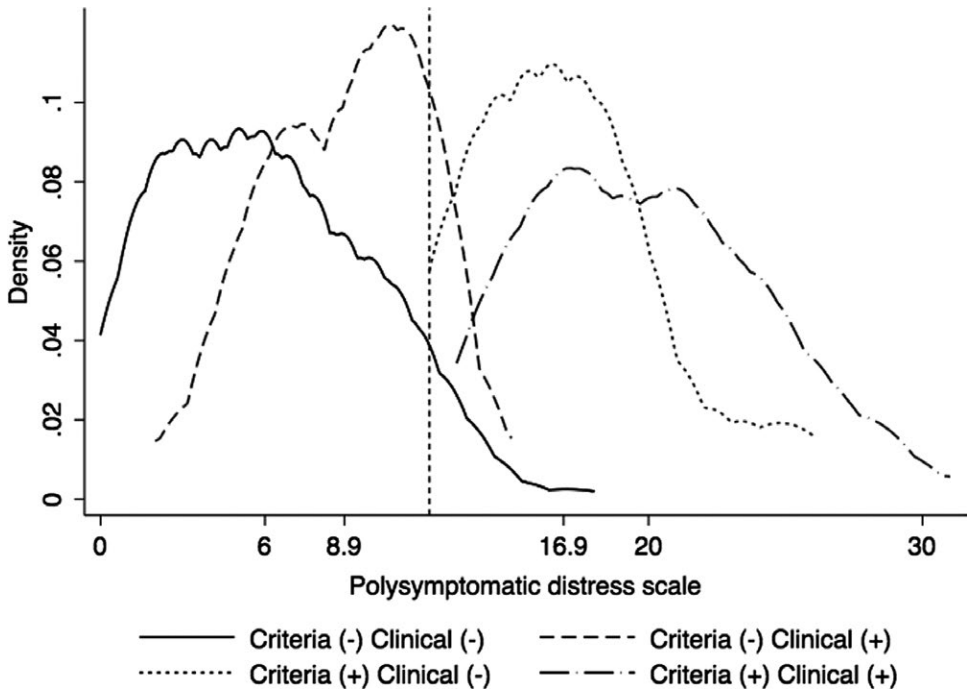


Figure 3. Kernel density estimations for polysymptomatic distress (PSD) in the 4 groups shown in Table 2.

to be wrong. Using surrogate criteria, they found that 73.5% of the 1.8% of respondents who reported doctor-diagnosed fibromyalgia did not satisfy criteria for fibromyalgia because their symptoms were not sufficiently severe (18). With their criteria, Walitt et al also identified fibromyalgia in 1.8% (95% confidence interval 1.42–2.07) of the population. However, 73% of identified patients self-reported a physician's diagnosis other than fibromyalgia. Our data support the observation of Walitt et al of misclassification. In addition, a recent study involving 42,394 female nurses evaluated diagnosis in chronic fatigue syndrome, a condition with similarities and overlap with fibromyalgia, and found exactly what we and Walitt et al found—an excess of underdiagnosis and overdiagnosis (35).

Other evidence of misdiagnosis comes from recent epidemiology studies that have found that women constitute 60% of cases of fibromyalgia compared with clinical studies in which the proportion of women can be ≥90% (36). This suggests that a bias exists toward overidentification of fibromyalgia in women or in those with psychological distress, and that such a bias may distort case identification. We found evidence of the bias in the current study. We have also suggested that increased diagnoses of fibromyalgia in patients who do not satisfy criteria may be a response to extensive direct-to-patient pharmaceutical advertising (21), because the prevalence almost doubled in the US military "following advent of federally approved drugs for FMS in concert with pharmaceutical industry marketing (12)."

Although the questions of diagnosis and diagnostic accuracy seem simple, they are in fact complex. Considering the 2010 and the self-reported 2011 criteria, fibromyalgia is diagnosed by criteria that are based on severity as measured by the number of pain locations—the WPI score—and the SSS score, as shown in Figure 1. When added together, they form the 0–31 PSD scale (Figures 2 and 3). Patients with PSD scores <12 cannot by definition satisfy fibromyalgia criteria. Depending on the characteristics of the study sample, 70% to >95% of those with PSD scores ≥12 will satisfy criteria (5,37).

Symptom severity, however, is not fixed. A patient who improves by decreasing severity measured by the PSD score to 11 no longer satisfies fibromyalgia criteria. A patient who decreases the PSD score to 11 and stays there is clinically quite different from one who decreases to 2 and stays there. However, both fail to satisfy fibromyalgia criteria. There is no general consensus as to how to characterize fibromyalgia in individuals such as these. Some clinicians feel that fibromyalgia is a trait disorder—once fibromyalgia always fibromyalgia (38). Criteria-based diagnosis treats fibromyalgia as a state disorder—present or absent. Clinicians whose patients with fibromyalgia decrease their PSD scores to 9 during treatment may rightly be considered to have improved, while patients who appear in the clinic for the first time with a score of 9 will likely be considered not to have fibromyalgia. In the current study, we define fibromyalgia presence and absence based on current fibromyalgia criteria status. While this standard may be too strict (or not), it provides a measurable yardstick.

There has been no prospective, unbiased study of fibromyalgia diagnostic accuracy in the community of primary care physicians when fibromyalgia criteria are used as the diagnostic standard. The current study is the first to approach this issue, but in a university rheumatology specialty clinic. We found that physician-diagnosed fibromyalgia had only a fair level of agreement with criteria-based diagnosis ($\kappa = 0.44$ [probabilistic benchmark interval 0.20–0.40]). A total of 41.3% of physician-diagnosed fibromyalgia did not meet the 2011 criteria, and 15.3% of patients who did not meet criteria were diagnosed with fibromyalgia by physicians. Patients with physician-diagnosed fibromyalgia had less severe symptoms than those with criteria-diagnosed fibromyalgia, even when a physician diagnosis and criteria agreed. The physicians in our study were experts in rheumatology—faculty and fellows who had training in fibromyalgia and fibromyalgia criteria. However, not only did their diagnoses not agree with criteria, but examiners subsequently indicated that they did not pay attention to criteria in assessing patients. We also found that clinicians were influenced by patients' sex and symptom severity more than by pain extent. Figure 2 emphasizes how wide is the diagnostic capture when no criteria are used. As noted above, however, another contributing reason for clinician–criteria difference might be improvement in symptom severity from when fibromyalgia was first diagnosed by the clinician, measurements that we do not have.

It is likely that diagnosis in the community by general physicians, for whom the criteria are not as well known, will be even more inaccurate than what we observed in this university clinic. Primary care physicians find fibromyalgia difficult to diagnose. Kumbhare et al reported that "Physician knowledge overall was not comprehensive and was very poor for the 1990 and 2010 criteria." Fifty-one percent of physicians used a set of criteria in their practice, and 49% used their clinical acumen (19). Even rheumatology specialists have problems. In one survey of 206 rheumatologists, 31% found fibromyalgia difficult to diagnose, as did 61% of 809 primary care physicians (20).

The observation of Walitt et al of overdiagnosis and their data indicating that demographic and social factors influence misdiagnosed fibromyalgia are essential to understanding diagnostic problems (17,18). Neurobiologic mechanisms may play an important role in fibromyalgia pain; however, the diagnosis of fibromyalgia is subject to social and cultural factors, and a diagnosis or non-diagnosis can influence social status, the availability of medical care, and access to private and government disability payments.

We have raised the issue previously of whether the criteria-negative "pain and distress" type of fibromyalgia patient identified by the Walitt group in the general population should be considered de facto as part of the fibromyalgia definition (21). Perhaps. Maybe the disorder that "has no binding definition … and no way of objectively testing for it" (40) cannot do better in the community. Perhaps a clinician familiar with the patient's total history and

a thorough understanding of the psychosocial aspects of the patient's presentation (through doctor–patient interaction) may have a different threshold for diagnosing.

It is not clear why the physicians in this study who misclassified fibromyalgia and missed patients with the diagnosis did not do better. They may have simply misdiagnosed the condition or not accepted the use of the formal diagnostic system as necessary or clinically useful for identifying fibromyalgia in routine care. Other important reasons for discordance include not accepting the fibromyalgia concept, not considering fibromyalgia, or not needing a fibromyalgia diagnosis for patient care. The 2016 criteria recommendations use the phrase "*may* be diagnosed," indicating that it is the clinicians' judgment that should guide the need for the diagnosis. We would add that clinicians *ought* to consider fibromyalgia symptoms, if not a diagnosis, in patients with rheumatic pain disorders such as RA.

One solution to the application of a dichotomous diagnosis of fibromyalgia to what is really a dimensional disorder (39) is the use of the PSD scale, as shown in Figures 1 and 2. It avoids the issue of crossing back and forth across the diagnostic border. More importantly, it allows clinicians and researchers to measure the degree of PSD—what has been called "fibromyalgianess" (40). With it one can follow the course of fibromyalgia symptoms without the need to categorize patients into sick and well categories.

A limitation to this study is that it was performed in a rheumatology clinic. It is likely that the discordance we observed may be much more severe in the community of primary care (17–20,35). However, having a rheumatology "standard" is likely to be useful in the ongoing assessment of this problem. A second limitation might be that rheumatology examiners may have been motivated to be on the alert for fibromyalgia patients because of knowledge of the study. Such a bias, if present, could have influenced the results. The use of ICD codes raises additional issues, although they are now widely used in fibromyalgia studies (22–26). Selection of ICD categories can be arbitrary. Although selection of ICD categories is accepted in many studies, it seems clear from our data that disorders such fibromyalgia may be inaccurately categorized in study reports (22–26).

In summary, we found that expert physicians in a university clinic often misdiagnose fibromyalgia when compared with published criteria, probably for a variety of reasons. If, as we believe, these data offer support for the observation of Walitt et al of substantial misdiagnosis in the general population (17,18), it is likely that misdiagnosis is a public health problem and one that can lead to overdiagnosis and overtreatment (1,41–45), as well as to inappropriate treatment of individuals not recognized to have fibromyalgia symptoms (46).

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Wolfe 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. Wolfe, Gibson, Pincus.
Acquisition of data. Wolfe, Schmukler, Jamal, Castrejon, Pincus.
Analysis and interpretation of data. Wolfe, Schmukler, Jamal, Castrejon, Gibson, Srinivasan, Häuser, Pincus.

REFERENCES

1. Barker KK. Listening to Lyrica: contested illnesses and pharmaceutical determinism. Soc Sci Med 2011;73:833–42.
2. Dew K, Scott A, Kirkman A. Medicalization and contested illnesses. In: Social, political and cultural dimensions of health. New York: Springer; 2016. p. 95–110.
3. Armentor JL. Living with a contested, stigmatized illness: experiences of managing relationships among women with fibromyalgia. Qual Health Res 2017;27:462–73.
4. Briones-Vozmediano E, Öhman A, Goicolea I, Vives-Cases C. "The complaining women": health professionals' perceptions on patients with fibromyalgia in Spain. Disabil Rehabil 2018:1679–86.
5. Ablin JN, Wolfe F. A comparative evaluation of the 2011 and 2016 criteria for fibromyalgia. J Rheumatol 2017;44:1271–6.
6. Hench PK. Nonarticular rheumatism. Arthritis Rheum 1976;19:1081–8.
7. Yunus M, Masi AT, Calabro JJ, Miller KA, Feigenbaum SL. Primary fibromyalgia (fibrositis): clinical study of 50 patients with matched normal controls. Semin Arthritis Rheum 1981;11:151–71.
8. Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL, et al. The American College of Rheumatology 1990 criteria for the classification of fibromyalgia: report of the Multicenter Criteria Committee. Arthritis Rheum 1990;33:160–72.
9. Wolfe F, Clauw D, Fitzcharles MA, Goldenberg D, Katz RS, Mease P, et al. The American College of Rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity. Arthritis Care Res 2010;62:600–10.
10. Wolfe F, Clauw D, Fitzcharles MA, Goldenberg D, Häuser W, Katz RS, et al. Fibromyalgia criteria and severity scales for clinical and epidemiological studies: a modification of the ACR preliminary diagnostic criteria for fibromyalgia. J Rheumatol 2011;38:1113–22.
11. Wolfe F, Clauw DJ, Fitzcharles MA, Goldenberg DL, Häuser W, Katz RL, et al. 2016 revisions to the 2010/2011 fibromyalgia diagnostic criteria. Semin Arthritis Rheum 2016;46:319–29.
12. Jeffery DD, Bulathsinhala L, Kroc M, Dorris J. Prevalence, health care utilization, and costs of fibromyalgia, irritable bowel, and chronic fatigue syndromes in the military health system, 2006-2010. Mil Med 2014;179:1021–9.
13. Moynihan R, Doran E, Henry D. Disease mongering is now part of the global health debate. PLoS Med 2008;5:e106.
14. Moynihan RN, Cooke GP, Doust JA, Bero L, Hill S, Glasziou PP. Expanding disease definitions in guidelines and expert panel ties to industry: a cross-sectional study of common conditions in the United States. PLoS Med 2013;10:e1001500.
15. Reed M, Herrmann M. The difficulties in developing and implementing fibromyalgia guidelines: new insights into fibromyalgia. InTech.com: InTech; 2012.
16. Hadler NM, Greenhalgh S. Labeling woefulness: the social construction of fibromyalgia [abstract]. Spine (Phila PA 1976) 2005; 30:1–4.
17. Walitt B, Nahin RL, Katz RS, Bergman MJ, Wolfe F. The prevalence and characteristics of fibromyalgia in the 2012 National Health Interview Survey. PLoS One 2015;10:e0138024.
18. Walitt B, Katz RS, Bergman MJ, Wolfe F. Three-quarters of persons in the US population reporting a clinical diagnosis of fibromyalgia do

not satisfy fibromyalgia criteria: the 2012 national health interview survey. PLoS One 2016;11:e0157235.

19. Kumbhare D, Ahmed S, Sander T, Grosman-Rimon L, Srbely J. A survey of physicians' knowledge and adherence to the diagnostic criteria for fibromyalgia. Pain Med 2018;19:1254–64.
20. Perrot S, Choy E, Petersel D, Ginovker A, Kramer E. Survey of physician experiences and perceptions about the diagnosis and treatment of fibromyalgia. BMC Health Serv Res 2012;12:356.
21. Wolfe F. Criteria for fibromyalgia? What is fibromyalgia? Limitations to current concepts of fibromyalgia and fibromyalgia criteria. Clin Exp Rheumatology 2017;35 Suppl 105:3–5.
22. Doss J, Mo H, Carroll RJ, Crofford LJ, Denny JC. Phenome-wide association study of rheumatoid arthritis subgroups identifies association between seronegative disease and fibromyalgia. Arthritis Rheumatol 2017;69:291–300.
23. Gostine M, Davis F, Roberts B, Risko R, Asmus M, Cappelleri JC, et al. Clinical characteristics of fibromyalgia in a chronic pain population. Pain Pract 2018;18:67–8.
24. Collin SM, Bakken IJ, Nazareth I, Crawley E, White PD. Trends in the incidence of chronic fatigue syndrome and fibromyalgia in the UK, 2001-2013: a Clinical Practice Research Datalink study. J R Soc Med 2017;110:231–44.
25. Marschall U, Arnold B, Häuser W. [Treatment and healthcare costs of fibromyalgia syndrome in Germany: analysis of the data of the Barmer health insurance (BEK) from 2008-2009]. Schmerz 2011;25:402–4, 6–10. In German.
26. Sauer K, Kemper C, Glaeske G. Fibromyalgia syndrome: Prevalence, pharmacological and non-pharmacological interventions in outpatient health care. An analysis of statutory health insurance data. Joint Bone Spine 2011;78:80–4.
27. Pincus T, Swearingen C, Wolfe F. Toward a multidimensional Health Assessment Questionnaire (MD-HAQ): assessment of advanced activities of daily living and psychological status in the patient-friendly health assessment questionnaire format. Arthritis Rheum 1999;42:2220–30.
28. Wolfe F. Pain extent and diagnosis: development and validation of the regional pain scale in 12,799 patients with rheumatic disease. J Rheumatol 2003;30:369–78.
29. Stucki G, Liang MH, Stucki S, Bruhlmann P, Michel BA. A self-administered rheumatoid arthritis disease activity index (RADAI) for epidemiologic research. Psychometric properties and correlation with parameters of disease activity. Arthritis Rheum 1995; 38:795–8.
30. Chua J, Pincus T, Castrejon I, Gibson K, Block J, Malfait A. Self-report of painful joints in patients with osteoarthritis: symmetrical involvement in similar numbers and distribution as patients with rheumatoid arthritis. Osteoarthritis Cartilage 2017;25:S221.
31. Stata statistical software: release 15.0. College Station (TX): StataCorp; 2017.
32. Landis J, Koch G. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
33. Klein D. KAPPAETC: Stata module to evaluate interrater agreement. Statistical software components. URL: https://EconPapers.repec.org; 2018.
34. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. Br J Math Stat Psychol 2008;61:29–48.
35. Palacios N, Fitzgerald KC, Komaroff AL, Ascherio A. Incidence of myalgic encephalomyelitis/chronic fatigue syndrome in a large prospective cohort of US nurses. Fatigue: Biomedicine, Health & Behavior 2017:1–8.
36. Wolfe F, Walitt B, Perrot S, Rasker JJ, Häuser W. Fibromyalgia diagnosis and biased assessment: sex, prevalence and bias. PLoS One 2018;13:e0203755.
37. Wolfe F, Brähler E, Hinz A, Häuser W. Fibromyalgia prevalence, somatic symptom reporting, and the dimensionality of polysymptomatic distress: results from a survey of the general population. Arthritis Care Res 2013;65:777–85.
38. Katz RS, Wolfe F, Michaud K. Fibromyalgia diagnosis: a comparison of clinical, survey, and American College of Rheumatology criteria. Arthritis Rheum 2006;54:169–76.
39. Wolfe F, Walitt B. Culture, science and the changing nature of fibromyalgia. Nat Rev Rheumatol 2013;9:751–5.
40. Brummett CM, Clauw DJ. Fibromyalgia: a primer for the anesthesia community. Curr Opin Anaesthesiol 2011;24:532–9.
41. Gilbody S, Wilson P, Watt I. Benefits and harms of direct to consumer advertising: a systematic review. Qual Saf Health Care 2005;14:246–50.
42. Moynihan R, Henry D. The fight against disease mongering: generating knowledge for action. PLoS Med 2006;3:e191.
43. Summerfield D, Veale D. Proposals for massive expansion of psychological therapies would be counterproductive across society. Br J Psychiatry 2008;192:326–30.
44. Tikkinen KA, Leinonen JS, Guyatt GH, Ebrahim S, Järvinen TL. What is a disease? Perspectives of the public, health professionals and legislators. BMJ Open 2012;2:e001632.
45. Doran E, Henry D. Disease mongering: expanding the boundaries of treatable disease. Intern Med J 2008;38:858–61.
46. Choy E, Perrot S, Leon T, Kaplan J, Petersel D, Ginovker A, et al. A patient survey of the impact of fibromyalgia and the journey to diagnosis. BMC Health Serv Res 2010;10:102.

AMERICAN COLLEGE OF RHEUMATOLOGY
EDUCATION • TREATMENT • RESEARCH

Arthritis Care & Research
Vol. 71, No. 3, March 2019, pp 352–361
DOI 10.1002/acr.23602

Neuromuscular Electrical Stimulation Compared to Volitional Exercise for Improving Muscle Function in Rheumatoid Arthritis: A Randomized Pilot Study

Sara R. Piva,[1] Samannaaz S. Khoja,[1] Frederico G. S. Toledo,[1] Mary Chester-Wasko,[2] G. Kelley Fitzgerald,[1] Bret H. Goodpaster,[3] Clair N. Smith,[1] and Anthony Delitto[1]

Objective. The aim of this study was to compare the feasibility and effectiveness of neuromuscular electrical stimulation (NMES) with that of high-intensity volitional resistance training for improving muscle structure and function and physical function in patients with rheumatoid arthritis (RA). We also compared pre-intervention and post-intervention values of myocyte characteristics.

Methods. In this 2-group, single-blind, randomized pilot study, adult patients with RA were assigned to 36 sessions of NMES (n = 31 patients) or volitional training (n = 28 patients) over 16 weeks. Outcome measures included muscle structure and function (quadriceps muscle area, density, and strength), physical function (performance-based and patient-reported), feasibility (increased pain, increased disease activity, attrition, and adherence), and myocyte characteristics (area, proportion of type I or II muscle fibers, and intramyocellular lipid content). Analysis of covariance was used to compare groups.

Results. The intervention intensity in the NMES group was less than half that in the volitional exercise group (31% versus 77% of maximum effort). Both groups experienced significant improvements in muscle structure and function ($P < 0.001$ to 0.019). Improvements in muscle characteristics and physical function were not different between groups. Exercise did not result in serious adverse events or increases in pain and disease activity. Attrition was 29% in the NMES group and 7% in the volitional exercise group.

Conclusion. Both NMES and high-intensity volitional resistance training can be used as effective approaches to improving muscle structure and function in patients with RA. NMES may be a viable alternative for improving muscle function in patients in whom high-intensity resistance exercise may not be tolerated or is contraindicated, but attrition must be considered when using this approach.

INTRODUCTION

Nearly two-thirds of patients with rheumatoid arthritis (RA) experience some level of muscle atrophy (1). Reversing muscle atrophy in these patients is important not only to improve performance during daily tasks but also to increase the capacity of muscles to store protein. Muscle protein plays a crucial role in enabling adaptation to metabolic stresses and fighting infections (1–4). Serious infections are a major concern in patients with RA and contribute to overall mortality (5). Therefore, promoting muscle hypertrophy may ultimately contribute to improved health in RA (4).

Volitional resistance exercise is typically the first countermeasure against muscle atrophy. Studies on resistance exercise in patients with RA have demonstrated significant improvements in muscle strength and physical function without parallel increases in muscle size (6–12), suggesting that this population of patients might be resistant to the anabolic effects of exercise (10). The few studies that demonstrated muscle hypertrophy in patients with RA used resistance training at high intensity, representing ~80% of the maximum load an individual can tolerate (13,14). In patients with RA, pain and disability due to disease activity may reduce tolerance

ClinicalTrials.gov Identifier: NCT00924625.

Supported by the NIH (grant 1-K01-HD058035 from the National Center for Medical Rehabilitation Research and grant KL2-RR-024154-02 from NIH Roadmap for Medical Research), and an American College of Rheumatology Research Education Foundation New Investigator Award.

[1]Sara R. Piva, PT, PhD, Samannaaz S. Khoja, PT, PhD, Frederico G. S. Toledo, MD, G. Kelley Fitzgerald, PT, PhD, Clair N. Smith, MS, Anthony Delitto, PT, PhD: University of Pittsburgh, Pittsburgh, Pennsylvania; [2]Mary Chester-Wasko, MD, MSc: Allegheny Health Network–Lupus Center of Excellence, Pittsburgh, Pennsylvania; [3]Bret H. Goodpaster, PhD: Translational Research Institute for Metabolism and Diabetes, Orlando, Florida.

Address correspondence to Sara R. Piva, PT, PhD, Bridgeside Point 1, 100 Technology Drive, Suite 210, Pittsburgh, PA 15219-3130. E-mail: spiva@pitt.edu.

Submitted for publication August 8, 2017; accepted in revised form May 15, 2018.

SIGNIFICANCE & INNOVATIONS

- Muscle atrophy is prevalent and associated with morbidity in patients with rheumatoid arthritis (RA). While high-intensity volitional exercise is typically the first countermeasure against muscle atrophy, patients with RA may be more resistant to the anabolic effects of exercise. Neuromuscular electrical stimulation (NMES) appears to activate all muscle fibers at relatively lower contraction intensities and could be an alternative to high-intensity volitional resistance training.
- This is the first randomized study to compare NMES with high-intensity resistance training in patients with RA. The results showed that both NMES and high-intensity volitional resistance training can be used as effective approaches to improve muscle structure and function in patients with RA.
- NMES improves muscle structure and function at lower training intensity compared to high-intensity resistance training and may be a viable alternative for patients in whom high-intensity resistance exercise may not be tolerated or is contraindicated.

to high-intensity resistance exercises, and use of alternative approaches to promote muscle hypertrophy should be explored.

Neuromuscular electrical stimulation (NMES) is one such alternative approach. NMES incorporates the use of electrical current to activate skeletal muscles and produce contractions (15). NMES recruits motor units in a nonselective pattern of activation, in which both slow-twitch and fast-twitch fibers are activated simultaneously, unlike volitional contractions, in which slow-twitch fibers are activated first and fast-twitch fibers are recruited only with near-maximal effort (16–19). Thus, a clinical advantage of NMES over volitional exercise is the potential to activate all muscle fibers at relatively lower contraction intensities. Use of this approach may be particularly effective for reversing muscle atrophy, which seems to predominantly affect type II fibers, in patients with RA (20,21).

NMES has been shown to promote muscle hypertrophy in a variety of patient populations (22–25). Our preliminary work indicated that NMES is a promising intervention to enhance muscle mass and strength in RA patients. However, NMES was previously tested in a small series of patients, and randomized studies are needed to assess the effectiveness of NMES in patients with RA (26). To that end, the primary objective of this study was to compare the effects of NMES and volitional resistance training on muscle structure and function (hypertrophy, strength, and density) and physical function (performance-based and patient-reported outcomes) in patients with RA. The secondary objective was to determine the safety and feasibility of using NMES. We also characterized and compared the pre-intervention and post-intervention myocyte area, the proportion of type I or II muscle fibers, and intramyocellular lipid content.

PATIENTS AND METHODS

This was a 2-group, parallel-design, single-blind randomized pilot study implemented from December 2009 to December 2013 at the Department of Physical Therapy, University of Pittsburgh. Participants were enrolled if they were older than age 21 years, had rheumatologist-diagnosed RA for >5 years (according to the American College of Rheumatology criteria (27), were fluent in English, and were able to ambulate independently without an assistive device. Exclusion criteria were cardiovascular disease, uncontrolled hypertension, neurologic or muscular conditions, history of quadriceps/patellar tendon rupture or adverse reaction to electrical stimulation, participation in resistance training ≥2 times/week, change in RA medication in the previous month, use of cholesterol-lowering medication, malignancy, and pregnancy. Individuals with <70 degrees of knee flexion were excluded, to enable proper positioning on the isokinetic dynamometer.

Patients were recruited through letters sent by their rheumatologists, the University of Pittsburgh Medical Center Arthritis Network Registry, and public announcements. The University of Pittsburgh Institutional Review Board approved the study protocol (PRO08090141). All participants provided written informed consent.

Randomization and interventions. A statistician (CNS) generated the randomization sequence, and the research coordinator (SSK) (who was not involved in testing or treatment) randomized the participants through a web-based computer system after the baseline visit. Participants were allocated 1:1 to receive NMES or volitional exercise training. The testers (those who administered the questionnaires and physical examinations) were blinded to group assignment. Although participants could not be blinded to treatment assignment, they were instructed not to discuss treatment with the testers. Physical therapists blinded to participants' performance on outcome measures delivered the interventions.

The groups participated in 2–3 exercise sessions/week, for a total of 36 exercise sessions over 16 weeks. Both legs were targeted for training. The interventions adhered to the best evidence-based practice, meaning that the NMES was administered at maximum tolerance, while volitional exercise training was applied at high intensity (15,28,29). Each exercise session started with a 5-minute warm-up period of stationary cycling without resistance and 1–2 bouts of quadriceps stretching for 30 seconds.

NMES training was administered using an Empi Infinity Plus portable NMES unit. Stimulus parameters were based on previous research and included a pulse rate of 75 pulses/second and pulse duration of 450 microseconds (15,26). The stimulus

on/off time was 12 seconds on (3 seconds of ramp-up, 6 seconds of full contraction, 3 seconds of ramp-down) and 48 seconds off to minimize muscle fatigue (1-minute cycle). The skin areas on which electrodes would be placed were rubbed with alcohol. Two 6.9 cm × 12.7 cm self-adhesive electrodes (Dura-Stick) were placed on the skin, one proximally, over the vastus lateralis, and one distally, over the vastus medialis. Participants were instructed not to voluntarily contract their muscles and to let the stimulator do the work. A total of 15 electrically elicited contractions were applied. The amperage of the NMES unit was gradually increased as tolerated during each session. To help participants tolerate the NMES, we used coping strategies such as ignoring pain and distraction (e.g., engaging in conversation or asking the patient to visualize something pleasant) (30). The intensity of NMES was calculated in percentage as the torque generated during the NMES contraction divided by the torque generated during a maximum voluntary isometric contraction (MVIC). The MVIC torque output (expressed in Nm) of the quadriceps muscles was assessed using an isokinetic dynamometer (Biodex System 3 Pro) as described previously (26). NMES torque output during training was determined with participants seated on the same dynamometer chair that was used to test MVIC. LabVIEW software was used to determine the NMES intensity, which was calculated for each exercise session and averaged across all sessions.

Volitional exercise was performed using exercise equipment and was based on best evidence (9–11). The exercises targeted mainly the quadriceps muscles and included leg extension and leg press exercises on the respective machines. Prior to training, the tester assessed the participants' 1-repetition maximum (1-RM) for leg extension and leg press. Volitional training was performed in a pain-free range and consisted of 15 repetitions of each exercise with 40% load, followed by 3 sets of 8 repetitions with a load of 80% of 1-RM. Each contraction was timed for a 2-second concentric load, a 2-second eccentric load, and a 3-second interval between contractions. A 2-minute rest was provided between each set. The intensity of volitional training was calculated in percentage as the force produced during each repetition of the exercises divided by the force produced during 1-RM testing. The intensity was calculated for each exercise session and averaged across all sessions for each participant.

To maintain exercise intensity in the volitional exercise group at 80% of 1-RM and concurrently consider the expected increase in muscle strength during physical training, participants were asked to repeat the 1-RM test every 4 visits, and the exercise dose was re-calculated based on the new 1-RM value. In the NMES group, MVIC testing was also performed every 4 visits. Although exercise was prescribed at maximal tolerance in the NMES group irrespective of maximal effort, the MVIC test was repeated during the same time interval as that in the volitional exercise group in order to balance both groups for the number of tests of maximum strength.

The NMES and volitional exercise training programs were designed to have comparable time (attention) demands. NMES sessions (without MVIC testing) took ~45 minutes (15 minutes of NMES on each side plus 15 minutes of set-up time for the dynamometer and NMES devices). The volitional training sessions (without 1-RM testing) were also 45 minutes in length (11 minutes for the leg extension machine for each leg, 6 minutes in the leg press for each leg, and 11 minute set-up time).

Outcome measures were assessed at baseline and 4 months after randomization and included quadriceps muscle structure and function, physical function, feasibility, and myocyte characteristics. Primary outcomes included muscle area and performance-based physical function. The quadriceps muscle structure was assessed by muscle area and density, using mid-thigh computerized tomography (CT), as previously described (31).

Briefly, a 10-mm–thick axial image of the mid-thigh region was obtained at the femoral midpoint. Scanning parameters were 120 kVp and 200–250 mA, and subjects were positioned the supine neutral position. CT images were processed using Slice-O-Matic software, which differentiates between muscle, fat, and bone tissue based on their density measured in Hounsfield units (HU), with water as the reference at 0 HU. Muscle density ranges from 0 HU to 100 HU (higher values correspond to lower lipid content and better muscle quality,) while adipose tissue density ranges from −190 to −30 HU. The average muscle density (in HU) and muscle cross-sectional area (cm^2) were obtained. This method is reliable and valid (32). Quadriceps muscle function was assessed as muscle strength, using the MVIC test (33). Briefly, participants sat on an isokinetic dynamometer (Biodex) with the knee at 70 degrees of flexion and performed 5 trials of isometric knee extension. The 3 trials with the highest torque were averaged.

Physical function was measured using a battery of performance-based tests and patient-reported outcomes. Performance-based tests included 1) self-selected gait speed performed in a 4-meter pathway; 2) timed chair stand, which involved timing 5 repetitions of rising to a full upright position and sitting back down in a chair; 3) single-leg stance, which entails timing participants balancing on 1 leg while keeping their hands on the waist (average of 3 trials per leg); and 4) stair climbing, during which subjects are timed while climbing up 11 steps using the handrail on the preferred side. These tests are reliable and responsive to exercise interventions (33–37). Patient-reported outcomes included the Lower Extremity Functional Scale (LEFS) and the Health Assessment Questionnaire Disability Index (HAQ DI). The LEFS is a valid questionnaire containing questions regarding an individual's performance in 20 general daily activities. The LEFS score ranges from 0 to 80, with higher scores representing better function (38). The HAQ DI is valid and widely used in RA. The HAQ DI is used to assess 20 activities of daily living, with scores ranging from 0 (functional disability) to 3 (severe functional disability) (39).

Measures of feasibility were leg pain (on a 0–10 numeric rating scale), disease activity (assessed by the Disease Activity Score

in 28 joints [DAS28], and the erythrocyte sedimentation rate) (40), attrition (the number of participants who completed the 4-month assessment divided by the number of participants randomized to each group), adherence (the number of intervention sessions attended divided by the 36 sessions prescribed), and adverse events. During the 4-month assessment, participants were asked about the occurrence of adverse events during participation in the study. The interventionists also documented discomfort or change in health status during each intervention session.

In a subset of patients who agreed to undergo quadriceps muscle biopsy, the following outcome measures were assessed for descriptive purposes: 1) myocyte area (for all muscle fibers, for high oxidative type I fibers, and for low oxidative type II fibers); 2) intramyocellular lipid (IMCL) content (for all muscle fibers and for fibers types I and II); and 3) proportion of type I and type II muscle fibers. A description of biopsy and histochemical methods is shown in Appendix 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23602/abstract).

Sample size and statistical analysis. We predetermined that with a sample size of 27 subjects per group, the study would have 80% power (2-sided $\alpha = 0.05$) to detect a 2.1 cm^2 between groups in muscle area at 4 months. We used the SD for muscle area (2.74 cm^2 from a previous publication) (26). When we accounted for 10% attrition, the target recruitment size was 30 subjects per group.

We planned a priori to adjust the baseline characteristics of the participants in the analysis if they were considered different according to clinical relevance rather than statistical significance. Within-group pre-intervention to post-intervention changes were assessed by paired *t*-tests. To obtain adjusted comparison of continuous outcomes between treatment arms, we fitted an analysis of covariance model using change from baseline to follow-up for each outcome as the response variable and treatment arm as the main factor of interest, and adjusted for potential baseline characteristics. All tests were 2-sided with an alpha level of 0.05 and were not corrected for multiple comparisons to minimize Type 2 error. Primary analysis used intent-to-treat and per-protocol approaches. For the descriptive aim, box plots were used to characterize pre-intervention and post-intervention values of myocyte area, proportion of muscle fibers according to fiber type, and IMCL content by intervention. IBM SPSS version 23 was used for all plots and calculations.

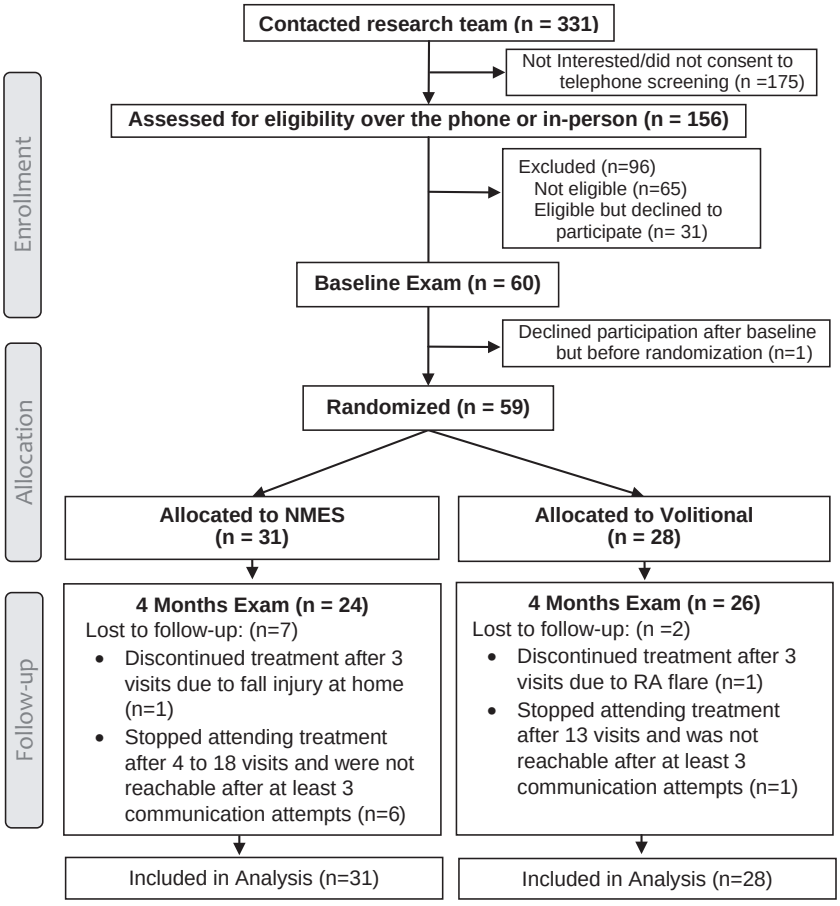


Figure 1. Study flow diagram. NMES = neuromuscular electrical stimulation; RA = rheumatoid arthritis.

RESULTS

Among the 332 individuals who contacted the research group, 156 expressed interest in participating and were screened for eligibility over the phone (Figure 1). Among those subjects, 60 were eligible for participation and were invited to undergo baseline testing. Because one of these subjects subsequently declined participation, 59 subjects were randomized (31 to the NMES group and 28 to volitional exercise group). In the NMES and volitional exercise groups, 24 and 26 participants, respectively, completed the study. The reasons for withdrawal are shown in Figure 1). The baseline characteristics showed clinically important differences between groups for age and body weight (Table 1). Clinically important differences were also observed for some outcome measures at baseline (Table 2). Thus, between-group comparisons were adjusted for baseline values for outcomes, age, and weight.

The within-group changes (from pre-intervention to post-intervention) for muscle structure and function demonstrated statistically significant improvements in all outcome measures in both groups ($P < 0.001$ to 0.019) (Table 2). The quadriceps muscle area increased 3.60 cm^2 on average in both legs (right leg, 3.9 cm^2; left leg, 3.3 cm^2) in the NMES group and 2.65 cm^2 (right leg, 2.8 cm^2; left leg, 2.5 cm^2) in the volitional exercise group. Muscle density increased 1.55 HU (right leg, 1.5 HU; left leg, 1.6 HU left) in the NMES group and 1.65 HU (right leg, 1.8 HU; left leg, 1.5 HU) in the volitional exercise group. Muscle strength increased on average 14 Nm in both the NMES and volitional exercise groups. The between-group differences were not statistically significant.

Table 1. Baseline demographics and clinical characteristics of the participants*

	NMES (n = 31)	Volitional (n = 28)
Age, mean ± SD years	57.2 ± 8.6	61.0 ± 11.0
Female sex	25 (81)	23 (82)
Height, mean ± SD cm	165.2 ± 8.5	164.5 ± 8.1
Weight, mean ± SD kg	88.1 ± 20.4	81.1 ± 22.1
BMI, mean ± SD kg/m^2	32.4 ± 7.1	29.8 ± 7.2
Race		
White	27 (87)	22 (79)
Black	4 (13)	6 (21)
Hispanic	0	1 (4)
Education		
High school or less	9 (29)	6 (21)
Attended college	22 (71)	22 (79)
Employed	12 (39)	13 (46)
Smoker	4 (13)	5 (18)
Disease duration, mean ± SD years	14.2 ± 10.7	17.1 ± 11.0
Disease activity, mean ± SD		
DAS28	4.1 ± 1.3	3.8 ± 1.1
ESR	23.2 ± 16.1	28.4 ± 19.7
Leg pain, mean ± SD (0–10 scale)	2.8 ± 2.7	3.2 ± 2.7

* Except where indicated otherwise, values are the number (%). NMES = neuromuscular electrical stimulation; BMI = body mass index; DAS28 = Disease Activity Score in 28 joints; ESR = erythrocyte sedimentation rate.

The within-group changes for performance-based physical function were significant only for the stair climbing (shortened by 0.3 second; $P = 0.034$) and chair stand (shortened by 1.2 seconds; $P = 0.024$) tests in the volitional exercise group (Table 2). For the patient-reported outcomes, the only significant within-group change was an increase of 4.2 points in the LEFS ($P = 0.001$) in the NMES group. None of the between-group differences in physical function were statistically significant.

Feasibility analysis demonstrated that changes in leg pain (–0.2 in the NMES group and –0.6 in the volitional group; $P = 0.620$) and disease activity according to the DAS28 (–0.6 in the NMES group and –0.3 in the volitional exercise group; $P = 0.537$) were not different between groups. There were no serious adverse events related to study participation. Minor adverse events due to exercise participation included muscle soreness and joint pain, but these resolved within days or weeks without requiring medical attention. Participants attended a similar number of exercise sessions, with a median adherence (Q25, Q75) of 36 (18,37) for NMES and 36 (35,37) for volitional training ($P = 0.100$ by Wilcoxon's test). The attrition rate was 29% in the NMES group and 7% in the volitional exercise group, and the difference was not significant ($\chi^2 = 2.71$, $P = 0.100$). The average intervention intensity across all treatment sessions for the NMES group was 30 ± 19% of MVIC on the right leg and 31 ± 18% of MVIC on the left leg. In the volitional exercise group, exercise intensity was 77 ± 5% of the 1-RM on the right leg and 76 ± 4% of the 1-RM on the left leg.

For the descriptive aim, 17 participants in each group agreed to undergo quadriceps muscle biopsy pre-intervention and post-intervention. The average number of muscle fibers examined pre-intervention was 127 (95% confidence interval [95% CI] 108–145), and the average number examined post-intervention was 118 (95% CI 105–132). Changes in the proportion of muscle fibers from pre-intervention to post-intervention were negligible. Before intervention, the proportions of type I and type II muscle fibers in the NMES group were 63% and 37%, respectively. Post-intervention, the proportions were 60% and 40%, respectively. In the volitional exercise group, the proportions of type I and type II muscle fibers before intervention were 53% and 47%, respectively, and post-intervention the proportions were 52% and 48%, respectively.

The myocyte area in type I and type II fibers combined increased slightly in both groups. Although the interquartile ranges (IQRs) remained essentially unchanged in the NMES group (with a slight upward shift in the volitional exercise group), the increase in median values from pre-intervention to post-intervention did not appear to be different between groups (276 [3,700 minus 3,424] μm^2 in the NMES group and 273 [3,103 minus 2830] μm^2 in the volitional training group) (Figure 2). The changes in the NMES

Table 2. Changes in outcome over time in the NMES and volitional exercise training groups*

	NMES (n = 24)			Volitional exercise (n = 26)			Non-imputed		Imputed†	
	Pre-intervention	Post-intervention	Change	Pre-intervention	Post-intervention	Change	Adjusted between-group difference (mean ± SE)‡	*P*	Adjusted between-group difference (mean ± SE)‡	*P*
Muscle structure and function										
Right quadriceps area, cm^2	49.0 ± 8.2	52.9 ± 8.6	3.9 ± 2.4§	45.2 ± 5.7	48.0 ± 16.3	2.8 ± 3.4§	1.1 ± 0.9	0.209	1.2 ± 0.9	0.168
Left quadriceps area, cm^2	48.6 ± 9.0	51.9 ± 8.6	3.3 ± 3.2§	46.4 ± 15.6	48.9 ± 16.9	2.5 ± 3.2§	0.6 ± 0.9	0.482	0.7 ± 0.9	0.445
Right quadriceps attenuation, HU	45.7 ± 4.6	47.2 ± 5.4	1.5 ± 2.3¶	45.4 ± 5.4	47.2 ± 4.9	1.8 ± 2.0§	−0.2 ± 0.6	0.745	−0.2 ± 0.6	0.770
Left quadriceps attenuation, HU	46.1 ± 4.7	47.7 ± 4.9	1.6 ± 2.4¶	46.4 ± 4.6	47.9 ± 4.9	1.5 ± 3.1#	−0.3 ± 0.7	0.674	0.1 ± 0.8	0.860
Right MVIC, Nm	145 ± 32	160 ± 35	15 ± 20¶	143 ± 65	158 ± 68	15 ± 14§	−0.5 ± 5.2	0.924	−0.3 ± 5.1	0.953
Left MVIC, Nm	144 ± 37	157 ± 39	13 ± 20¶	148 ± 64	161 ± 70	13 ± 19¶	−2.1 ± 5.8	0.719	−1.2 ± 5.9	0.838
Performance-based physical function										
Stair climb, sec	6.6 ± 3.2	6.2 ± 2.4	−0.4 ± 2.1	7.0 ± 2.7	6.7 ± 2.4	−0.3 ± 0.8#	−0.3 ± 0.4	0.470	−0.2 ± 0.4	0.567
Timed chair stand, sec	14.7 ± 8.0	13.3 ± 7.3	−1.4 ± 5.3	12.6 ± 3.3	11.4 ± 3.2	−1.2 ± 2.4#	0.5 ± 1.0	0.620	0.4 ± 1.0	0.657
Gait speed, meters/sec	1.07 ± 0.28	1.07 ± 0.23	0.00 ± 0.19	1.09 ± 0.23	1.10 ± 0.21	0.01 ± 0.19	0.004 ± 0.05	0.942	−0.0001 ± 0.05	0.998
Right single-leg stance, sec	15.3 ± 2.1	16.1 ± 11.4	0.8 ± 4.7	10.6 ± 9.8	12.4 ± 11.0	1.8 ± 4.8	−0.4 ± 1.3	0.770	0.05 ± 1.4	0.972
Left single-leg stance, sec	17.8 ± 2.1	17.7 ± 8.7	−0.1 ± 6.7	11.7 ± 11.3	14.2 ± 9.7	2.5 ± 6.5	−0.2 ± 1.8	0.896	−0.6 ± 1.7	0.752
Patient-reported outcomes of physical function										
LEFS	54.0 ± 18.0	58.2 ± 17.0	4.2 ± 5.5§	58.8 ± 2.2	58.1 ± 13.2	−0.7 ± 10.3	4.2 ± 2.4	0.087	4.3 ± 2.4	0.078
HAQ	0.86 ± 0.59	0.80 ± 0.63	−0.05 ± 0.22	0.74 ± 0.61	0.69 ± 0.64	−0.05 ± 0.40	−0.01 ± 0.1	0.926	−0.03 ± 0.09	0.754

* Except where indicated otherwise, values are the mean ± SD. In the non-imputed analyses (complete cases only), there were 24 observations in the neuromuscular electrical stimulation (NMES) group and 26 observations in the volitional exercise group. In the imputed analysis, the numbers of observations were 31 and 28, respectively, in the NMES and volitional exercise groups. HU = Hounsfeld unit; MVIC = maximum voluntary isometric contraction; LEFS = Lower Extremity Functional Scale; HAQ = Health Assessment Questionnaire.
† The multiple imputation procedure was performed using the Markov chain Monte Carlo expectation maximization method (SAS version 9.4).
‡ Adjusted for age, weight, and baseline values for the outcome.
§ Significant at $\alpha < 0.001$.
¶ Significant at $\alpha < 0.05$.
# Significant at $\alpha < 0.01$.

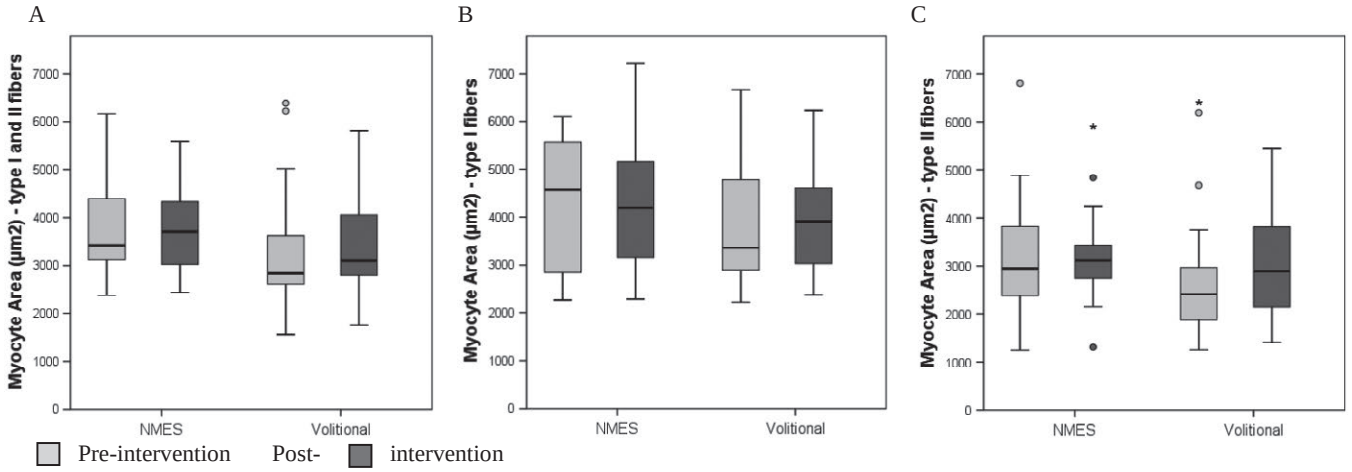


Figure 2. Myocyte areas of type I and type II fibers in the neuromuscular electrical stimulation (NMES) group (n = 17 subjects) and the volitional exercise group (n = 17 subjects), pre-intervention and post-intervention. **A,** Type I and type II fibers combined. **B,** Type I fibers. **C,** Type II fibers. Data are shown as box plots. Each box represents the upper and lower interquartile range (IQR). Lines inside the boxes represent the median. Whiskers represent the highest and lowest values no greater than 1.5 times the IQR. Circles and stars indicate outliers. Circles represent cases with values 1.5–3 times the IQR; stars represent cases with values more than 3 times the IQR.

group appeared to be mainly attributable to increases in the area of type II fibers, whereas in the volitional exercise group, the changes seemed to be attributable to increases the number of both type I and type II fibers (Figure 2).

The IMCL content in combined type I and type II muscle fibers decreased slightly in both groups. The pre-intervention to post-intervention change in IQRs for IMCL content appeared to be more pronounced in the NMES group compared to the volitional exercise group, with changes of –712 (4,740 minus 5,452) arbitrary units (AU) in the NMES group and –31 (4,786 minus 4,817) AU in the volitional exercise group (Figure 3). These changes appeared to be driven by different effects of each intervention on fiber phenotype. In the volitional exercise group, decreases in IMCL content were mainly observed in type I fibers, whereas in the NMES group they were observed in both type I and type II fibers (Figure 3).

DISCUSSION

This is the first study in RA to demonstrate no differences in the effectiveness of NMES and volitional exercise training on quadriceps muscle area, strength, and density. The increases in muscle area were 7% in the NMES group and 6% in the voli-

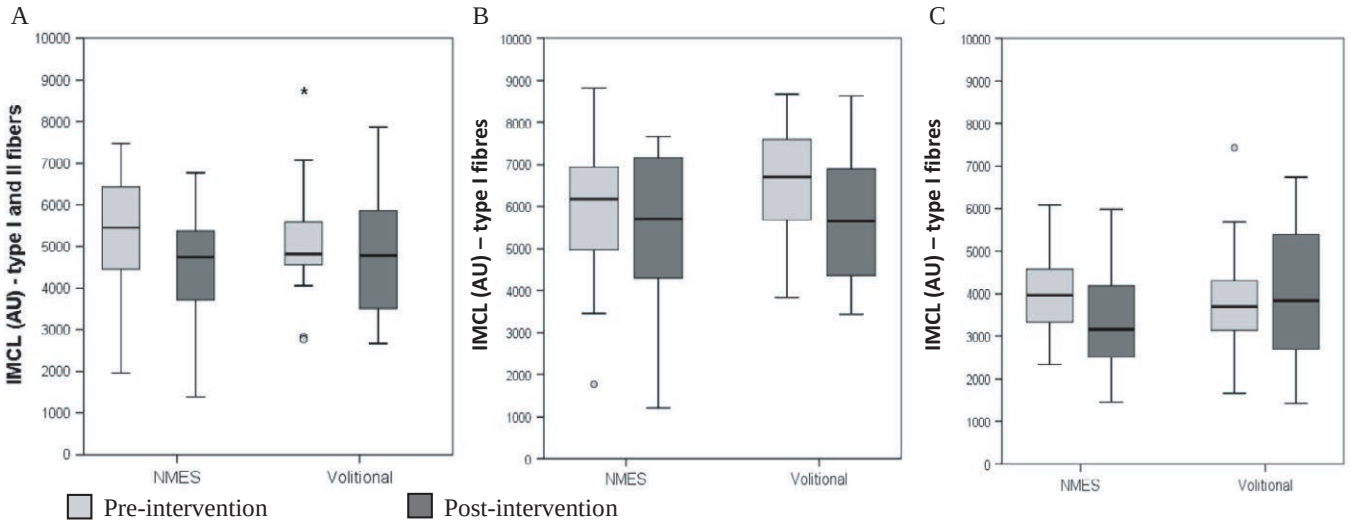


Figure 3. Intramyocellular lipid (IMCL) content in type I and type II muscle fibers in the neuromuscular electrical stimulation (NMES) group (n = 17 subjects) and the volitional exercise group (n = 17 subjects), pre-intervention and post-intervention. **A,** Type I and II fibers combined. **B,** Type I fibers. **C,** Type II fibers. Data are shown as box plots. Each box represents the upper and lower interquartile range (IQR). Lines inside the boxes represent the median. Whiskers represent the highest and lowest values no greater than 1.5 times the IQR. Circles and the star indicate outliers. Circles represent cases with values 1.5–3 times the IQR; the star represents cases with values more than 3 times the IQR. AU = arbitrary unit.

tional exercise group. These increases were similar to improvements observed in previous studies in RA (6–8%) in which comparable high-intensity volitional exercise training was used (13,14,29). The improvements in this study were slightly less than those in studies in older and young healthy individuals (10–11%) that used comparable volitional exercise training (41,42). The changes in muscle quality of ~3% in both groups were slightly less than those in healthy older adults (5%) (43). The lower gains in muscle area and quality in patients with RA compared to healthy subjects were probably due to the hypermetabolic state that renders RA patients more resistant to the anabolic effects of exercise (10,29). The gains in muscle strength in the NMES group and the volitional exercise group were 10% and 9%, respectively; these improvements are similar to those observed in healthy adults (10%) (41,42) but lower than those observed in other studies in RA patients in which high-intensity training was used (25%) (14,29). The greater gains in muscle strength observed in previous studies in RA patients are probably because the duration of exercise training was 24 weeks (14,29) versus 16 weeks in the current study and 12 weeks in studies in healthy adults (41,42). Of note, the improvements in muscle function in our study should be put in the context that patients with RA who do not exercise lose ~0.7% of lean muscle mass and 2% of muscle strength annually (28,29). Thus, both NMES and high-intensity volitional resistance training can be used as effective approaches to improve muscle structure and function in patients with RA.

Despite similar improvement in muscle function in both groups, the average intensity of the NMES was less than half the intensity of volitional training (30.5% and 76.5% of maximum effort, respectively), which is a novel finding. This observation supports the notion that NMES affects muscle function at relatively lower contraction intensities compared to volitional exercise (17), which is probably due to their different patterns of muscle activation. Volitional contractions follow the Henneman's size principle, which states that under load, the smallest motoneurons (supplying type I fibers) are activated first, and type II fibers are activated last, and only reach their maximum force production near maximal contractions (44). In contrast, NMES is generally delivered at frequencies higher than the firing frequencies of muscle fibers, which overrides the central nervous system recruitment patterns and causes synchronous activation of all motor units within the field of current (16,17,45,46). Because NMES improves muscle function at lower intensities than volitional training, it could be a viable alternative for promoting muscle function in patients for whom high-intensity resistance exercise may not be tolerated or is contraindicated.

An unexpected finding was that although both interventions improved muscle structure and function, those improvements barely influenced physical function. The explanation for this finding is 2-fold. First, the exercise training in both intervention arms mainly targeted the knee extensor muscles. However, the knee extensors contribute only to a small-to-moderate degree to the functional activities tested in this study (47,48). Second, the exercise programs lacked specificity. According to the principle of training specificity, in order to become better at a skill, one must perform that skill. To that end, the only exercises in which skills similar to those used in the performance-based tests were those in the volitional exercise group (i.e., leg press and knee extension), which aligned with the skills required for climbing stairs and rising from a chair.

Both exercise training regimens did not result in adverse events or increased pain or disease activity. On the contrary, there was a small decrease in pain (–0.4 points [14%]) and disease activity (–0.5 [16%]) in both groups. While the attrition rate was 3 times (22%) higher in the NMES group compared with that in the volitional exercise group, the majority (71%) of patients in the NMES group tolerated the intervention well. Because both interventions were administered under similar conditions (i.e., setting, interventionist, time commitment), and the groups were similar regarding characteristics related to attrition (education and employment status [30]), we speculate that the higher attrition rate in the NMES group was not attributable to study or participant factors but rather to the discomfort associated with the noxious electrical stimulation.

Based on previous suggestions that NMES could have a preferred effect on type II muscle fibers (45,49), in the current study we also sought to describe the effects of interventions on muscle phenotype. This is particularly relevant in RA patients, in whom muscle atrophy seems to predominantly affect type II fibers (20,21). Our findings suggested that increases in muscle area in the NMES group were predominantly in type II fibers, whereas in the volitional training group, the size of both type I and type II muscle fibers increased similarly. These findings need further validation in larger samples along with assessment of the clinical relevance of selective hypertrophy of muscle phenotypes.

Limitations of this pilot study include the small sample size. Although between-group differences in muscle and physical function were very small and would likely not be significant in a larger sample, the attrition rate was 3 times higher in the NMES group, and differences would be significant in a larger sample. Moreover, patients who withdrew also attended few sessions of NMES. Thus, attrition and adherence are concerns when using NMES. The small sample also resulted in imbalances between groups in some baseline characteristics, which, despite adjustments in the analyses, could still have influenced the results. Additionally, the patients were not blinded to group assignment, because such blinding would be difficult given the very different natures of the interventions. Last, this study included patients with moderate disease activity and mild functional limitations, and the results may not be generalizable to populations of RA patients who have severe disease and are more debilitated.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version

to be published. Dr. Piva 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. Piva, Fitzgerald, Goodpaster, Delitto.
Acquisition of data. Khoja, Toledo, Chester-Wasko.
Analysis and interpretation of data. Piva, Toledo, Smith.

REFERENCES

1. Walsmith J, Roubenoff R. Cachexia in rheumatoid arthritis. Int J Cardiol 2002;85:89–99.
2. Kotler DP. Cachexia. Ann Intern Med 2000;133:622–34.
3. Rall LC, Roubenoff R. Rheumatoid cachexia: metabolic abnormalities, mechanisms and interventions. Rheumatology (Oxford) 2004;43:1219–23.
4. Roubenoff R, Walsmith J, Lundgren N, Snydman L, Dolnikowski GJ, Roberts S. Low physical activity reduces total energy expenditure in women with rheumatoid arthritis: implications for dietary intake recommendations. Am J Clin Nutr 2002;76:774–9.
5. Wolfe F, Mitchell DM, Sibley JT, Fries JF, Bloch DA, Williams CA, et al. The mortality of rheumatoid arthritis. Arthritis Rheum 1994;37:481–94.
6. Bearne LM, Scott DL, Hurley MV. Exercise can reverse quadriceps sensorimotor dysfunction that is associated with rheumatoid arthritis without exacerbating disease activity. Rheumatology 2002;41:157–66.
7. Ekdahl C, Andersson SI, Moritz U, Svensson B. Dynamic versus static training in patients with rheumatoid arthritis. Scand J Rheumatol 1990;19:17–26.
8. Hakkinen A, Hannonen P, Nyman K, Lyyski T, Hakkinen K. Effects of concurrent strength and endurance training in women with early or longstanding rheumatoid arthritis: comparison with healthy subjects. Arthritis Rheum 2003;49:789–97.
9. Hakkinen A, Sokka T, Kotaniemi A, Hannonen P. A randomized two-year study of the effects of dynamic strength training on muscle strength, disease activity, functional capacity, and bone mineral density in early rheumatoid arthritis. Arthritis Rheum 2001;44:515–22.
10. Rall LC, Meydani SN, Kehayias JJ, Dawson-Hughes B, Roubenoff R. The effect of progressive resistance training in rheumatoid arthritis: increased strength without changes in energy balance or body composition. Arthritis Rheum 1996;39:415–26.
11. Van den Ende CH, Breedveld FC, le Cessie S, Dijkmans BA, de Mug AW, Hazes JM. Effect of intensive exercise on patients with active rheumatoid arthritis: a randomised clinical trial. Ann Rheum Dis 2000;59:615–21.
12. Van den Ende CH, Hazes JM, le Cessie S, Mulder WJ, Belfor DG, Breedveld FC, et al. Comparison of high and low intensity training in well controlled rheumatoid arthritis: results of a randomised clinical trial. Ann Rheum Dis 1996;55:798–805.
13. Marcora SM, Lemmey AB, Maddison PJ. Can progressive resistance training reverse cachexia in patients with rheumatoid arthritis? Results of a pilot study. J Rheumatol 2005;32:1031–9.
14. Hakkinen A, Pakarinen A, Hannonen P, Kautiainen H, Nyman K, Kraemer WJ, et al. Effects of prolonged combined strength and endurance training on physical fitness, body composition and serum hormones in women with rheumatoid arthritis and in healthy controls. Clin Exp Rheumatol 2005;23:505–12.
15. Bax L, Staes F, Verhagen A. Does neuromuscular electrical stimulation strengthen the quadriceps femoris? A systematic review of randomised controlled trials. Sports Med 2005;35:191–212.
16. Delitto A, Snyder-Mackler L. Two theories of muscle strength augmentation using percutaneous electrical stimulation. Phys Ther 1990;70:158–64.
17. Gregory CM, Bickel CS. Recruitment patterns in human skeletal muscle during electrical stimulation. Phys Ther 2005;85:358–64.
18. Milner-Brown HS, Stein RB, Yemm R. The orderly recruitment of human motor units during voluntary isometric contractions. J Physiol 1973;230:350–70.
19. Maffiuletti NA, Cometti G, Amiridis IG, Martin A, Pousson M, Chatard JC. The effects of electromyostimulation training and basketball practice on muscle strength and jumping ability. Int J Sports Med 2000;21:437–43.
20. Edstrom L, Nordemar R. Differential changes in type I and type II muscle fibres in rheumatoid arthritis: a biopsy study. Scand J Rheumatol 1974;3:155–60.
21. Fiori MG, Andreola S, Ladelli G, Scirea MR. Selective atrophy of type IIb muscle fibers in rheumatoid arthritis and progressive systemic sclerosis (scleroderma): a biopsy histochemical study. Eur J Rheumatol Inflamm 1983;6:168–81.
22. Neder JA, Sword D, Ward SA, Mackay E, Cochrane LM, Clark CJ. Home based neuromuscular electrical stimulation as a new rehabilitative strategy for severely disabled patients with chronic obstructive pulmonary disease (COPD). Thorax 2002;57:333–7.
23. Snyder-Mackler L, Delitto A, Bailey SL, Stralka SW. Strength of the quadriceps femoris muscle and functional recovery after reconstruction of the anterior cruciate ligament: a prospective, randomized clinical trial of electrical stimulation. J Bone Joint Surg Am 1995;77:1166–73.
24. Quittan M, Wiesinger GF, Sturm B, Puig S, Mayr W, Sochor A, et al. Improvement of thigh muscles by neuromuscular electrical stimulation in patients with refractory heart failure: a single-blind, randomized, controlled trial. Am J Phys Med Rehabil 2001;80:206–14.
25. Vaquero AF, Chicharro JL, Gil L, Ruiz MP, Sanchez V, Lucia A, et al. Effects of muscle electrical stimulation on peak VO2 in cardiac transplant patients. Int J Sports Med 1998;19:317–22.
26. Piva SR, Goodnite EA, Azuma K, Woollard JD, Goodpaster BH, Wasko MC, et al. Neuromuscular electrical stimulation and volitional exercise for individuals with rheumatoid arthritis: multiple-patient case report. Phys Ther 2007;87:1064–77.
27. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315–24.
28. Lemmey AB, Williams SL, Marcora SM, Jones J, Maddison PJ. Are the benefits of a high-intensity progressive resistance training program sustained in rheumatoid arthritis patients? A 3-year followup study. Arthritis Care Res (Hoboken) 2012;64:71–5.
29. Lemmey AB, Marcora SM, Chester K, Wilson S, Casanova F, Maddison PJ. Effects of high-intensity resistance training in patients with rheumatoid arthritis: a randomized controlled trial. Arthritis Rheum 2009;61:1726–34.
30. Piva SR, Lasinski S, Almeida GJ, Fitzgerald GK, Delitto A. Association between disability and psychological factors and dose of neuromuscular electrical stimulation in subjects with rheumatoid arthritis. Physiother Pract Res 2013;34:57–66.
31. Goodpaster BH, Carlson CL, Visser M, Kelley DE, Scherzinger A, Harris TB, et al. Attenuation of skeletal muscle and strength in the elderly: the Health ABC study. J Appl Physiol 2001;90:2157–65.
32. Goodpaster BH, Kelley DE, Thaete FL, He J, Ross R. Skeletal muscle attenuation determined by computed tomography is associated with skeletal muscle lipid content. J Appl Physiol 2000;89:104–10.
33. Piva SR, Teixeira PE, Almeida GJ, Gil AB, DiGioia AM, Levison TJ, et al. Contribution of hip abductor strength to physical

function in patients with total knee arthroplasty. Phys Ther 2011;91:225–33.
34. Curb JD, Ceria-Ulep CD, Rodriguez BL, Grove J, Guralnik J, Willcox BJ, et al. Performance-based measures of physical function for high-function populations. J Am Geriatr Soc 2006;54:737–42.
35. Guralnik JM, Ferrucci L, Simonsick EM, Salive ME, Wallace RB. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. N Engl J Med 1995;332:556–61.
36. Simonsick EM, Newman AB, Nevitt MC, Kritchevsky SB, Ferrucci L, Guralnik JM, et al. Measuring higher level physical function in well-functioning older adults: expanding familiar approaches in the Health ABC study. J Gerontol A Biol Sci Med Sci 2001;56:M644–9.
37. Almeida GJ, Schroeder CA, Gil AB, Fitzgerald GK, Piva SR. Interrater reliability and validity of the stair ascend/descend test in subjects with total knee arthroplasty. Arch Phys Med Rehabil 2010;91:932–8.
38. Binkley JM, Stratford PW, Lott SA, Riddle DL. The lower extremity functional scale (LEFS): scale development, measurement properties, and clinical application. Phys Ther 1999;79:371–83.
39. Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: a review of its history, issues, progress, and documentation. J Rheumatol 2003;30:167–78.
40. Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight–joint counts: development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum 1995;38:44–8.
41. Luthi JM, Howald H, Claassen H, Rosler K, Vock P, Hoppeler H. Structural changes in skeletal muscle tissue with heavy-resistance exercise. Int J Sports Med 1986;7:123–7.
42. Frontera WR, Meredith CN, O'Reilly KP, Knuttgen HG, Evans WJ. Strength conditioning in older men: skeletal muscle hypertrophy and improved function. J Appl Physiol (1985) 1988;64:1038–44.
43. Taaffe DR, Henwood TR, Nalls MA, Walker DG, Lang TF, Harris TB. Alterations in muscle attenuation following detraining and retraining in resistance-trained older adults. Gerontology 2009;55:217–23.
44. Kanosue K, Yoshida M, Akazawa K, Fujii K. The number of active motor units and their firing rates in voluntary contractions of the human brachialis muscle. Jpn J Physiol 1979;29:427–43.
45. Sinacore DR, Delitto A, King DS, Rose SJ. Type II fiber activation with electrical stimulation: a preliminary report. Phys Ther 1990;70:416–22.
46. Binder-Macleod SA, Halden EE, Jungles KA. Effects of stimulation intensity of the physiological responses of human motor units. Med Sci Sports Exerc 1995;27:556–65.
47. Brown M, Sinacore DR, Host HH. The relationship of strength to function in the older adult. J Gerontol A Biol Sci Med Sci 1995;50 Spec No:55–9.
48. Khoja SS, Moore CG, Goodpaster BH, Delitto A, Piva SR. Skeletal muscle fat and its association with physical function in rheumatoid arthritis. Arthritis Care Res (Hoboken) 2018;70:333–42.
49. Cabric M, Appell HJ, Resic A. Fine structural changes in electrostimulated human skeletal muscle: evidence for predominant effects on fast muscle fibres. Eur J Appl Physiol Occup Physiol 1988;57:1–5.

AMERICAN COLLEGE OF RHEUMATOLOGY
EDUCATION • TREATMENT • RESEARCH

Arthritis Care & Research
Vol. 71, No. 3, March 2019, pp 362–366
DOI 10.1002/acr.23467

BRIEF REPORT

Arthritis After Cancer Immunotherapy: Symptom Duration and Treatment Response

Melanie H. Smith[1] and Anne R. Bass[2]

Objective. Musculoskeletal manifestations of immune-related adverse events (irAEs) after checkpoint inhibitor immunotherapy for cancer remain incompletely characterized and poorly understood. A recently published case series suggested that immunotherapy-induced arthritis is an aggressive process requiring high-dose corticosteroids.

Methods. This was a retrospective chart review of all patients with musculoskeletal irAEs first seen by one of the authors between 2014 and 2016. All patients had been treated for a malignancy with immune checkpoint inhibitors targeting PD-1 (nivolumab, pembrolizumab), PD-L1 (durvalumab), and/or CTLA-4 (ipilimumab, tremelimumab) at Memorial Sloan Kettering Cancer Center.

Results. We identified 10 patients with a mean ± SD age of 63.2 ± 9.7 years. Seven were treated with a combination of checkpoint inhibitors and 3 with nivolumab monotherapy. Four patients developed inflammatory polyarthritis, 4 oligoarthritis, and 2 tenosynovitis. Six were antinuclear antibody positive and 2 had anti–cyclic citrullinated peptide antibodies. Mean ± SD time from the first dose of immunotherapy until joint involvement was 6.3 ± 4.3 months. All 10 patients were treated with systemic corticosteroids, but 6 of 10 required ≤20 mg per day of prednisone. Five patients received steroid-sparing agents. Mean ± SD time until resolution of joint symptoms after the last dose of immunotherapy was 9.2 ± 6.1 months.

Conclusion. Musculoskeletal irAEs can manifest as a rheumatoid arthritis–like polyarthritis, oligoarthritis, tenosynovitis, or polymyalgia rheumatica. Musculoskeletal symptoms can last more than a year, but they can generally be managed with low to moderate doses of corticosteroids.

INTRODUCTION

The introduction of checkpoint inhibitor immunotherapy has fundamentally changed the treatment for multiple cancers. Through the inhibition of PD-1, its ligand PD-L1, or CTLA-4, the immune checkpoint inhibitors remove inherent inhibition built into our immune system and allow for more robust antitumor immune responses (1). Since the Food and Drug Administration (FDA) approval of ipilimumab, a monoclonal antibody targeting CTLA-4 in 2011 (2), the number of cancers for which checkpoint inhibitors have shown a survival benefit has grown rapidly, and there are multiple clinical trials ongoing for the use of these treatments in a variety of additional cancers.

By blocking the negative regulators of T cell activation that are responsible for immune homeostasis, the major unwanted effect of the immune checkpoint inhibitors is the risk of immune over-activation and damage to healthy tissues. The resulting toxicities, termed immune-related adverse events (irAEs), vary widely in terms of the tissue affected and the severity of the reaction (mild/not requiring treatment to life-threatening) (3). The most common include dermatologic (rashes), colitis, hepatitis, endocrinopathies, and pneumonitis. Rheumatic manifestations are reported in 6–8% of patients treated with single-agent immunotherapy and 10% treated with dual checkpoint inhibition (4), and can manifest as polyarthritis, tenosynovitis, polymyalgia rheumatica (PMR), or sicca (5–7). Here we present a series of 10 patients that experienced musculoskeletal irAEs with long-term followup to illustrate the required duration of immunosuppressive treatment.

PATIENTS AND METHODS

This was a retrospective chart review of all patients with musculoskeletal irAEs first seen by one of the authors between

[1]Melanie H. Smith, MD, PhD: New York-Presbyterian Hospital/Weill Cornell Medicine, New York; [2]Anne R. Bass, MD: Weill Cornell Medicine and Hospital for Special Surgery, New York, New York.

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

Submitted for publication July 22, 2017; accepted in revised form November 7, 2017.

SIGNIFICANCE & INNOVATIONS

- Musculoskeletal immune-related adverse events (irAEs) occur later after initiation of immunotherapy than other irAEs.
- Musculoskeletal irAEs often require treatment for >6 months after discontinuation of immunotherapy.
- Musculoskeletal irAEs can often be managed with low to moderate dose systemic corticosteroids.

February 2014 and October 2016 and was approved by our local institutional review board. All patients had been treated for a malignancy with immune checkpoint inhibitors targeting PD-1 (nivolumab, pembrolizumab), PD-L1 (durvalumab), and/or CTLA-4 (ipilimumab, tremelimumab) at Memorial Sloan Kettering Cancer Center and were referred for management of joint pain through a "fast track" program created by the senior author (ARB) that enabled them to be seen within 1 week of referral. These patients represent the large majority of the patients with musculoskeletal irAEs seen at our institution during this period. Patients with irAEs but without musculoskeletal complaints were excluded from this case series. The diagnoses of inflammatory arthritis and tenosynovitis were made based on clinical examination of the joints and/or tendons. Oligoarthritis was defined as involvement of 4 or fewer joints, and polyarthritis by involvement of 5 or more joints. Autoimmune serologies were available for 8 patients, but no patient had serologic testing prior to immunotherapy. Cancer outcomes were reported by the patients' primary oncologist based on clinical assessment and cross-sectional imaging.

RESULTS

Patient demographics and clinical presentation are shown in Table 1. The mean ± SD age was 63.2 ± 9.7 years and 50% were female. Four patients received treatment for melanoma, 2 for non–small cell lung cancer, and 1 each for anal, cervical, Merkel, and renal cell carcinoma. Four patients received combination therapy with both PD-1/PD-L1 and CTLA-4 inhibition. Four patients were treated with a sequential therapy of different checkpoint inhibitors, and 3 received nivolumab monotherapy. Patient 2 was also treated with an anti–killer cell immunoglobulin-like receptor (KIR) therapy (lirilumab). Patient history was notable for mild symptoms of PMR prior to the diagnosis of cancer in patient 5. Two patients had first-degree relatives with autoimmune diseases: patient 4 had a daughter with lupus and patient 8 had a sister with rheumatoid arthritis (RA).

Four patients developed inflammatory polyarthritis on immunotherapy, 4 oligoarthritis, and 2 tenosynovitis. Patient 1 had a grade 3 musculoskeletal irAE; all other irAEs observed were grade 1–2. Of the 10 patients, 1 had features of PMR and 2 patients had concomitant keratoconjunctivitis sicca (Sjögren's). None of the patients had conjunctivitis, uveitis, or iritis. Seven patients had other irAEs: 3 with hypophysitis, 2 each with colitis, hepatitis, and hypothyroidism, as well as individual cases of peripheral neuropathy and prostatitis. Five patients had more than 1 additional irAE. Of the 8 patients with serologic data, 6 were antinuclear antibody (ANA) positive (low titer), and 2 had anti–cyclic citrullinated peptide (anti-CCP) antibodies (1 of which was high titer).

Time of onset, treatment, and duration can be seen in Table 2. The mean ± SD time from first dose of immunotherapy until any irAE

Table 1. Patient demographics and clinical presentation*

Case	Age	Sex	Malignancy	CTLA4	PD-1	Poly.	Oligo.	Tenosyn.	Joints	Comments	Other
1	57	M	NSCLC	X	X			X	Wrists	De Quervain's, ANA+	Hypophysitis, dermatitis
2	74	F	Melanoma		X			X	Wrist, shoulder	ANA+	Neuropathy
3	67	M	Melanoma	X	X	X				RA, Sjögren's	
4	56	F	Melanoma	X	X		X		Knees, ankles	ANA+	Hepatitis, colitis,
5	63	F	Anal	X	X		X			PMR, ANA+, CCP+	Hypothyroid
6	47	F	Cervical	X	X		X		Knees, ankles		
7	68	F	Melanoma	X	X	X					Colitis, hypothyroid, hypophysitis
8	60	M	NSCLC		X	X				RA, ANA+	
9	81	M	Merkel		X	X				RA, CCP+	Psoriasis, Parkinsonism
10	59	M	Renal	X	X		X		Knee	Sjögren's, ANA+	Hepatitis, prostatitis, hypophysitis

* Poly. = polyarthritis; Oligo. = oligoarthritis; Tenosyn. = tenosynovitis; NSCLC = non–small cell lung cancer; ANA = antinuclear antibody; RA = rheumatoid arthritis; PMR = polymyalgia rheumatica; CCP = cyclic citrullinated peptide.

Table 2. Onset, treatment, and duration of musculoskeletal irAE*

Case	Max. prednisone dose (mg)	Other medications	Onset, months	Duration of IT, months	Reason IT stopped	Months to resolution after IT stopped	Cancer outcome
1	10 daily, TSI	HCQ	11	20	Persistent de Quervain's	11	Progression
2	10 daily, IA		4	14	New breast cancer	13	Stable
3	20 twice daily, IA	Infliximab	7	9	Progression	3	Progression
4	20 in AM, 10 in PM	Celecoxib	4	2	Fevers (IT reaction)	Ongoing (3 mos)	Recurrence
5	15 daily		1	8	Progression	Ongoing (9 mos)	Progression
6	60 daily		14	16	Progression	Ongoing (3 mos)	Progression
7	20 daily, IA		9	8	NED	18	Remission
8	10 daily	HCQ, SSZ	1	11	Progression	Ongoing (15 mos)	Stable
9	20 daily, IA	MTX	4	8	NED	8	Remission
10	50 daily	MMF	8	8	Nonmusculoskeletal irAE	2	Remission

* irAE = immune-related adverse events; IT = immunotherapy; TSI = tendon sheath injection; HCQ = hydroxychloroquine; IA = intraarticular steroid injection; NED = no evidence of disease ; SSZ = sulfasalazine; MTX = methotrexate; MMF = mycophenolate mofetil.

was 4.8 ± 3.8 months, whereas mean ± SD time from first dose of immunotherapy until joint involvement was 6.3 ± 4.3 months, median 5.5 months (range 1–14 months). The 2 patients with the shortest interval between immunotherapy and the onset of musculoskeletal irAEs were patient 5, who had mild PMR-like symptoms prior to immunotherapy, and patient 8, who had a sister with RA.

All 10 patients were treated with systemic corticosteroids for their arthritis or tenosynovitis. Patients generally advocated for the lowest dose of steroid possible, both to minimize interference with trial protocols and to maximize the efficacy of immunotherapy. Six of the 10 patients had good symptom control with ≤20 mg prednisone daily. Three patients were started on disease-modifying antirheumatic drugs (DMARDs), and 1 patient required a tumor necrosis factor inhibitor (infliximab) to allow tapering of steroids. Patient 10 had onset of prostatitis, followed in a matter of weeks by severe knee arthritis, and Sjögren's with periorbital edema and pain in bilateral parotid glands. These irAEs were initially treated with high-dose corticosteroids, and mycophenolate mofetil was started as a steroid-sparing agent, but he then developed autoimmune hepatitis that required additional high-dose steroids.

Six of the patients had resolution of musculoskeletal symptoms and discontinued treatment (generally prednisone) an average of 9.2 ± 6.1 months after the last dose of immunotherapy. A single patient stopped immunotherapy due to the severity of the musculoskeletal irAE (bilateral de Quervain's tenosynovitis). Four patients continued to be treated for their arthritis at the time of last rheumatology followup. Patient 10 required ongoing treatment for nonmusculoskeletal irAEs.

All patients were alive at the time of manuscript submission. Five patients had progression or recurrence of their cancers requiring conventional chemotherapy, surgery, and/or radiation. Two patients had stable cancer and 3 were in remission.

DISCUSSION

Our case series demonstrates that patients experiencing musculoskeletal irAEs can have a wide variety of manifestations from oligoarthritis and tenosynovitis, to PMR, to a polyarthritis consistent with RA. These irAEs had a delayed onset after initiation of immunotherapy, and the majority of patients required prednisone 10–20 mg daily for at least 6 months to control their symptoms. Six patients had a positive ANA without evidence of systemic lupus erythematosus, and 2 patients were CCP positive, 1 with polyarthritis and 1 with oligoarthritis.

Compared to other common irAEs, musculoskeletal irAEs appear to have a later onset. For example, skin involvement is typically observed first, 2–4 weeks after initiation of immunotherapy, followed by colitis at 4–6 weeks, and finally endocrinopathies and hepatitis, which begin at 6–8 weeks, but may not become clinically apparent until later (8). For rheumatologic irAEs, the removal of immune inhibition may unmask native disease, and the delayed onset of symptoms may reflect a preclinical phase similar to that seen in non–immunotherapy-related rheumatic disease. The 2 patients with the shortest time until onset of musculoskeletal symptoms (patients 5 and 8) either had symptoms prior to initiating immunotherapy or a strong genetic predisposition, and could have already had preclinical disease prior to immunotherapy. Alternatively, it is also possible that early musculoskeletal symptoms were masked by steroids that were used to treat earlier life-threatening irAEs (as was required in patients 4, 7, and 10).

IrAEs commonly require treatment with immunosuppression; almost one-third of patients treated with ipilimumab required systemic corticosteroids in one study (9). In contrast to an earlier case series that suggested that high-dose steroids are required for arthritis control (5), we found that lower doses of prednisone were usually adequate. More aggressive treatments were generally given to patients with other, more life-threatening irAEs and were not often given exclusively for the treatment of musculoskeletal symptoms.

With treatment and cessation of immunotherapy, resolution of most irAEs has been reported to be on the order of 1–2 months (10). The exception is endocrinopathies, which may not resolve and may require life-long hormone replacement. Based on our case series, the duration of musculoskeletal irAEs appears to be much longer: the mean symptom duration after stopping immunotherapy was 9 months. Although this is much longer than would be expected based on a serum half-life of 12–20 days, pharmacokinetic studies suggest that PD-1 receptor occupancy is ≥2 months in patients receiving nivolumab (11). Alternatively, the long duration of symptoms may be due to the unmasking or stimulation of native rheumatologic disease. A significant strength of our study is the length of followup.

Increasingly, patients are being treated with combination immunotherapy, and over one-third must stop therapy due to severe irAEs (4,12). In our study, only 3 patients stopped therapy due to adverse reactions (only 1 of these was a musculoskeletal irAE and 1 was due to fevers). This difference is likely because this case series focuses on musculoskeletal irAEs that are most commonly grade 1–2. In addition, this series may have been biased toward cases of milder arthritis because all patients were seen in the outpatient setting. Future research should focus on differentiating the pathways responsible for antitumor immune activity versus irAEs to develop more targeted therapies. Data extracted from recent clinical trials suggest that this may be possible: treatment of irAEs does not decrease the antitumor effect (9), and patients with preexisting autoimmune diseases do not appear to have a higher risk of irAEs than other patients (13).

IrAEs such as hypophysitis, which are common following immunotherapy but rare in other contexts, suggest a pathogenesis that is treatment specific (hypophysitis is most commonly seen with ipilimumab, possibly due to pituitary expression of the CTLA-4 antigen [14], but it has also been reported in patients receiving PD-1 inhibitors [15]). In contrast, musculoskeletal irAEs are heterogeneous (polyarthritis, monoarthritis, tenosynovitis) and suggest that patient-specific factors are critical in defining phenotype while their occurrence is the result of increased immune activity in general. One study that described irAEs in 30 patients with preexisting autoimmune disease who received ipilimumab for advanced melanoma found that only half developed irAEs, but 5 of 6 patients with RA had a flare (13). This suggests that patients with rheumatic autoimmune conditions have an increased risk of disease flare in the setting of checkpoint inhibition. However, these flares can often be managed medically and do not have to stand in the way of life-saving cancer treatment.

In summary, musculoskeletal irAEs tend to occur relatively late after initiation of immunotherapy, and symptoms typically require treatment for greater than 6 months. They may represent unmasking or stimulation of native disease and patient-related genetic factors likely play a role in determining the form that immunotherapy-induced arthritis takes. Patients treated with checkpoint inhibitors may be useful models for the study of early rheumatic disease.

AUTHOR CONTRIBUTIONS

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

Study conception and design. Smith, Bass.
Acquisition of data. Smith, Bass.
Analysis and interpretation of data. Smith, Bass.

REFERENCES

1. Leach DR, Krummel MF, Allison JP. Enhancement of antitumor immunity by CTLA-4 blockade. Science 1996;271:1734–16.
2. Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, et al. Improved survival with ipilimumab in patients with metastatic melanoma. N Engl J Med 2010;363:711–23.
3. Naidoo J, Page DB, Li BT, Connell LC, Schindler K, Lacouture ME, et al. Toxicities of the anti-PD-1 and anti-PD-L1 immune checkpoint antibodies. Ann Oncol 2015;26:2375–91.
4. Larkin J, Chiarion-Sileni V, Gonzalez R, Grob JJ, Cowey CL, Lao CD, et al. Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. N Engl J Med 2015;373:23–34.
5. Cappelli LC, Gutierrez AK, Baer AN, Albayda J, Manno RL, Haque U, et al. Inflammatory arthritis and sicca syndrome induced by nivolumab and ipilimumab. Ann Rheum Dis 2017;76:43–50.
6. Cappelli LC, Gutierrez AK, Bingham CO, Shah AA. Rheumatic and musculoskeletal immune-related adverse events due to immune checkpoint inhibitors: a systematic review of the literature. Arthritis Care Res 2017;69:1751–63.
7. Calabrese L, Velcheti V. Checkpoint immunotherapy: good for cancer therapy, bad for rheumatic diseases. Ann Rheum Dis 2017;76:1–3.
8. Weber JS, Kähler KC, Hauschild A. Management of immune-related adverse events and kinetics of response with ipilimumab. J Clin Oncol 2012;30:2691–7.
9. Horvat TZ, Adel NG, Dang TO, Momtaz P, Postow MA, Callahan MK, et al. Immune-related adverse events, need for systemic immunosuppression, and effects on survival and time to treatment failure in patients with melanoma treated with ipilimumab at Memorial Sloan Kettering Cancer Center. J Clin Oncol 2015;33:3193–8.
10. Weber JS, Dummer R, Pril V de, Lebbé C, Hodi FS, for the MDX010-20 Investigators. Patterns of onset and resolution of immune-related adverse events of special interest with ipilimumab. Cancer 2013;119:1675–82.
11. Brahmer JR, Drake CG, Wollner I, Powderly JD, Picus J, Sharfman WH, et al. Phase I study of single-agent anti-programmed death-1 (MDX-1106) in refractory solid tumors: safety, clinical activity, pharmacodynamics, and immunologic correlates. J Clin Oncol 2010;28:3167–75.

12. Postow MA, Chesney J, Pavlick AC, Robert C, Grossmann K, McDermott D, et al. Nivolumab and ipilimumab versus ipilimumab in untreated melanoma. N Engl J Med 2015;372:2006–17.
13. Johnson DB, Sullivan RJ, Ott PA, Carlino MS, Khushalani NI, Ye F, et al. Ipilimumab therapy in patients with advanced melanoma and preexisting autoimmune disorders. JAMA Oncol 2016;2:234–40.
14. Iwama S, Remigis AD, Callahan MK, Slovin SF, Wolchok JD, Caturegli P. Pituitary expression of CTLA-4 mediates hypophysitis secondary to administration of CTLA-4 blocking antibody. Sci Transl Med 2014;6:230ra45.
15. Okano Y, Satoh T, Horiguchi K, Toyoda M, Osaki A, Matsumoto S, et al. Nivolumab-induced hypophysitis in a patient with advanced malignant melanoma. Endocr J 2016;63:905–12.

Arthritis Care & Research
Vol. 71, No. 3, March 2019, pp 367–378
DOI 10.1002/acr.23738

AMERICAN COLLEGE OF RHEUMATOLOGY
EDUCATION • TREATMENT • RESEARCH

Safety of Ixekizumab in Patients With Psoriatic Arthritis: Results From a Pooled Analysis of Three Clinical Trials

Philip Mease,[1] Euthalia Roussou,[2] Gerd-Rüdiger Burmester,[3] Philippe Goupille,[4] Alice Gottlieb,[5] Susan R. Moriarty,[6] Olivier Benichou,[6] David H. Adams,[6] Wen Xu,[6] and Peter Nash[7]

Objective. To evaluate ixekizumab safety in adults with psoriatic arthritis (PsA).

Methods. Safety data from 2 integrated data sets are presented: 1) 24-week, double-blind, placebo-controlled period of SPIRIT-P1 and SPIRIT-P2; and 2) all ixekizumab-treated patients of SPIRIT-P1 and SPIRIT-P2 plus SPIRIT-P3 open-label period. We report adverse event (AE) frequency and exposure-adjusted incidence rates per 100 patient-years at 12-week intervals to week 96.

Results. The placebo-controlled period had 678 patients (safety population): 224 placebo, 229 ixekizumab every 4 weeks, and 225 ixekizumab every 2 weeks. Overall, 1,118 patients received ixekizumab (total exposure 1,373.4 patient-years). In the placebo-controlled period, the frequencies of ixekizumab-treated patients experiencing ≥1 treatment-emergent AE (TEAE) and those experiencing serious AEs were 68.1% (56.7% placebo) and 4.4% (2.7% placebo), respectively. Injection site reactions (ISRs) were very common (21.4% ixekizumab [4.5% placebo]), with ISR discontinuation rates of 1.1% (ixekizumab) and 0.4% (placebo). Through week 96, the incidence rates of ISRs decreased with increasing ixekizumab exposure. The frequencies of AEs of special interest were 32.8% (ixekizumab) and 27.7% (placebo); for serious infections, the frequencies were 1.3% and 0%, respectively; *Candida* infections, 2.6% and 0.4%; confirmed major adverse cardiac events, 0% and 0%; malignancy, 0.4% and 0%; hypersensitivities, 5.3% and 1.8%; and depression-related, 1.8% and 1.3%. The frequency of Crohn's disease and ulcerative colitis (investigator-reported) was 0% in both groups, and the frequencies of sponsor-determined inflammatory bowel disease were 0.2% in the ixekizumab group and 0% in the placebo group. Overall, no active tuberculosis, invasive *Candida* infections, anaphylaxis, or suicide/self-injury behaviors were reported.

Conclusion. The PsA ixekizumab safety integrated data set reached 1,373.4 patient-years total exposure. Ixekizumab-treated patients had higher rates of overall TEAEs, serious infections, mucocutaneous *Candida*, hypersensitivities (non-anaphylactic), and ISRs than placebo-treated patients. No unexpected safety outcomes were reported.

INTRODUCTION

Interleukin-17A (IL-17A) is a therapeutic target for psoriasis and psoriatic arthritis (PsA) given its role in pathological immune responses associated with those conditions (1). Secukinumab was the first IL-17A inhibitor approved for the treatment of PsA (2,3). Ixekizumab is a high-affinity monoclonal antibody that selectively targets IL-17A (4). In the SPIRIT-P1

Supported by Eli Lilly and Company.

[1]Philip Mease, MD: Swedish Medical Center and University of Washington, Seattle; [2]Euthalia Roussou, MD, PhD: King George Hospital, Goodmayes Essex, UK; [3]Gerd-Rüdiger Burmester, MD: Charité – University Medicine Berlin Free University and Humboldt University Berlin, Germany; [4]Philippe Goupille, MD: University of Tours, Tours, France; [5]Alice Gottlieb, MD, PhD: New York Medical College, New York; [6]Susan R. Moriarty, MD, Olivier Benichou, MD, PhD, David H. Adams, PhD, Wen Xu, PhD: Eli Lilly and Company, Indianapolis, Indiana; [7]Peter Nash, MBBS (Hons), FRACP: University of Queensland Brisbane, Rheumatology Research Unit, Sunshine Coast, Brisbane, Queensland, Australia.

Dr. Mease has received research grants from AbbVie, Amgen, Bristol-Myers Squibb, Celgene, Janssen, Eli Lilly and Company, Novartis, Pfizer, Sun Pharmaceutical and UCB Pharma and consulting fees, speaking fees, and/or honoraria from AbbVie, Amgen, Eli Lilly and Company, and Novartis (more than $10,000 each). Dr. Roussou has received research grants from Pfizer, UCB Pharma, Merck, and Eli Lilly and Company. Dr. Burmester has received honoraria (less than $10,000 each) and research support from AbbVie, Pfizer, Roche, UCB Pharma, Merck Sharp & Dohme, Bristol-Myers Squibb, Sanofi, Novartis, Eli Lilly, and Company and Janssen. Dr. Goupille has received consulting fees from AbbVie, Biogaran, Bristol-Myers Squibb, Celgene, Eli Lilly and Company, Janssen, MSD, Novartis, Pfizer, Roche-Chugai, and UCB Pharma (less than $10,000 each). Dr. Gottlieb has received consulting fees from Janssen, AbbVie, UCB Pharma, and Eli Lilly and Company (more than $10,000 each), and from Celgene, Bristol-Myers Squibb, Beiersdorf, Novartis, Incyte, Dr. Reddy's Laboratories, Valeant, Demira, Allergan, and Sun Pharmaceutical (less than $10,000 each). She has received research/educational grants from UCB Pharma, Eli Lilly and Company, Novartis, Janssen, and Incyte. Drs. Moriarty, Benichou, Adams, and Xu own stock in Eli Lilly and Company. Dr. Nash has received honoraria (less than $10,000 each) and research support from Pfizer, Roche, UCB Pharma, Merck Sharp & Dohme, Bristol-Myers Squibb, Sanofi, Novartis, Eli Lilly, AbbVie, and Janssen.

Address correspondence to Philip Mease, MD, Seattle Rheumatology Associates, 601 Broadway, Suite 600, Seattle, WA 98122. E-mail: pmease@philipmease.com.

Submitted for publication January 23, 2018; accepted in revised form August 21, 2018.

SIGNIFICANCE & INNOVATIONS

- Results from 1,373.4 patient-years of ixekizumab exposure in patients with psoriatic arthritis (PsA) are reported.
- A major strength of this study is that it combines 2 data sets, each of which has its own advantages. The placebo-controlled period data set allows comparisons with placebo in a data set combining all data available to date. The all ixekizumab-treated data set is larger, with increased numbers of patients and longer durations of exposure, allowing detection of adverse events that occur less frequently than in the placebo-controlled period data set, and enabling assessment of how the incidence rates of adverse events of special interest evolve over time.
- Overall, treatment-emergent adverse events, injection site reactions, serious infections, mucocutaneous (but not systemic) *Candida* infection, and hypersensitivities (non-anaphylactic) were observed more frequently in the ixekizumab group than in the placebo group.
- The safety profile of ixekizumab for the treatment of PsA was consistent with the known safety profile of ixekizumab for the treatment of patients with moderate-to-severe plaque psoriasis, and no unexpected safety signals were observed.

trial, ixekizumab was shown to be superior to placebo for reducing disease activity and radiographic disease progression and for improving physical function and patient-reported quality of life in biologic-naive patients with active PsA (5). In SPIRIT-P2, in which patients with active PsA who previously had an inadequate response to a tumor necrosis factor inhibitor (TNFi) were enrolled, ixekizumab was also effective for improving the signs and symptoms of PsA (6), with a safety profile consistent with those seen in studies involving ixekizumab treatment in patients with PsA and psoriasis (5,7–9).

The benefit/risk profile is an important consideration for any drug. Given the role of IL-17A in host immunity, safety considerations for IL-17A inhibitors include an increased risk of certain types of infections, including mucocutaneous *Candida* and upper respiratory tract infections (10–13). Inflammatory bowel disease (IBD) is also a potential concern with regard to IL-17 inhibitors, based on unexpected findings in studies in which an IL-17 inhibitor was used (14,15). General concerns more broadly for immunomodulatory agents, such as a TNFi, include serious infections (active tuberculosis [TB]), malignancies, and major adverse cardiovascular events (MACE) (16–18). Monoclonal antibody treatment may cause hypersensitivity, including anaphylaxis (16). Short- and long-term safety analyses using integrated data sets from clinical trials were reported for ixekizumab and secukinumab in patients with plaque psoriasis (9,19). In the current study, we report an integrated safety analysis of ixekizumab in patients with active PsA, using data pooled from phase III trials.

PATIENTS AND METHODS

Patients and study design. Data were derived from SPIRIT-P1 (ClinicalTrials.gov identifier: NCT01695239) (5), SPIRIT-P2 (ClinicalTrials.gov identifier: NCT02349295) (6), and SPIRIT-P3 (ClinicalTrials.gov identifier: NCT02584855) (Table 1). Supplementary Listing 1 (available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23738/abstract) shows key enrollment criteria. SPIRIT-P1 and SPIRIT-P2 are randomized, double-blind, placebo-controlled, phase III trials involving patients with active PsA (5,6) (for details, see Supplementary Text 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23738/abstract). SPIRIT-P3 is a phase III study with a 36-week to 64-week open-label treatment period during which the effects of treatment with ixekizumab administered every 2 weeks were examined, followed by a randomized withdrawal period in patients with active PsA who have an inadequate response to a conventional disease-modifying antirheumatic drug (cDMARD) and also are biologic DMARD (bDMARD)–naive. SPIRIT-P3 is ongoing; therefore, only data from the open-label period are included. SPIRIT-P2 is also ongoing.

Integrated data sets. The placebo-controlled period data set contains data from the placebo-controlled periods (week 0 through week 24) of SPIRIT-P1 and SPIRIT-P2 but excludes data observed after week 16 for week-16 inadequate responders. Adalimumab was used during the placebo-controlled period of SPIRIT-P1 only. Results for adalimumab are not presented herein, because the adalimumab safety profile in SPIRIT-P1 was previously reported (5). The all ixekizumab-treated data set consists of available ixekizumab safety data for ixekizumab-exposed patients during all treatment periods of SPIRIT-P1 and SPIRIT-P2 and the open-label period of SPIRIT-P3.

The interim database locks used in this study for the placebo-controlled period data set were September 15, 2016 (SPIRIT-P1) and September 30, 2016 (SPIRIT-P2). For the all ixekizumab-treated data set, the database locks were February 22, 2017 (SPIRIT-P1 and SPIRIT-P3) and April 26, 2017 (SPIRIT-P2).

Safety evaluations. For the safety analyses, we used data from the PsA Safety Population, defined as all randomized patients who received ≥1 dose of study drug (placebo-controlled period data set) and all patients who received ≥1 dose of ixekizumab (all ixekizumab-treated data set). Adverse events (AEs) are presented based on Medical Dictionary for Regulatory Activities (MedDRA) versions 19.0 and 19.1. A treatment-emergent AE (TEAE) was defined as an event that

Table 1. Overview of the clinical trials*

Study	Study description	Population	Treatments according to study period†	Status‡
SPIRIT-P1 (ClinicalTrials.gov identifier: NCT01695239), pivotal, phase III	Multicenter, randomized, double-blind, active and placebo-controlled, parallel-group study followed by long-term extension; primary end point at week 24	Adult bDMARD-naive patients with active PsA (meet CASPAR criteria and ≥3 of 68 tender and ≥3 of 66 swollen joints, ≥1 disease-related hand or foot joint erosion, or a CRP level >6 mg/liter)	Double-blind placebo-controlled treatment period (week 0 to week 24) IXE 80 mg Q4W IXE 80 mg Q2W ADA 40 mg Q2W Placebo Extended and long-term treatment period (week 24 to week 156) IXE 80 mg Q4W IXE 80 mg Q2W	24- and 52-week analyses completed; ongoing long-term extension treatment period
SPIRIT-P2 (ClinicalTrials.gov identifier: NCT02349295), pivotal, phase III	Multicenter, randomized, double-blind, placebo-controlled, parallel-group study followed by long-term extension; primary end point at week 24	Adult patients who are both bDMARD-and cDMARD-experienced and have active PsA (meet CASPAR criteria, and ≥3 of 68 tender and ≥3 of 66 swollen joints)	Double-blind placebo-controlled treatment period (week 0 to week 24) IXE 80 mg Q4W IXE 80 mg Q2W Placebo Extension treatment period (week 24 to week 156) IXE 80 mg Q4W IXE 80 mg Q2W	24-week analysis completed; ongoing extension treatment period (week 52 database lock and analysis completed)
SPIRIT-P3 (ClinicalTrials.gov identifier: NCT02584855), support, phase III§	Multicenter, randomized, double-blind withdrawal study, preceded by 36-week open-label treatment period	Adult patients with active PsA (meet CASPAR criteria, ≥3 of 68 tender, and ≥3 of 66 swollen joints) who are cDMARD-inadequate responders and also bDMARD-naive	Open-label treatment period (week 0 to weeks 36–64) IXE 80 mg Q2W Randomized double-blind withdrawal period (after week 36) Patients meeting MDA for 3 months and had received IXE 80 mg Q2W for ≥6 months¶ IXE 80 mg Q2W Placebo Patients who no longer met MDA and had received IXE 80 mg Q2W¶ Patients who did not meet MDA¶ IXE 80 mg Q2W	Ongoing

* bDMARD = biologic disease-modifying antirheumatic drug; PsA = psoriatic arthritis; CASPAR = Classification Criteria for Psoriatic Arthritis; CRP = C-reactive protein; cDMARD = conventional DMARD; MDA = minimal disease activity.
† During week 16 in SPIRIT-P1 and SPIRIT-P2, inadequate responders received rescue therapy as defined in the protocols, which is a modification of allowed concomitant medications. At week 16, inadequate responders who were originally assigned to adalimumab (ADA) or placebo were re-randomized 1:1 to receive either ixekizumab (IXE) 80 mg every 2 weeks (Q2W) or every 4 weeks (Q4W), and those who were originally assigned to IXE 80 mg Q2W or Q4W continued their originally assigned dosing regimen up to week 24.
‡ Status at the time of the data cutoff date for the safety summary described in this report.
§ Safety data from the open-label treatment period of SPIRIT-P3 are included in the analyses presented in the safety summary described in this report.
¶ According to a report by Coates et al (27), "A patient is classified as achieving minimal disease activity (MDA) when meeting 5 of the 7 following criteria: tender joint count ≤1; swollen joint count ≤1; Psoriasis Activity and Severity Index ≤1 or body surface area ≤3; patient pain visual analog score (VAS) score ≤15; patient global disease activity VAS ≤20; Health Assessment Questionnaire ≤0.5; tender entheseal points ≤1."

first occurred or worsened in severity after the start of study treatment (baseline) and on or before the date of the last visit within the treatment period. The maximum severity for each TEAE during the baseline period was used as baseline severity. TEAEs were assigned to the treatment period during which they were considered treatment-emergent.

A serious AE (SAE) was defined as an event causing death, initial or prolonged inpatient hospitalization, a life-threatening experience, persistent or significant disability/incapacity, congenital anomaly/birth defect, or any other event considered by the investigator to be significant. AEs of special interest (AESIs) were prespecified, and selected AESIs are described in the current study. Injection site reactions (ISRs), in the plural form, refers to a grouping of injection site-related TEAEs based on a prespecified list of terms, whereas ISR, in its singular form, refers to the individual preferred term (PT). The grouping of ISRs included all PTs defined by ISR-related high-level terms in the MedDRA dictionary and excluded 10 joint-related PTs. The number of the PTs in ISR grouping depends on the MedDRA version. In MedDRA version 19.1, the ISRs high-level term includes 78 PTs; therefore, the ISRs grouping included 68 PTs.

For neutrophils, the lower limit of normal (2.0×10^9/liter) was defined conservatively as an increment above Common Terminology Criteria for Adverse Events grade 1, equal to increments defining grades 1 through 4. Shift tables were produced for neutrophils: normal $\geq 2.0 \times 10^9$/liter; grade 1 $<2.0 \times 10^9$/liter to $\geq 1.5 \times 10^9$/liter; grade 2 $<1.5 \times 10^9$/liter to $\geq 1.0 \times 10^9$/liter; grade 3 $<1.0 \times 10^9$/liter to $\geq 0.5 \times 10^9$/liter; and grade 4 $<0.5 \times 10^9$/liter.

Infections temporally associated with neutropenia were defined as infections with an onset date ≤14 days before or after the date of laboratory-identified neutropenia. Cerebrocardiovascular events, including MACE, were externally adjudicated by the Clinical Events Committee (CEC) at the Cleveland Clinic.

TB testing was conducted at baseline and then at yearly intervals in patients with no history of TB infection. In SPIRIT-P1, patients who had a post-baseline positive TB test during yearly testing were required to discontinue ixekizumab treatment. In SPIRIT-P2 and SPIRIT-P3, patients were allowed to continue if active TB was excluded and if they received a full course of latent TB treatment with no evidence of hepatotoxicity.

Statistical analysis. Overall exposure was summarized in total patient-years, calculated as follows: patient-year = sum of duration of ixekizumab exposure (days) for all patients in the treatment group/365.25. TEAEs were summarized by frequencies and incidence rates (IRs). For AESIs that maintained a constant hazard rate over the treatment period, the exposure-adjusted IRs of the entire treatment period were also analyzed. The slope of the hazard rate plot for each AESI was visually assessed. If the line was approximately flat, the hazard rate was considered constant. IRs were calculated by dividing the total number of patients experiencing the TEAE for each PT by the sum of all patients' time (in 100 years) of exposure during the treatment period. The entire time on study during the treatment period was used for the calculations. IRs are expressed as the IR (95% confidence interval [95% CI])/100 patient-years.

For the placebo-controlled period data set, treatment comparisons for categorical data (frequency) were analyzed using the Cochran-Mantel-Haenszel test, stratified by study. For the all ixekizumab-treated data set, the number and percentage of patients and the exposure-adjusted IRs are presented by 12-week intervals from week 0 to week 96 for patients with ≥1 TEAE, patients with ≥1 TEAE in each system organ class group, and for individual AESIs.

RESULTS

Baseline characteristics and exposure. Overall, 454 ixekizumab-treated patients (193.8 patient-years) and 1,118 ixekizumab-treated patients (1,373.4 patient-years) were included in the placebo-controlled period data set (Table 2) and the all ixekizumab-treated safety data set, respectively (see Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23738/abstract); between-group demographics were similar in the placebo-controlled period data set. The median numbers of ixekizumab injections were 7 (range 2–14) during the placebo-controlled period and 19 (range 1–79) among all ixekizumab-treated patients. Supplementary Table 2 (available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23738/abstract) shows study drug exposure.

Safety overview. *Placebo-controlled period.* The overall incidence of TEAEs in both ixekizumab groups was higher than that in the placebo group (Table 3). The most common TEAEs in the total ixekizumab group were ISRs, upper respiratory tract infection, and erythema at the injection site (Table 3). The frequencies of patients experiencing ≥1 SAE tended to be higher in the ixekizumab groups compared with those in the placebo group (Table 3). No SAE term was reported for >1 patient. Similar numbers of patients in the total ixekizumab and placebo groups discontinued treatment because of AEs. There were no clinically meaningful (author assessment) differences between ixekizumab doses for overall TEAEs, SAEs, or treatment discontinuation (Table 3). There were no deaths.

All ixekizumab-treated. When AEs were analyzed in 12-week intervals (week 0 to week 96), the IR of overall TEAEs decreased with longer ixekizumab exposure (Figure 1A). The most common TEAEs were ISRs, upper respiratory tract infection, and nasopharyngitis. When overall SAEs were analyzed at 12-week intervals, there was some variation in IRs, but there was no clinically meaningful increase in the exposure-adjusted IR with longer ixekizumab exposures (Figure 1B). SAEs that occurred in >1 patient were cholelithiasis (IR 0.3 [95% CI 0.1–0.8]/100 patient-years); pneumonia (IR 0.2 [95% CI 0.1–0.7]/100 patient-years); and lower respiratory tract infection, acute cholecystitis, latent TB, esophageal candidiasis, carotid artery stenosis, cerebrovascular accident,

Table 2. Demographic and baseline characteristics of patients included in the placebo-controlled period data set (SPIRIT-P1 and SPIRIT-P2), according to treatment group*

Characteristic	Placebo (N = 224)	IXEQ4W (N = 229)	IXEQ2W (N = 225)	Total IXE (N = 454)
Age, mean ± SD years	51.1 ± 11.33	50.9 ± 12.16	50.9 ± 12.22	50.9 ± 12.18
≥65 years, no. (%)	25 (11.2)	34 (14.8)	35 (15.6)	69 (15.2)
Sex, no. (%)				
Male	104 (46.4)	108 (47.2)	97 (43.1)	205 (45.2)
Female	120 (53.6)	121 (52.8)	128 (56.9)	249 (54.8)
White race, no. (%)†	207 (92.4)	213 (93.0)	208 (92.9)	421 (92.9)
Weight, mean ± SD kg	88.0 ± 21.58	87.8 ± 22.54	83.5 ± 19.25	85.7 ± 21.06
BMI, mean ± SD kg/m^2†	30.6 ± 7.25	30.6 ± 7.73	29.4 ± 6.70	30.0 ± 7.26
Tobacco use, no. (%)‡	72 (32.3)	89 (38.9)	86 (38.1)	175 (38.5)
cDMARD-experienced, no. (%)§	121 (54.0)	128 (55.9)	136 (60.2)	264 (58.0)
MTX at baseline, no. (%)§	99 (44.2)	105 (45.9)	114 (50.4)	219 (48.1)
Corticosteroids at baseline, no. (%)	31 (13.8)	29 (12.7)	44 (19.6)	73 (16.1)
Time since PsA onset, mean ± SD years§	10.4 ± 8.55	11.7 ± 10.21	10.9 ± 9.04	11.3 ± 9.65
Time since psoriasis onset, mean ± SD years§	17.5 ± 12.71	18.3 ± 13.17	18.6 ± 13.68	18.5 ± 13.41
Active psoriasis with BSA ≥3%, no. (%)¶	134 (65.0)	141 (67.1)	127 (63.8)	268 (65.5)
No. of tender joints (68 assessed), mean ± SD‡	21.2 ± 14.88	21.3 ± 13.88	23.4 ± 15.97	22.4 ± 14.98
No. of swollen joints (66 assessed), mean ± SD‡	10.4 ± 7.29	12.3 ± 9.90	12.9 ± 9.79	12.6 ± 9.84
CRP >6 mg/liter, no. (%)‡	122 (55.0)	129 (57.1)	107 (47.3)	236 (52.2)

* Unless indicated, data are from the safety population. BMI = body mass index; BSA = body surface area; MTX = methotrexate (see Table 1 for other definitions).
† Data are missing for some patients; the actual denominator of a particular baseline measure is the number of patients with no missing data for baseline measures.
‡ Baseline data are missing for some patients; the actual denominator of a particular baseline measure is the number of patients with no missing data for baseline measures. Data are from the intent-to-treat (ITT) population (226 for IXEQ2W and 455 for total IXE).
§ Data are from the ITT population (226 for IXEQ2W and 455 for total IXE).
¶ Among patients with plaque psoriasis in the ITT population (206 for placebo, 210 for IXEQ4W, 199 for IXEQ2W, and 409 for total IXE).

transient ischemic attack, fall, meniscus injury, acute myocardial infarction, unstable angina, coronary artery disease, lumbar spine stenosis, and osteoarthritis (IR 0.1 [95% CI 0.0–0.6]/100 patient-years each). Four deaths (IR 0.3 [95% CI 0.1–0.8]/100 patient-years) were reported (pneumonia [pathogen not identified], cerebrovascular accident, cardiorespiratory arrest, and drowning). TEAEs causing ixekizumab discontinuation included positive results of an interferon-γ release assay (IR 0.7 [95% CI 0.4–1.4]/100 patient-years); latent TB (IR 0.4 [95% CI 0.2–1.0]/100 patient-years); ISRs (IR 0.3 [95% CI 0.1–0.8]/100 patient-years); positive tuberculin test (IR 0.2 [95% CI 0.1–0.7]/100 patient-years); and pneumonia, myalgia, and cerebrovascular accident (IR 0.1 [95% CI 0.0–0.6]/100 patient-years each).

AEs of special interest. *Infections.* Placebo-controlled period. The frequency of patients with ≥1 infection-related TEAE was slightly higher in the total ixekizumab group than in the placebo group and was similar in the 2 ixekizumab dosage groups (Table 3). The most common infections occurring in ixekizumab-treated patients included upper respiratory tract infection, nasopharyngitis, and sinusitis. Six ixekizumab-treated patients (5 in the group that received ixekizumab every 2 weeks and 1 in the group that received ixekizumab once every 4 weeks) experienced ≥1 serious infection (none occurred in the placebo group); no serious infection type was experienced by >1 patient. Three patients discontinued ixekizumab because of a nonserious infection (folliculitis, subcutaneous abscess, and urinary tract infection, respectively).

The frequency of mucocutaneous *Candida* infections was higher in the ixekizumab treatment groups compared with that in the placebo group (Table 3). No *Candida* infection led to treatment discontinuation. There were no deep organ or systemic fungal infections. Predefined opportunistic infections (see Supplementary Listing 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23738/abstract) in the ixekizumab groups were oral candidiasis (0.4% in the group receiving ixekizumab every 4 weeks and 1.8% in the group receiving ixekizumab every 2 weeks), esophageal candidiasis (0.4% in the group receiving ixekizumab every 2 weeks), and unidermatomal herpes zoster (0.4% in the group receiving ixekizumab every 2 weeks).

Table 3. Frequency of adverse events in patients included in the placebo-controlled data set (SPIRIT-P1 and SPIRIT-P2), according to treatment group*

Event type or category	Placebo (n = 224)	IXEQ4W (n = 229)	IXEQ2W (n = 225)	Total IXE (n = 454)
Patients with ≥1 TEAE	127 (56.7)	153 (66.8)†	156 (69.3)†	309 (68.1)†
Mild‡	60 (26.8)	91 (39.7)	81 (36.0)	172 (37.9)
Moderate‡	63 (28.1)	54 (23.6)	61 (27.1)	115 (25.3)
Severe	4 (1.8)	8 (3.5)	14 (6.2)†	22 (4.8)
Patients discontinuing study drug because of an AE	8 (3.6)	7 (3.1)	12 (5.3)	19 (4.2)
Patients with ≥1 SAE	6 (2.7)	9 (3.9)	11 (4.9)	20 (4.4)
Deaths	0	0	0	0
Patients with ≥1 most frequent TEAE (preferred term)§				
Injection site reaction	1 (0.4)	22 (9.6)†	32 (14.2)†	54 (11.9)†
Upper respiratory tract infection	16 (7.1)	16 (7.0)	15 (6.7)	31 (6.8)
Injection site erythema	0	9 (3.9)†	17 (7.6)†	26 (5.7)†
Nasopharyngitis	9 (4.0)	15 (6.6)	7 (3.1)	22 (4.8)
Diarrhea	6 (2.7)	7 (3.1)	10 (4.4)	17 (3.7)
Headache	4 (1.8)	10 (4.4)	6 (2.7)	16 (3.5)
Sinusitis	5 (2.2)	9 (3.9)	6 (2.7)	15 (3.3)
Patients with ≥1 AE of special interest				
Hepatic¶	10 (4.5)	7 (3.1)	11 (4.9)	18 (4.0)
Infections	62 (27.7)	77 (33.6)	72 (32.0)	149 (32.8)
Serious infection	0	1 (0.4)	5 (2.2)†	6 (1.3)
Candida infection#	1 (0.4)	4 (1.7)	8 (3.6)†	12 (2.6)
Esophageal candidiasis	0	0	1 (0.4)	1 (0.2)
Active tuberculosis	0	0	0	0
Latent tuberculosis**	0	0	0	0
Injection site reactions‡‡	10 (4.5)	40 (17.5)†	57 (25.3)§§	97 (21.4)†
Allergic reaction/hypersensitivity	4 (1.8)	10 (4.4)	14 (6.2)†	24 (5.3)†
Confirmed cerebrocardiovascular event	2 (0.9)	0	0	0†
Confirmed MACE event	0	0	0	0
Malignancy	0	2 (0.9)	0	2 (0.4)
Depression-related	3 (1.3)	4 (1.7)	4 (1.8)	8 (1.8)
Inflammatory bowel disease (narrow and broad terms) ¶¶	0	0	1 (0.4)	1 (0.2)
Crohn's disease	0	0	0	0
Ulcerative colitis	0	0	0	0

* Values are the number (%). TEAE = treatment-emergent adverse event (AE); SAE = serious AE; MACE = major adverse cardiovascular event (see Table 1 for other abbreviations).
† $P \leq 0.05$ vs. placebo, by Cochran-Mantel-Haenszel test.
‡ Comparisons between treatment and placebo were not performed.
§ AEs are listed according to the preferred terms (PTs) in Medical Dictionary for Regulatory Activities (MedDRA). Shown are AEs occurring in ≥3% of patients in the combined (total) ixekizumab group.
¶ Potentially drug-related hepatic disorders using the MedDRA PTs contained any of the following MedDRA Queries: broad and narrow terms in the liver-related investigations, signs and symptoms; broad and narrow terms in the cholestasis and jaundice of hepatic origin; broad and narrow terms in the hepatitis, non-infectious; broad and narrow terms in the hepatic failure, fibrosis, cirrhosis, and other liver damage; narrow terms in the liver-related coagulation and bleeding disturbances.
# Patients with ≥1 TEAE of *Candida* infections (high-level plus clinical terms). The esophageal candidiasis case was an SAE.
** Two ixekizumab-treated patients had positive results of an interferon-γ (IFNγ) release assay. At screening, these patients had negative results using a purified protein derivative test, which later appeared to be poorly documented; this prompted the study site to perform an IFNγ release assay, which appeared to be positive (in the absence of any baseline IFNγ release assay).
‡‡ Composite of several injection site reaction–related terms.
§§ $P \leq 0.05$, IXE vs. placebo and IXEQ2W vs. IXEQ4W, by Cochran-Mantel-Haenszel test.
¶¶ Composite of inflammatory bowel disease (IBD; narrow terms) and events that can occur with IBD (broad terms).

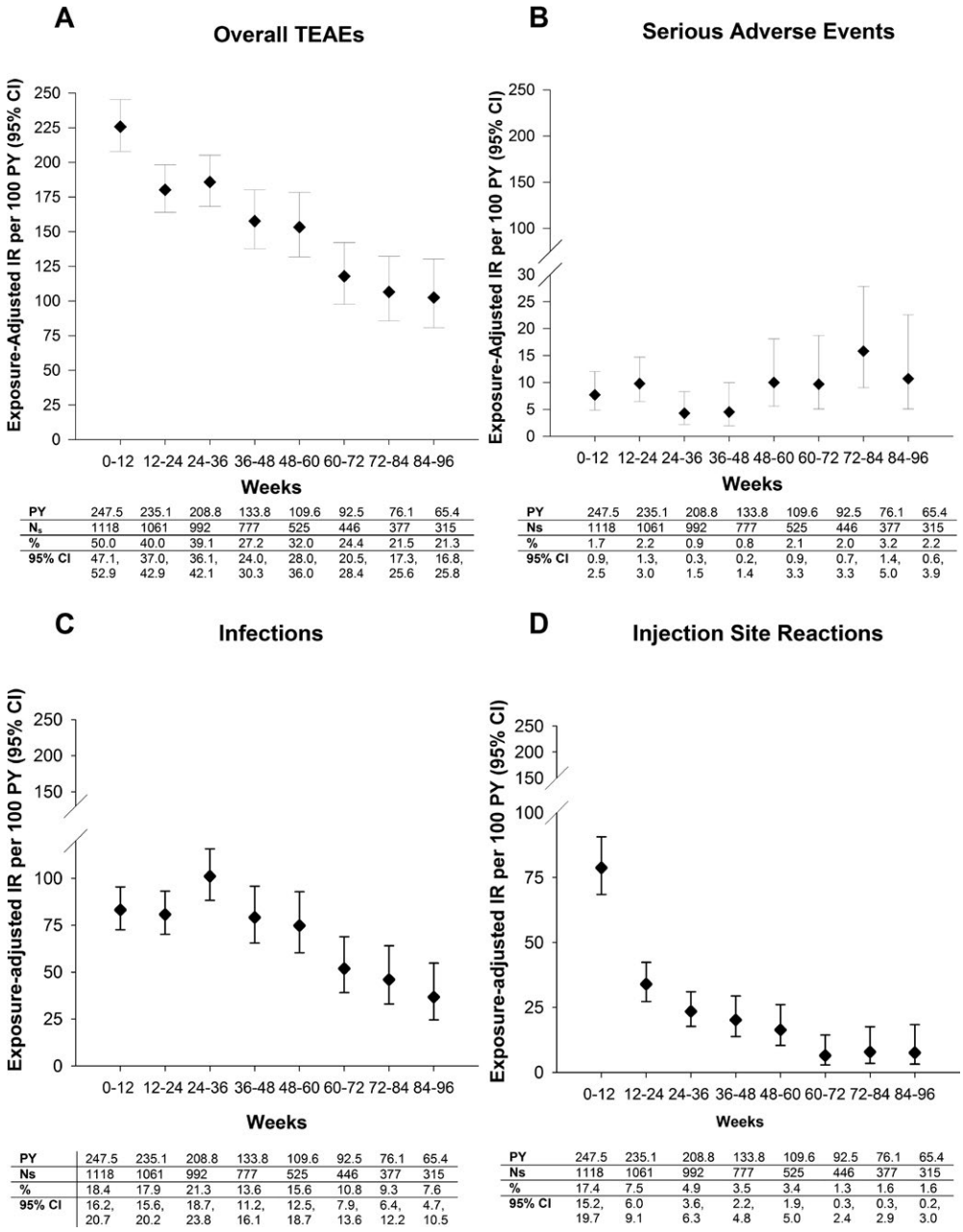


Figure 1. Exposure-adjusted incidence rates (IRs) of treatment-emergent adverse events (TEAEs) at 12-week intervals (week 0 to week 96) in the all ixekizumab-treated data set (SPIRIT-P1, SPIRIT-P2, and SPIRIT-P3). **A,** Overall TEAEs. **B,** Serious AEs. **C,** Infections (System Organ Class). **D,** Injection site reactions (this is a composite of several injection site reaction–related preferred terms). The 95% confidence intervals (95% CIs) for the IRs are from likelihood ratio test of treatment effect from the Poisson regression model. AEs were coded using Medical Dictionary for Regulatory Activities version 19.1. Values are from a binomial model. PY = patient-year; Ns = number of patients entered in each time interval.

A 65-year old woman (in the group receiving ixekizumab every 2 weeks) who had been taking oral prednisone for >2 years was hospitalized on day 7 of ixekizumab treatment because of esophageal candidiasis. She was successfully treated with a 16-day course of oral fluconazole. Endoscopy was not performed. Ixekizumab dosing in this patient was interrupted and resumed after resolution of the infection.

All ixekizumab-treated. There was no increase in the IRs of infection-related TEAEs with increasing durations of ixekizumab exposure (Figure 1C). Ten TEAEs (IR 0.7 [95% CI 0.4–1.4]/100 patient-years) of latent TB were reported. The incidence of serious infections was low (IR 1.2 [95% CI 0.7–1.9]/100 patient-years). Serious infections occurring in >1 patient were pneumonia (IR 0.2 [95% CI 0.1–0.7]/100 patient-years) and latent TB (considered

serious due to testing), lower respiratory tract infection, and esophageal candidiasis (IR 0.1 [95% CI 0.0–0.6]/100 patient-years each).

Two patients reported SAEs of esophageal candidiasis (one reported during the placebo-controlled period and described above). In the other patient (a 48-year-old man), esophageal candidiasis led to hospitalization. Endoscopy showed esophageal mucosa covered in its majority by whitish exudate. The patient was treated with fluconazole. At a follow-up visit after hospital discharge, the investigator maintained treatment with ixekizumab and methotrexate without lowering the dosage; the outcome of the infection had not been reported at the time of database lock. The 2 esophageal candidiasis cases were the only serious *Candida* infections in the all ixekizumab-treated data sets. No treatment-emergent *Candida* infection resulted in ixekizumab discontinuation. There were no reports of deep organ or bloodstream *Candida* infections. Among predefined opportunistic infections (see Supplementary Listing 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23738/abstract), only mucocutaneous *Candida* infections and localized herpes zoster infection were reported. There were no disseminated or central nervous system herpes simplex infections, endemic mycoses, invasive aspergillosis, or other deep fungal infections.

Eighteen patients (IR 1.3 [95% CI 0.8–2.1]/100 patient-years) discontinued ixekizumab because of infections. Of these, 6 patients (IR 0.4 [95% CI 0.2–1.0]/100 patient-years) and 2 patients (IR 0.1 [95% CI 0.0–0.6]/100 patient-years) discontinued because of latent TB and pneumonia (1 case of pneumonia was fatal), respectively. For all patients who discontinued due to latent TB, the screening test was negative and patients tested positive during protocol-required yearly testing. There were no cases of active TB. One patient with a history of hepatitis B virus (HBV) infection discontinued the study drug, as required by protocol, when a protocol-required test for HBV DNA was positive according to the local laboratory despite a result below the level of quantification. Three weeks later, testing was negative for HBV DNA as well as on 3 subsequent samples. The remaining treatment discontinuations were due to bacterial arthritis, bronchitis, nasopharyngitis, staphylococcal infection, tooth abscess, and tonsillitis, in addition to the folliculitis, subcutaneous abscess, and urinary tract infection described above.

Neutropenia. Placebo-controlled period. The percentage of ixekizumab-treated patients with a treatment-emergent low absolute neutrophil count (<2.0 × 10^9/liter) was 9.7% (43 of 444) versus 2.7% (6 of 219) of placebo-treated patients (P = 0.001). The frequencies of grade 1 or worse and grade 2 or worse neutropenia were higher in ixekizumab-treated patients than those in placebo-treated patients. (see Supplementary Table 3, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23738/abstract). No grade 3 or grade 4 neutropenia was reported. Approximately 50% of ixekizumab-treated patients (and 83% of placebo-treated patients) who had a treatment-emergent low absolute neutrophil count at an earlier visit had values within the normal range at all subsequent visits.

All ixekizumab-treated. The percentage of patients with ≥1 treatment-emergent low neutrophil count (<2.0 × 10^9/liter) was 14.5% (158 of 1,092). Fifty percent of ixekizumab-treated patients with treatment-emergent low absolute neutrophil counts at an earlier visit had values within the normal range at all subsequent visits. Four patients had grade 3 neutropenia that improved to a normal or baseline grade at subsequent visits during which patients were receiving treatment, without reoccurrence of grade 3 neutropenia during the observation period. No patient developed grade 4 neutropenia. There was no tendency for an increase or a decrease in the IR of treatment-emergent neutropenia with increased ixekizumab exposure (data not shown).

The frequency of infection-related TEAEs was comparable in patients with neutrophil counts <2.0 × 10^9/liter (54.4%) versus patients whose counts were above this cutoff (49.0%) and in patients with neutrophil counts <1.5 × 10^9/liter (42.0%) versus patients whose counts were above the latter (<1.5 × 10^9/liter) cutoff (50.1%). Seventeen patients (IR 1.2 [95% CI 0.8–2.0]/100 patient-years) had ≥1 infection-related TEAE that was temporally associated with grade 1 neutropenia; the infection-related TEAEs occurring in >1 patient were nasopharyngitis (IR 0.3 [95% CI 0.1–0.8]/100 patient-years), oral herpes (IR 0.1 [95% CI 0.0–0.6]/100 patient-years), and upper respiratory tract infection (IR 0.1 [95% CI 0.0–0.6]/100 patient-years). No patients had infections that were temporally associated with grade 2, 3, or 4 neutropenia.

ISRs. Placebo-controlled period. ISRs were common in ixekizumab-treated patients and occurred less frequently in the group receiving ixekizumab every 4 weeks than in the group receiving ixekizumab every 2 weeks (Table 3). The most common types of reported ISRs were injection site reaction, injection site erythema, and injection site hypersensitivity. Most ISRs were mild or moderate in severity. No serious ISRs were reported. Five ixekizumab-treated patients (1.1% versus 0.4% of placebo-treated patients) discontinued treatment because of ISRs (4 in the group receiving ixekizumab every 2 weeks and 1 in the group receiving ixekizumab every 4 weeks).

All ixekizumab-treated. When the incidence of ISRs was analyzed at 12-week intervals (week 0 to week 96), the incidence was shown to decrease substantially over time (Figure 1D). The most common types of reported ISRs were injection site reaction, injection site erythema, and injection site pain. There were 4.0 ISRs per 100 active injections. Injection site pain, injection site swelling, and injection site discoloration were the first types of ISRs to appear, with median onset times of 1, 2, and 2 days, respectively. The median duration of ISRs was 3 days. Injection site pain and injection site discoloration had the shortest median duration (1 day), whereas injection site hematoma (7 days) and injection site warmth (13 days) had the longest median duration.

There were no serious ISRs; 6 patients (IR 0.4 [95% CI 0.2–1.0]/100 patient-years) discontinued treatment because of ISRs. There was no clear association between the development of treatment-emergent antidrug antibodies and ISRs (see Supplementary Table 4 and Supplementary Text 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23738/abstract).

Hypersensitivity events. Placebo-controlled period. The frequencies of hypersensitivity events in the ixekizumab and placebo treatment groups were 5.3% and 1.8%, respectively (see Supplementary Listing 3, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23738/abstract). No SAEs of hypersensitivity or anaphylaxis were reported. Two patients treated with ixekizumab every 4 weeks discontinued treatment because of hypersensitivity (hypersensitivity and pruritic rash).

All ixekizumab-treated. When hypersensitivity was analyzed at 12-week intervals (week 0 to week 96, the IR of hypersensitivity events did not increase with increasing duration of ixekizumab exposure (Figure 2A). No anaphylaxis was reported, but 1 patient (IR 0.1 [95% CI 0.0–0.5]/100 patient-years) experienced an SAE of angioedema. Six patients (IR 0.4 [95% CI 0.2–1.0]/100 patient-years) (including the patient with angioedema) discontinued treatment due to hypersensitivity (drug eruption, rash, pruritic rash, solar urticaria, hypersensitivity). There was no clear association between the development of treatment-emergent antidrug antibodies and hypersensitivity events (see Supplementary Table 5 and Supplementary Listing 3, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23738/abstract).

Inflammatory bowel disease. Placebo-controlled. No cases of IBD, ulcerative colitis, or Crohn's disease were explicitly reported (Table 3). Supplementary Listing 4, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23738/abstract, shows preferred terms for inflammatory bowel disease. However, in reviewing other reported terms potentially representing IBD, 1 patient (in the group receiving ixekizumab every 2 weeks) who had no history of IBD developed an anal abscess and an anal fistula. These were SAEs but did not lead to discontinuation of ixekizumab treatment. The abscess resolved, and the patient underwent surgery for the fistula. The sponsor assessed that this patient had IBD.

All ixekizumab-treated. One case (IR 0.1 [95% CI 0.0, 0.5]/100 patient-years) each of CD and UC were reported. These were SAEs but did not lead to discontinuation of ixekizumab treatment. Although the patient with Crohn's disease was reported as a new case, that patient had a history of irritable bowel syndrome. Of the 12 patients with preexisting conditions or a history of illness suggestive of IBD, none experienced disease exacerbation during ixekizumab treatment, and none used steroids during these studies for prior conditions related to IBD, but 1 patient used mesalazine.

Other adverse events of special interest. Placebo-controlled. No cases of CEC-confirmed MACE were reported in any treatment group. Malignancies developed in 2 ixekizumab-treated patients (0.4%) and 0 placebo-treated patients (Table 3), including 1 basal cell carcinoma and 1 prostate cancer (SAE causing discontinuation of treatment). Eight (1.8%) ixekizumab-treated patients (3 [1.3%] placebo) experienced ≥1 depression-related event (Table 3). None were SAEs. One ixekizumab-treated patient discontinued treatment because of depression; this patient (in the group receiving ixekizumab every 2 weeks) had a history of depression and had been treated with venlafaxine.

All ixekizumab-treated. No TEAEs of uveitis were reported. Nine patients (IR 0.7 [95% CI 0.3–1.3]/100 patient-years) had CEC-confirmed MACE (2 vascular deaths, 3 nonfatal myocardial infarctions, 4 nonfatal strokes). The MACE incidence did not increase with longer ixekizumab exposure (Figure 2B). Nine patients (IR 0.7 [95% CI 0.3–1.3]/100 patient-years) developed ≥1 malignancy; among these 9 patients, 5 had non-melanoma skin cancer. The remaining 4 patients had breast cancer (n = 2), prostate cancer (n = 1), and papillary thyroid cancer (n = 1); these were SAEs causing discontinuation of treatment. There was no increase in the malignancy rate with longer ixekizumab exposure (Figure 2C). There was no evidence of an increase in depression-related events over time (Figure 2D). The incidence of depression-related events was 1.7 (95% CI 1.2–2.6)/100 patient-years. None of these events was serious, and only the ixekizumab-treated patient discontinued treatment due to depression. No suicides or self-injury–related behaviors were reported. No patient met the laboratory criteria for potential drug-induced liver injury.

DISCUSSION

In the current study, we report safety results for 1,373.4 patient-years of ixekizumab exposure in 1,118 patients with PsA. The rates of TEAEs and SAEs did not increase with longer ixekizumab exposure. During the placebo-controlled period, patients in the ixekizumab group experienced higher rates of overall TEAEs and individual types of TEAEs (serious infections, mucocutaneous *Candida* infection, hypersensitivities [non-anaphylactic], and ISRs) than the placebo group. These results are consistent with the known safety profile of ixekizumab in patients with moderate-to-severe psoriasis (9).

The use of immunomodulation therapies may be associated with an increased risk of infection (20–22). As shown in the current study, during ixekizumab treatment, the rates of patients with infections did not increase with longer ixekizumab exposure, and the rate of serious infections was low (IR 1.2 [95% CI 0.7–1.9]/100 patient-years). For comparison, among patients with confirmed PsA, the rates of serious infection per 100 patient-years were 1.06 in those treated with ustekinumab, 2.83 in patients treated with infliximab, 2.58 in those treated with adalimumab/etanercept, and 1.63 in those treated with nonbiologics (22). TEAEs of infections were upper respiratory tract and other common types of infection.

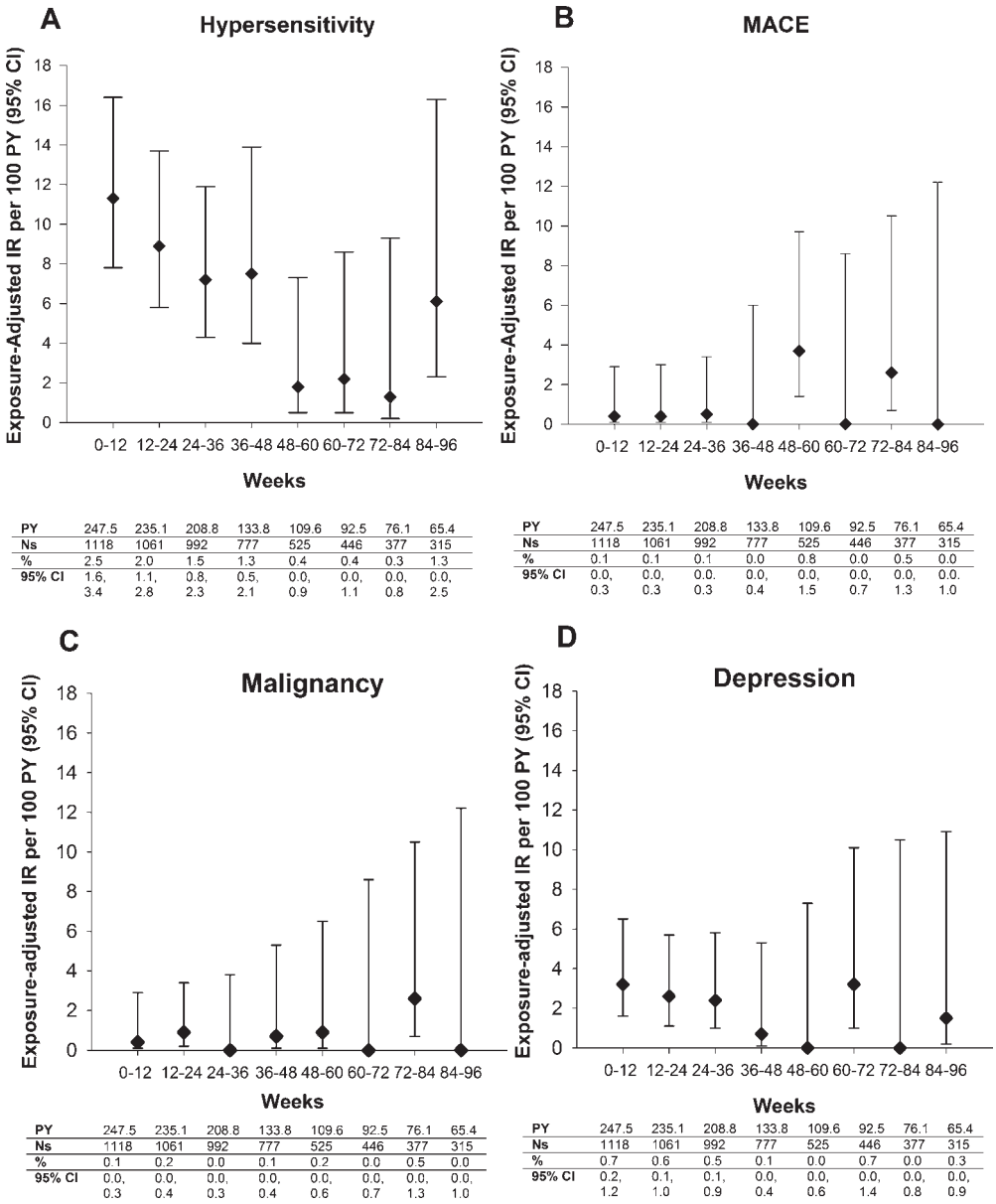


Figure 2. Exposure-adjusted incidence rate of TEAEs at 12-week intervals (week 0 to week 96) in the all ixekizumab-treated data set (SPIRIT-P1, SPIRIT-P2, SPIRIT-P3). **A,** Hypersensitivity. **B,** Major adverse cardiovascular events (MACE) (Clinical Events Committee–adjudicated). **C,** Malignancy. **D,** Depression-related. The 95% CIs for the IRs are from likelihood ratio test of treatment effect from the Poisson regression model. AEs were coded using Medical Dictionary for Regulatory Activities version 19.1. Values are from a binomial model. See Figure 1 for other definitions.

Opportunistic infections were limited to oral and esophageal *Candida* infections (as expected, based on the known role of IL-17A in host defense against these infections) (10–13) and localized herpes zoster. The lack of deep organ or bloodstream *Candida* infections is consistent with published cases of IL-17 deficiency (10,11,13). Two serious *Candida* infections (esophageal) were reported, but these events did not lead to ixekizumab discontinuation, nor did any other *Candida* infection. *Candida* infections resolved or were being treated at the time of database locking. There were no reports of endemic mycoses, invasive aspergillosis, or other deep fungal infections. Several patients had a TEAE of latent TB, as detected through protocol-required yearly testing; there were no cases of active TB.

Because IL-17A antagonism is involved in neutrophil trafficking and granulopoiesis (23), the observed decreases in neutrophil counts with ixekizumab versus placebo were expected; however, no patient had grade 4 neutropenia. Grade 2 and grade 3 neutropenia were not temporally associated with infections. Neutropenia rates did not increase with longer ixekizumab exposure. Patients with PsA who were enrolled in ixekizumab treatment studies were allowed to be treated with concomitant cDMARDs. Conventional DMARDs such as methotrexate can cause bone

marrow suppression (24), which may have contributed to the neutropenia observed in ixekizumab-treated patients; however, the presence of cDMARDs does not explain differences between the ixekizumab-treated patients and the placebo-treated patients.

The higher frequency of TEAEs in the ixekizumab group compared with the placebo group was mainly attributable to a higher frequency of patients with ≥1 ISR. Although ISRs were common during the placebo-controlled period, the frequencies of patients with ISRs decreased substantially with longer ixekizumab exposure. ISRs were well tolerated and did not typically lead to discontinuation. Likewise, the incidence of discontinuations due to hypersensitivity was only 0.4 (95% CI 0.2–1.0)/100 patient-years; anaphylaxis was not reported. Aggregate findings for patients with treatment-emergent antidrug antibody positivity do not support a clear relationship between hypersensitivity and immunogenicity and between ISRs and immunogenicity.

There are potential concerns regarding new or exacerbated IBD in patients treated with IL-17A blockers (14,15). In the current study, the incidences of IBD (including Crohn's disease and ulcerative colitis) were low (IR 0.1 [95% CI 0.0–0.5]/100 patient-years) for both Crohn's disease and ulcerative colitis and IR 0.1 (95% CI 0.0–0.6)/patient years for unspecified IBD. Similarly, the incidences of patients with other AESIs, including depression, confirmed MACE, and malignancies, were low and did not increase substantially with longer ixekizumab exposure. Additionally, the types of MACE did not change with longer ixekizumab exposure, and the incidence of MACE was consistent with background rates among patients with PsA (25,26). No self-injury–related events or suicides were reported.

A strength of this study is inclusion of 2 data sets. The placebo-controlled period data set allows direct comparison of ixekizumab with placebo for 24 weeks, whereas the all ixekizumab-treated data set is larger with an aggregate exposure of 1,373.4 patient-years, thus allowing detection of less frequent AEs and allowing assessment of IRs of AEs over time. Limitations include the lack of a comparator group past 24 weeks and incomplete patient follow-up at the database locks used for the purpose of this study; updated data will be reported periodically.

In conclusion, ixekizumab-treated patients had higher rates of overall TEAEs, serious infections, mucocutaneous *Candida* infection, hypersensitivities (non-anaphylactic), and ISRs compared with patients treated with placebo. However, the safety profile of ixekizumab in patients with PsA is consistent with the mechanism of IL-17A antagonism and with the known safety profile of ixekizumab in the treatment of patients with plaque psoriasis (7–9).

ACKNOWLEDGMENTS

The authors would like to thank the patients, their families, and the study personnel who participated in these clinical trials. In addition, the authors would like to thank Lotus Mallbris, MD, PhD (Eli Lilly) for her critical review of this manuscript and Sandra Garces, MD, PhD (Eli Lilly) for reviewing the revised manuscript.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Mease 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. Mease, Moriarty, Adams, Xu.
Acquisition of data. Mease, Roussou, Burmester, Goupille, Nash.
Analysis and interpretation of data. Mease, Burmester, Goupille, Gottlieb, Moriarty, Benichou, Adams, Xu, Nash.

ROLE OF THE STUDY SPONSOR

Eli Lilly and Company facilitated the study design, performed the statistical analyses, provided writing assistance for the manuscript, and reviewed and approved the manuscript prior to submission. The authors independently collected the data, interpreted the results, and had the final decision to submit the manuscript for publication. Writing assistance was provided by Matthew Hufford, PhD (Eli Lilly and Company) and Lori Kornberg, PhD (Syneos Health, Raleigh, NC). Publication of this article was not contingent upon approval by Eli Lilly and Company.

REFERENCES

1. Raychaudhuri SP. Role of IL-17 in psoriasis and psoriatic arthritis. Clin Rev Allergy Immunol 2013;44:183–93.
2. McInnes IB, Mease PJ, Kirkham B, Kavanaugh A, Ritchlin CT, Rahman P, et al. Secukinumab, a human anti-interleukin-17A monoclonal antibody, in patients with psoriatic arthritis (FUTURE 2): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet 2015;386:1137–46.
3. Mease PJ, McInnes IB, Kirkham B, Kavanaugh A, Rahman P, van der Heijde D, et al. Secukinumab inhibition of interleukin-17A in patients with psoriatic arthritis. N Engl J Med 2015;373:1329–39.
4. Liu L, Lu J, Allan BW, Tang Y, Tetreault J, Chow CK, et al. Generation and characterization of ixekizumab, a humanized monoclonal antibody that neutralizes interleukin-17A. J Inflamm Res 2016;9:39–50.
5. Mease PJ, van der Heijde D, Ritchlin CT, Okada M, Cuchacovich RS, Shuler CL, et al. Ixekizumab, an interleukin-17A specific monoclonal antibody, for the treatment of biologic-naive patients with active psoriatic arthritis: results from the 24-week randomised, double-blind, placebo-controlled and active (adalimumab)-controlled period of the phase III trial SPIRIT-P1. Ann Rheum Dis 2017;76:79–87.
6. Nash P, Kirkham B, Okada M, Rahman P, Combe B, Burmester GR, et al. Ixekizumab for the treatment of patients with active psoriatic arthritis and an inadequate response to tumour necrosis factor inhibitors: results from the 24-week randomised, double-blind, placebo-controlled period of the SPIRIT-P2 phase 3 trial. Lancet 2017;389:2317–27.
7. Gordon KB, Blauvelt A, Papp KA, Langley RG, Luger T, Ohtsuki M, et al. Phase 3 trials of ixekizumab in moderate-to-severe plaque psoriasis. N Engl J Med 2016;28:345–56.
8. Griffiths CE, Reich K, Lebwohl M, van de Kerkhof P, Paul C, Menter A, et al. Comparison of ixekizumab with etanercept or placebo in moderate-to-severe psoriasis (UNCOVER-2 and UNCOVER-3): results from two phase 3 randomised trials. Lancet 2015;386:541–51.
9. Strober B, Leonardi C, Papp KA, Mrowietz U, Ohtsuki M, Bissonnette R, et al. Short- and long-term safety outcomes with ixekizumab from

7 clinical trials in psoriasis: etanercept comparisons and integrated data. J Am Acad Dermatol 2017;76:432–440.e.17.
10. Cypowyj S, Picard C, Maródi L, Casanova JL, Puel A. Immunity to infection in IL 17-deficient mice and humans. Eur J Immunol 2012;42:2246–54.
11. Gaffen SL, Hernández-Santos N, Peterson AC. IL-17 signaling in host defense against Candida albicans. Immunol Res 2011;50:181–7.
12. Peck A, Mellins ED. Precarious balance: Th17 cells in host defense. Infect Immun 2010;78:32–8.
13. Puel A, Cypowyj S, Bustamante J, Wright JF, Liu L, Lim HK, et al. Chronic mucocutaneous candidiasis in humans with inborn errors of interleukin-17 immunity. Science 2011;332:65–8.
14. Hueber W, Sands BE, Lewitzky S, Vandemeulebroecke M, Reinisch W, Higgins PD, et al. Secukinumab, a human anti-IL-17A monoclonal antibody, for moderate to severe Crohn's disease: unexpected results of a randomised, double-blind placebo-controlled trial. Gut 2012;61:1693–700.
15. Mozaffari S, Nikfar S, Abdollahi M. Inflammatory bowel disease therapies discontinued between 2009 and 2014. Expert Opin Investig Drugs 2015;24:949–56.
16. Boyman O, Comte D, Spertini F. Adverse reactions to biologic agents and their medical management. Nat Rev Rheumatol 2014;10:612–27.
17. Molinelli E, Campanati A, Ganzetti G, Offidani A. Biologic therapy in immune mediated inflammatory disease: basic science and clinical concepts. Curr Drug Saf 2016;11:35–43.
18. Ruderman EM. Overview of safety of non-biologic and biologic DMARDs. Rheumatology (Oxford) 2012;51:vi37–43.
19. Van de Kerkhof PC, Griffiths CE, Reich K, Leonardi CL, Blauvelt A, Tsai TF, et al. Secukinumab long-term safety experience: a pooled analysis of 10 phase II and III clinical studies in patients with moderate to severe plaque psoriasis. J Am Acad Dermatol 2016;75:83–98.
20. Murdaca G, Spano F, Contatore M, Guastalla A, Penza E, Magnani O, et al. Infection risk associated with anti-TNF-α agents: a review. Expert Opin Drug Saf 2015;14:571–82.
21. Selmi C, Ceribelli A, Naguwa SM, Cantarini L, Shoenfeld Y. Safety issues and concerns of new immunomodulators in rheumatology. Expert Opin Drug Saf 2015;14:389–99.
22. Ritchlin CT, Gottlieb AB, Menter A, Mease PJ, Kalia S, Kerdel F, et al. Serious infections in psoriasis patients with psoriatic arthritis in the Psoriasis Longitudinal Assessment and Registry Study [abstract]. Arthritis Rheumatol 2015;67 Suppl:S1692.
23. Krstic A, Mojsilovic S, Jovcic G, Bugarski D. The potential of interleukin-17 to mediate hematopoietic response. Immunol Res 2012;52:34–41.
24. Weinblatt ME. Toxicity of low dose methotrexate in rheumatoid arthritis. J Rheumatol Suppl 1985;12 Suppl 12:35–9.
25. Li L, Hagberg KW, Peng M, Shah K, Paris M, Jick S. Rates of cardiovascular disease and major adverse cardiovascular events in patients with psoriatic arthritis compared to patients without psoriatic arthritis. J Clin Rheumatol 2015;21:405–10.
26. Ogdie A, Yu Y, Haynes K, Love TJ, Maliha S, Jiang Y, et al. Risk of major cardiovascular events in patients with psoriatic arthritis, psoriasis and rheumatoid arthritis: a population-based cohort study. Ann Rheum Dis 2015;74:326–32.
27. Coates LC, Fransen J, Helliwell PS. Defining minimal disease activity in psoriatic arthritis: a proposed objective target for treatment. Ann Rheum Dis 2010;69:48–53.

Arthritis Care & Research
Vol. 71, No. 3, March 2019, pp 379–384
DOI 10.1002/acr.23605

AMERICAN COLLEGE OF RHEUMATOLOGY
EDUCATION • TREATMENT • RESEARCH

Impact of Information Presentation Format on Preference for Total Knee Replacement Surgery

Liana Fraenkel,[1] W. Benjamin Nowell,[2] Christine E. Stake,[2] Shilpa Venkatachalam,[2] Rachel Eyler,[3] George Michel,[4] and Ellen Peters[5]

Objective. Patients have a poor understanding of outcomes related to total knee replacement (TKR) surgery, with most patients underestimating the potential benefits and overestimating the risk of complications. In this study, we sought to compare the impacts of descriptive information alone or in combination with an icon array, experience condition (images), or spinner on participants' preference for TKR.

Methods. A total of 648 members of an online arthritis network were randomized to 1 of 4 outcome presentation formats: numeric only, numeric with an icon array, numeric with a set of 50 images, or numeric with a functional spinner. Preferences for TKR were measured before and immediately after viewing the outcome information using an 11-point numeric rating scale. Knowledge was assessed by asking participants to report the frequency of each outcome.

Results. Participants randomized to the icon array, images, and spinner had stronger preferences for TKR (after controlling for baseline preferences) compared to those viewing the numeric only format ($P < 0.05$ for all mean differences). Knowledge scores were highest in participants randomized to the icon array; however, knowledge did not mediate the association between format and change in preference for TKR.

Conclusion. Decision support at the point-of-care is being increasingly recognized as a vital component of care. Our findings suggest that adding graphic information to descriptive statistics strengthens preferences for TKR. Although experience formats using images may be too complex to use in clinical practice, icon arrays and spinners may be a viable and easily adaptable decision aid to support communication of probabilistic information.

INTRODUCTION

Total knee replacement (TKR) surgery is a cost-effective treatment option for patients with knee arthritis who continue to have pain and disability despite medical management (1). Most patients have an excellent clinical response; however, a significant number continue to have pain after the procedure (2,3). Patient satisfaction after TKR ranges from 75% to >90% (4,5) and is strongly associated with presurgery expectations in addition to postoperative pain and function (6). Patients are likely to first hear about TKR from their primary care providers, and ideally would then have gained an accurate understanding of their chances of having an excellent, moderate, or bad outcome following surgery. Both qualitative and quantitative studies suggest, however, that patients have a poor understanding of outcomes related to TKR, with most underestimating the potential benefits and overestimating the risk of complications (7–10).

Numerous studies have documented the difficulties associated with communicating probabilistic information (11–13). Graphics, such as bar graphs and icon arrays, have been shown to improve understanding of probabilities and are recommended to support patient decision-making (14). Yet even with these aids, comprehension remains poor among many patients (15). There is, therefore, a need to develop novel approaches to ensure effective communication and high-quality decision-making.

The contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health or the National Science Foundation.

Dr. Fraenkel's work was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases/NIH (grant AR-060231-06). Dr. Peters' work was supported by the National Science Foundation (SES-1558230).

[1]Liana Fraenkel, MD, MPH: Yale University School of Medicine, New Haven, and Veterans Administration Connecticut Healthcare System, West Haven, Connecticut; [2]W. Benjamin Nowell, PhD, Christine E. Stake, DHA, Shilpa Venkatachalam, PhD: CreakyJoints, Global Healthy Living Foundation, Upper Nyack, New York; [3]Rachel Eyler, PharmD: University of Connecticut School of Pharmacy, Storrs; [4]George Michel, MS: Yale University School of Medicine, New Haven, Connecticut; [5]Ellen Peters, PhD: Ohio State University, Columbus.

Address correspondence to Liana Fraenkel, MD, MPH, Yale University School of Medicine, Section of Rheumatology, 300 Cedar Street, TAC Building, RM #525, P.O. Box 208031, New Haven, CT 06520-8031. E-mail: liana.fraenkel@yale.edu.

Submitted for publication February 12, 2018; accepted in revised form May 22, 2018.

SIGNIFICANCE & INNOVATIONS

- Adding graphic information (in the form of icons, photographs, or a spinner) to descriptive statistics strengthens patients' preferences for total knee replacement surgery. However, the format of information did not influence patients who had a strong preference for total knee replacement at baseline.
- Visual aids may help improve communication of probabilistic information in clinical practice.

Although the most common method of communicating probabilistic information is through descriptive formats (i.e., using words or numbers to describe probabilities), risk information can also be communicated through experience. In a seminal study, Hertwig et al (16) found that participants learning about probabilities by receiving feedback while performing a sampling task underweighted the probability of a rare outcome compared to those receiving a description of the probability of each outcome, presumably because they are less likely to encounter rare outcomes. The potential use of experience as a decision aid has been recently examined in health care decisions. In a study by Tyszka and Sawicki (17), students were presented with a hypothetical scenario in which they were asked to consider prenatal genetic testing. Participants experienced risk by viewing a series of photographs of children with and those without Down syndrome. The authors found that worry about the genetic disease was lower among those randomized to the experience versus description format. Similarly, Wegier and Shaffer (18) found that participants' estimates of the positive predictive value of prenatal screening for Down syndrome were more accurate among those viewing simulated test results using a series of grids compared to those given explicit statistics. In contrast, our group (19) did not find any difference in preferences for lung cancer screening among participants in a pulmonary practice randomized to an experience task (composed of 250 computed tomography scan images representing the expected number of normal scans, false positive lung nodules, cancers leading to a life saved, and cancers leading to deaths despite screening) versus those receiving descriptive statistics.

Eyler et al (20) recently examined whether a spinner (an arrow in the center of a donut-shaped ring with a colored segment representing the risk of an adverse event) could be used to facilitate risk communication. Spinners might improve understanding of probabilistic events by displaying proportions in a continuous format (21). Unlike pie charts, spinners allow participants to experience chance by observing varying outcomes with each spin. In that study, the authors found that knowledge scores were higher among patients (recruited from outpatient medicine clinics) randomized to the spinner format compared to those receiving numeric information only. Participants also preferred the spinner format over the descriptive statistical information.

In the current experimental study, we sought to compare the impact of descriptive information alone or in combination with an icon array, experience condition, and spinner on participants' preference for TKR. Given previous data demonstrating that participants are likely to consider new information only when they do not have strong baseline preferences (either for or against a specific treatment) (22), we also examined the effect of the information format among subgroups of participants with varying baseline preferences for TKR.

SUBJECTS AND METHODS

English-speaking patients were recruited via email invitations with a unique survey link to CreakyJoints members. CreakyJoints (https://creakyjoints.org/) is an online community of more than 100,000 patients with arthritis that supports education, advocacy, and patient-centered research. CreakyJoints members were identified as eligible to participate if they were age ≥50 years, living in the US, had self-reported physician-diagnosed rheumatoid arthritis, psoriatic arthritis, and/or osteoarthritis involving 1 or both knees, and had not had previous total hip or knee replacement surgery. Individuals who received the invitation email were already CreakyJoints members or existing fans of the CreakyJoints Facebook page. To prevent patients from signing up to take the survey more than once, unique survey links were generated, each of which could be used only once by the individual who received the email invitation. We report both the open rate (i.e., opening the email invitation) and click rate (i.e., clicking on the link to open the survey). This study was classified as exempt from institutional board review by the Yale Human Subjects Research Program.

Risk formats. After collecting demographic data (Table 1), we presented participants with the following information: "Total knee replacement surgery is an option for patients with arthritis who continue to have significant knee pain despite having tried physical therapy, medications and injections. If you are (or were to become) someone with this type of knee pain, please indicate how you feel about this surgery on the scale below." Baseline preference for TKR was measured using an 11-point numeric rating scale, where 0 = I am certain that I would not want total knee replacement surgery, 5 = unsure, and 10 = I am certain that I would want total knee replacement surgery.

Numeric information (23,24) (viewed by all participants) read as follows: "One of the most common questions patients ask is what should I expect after total knee replacement surgery? There are 3 possibilities: 1) Most patients (about 42 in 50) do great. They have significant pain relief and are very

Table 1. Respondent characteristics*

	Numeric (n = 154)	Icon array (n = 176)	Images (n = 157)	Spinner (n = 161)	*P*
Age, mean ± SD years	61.5 ± 7.2	61.6 ± 6.9	61.6 ± 6.7	62.2 ± 6.6	0.8
Women	142 (92.2)	163 (92.6)	142 (90.5)	153 (95.0)	0.5
White	134 (87.0)	154 (87.5)	136 (86.6)	145 (90.1)	0.8
College graduate	64 (22.3)	74 (26.0)	73 (25.6)	74 (26.0)	0.7
Married	87 (56.5)	99 (56.3)	88 (56.1)	88 (54.7)	1.0
Employed	54 (35.1)	65 (36.9)	44 (28.0)	52 (32.3)	0.3
Private insurance	68 (44.2)	75 (42.6)	65 (41.4)	55 (34.2)	0.3
Fair or poor health status	50 (32.5)	57 (32.4)	55 (35.0)	56 (34.8)	0.9
Pain NRS, mean ± SD	6.27 ± 2.4	6.64 ± 2.4	6.39 ± 2.4	6.56 ± 2.2	0.5
Osteoarthritis	84 (54.6)	97 (55.1)	94 (59.9)	96 (59.6)	0.7
Know someone who did well	124 (81.1)	146 (85.4)	128 (82.6)	132 (82.5)	0.8
Know someone who did poorly	85 (55.6)	90 (52.6)	84 (54.2)	94 (58.8)	0.7
Baseline preference for TKR, mean ± SD	6.7 ± 2.7	6.8 ± 2.5	6.8 ± 2.6	6.8 ± 2.6	0.9

* Values are the number (%) unless indicated otherwise. NRS = numeric rating scale; TKR = total knee replacement.

satisfied with the surgery. These patients would have the surgery again without hesitation; 2) Some patients (about 7 in 50) don't do as well as they expected. They continue to have a fair amount of pain and are not very satisfied with the surgery. They don't think they would have this surgery again if they had bad arthritis in their other knee; 3) A few patients (about 1 in 50) have a serious complication after the surgery (such as an infection in the replaced knee). These patients regret having had the surgery."

Participants were then randomized to 1 of 4 information presentation formats: numeric only (the frequency information is described above), numeric with an icon array, numeric with a set of 50 images, or numeric with a functional spinner. Participants in the icon array group then viewed an icon array depicting people who do great in blue, people who do not do as well as expected in orange, and people who have a serious complication in black. Those in the images group viewed a set of images: 42 of active, happy people, representing people who do great, 7 of people with knee pain, representing people who do not do as well as expected, and 1 image of a red inflamed knee representing a serious complication. Each image was presented for 2 seconds in random order. Images were obtained from stock photo websites. Participants randomized to the spinner group viewed a donut-shaped figure with blue, orange, and black sections sized to represent the corresponding number of people who do great, do not do as well as expected, or have a serious complication, respectively. The spinner was programmed to rotate with a limited-range, randomly generated, initial speed after being clicked on by the participant. A constant damping factor was applied and the spinner gradually reached a minimum speed. When the spin was completed, the participant could spin again. The icon array, sample images, and spinner are shown in Supplementary Appendix A, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23605/abstract.

Preference for TKR was re-measured immediately after viewing the outcome information. We also measured knowledge using 3 questions designed for this study: How many people, out of every 50, 1) do really well after having total knee replacement surgery? 2) are not satisfied with the amount of pain relief they have after total knee replacement surgery? 3) experience a serious complication after total knee replacement surgery? Responses ranging from 40 to 44, 5 to 9, and 1 to 3 were considered correct for the first through third questions, respectively, and the number of correct responses was summed. We measured risk perceptions (perceived benefit and risk of TKR) and worry about the risk of a serious complication on 5-point scales (ranging from very to not at all). Risk/benefit expectation was measured by asking participants to choose 1 of the following 5 statements: the benefits of total knee replacement surgery greatly outweigh the risks; the benefits of total knee replacement surgery slightly outweigh the risks; the benefits of total knee replacement surgery are equal to the risks; the risks of total knee replacement surgery slightly outweigh the benefits; and the risks of total knee replacement surgery greatly outweigh the benefits. These items are frequently used to measure risk perceptions but have not been validated (12,13). Last, we measured overall health status on a 5-point scale (ranging from excellent to poor), knee pain over the past week on an 11-point numeric rating scale (where 0 = no pain at all and 10 = worst pain imaginable), and whether participants knew anyone who did really well or really poorly after having TKR surgery (yes/no).

Statistical analysis. We compared characteristics across formats using analysis of variance and chi-square tests for

continuous and categorical variables, respectively. We examined least squares mean (LSM) differences in TKR preferences across the 4 formats after controlling for baseline preference. We also performed subgroup analyses to examine the association between format and change in preference among participants with a baseline preference generally against TKR (score of 4 or lower), uncertainty (score of 5), generally in favor of TKR (score between 6 and 9), and a strong preference for TKR (score of 10).

RESULTS

Participant characteristics. Invitations to participate in the online survey were sent to 3,465 members of the Creaky-Joints community who had self-reported a doctor diagnosis of arthritis (32% open rate, 8% click rate). We followed up the initial invitation with up to 3 email reminders. The lead generation on Facebook yielded 2,227 email addresses to whom invitations were sent (55% open rate, 34% click rate). A total of 648 individuals completed the survey between April and June 2017. The mean ± SD age of the study population was 61.7 ± 6.9 years, 93% were female, 88% were white, and 44% were college graduates. The most common self-reported types of arthritis were osteoarthritis (57%), rheumatoid arthritis (12%), and both rheumatoid arthritis and osteoarthritis (25%). The remaining participants reported having psoriatic arthritis, with (4%) or without (2%) concomitant osteoarthritis. Mean ± SD knee pain, measured on a numeric rating scale, was 6.5 ± 2.4, and 34% of the participants reported having a fair or poor overall health status. Characteristics did not differ across formats (Table 1).

Impact of format on preference for TKR. Baseline preference for TKR did not differ across formats (Table 1). After controlling for baseline preference, the probability format was related to preference for TKR (F = 6.77; P = 0.0002). Preferences for TKR were higher in the icon array group (LSM ± SE 7.17 ± 0.10; P = 0.0002), images group (LSM ± SE 7.14 ± 0.10; P = 0.0005), and spinner group (LSM ± SE 7.19 ± 0.10; P = 0.0001), compared to the group receiving the numbers only format (LSM ± SE 6.66 ± 0.10), after controlling for baseline preference. Results remained unchanged after also controlling for age, insurance (private versus other), knowing someone who did poorly after TKR, and knee pain.

Among participants with preferences generally against TKR, LSM ± SE preferences were greater in participants randomized to the images compared to those viewing numbers only (P = 0.0376) (Table 2). The association between format and change in preference for TKR was accounted for primarily by participants with baseline preference scores generally in favor of TKR. Within this group, all formats had greater preference scores compared to the numbers only group (P < 0.05 for all) (Table 2). No significant differences in preferences were observed across formats for participants with an uncertain or very strong preference for TKR at baseline (Table 2).

Impact of format on knowledge. Format influenced knowledge (F = 13.62, P < 0.0001). The mean ± SD knowledge score (possible range 0–3) was higher in the icon array group (2.0 ± 1.1) compared to all other formats (numeric 1.4 ± 1.2, images 1.4 ± 1.1, and spinner 1.3 ± 1.1). The median knowledge score for participants randomized to the icon array was 2.0 compared to 1.0 for participants randomized to the other 3 formats. Knowledge was also related to preference (after controlling for baseline preference, F = 8.16, P = 0.0044). The estimates for both format and knowledge remained largely unchanged when included in the same multivariate model, indicating that knowledge did not mediate the association between format and change in preference for TKR.

Impact of format on risk perceptions and judgment. The mean ± SD worry, riskiness, and judgment scores were 2.95 ± 1.15, 2.85 ± 0.85, and 1.87 ± 1.13, respectively. We found no relationship between format and perceived risk of TKR, worry related to potential complications of TKR, or judgment related to the risk/benefit tradeoff associated with TKR (data not shown).

DISCUSSION

In this study, we found that the format used to communicate the probabilities of possible outcomes related to TKR affected patients' preferences for the procedure. Stronger preferences for TKR were seen in participants randomized to view an icon array, a set of images, or a spinner compared to the numbers only

Table 2. Preference for total knee replacement (TKR) compared with baseline preference*

Format	Leaning against TKR (n = 107)	Uncertain (n = 101)	Leaning toward TKR (n = 319)	Strongly favor TKR (n = 121)
Numeric	2.72 ± 0.39	5.72 ± 0.24	7.27 ± 0.17	9.41 ± 0.16
Icon array	3.78 ± 0.41	5.88 ± 0.23	7.74 ± 0.15†	9.71 ± 0.16
Images	3.92 ± 0.41†	5.56 ± 0.24	7.82 ± 0.17†	9.69 ± 0.15
Spinner	3.04 ± 0.42	5.68 ± 0.24	8.01 ± 0.16†	9.66 ± 0.16

* Values are the least squares mean ± SE.
† Significantly greater preference for TKR (P <0.05) compared to the numbers only format.

format. As expected, format did not influence participants who had a strong preference for TKR at baseline. These results contribute to the literature demonstrating that added graphic information influences preferences (14,25) and highlight the importance of accounting for baseline choice predisposition when evaluating the impact of specific interventions on patient preferences. We did not find any difference in risk perceptions or judgment across the 4 formats, indicating that these factors did not mediate the relationship between format and preference for TKR. This result is in line with previous research showing that direct effects of icon arrays do not always exist on perceived risk or worry (26). Consistent with previous research, the icon array improved knowledge scores; nevertheless, knowledge did not account for the effect of format on preference. Because the additional graphic information affected those whose baseline preferences favored TKR, the added information possibly reinforced positive beliefs or expectations related to TKR, leading to stronger preferences. However, we did not measure these variables, and this hypothesis should be tested in future research.

Describing outcomes through experience (images) is an appealing concept, because this approach may correct patients' overweighting of rare adverse events. Learning through experience is limited, however, by the need to attend to tasks that may be too long and/or complex to ensure adequate representation of all possible events. In this study, we constrained the denominator to 50 to simplify the task (as opposed to the frequently used 100) based on feedback we obtained while piloting the survey. Wegier and Shaffer developed a method that may overcome this limitation. In their study, students viewed grids representing the experiences of cohorts composed of 100 participants in rapid succession without compromising the impact of the task (18). Further research is needed to determine whether this approach is feasible and effective in patients facing personal health-related decisions.

Despite their intuitive appeal, spinners have not been rigorously tested as a decision aid. Eyler et al (20) demonstrated that they were at least as effective as icon arrays in communicating the risk of adverse events associated with a medication. The spinner used in this study was more complex in that it included the spectrum of expected outcomes. Thus, this format may be a possible approach to effectively communicate the range of outcomes experienced by patients undergoing a specific treatment. Whether the spinner offers any advantage over icon arrays, for example among patients with lower education or numeracy levels, requires further study.

This study has several limitations. Although we recruited a large number of participants with arthritis, our study population represents a narrow demographic segment, which limits generalizability of our results. Moreover, patients volunteered to take the survey and do not represent a population-based sample of arthritis patients. For example, the relatively low click rates may indicate that individuals who did choose to take part were more engaged and had higher disease knowledge and possibly higher health literacy than nonparticipants. In addition, eligibility criteria were ascertained based on self-report, and diagnoses were not confirmed by medical record or claims data. The high education level of the participants precluded examining whether this variable might modify the impact of format on preference and/or knowledge.

Decision support at the point-of-care is being increasingly recognized as a vital component of care. The Centers for Medicare and Medicaid Services, for example, mandate the use of 1 or more decision aids when counseling patients about lung cancer screening. Our findings suggest that adding graphic information to descriptive statistics strengthens preferences for TKR. Although experience formats using images may be too complex to use at the point-of-care, spinners may be a viable and easily adaptable decision aid to support communication of probabilistic information. Experience formats including images could, however, be used in conjunction with other decision aids at home in preparation for a surgical consultation. Further research is required to examine whether these tools increase the accuracy of patients' expectations in clinical practice.

ACKNOWLEDGMENTS

The authors thank the patients who participated in this study.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. Fraenkel 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. Fraenkel, Nowell, Stake, Venkatachalam, Eyler, Michel, Peters.

Acquisition of data. Fraenkel, Nowell, Stake, Venkatachalam, Michel.

Analysis and interpretation of data. Fraenkel, Nowell, Stake, Peters.

REFERENCES

1. Losina E, Walensky RP, Kessler CL, Emrani PS, Reichmann WM, Wright EA, et al. Cost-effectiveness of total knee arthroplasty in the United States: patient risk and hospital volume. Arch Intern Med 2009;169:1113–21.
2. Jones CA, Pohar S. Health-related quality of life after total joint arthroplasty: a scoping review. Clin Geriatr Med 2012;28:395–429.
3. Jones CA, Voaklander DC, Johnston DW, Suarez-Almazor ME. Health related quality of life outcomes after total hip and knee arthroplasties in a community based population. J Rheumatol 2000;27:1745–52.
4. Baker PN, Rushton S, Jameson SS, Reed M, Gregg P, Deehan DJ. Patient satisfaction with total knee replacement cannot be predicted from pre-operative variables alone: a cohort study from the National Joint Registry for England and Wales. Bone Jt J 2013;95:1359–65.
5. Choi YJ, Ra HJ. Patient satisfaction after total knee arthroplasty. Knee Surg Relat Res 2016;28:1–15.
6. Hamilton DF, Lane JV, Gaston P, Patton JT, Macdonald D, Simpson AH, et al. What determines patient satisfaction with surgery? A prospective cohort study of 4709 patients following total joint replacement. BMJ Open 2013;3:pii:e002525.

7. Barlow T, Griffin D, Barlow D, Realpe A. Patients' decision making in total knee arthroplasty: a systematic review of qualitative research. Bone Jt Res 2015;4:163–9.
8. Hudak PL, Clark JP, Hawker GA, Coyte PC, Mahomed NN, Kreder HJ, et al. "You're perfect for the procedure! Why don't you want it?" Elderly arthritis patients' unwillingness to consider total joint arthroplasty surgery: a qualitative study. Med Decis Making 2002; 22:272–8.
9. Riddle DL, Golladay GJ, Hayes A, Ghomrawi HM. Poor expectations of knee replacement benefit are associated with modifiable psychological factors and influence the decision to have surgery: a cross-sectional and longitudinal study of a community-based sample. Knee 2017;24:354–61.
10. Mota RE, Tarricone R, Ciani O, Bridges JF, Drummond M. Determinants of demand for total hip and knee arthroplasty: a systematic literature review. BMC Health Serv Res 2012;12:225.
11. Peters E, Hibbard J, Slovic P, Dieckmann N. Numeracy skill and the communication, comprehension, and use of risk-benefit information. Health Aff (Millwood) 2007;26:741–8.
12. Garcia-Retamero R, Galesic M. Who profits from visual aids: overcoming challenges in people's understanding of risks [corrected]. Soc Sci Med 2010;70:1019–25.
13. Reyna VF, Brainerd CJ. Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. Learn Individ Differ 2008;18:89–107.
14. Fagerlin A, Zikmund-Fisher BJ, Ubel PA. Helping patients decide: ten steps to better risk communication. J Natl Canc Inst 2011;103:1436–43.
15. Galesic M, Garcia-Retamero R. Graph literacy: a cross-cultural comparison. Med Decis Making 2011;31:444–57.
16. Hertwig R, Barron G, Weber EU, Erev I. Decisions from experience and the effect of rare events in risky choice. Psychol Sci 2004;15:534–9.
17. Tyszka T, Sawicki P. Affective and cognitive factors influencing sensitivity to probabilistic information. Risk Anal 2011;31:1832–45.
18. Wegier P, Shaffer VA. Aiding risk information learning through simulated experience (ARISE): using simulated outcomes to improve understanding of conditional probabilities in prenatal Down syndrome screening. Patient Educ Couns 2017;100:1882–9.
19. Fraenkel L, Peters E, Tyra S, Oelberg D. Shared medical decision making in lung cancer screening: experienced versus descriptive risk formats. Med Decis Making 2016;36:518–25.
20. Eyler RF, Cordes S, Szymanski BR, Fraenkel L. Utilization of continuous "spinners" to communicate risk. Med Decis Making 2017; 37:725–9.
21. Jeong Y, Levine SC, Huttenlocher J. The development of proportional reasoning: effect of continuous versus discrete quantities. J Cogn Dev 2007;8:237–56.
22. Fraenkel L, Seng EK, Cunningham M, Mattocks K. Understanding how patients (vs physicians) approach the decision to escalate treatment: a proposed conceptual model. Rheumatology (Oxford) 2015;54:278–85.
23. Hawker G, Wright J, Coyte P, Paul J, Dittus R, Croxford R, et al. Health-related quality of life after knee replacement. J Bone Jt Surg Am 1998;80:163–73.
24. Blom AW, Brown J, Taylor AH, Pattison G, Whitehouse S, Bannister GC. Infection after total knee arthroplasty. J Bone Joint Surg Br 2004;86:688–91.
25. Sinayev A, Peters E, Tusler M, Fraenkel L. Presenting numeric information with percentages and descriptive risk labels: a randomized trial. Med Decis Making 2015;35:937–47.
26. Han PK, Klein WM, Killam B, Lehman T, Massett H, Freedman AN. Representing randomness in the communication of individualized cancer risk estimates: effects on cancer risk perceptions, worry, and subjective uncertainty about risk. Patient Educ Couns 2012;86:106–13.

Arthritis Care & Research
Vol. 71, No. 3, March 2019, pp 385–389
DOI 10.1002/acr.23592

AMERICAN COLLEGE OF RHEUMATOLOGY
EDUCATION • TREATMENT • RESEARCH

BRIEF REPORT

Relationship Between Patient-Reported Swelling and Magnetic Resonance Imaging–Defined Effusion-Synovitis in Patients With Meniscus Tears and Knee Osteoarthritis

Lindsey A. MacFarlane,[1] Heidi Yang,[2] Jamie E. Collins,[1] Ali Guermazi,[3] Lisa A. Mandl,[4] Bruce A. Levy,[5] Robert G. Marx,[4] Clare E. Safran-Norton,[2] Elena Losina,[1] Jeffrey N. Katz,[1] and the Meniscal Tear in Osteoarthritis Research Investigator Group

Objective. Synovitis is a prevalent feature in patients with knee osteoarthritis (OA) and meniscal tear and is associated with pain and cartilage damage. Patient-reported swelling is also prevalent in this population. The aim of this study was to investigate the cross-sectional association between patient-reported swelling and effusion-synovitis detected by magnetic resonance imaging (MRI) in patients with OA and meniscal tear.

Methods. We used baseline data from a multicenter, randomized controlled trial, Meniscal Tear in Osteoarthritis Research (METEOR). MRI-identified effusion-synovitis, a proxy for effusion and synovitis on noncontrast MRIs, was graded as none/small versus medium/large. Using MRI-identified effusion-synovitis as the gold standard, we assessed the sensitivity, specificity, and positive predictive value of patient self-reported swelling in the previous week (none, intermittent, constant) to detect effusion and synovitis.

Results. We analyzed data from 276 patients. Twenty-five percent of patients reported no swelling, 40% had intermittent swelling, and 36% had constant swelling. Fifty-two percent had MRI-identified medium/large–grade effusion-synovitis. As compared with MRI-identified effusion-synovitis, any patient-reported swelling (versus none) had a sensitivity of 84% (95% confidence interval [95% CI] 77–89), a specificity of 34% (95% CI 26–43), and a positive predictive value of 57% (95% CI 54–61). A history of constant swelling (versus none or intermittent) showed a sensitivity of 46% (95% CI 37–54), a specificity of 75% (95% CI 67–82), and a positive predictive value of 66% (95% CI 58–74).

Conclusion. We found that the sensitivity and specificity of patient-reported swelling were modest when compared with effusion-synovitis detected by MRI. These data urge caution against using patient-reported swelling as a proxy of inflammation manifesting as effusion-synovitis.

INTRODUCTION

The pathogenesis of osteoarthritis (OA) is increasingly understood as a dynamic process involving multiple joint structures (e.g., cartilage, bone, meniscus), with an important role for intraarticular inflammation manifesting as effusion and synovitis. Recent studies have demonstrated that synovitis is associated with pain, incident knee OA, and progression of OA (1–4). Synovitis is also a prominent feature of meniscal tears, even in the absence of OA (5–7). Of the 14 million Americans with symptomatic knee OA,

Clinicaltrials.gov identifier: NCT00597012.

Supported by the NIH (grants P60-AR-047782, R01-AR-05557, T32-AR-055885,and K24-AR-057827).

[1]Lindsey A. MacFarlane, MD, MPH, Jamie E. Collins, PhD, Elena Losina, PhD, Jeffrey N. Katz, MD, MSc: Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; [2]Heidi Yang, MS, MPH, Clare E. Safran-Norton, PT, PhD: Brigham and Women's Hospital, Boston, Massachusetts; [3]Ali Guermazi, MD, PhD: Boston University School of Medicine, Boston, Massachusetts; [4]Lisa A. Mandl, MD, MPH, Robert G. Marx, MD: Weil Cornell Medicine, New York, New York; [5]Bruce A. Levy, MD: Mayo Clinic, Rochester, Minnesota.

Dr. MacFarlane has received consulting fees from Flexion Therapeutics (less than $10,000). Dr. Guermazi is a shareholder of Boston Imaging Core Lab and has received consulting fees from GE, AstraZeneca, Sanofi, TissueGene, OrthoTrophix (less than $10.000 each), and from Pfizer and Merck Serono (more than $10,000 each). Dr. Mandl is an associate editor at the Annals of Internal Medicine and receives royalties from UpToDate (less than $10,000). Dr. Levy has received consulting fees from Arthrex and Conemed (less than $10,000 each), and Smith and Nephew (more than $10,000), and has received research support from Arthrex, Biomet, and Stryker. Dr. Safran-Norton is a shareholder of Merck and Johnson and Johnson. Dr. Losina is the deputy editor of the Journal of Bone & Joint Surgery and has received consulting fees from Samumed (less than $10,000).

Address correspondence to Lindsey A. MacFarlane, MD, MPH, 60 Fenwood Road, Boston, MA 02115. E-mail: lmacfarlane@bwh.harvard.edu.

Submitted for publication January 9, 2018; accepted in revised form April 24, 2018.

SIGNIFICANCE & INNOVATIONS

- We found that the sensitivity and specificity of patient-reported swelling were modest when compared with effusion-synovitis detected by magnetic resonance imaging in patients with knee osteoarthritis and meniscal tear.
- This relationship did not differ despite stratification based on sex, radiographic severity of osteoarthritis, body mass index, or patient-reported pain.

up to 91% will have a meniscal tear detected by magnetic resonance imaging (MRI), highlighting the subset of patients with OA and meniscal tear as a clinically relevant and prevalent population (8,9).

Knee swelling is a key feature elicited in the clinical history and is commonly reported in patients with knee OA and meniscal tear. However, whether patient-reported swelling accurately predicts the presence of effusion-synovitis on imaging is unknown. Because effusion and synovitis identified on imaging may be associated with progression of intraarticular damage (1), a firmer understanding of the association between patient-reported swelling and the presence of effusion and synovitis on imaging could help clinicians stratify their patients for risk. This analysis aimed to assess the relationship between self-reported knee swelling and effusion-synovitis detected by MRI in patients with concurrent meniscal tear and knee OA. We hypothesized that patient-reported swelling would have a positive relationship with MRI-identified effusion-synovitis.

MATERIALS AND METHODS

Sample. We used baseline data from patients enrolled in the Meniscal Tear in Osteoarthritis Research (METEOR) trial (Appendix A shows members of the METEOR group), a randomized clinical trial of arthroscopic partial meniscectomy versus physical therapy for management of symptomatic meniscal tears (10). In total, 351 subjects were recruited from 7 academic referral centers between June 2008 and July 2011. Recruited men and women were ages ≥45 years, had a noncontrast MRI of the knee with evidence of meniscal tear extending to the meniscal surface, and had ≥4 weeks of meniscal symptoms (clicking, catching, popping, giving way, pain with pivot or torque, pain that is episodic, and pain that is acute and localized to 1 joint line). All enrolled patients had evidence of knee OA as determined by either osteophyte and joint space narrowing on plain radiographs, or osteophytes and/or cartilage defects identified on MRI. Patients with Kellgren/Lawrence (K/L) grade 4 knee OA were excluded, as were patients with inflammatory arthritis (i.e., rheumatoid arthritis, psoriatic arthritis, or spondyloarthritis) or who had prior surgery on the index knee. We focused the analysis on data from the index knee. All data domains corresponded to the index knee.

For the current analysis, we used data from the subsample of METEOR patients with baseline MRIs that were available for central re-reading, reducing the sample to 281 patients. An additional 5 patients were excluded because they had missing data for the patient-reported swelling scale; thus, we had a final cohort of 276 patients. The METEOR study was approved by the Partners HealthCare Human Research Committee.

Assessments. Demographics, including age, sex, and body mass index (BMI; kg/m^2), were collected at baseline. Patient-reported swelling was assessed via questionnaire and classified in numbered categories as 1) no swelling, 2) once in the last week, 3) 2–6 times in the last week, 4) 1–2 times/day, or 5) several times/day. Because some of these categories were sparsely populated, the variable was recategorized into 0) no swelling, 1) intermittent swelling (combining categories 2 and 3), and 3) constant swelling (combining categories 4 and 5). Our group previously reported strong test–retest reliability of using the word "swelling" on questionnaires (κ = 0.75) versus providing an expanded description of swelling and found strong agreement between the 2 question stems (11). Assessing the frequency of swelling in patients with knee OA and meniscal tear has been deemed an important assessment based on input from patients and clinicians (11,12). Noncontrast MRIs are frequently used in larger studies due to lower cost than with contrast MRI and less risk of adverse events (13). However, noncontrast MRIs cannot distinguish between effusion (fluid within the joint cavity) and synovitis (thickening and enhancement of the synovium); thus, effusion-synovitis is a frequently used proxy for synovitis on noncontrast MRIs (13). Using T2/intermediate weighted or proton density–weighted fat-suppressed MRI images, we assessed baseline effusion-synovitis according to the MRI OA Knee Score (MOAKS), a semiquantitative scoring method for knee OA (13). All MRIs were re-read centrally by a single reader (AG). The intra- and interrater weighted kappa statistics for effusion-synovitis in MOAKS are 0.9 and 0.72, respectively (13). Effusion-synovitis was categorized as 0) none (physiologic), 1) small (fluid continuous in the retropatellar space), 2) medium (slight convexity of suprapatellar bursa), or 3) large (capsular distention) (13). We dichotomized effusion-synovitis into none to small (grade 0–1) and medium to large (grade 2–3), because some categories were sparsely populated and for ease of interpretation.

The K/L radiographic grade was used to determine OA severity. The K/L grade was categorized as 0 = normal, 1 = questionable osteophyte, 2 = definite osteophyte, and 3 = <50% joint space narrowing. The Knee Injury and Osteoarthritis Outcome Score (KOOS) pain scale was used to assess overall patient-reported knee pain in the previous week and was transformed to a 0–100 scale, where 0 = the least amount of pain and 100 = the greatest amount of pain (12). We assessed the relationships between KOOS pain and patient-reported swelling and between KOOS pain and MRI-identified effusion-synovitis.

Sensitivity, specificity, and positive predictive value (PPV). Sensitivity, specificity, PPV, and positive likelihood ratio were calculated using MRI-identified effusion-synovitis as the criterion standard. We performed analyses using 2 distinct cut points of patient-reported swelling for a positive test: history of any swelling (intermittent and constant) and presence of constant swelling.

We performed sensitivity analyses to further investigate the relationship between patient-reported swelling and MRI-identified effusion-synovitis. First, we examined the relationship between patient-reported swelling and effusion-synovitis after stratifying by OA severity, based on the K/L grade (those with K/L grades 0 and 1 versus grades 2 and 3). We also stratified by patient-reported pain, based on KOOS pain dichotomized at the cohort mean (≤46 for less pain and >46 for greater pain), as well as BMI dichotomized to obese (BMI ≥30 kg/m^2) and nonobese (BMI <30 kg/m^2). Last, we stratified by sex. We additionally dichotomized MRI-identified effusion-synovitis as none (grade 0) versus any (grades 1–3) and reinvestigated the relationship between patient-reported swelling and MRI-defined effusion-synovitis. The Breslow-Day test was used to assess for homogeneity between the strata.

RESULTS

The study sample consisted of 276 patients (1 affected knee per patient). Twenty-five percent of patients reported no swelling, 40% reported intermittent swelling, and 36% reported constant swelling. The mean age was similar among the swelling categories, ranging from 58 to 59 years. Women constituted 69% of patients reporting constant swelling, 51% of patients reporting intermittent swelling, and 47% of patients reporting no swelling. The KOOS pain score was greater in patients with constant swelling (54 points) than in those with intermittent swelling (45 points) and no swelling (38 points) ($P < 0.01$) (Table 1).

Table 1. Baseline characteristics and association of patient-reported swelling and magnetic resonance imaging–identified effusion-synovitis*

	None (n = 68)	Intermittent (n = 110)	Constant (n = 98)	*P*
Age, years	59 ± 9	58 ± 7	58 ± 6	0.53
Women, no. (%)	32 (47)	56 (51)	68 (69)	0.005
BMI, kg/m^2	30 ± 6	29 ± 5	31 ± 7	0.04
K/L grade, no. (%)				
0	22 (32)	22 (20)	18 (18)	0.21
1	13 (19)	24 (22)	28 (29)	0.21
2	16 (24)	35 (32)	21 (21)	0.21
3	17 (25)	29 (26)	31 (32)	0.21
KOOS pain	38 ± 15	45 ± 14	54 ± 15	<0.0001

* Values are the mean ± SD unless indicated otherwise. BMI = body mass index; K/L = Kellgren/Lawrence; KOOS = Knee Injury and Osteoarthritis Outcome Score.

Table 2. Severity of effusion-synovitis listed by frequency of patient-reported swelling*

	None/small	Medium/large
No swelling	45	23
Intermittent swelling	55	54
Constant swelling	33	65

* Values are the number.

A total of 133 patients (48%) had no or small effusion-synovitis (grades 0–1), and 142 patients (52%) had medium or large effusion-synovitis (grades 2–3) (1 patient was missing data on effusion-synovitis). Medium or large effusion-synovitis was identified in 34% of patients with no swelling, in 50% with intermittent swelling, and in 66% with constant swelling ($P < 0.01$) (Table 2). The KOOS pain score was 48 points in patients with medium or large effusion-synovitis and 45 in those with no or small effusion-synovitis ($P = 0.07$).

Using data from patients who reported any swelling (those with intermittent and constant swelling) to calculate sensitivity, specificity, and PPV, we found that swelling had a sensitivity of 84% (95% confidence interval [95% CI] 77–89), a specificity of 34% (95% CI 26–43), a PPV of 57% (95% CI 54–61), and a positive likelihood ratio of 1.3 (95% CI 1.1–1.5) as compared with the gold standard, MRI-identified effusion-synovitis. Using a more stringent cut point of constant swelling (versus intermittent and no swelling), sensitivity was 46% (95% CI 37–54), specificity was 75% (95% CI 67–82), PPV was 66% (95% CI 58–74), and positive likelihood ratio was 1.8 (95% CI 1.3–2.6) for effusion-synovitis identified on MRI (Table 3).

In sensitivity analyses, the associations between self-reported swelling and MRI-documented effusion-synovitis were largely similar after stratifying by K/L grade, patient-reported KOOS pain score, BMI, or sex (test of homogeneity $P > 0.05$ for all). Dichotomizing effusion-synovitis as none versus any led to qualitatively similar results, though some categories were underpopulated.

DISCUSSION

Our findings indicate that patient-reported swelling did not perform well as a proxy for MRI-identified effusion-synovitis. Any reported swelling had a sensitivity of 84% for MRI-defined effusion-synovitis but lacked specificity (34%). Conversely, more frequent swelling (constant) was more specific for effusion-synovitis (75%) but lacked sensitivity (46%). Neither cut point for patient-reported swelling resulted in a strong PPV for MRI-defined effusion-synovitis (range 57–66%). Positive likelihood ratios were small. These results suggest that patient-reported swelling is not an especially useful clinical marker of MRI-identified effusion-synovitis.

Table 3. Sensitivity, specificity, and positive predictive value (PPV) of patient-reported swelling for effusion-synovitis*

	Sensitivity	Specificity	PPV	Positive likelihood ratio
Any swelling†	84 (77–89)	34 (26–43)	57 (54–61)	1.3 (1.1–1.5)
Constant swelling	46 (37–54)	75 (67–82)	66 (58–74)	1.8 (1.3–2.6)

* Values are the percentage (95% confidence interval) unless indicated otherwise.
† Intermittent and constant swelling.

Weak associations between patient-reported symptoms and structural abnormalities have been identified in other studies and conditions (14). An evaluation of hip pain and radiographic hip OA using the Framingham Osteoarthritis Study showed that radiographic hip OA had a sensitivity of 27% and specificity of 91% for groin or anterior thigh pain (14). Similar results were reported in the Osteoarthritis Initiative (OAI), with frequent hip pain having a sensitivity of 9% and specificity of 94% for radiographic hip OA (14). One study also using data from the OAI demonstrated an association of patient-reported swelling and knee pain with extension and effusion-synovitis identified on MRI (15). However, that study did not assess the performance characteristics of patient-reported symptoms and MRI findings, precluding a direct comparison with our work. We stratified by K/L grade, KOOS pain score, BMI, and sex but found that patient-reported swelling was not strongly associated with effusion-synovitis on imaging in any of these distinct strata. Our study demonstrated a significant difference in pain among the patient-reported swelling categories, with those that indicated more frequent swelling also reporting greater pain. This relationship was not seen with MRI-identified effusion-synovitis, for which there was no difference in pain levels between those with no or small effusion-synovitis and those with medium or large effusion-synovitis. This finding suggests that the relationship between various patient-reported measures may be stronger than the association between patient-reported symptom measures and structural findings on imaging.

This study has several limitations. Effusion-synovitis identified on noncontrast MRI is a frequently used proxy for effusion and synovitis in clinical studies, given the lower cost and the lower risk of adverse effects with noncontrast scans. However, noncontrast MRI cannot differentiate between effusion and synovitis (13). Clinician assessment of effusion on physical examination was not recorded as part of the trial. Thus, we cannot evaluate the relationship between clinical examination evidence of effusion, MRI-identified effusion-synovitis, and patient-reported swelling. We recognize that it would be useful to minimize the role of patient subjectivity. Further study using clinical examination or musculoskeletal ultrasound could leverage safe bedside modalities for assessing synovitis. Our patient population had concurrent OA and meniscal tear; thus, we are unable to discern whether OA, meniscal tear, or a combination of both drives patient-reported swelling or effusion-synovitis noted on MRI. MRI and assessment of patient-reported swelling were not done on the same day. Because the presence of effusion-synovitis is dynamic, the possibility of change may have weakened the associations. Our study is strengthened by being a large multicenter study with rich demographic and imaging data, focusing on a prevalent population of patients with both meniscal tear and knee OA.

In conclusion, in a population of patients with OA and meniscal tear, we found that the sensitivity and specificity of patient-reported swelling were modest when compared with effusion-synovitis detected by MRI. We caution against clinical use of patient-reported swelling as a measure of proxy for synovitis in this population.

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. MacFarlane 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. MacFarlane, Yang, Collins, Guermazi, Losina, Katz.

Acquisition of data. Yang, Collins, Guermazi, Losina, Katz.

Analysis and interpretation of data. MacFarlane, Yang, Collins, Guermazi, Mandl, Levy, Marx, Safran-Norton, Losina, Katz.

REFERENCES

1. Felson DT, Niu J, Neogi T, Goggins J, Nevitt MC, Roemer F, et al. Synovitis and the risk of knee osteoarthritis: the MOST study. Osteoarthritis Cartilage 2016;24:458–64.
2. Collins JE, Losina E, Nevitt MC, Roemer FW, Guermazi A, Lynch JA, et al. Semiquantitative imaging biomarkers of knee osteoarthritis progression: data from the Foundation for the National Institutes of Health Osteoarthritis Biomarkers Consortium. Arthritis Rheumatol 2016;68:2422–31.
3. Baker K, Grainger A, Niu J, Clancy M, Guermazi A, Crema M, et al. Relation of synovitis to knee pain using contrast-enhanced MRIs. Ann Rheum Dis 2010;69:1779–83.
4. Neogi T, Guermazi A, Roemer F, Nevitt MC, Scholz J, Arendt-Nielsen L, et al. Association of joint inflammation with pain sensitization in knee osteoarthritis: the Multicenter Osteoarthritis Study. Arthritis Rheumatol 2016;68:654–61.
5. Scanzello CR, McKeon B, Swaim BH, DiCarlo E, Asomugha EU, Kanda V, et al. Synovial inflammation in patients undergoing arthroscopic meniscectomy: molecular characterization and relationship to symptoms. Arthritis Rheum 2011;63:391–400.
6. Troupis JM, Batt MJ, Pasricha SS, Saddik D. Magnetic resonance imaging in knee synovitis: clinical utility in differentiating asymptomatic and symptomatic meniscal tears. J Med Imaging Radiat Oncol 2015;59:1–6.

7. Roemer FW, Felson DT, Yang T, Niu J, Crema MD, Englund M, et al. The association between meniscal damage of the posterior horns and localized posterior synovitis detected on T1-weighted contrast-enhanced MRI: the MOST study. Semin Arthritis Rheum 2013;42:573–81.
8. 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.
9. 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.
10. 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.
11. Niu NN, Losina E, Martin SD, Wright J, Solomon DH, Katz JN. Development and preliminary validation of a meniscal symptom index. Arthritis Care Res (Hoboken) 2011;63:208–15.
12. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS): development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998;28:88–96.
13. 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.
14. Kim C, Nevitt MC, Niu J, Clancy MM, Lane NE, Link TM, et al. Association of hip pain with radiographic evidence of hip osteoarthritis: diagnostic test study. BMJ 2015;351:h5983.
15. Berlinberg A, Westra J, Ashbeck EL, Trost J, Roemer F, Guermazi A, et al. Knee physical exam findings and self-reported symptoms are associated with MRI-detected effusion-synovitis among participants with or at risk for knee osteoarthritis: data from the Osteoarthritis Initiative (OAI) [abstract]. Arthritis Rheumatol 2016;68 Suppl 10.

Appendix A: Members of the Meniscal Tear in Osteoarthritis Research Investigator Group

Members of the Meniscal Tear in Osteoarthritis Research Investigator group include Robert H. Brophy, MD, Brian J. Cole, MD, MBA, Mathew Matava, MD, Kurt P. Spindler, MD, Michael Stuart, MD, and Rick Wright, MD.

DOI 10.1002/acr.23849

7. Roemer FW, Felson DT, Yang T, Niu J, Crema MD, Englund M, et al. The association between meniscal damage of the posterior horns and localized posterior synovitis detected on T1-weighted contrast-enhanced MRI: the MOST study. Semin Arthritis Rheum 2013;42:573–81.
8. 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.
9. 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.
10. 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.
11. Niu NN, Losina E, Martin SD, Wright J, Solomon DH, Katz JN. Development and preliminary validation of a meniscal symptom index. Arthritis Care Res (Hoboken) 2011;63:208–15.
12. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS): development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998;28:88–96.
13. 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.
14. Kim C, Nevitt MC, Niu J, Clancy MM, Lane NE, Link TM, et al. Association of hip pain with radiographic evidence of hip osteoarthritis: diagnostic test study. BMJ 2015;351:h5983.
15. Berlinberg A, Westra J, Ashbeck EL, Trost J, Roemer F, Guermazi A, et al. Knee physical exam findings and self-reported symptoms are associated with MRI-detected effusion-synovitis among participants with or at risk for knee osteoarthritis: data from the Osteoarthritis Initiative (OAI) [abstract]. Arthritis Rheumatol 2016;68 Suppl 10.

Appendix A: Members of the Meniscal Tear in Osteoarthritis Research Investigator Group

Members of the Meniscal Tear in Osteoarthritis Research Investigator group include Robert H. Brophy, MD, Brian J. Cole, MD, MBA, Mathew Matava, MD, Kurt P. Spindler, MD, Michael Stuart, MD, and Rick Wright, MD.

DOI 10.1002/acr.23849

AMERICAN COLLEGE OF RHEUMATOLOGY
EDUCATION • TREATMENT • RESEARCH

Arthritis Care & Research
Vol. 71, No. 3, March 2019, pp 390–397
DOI 10.1002/acr.23621

Patient-Reported Disease Activity and Adverse Pregnancy Outcomes in Systemic Lupus Erythematosus and Rheumatoid Arthritis

Nathaniel Harris, Amanda Eudy, and Megan Clowse

Objective. While increased rheumatic disease activity during pregnancy has been associated with adverse pregnancy outcomes, this disease activity is typically assessed by physicians. Little is known, however, about the association between patient-reported measures of disease activity and pregnancy outcomes. The aim of our study was to evaluate this association.

Methods. Univariate and multivariable regression models were used to assess the relationship between patient- and physician-reported measures of disease activity and adverse pregnancy outcomes in 225 patients with systemic lupus erythematosus (SLE) or rheumatoid arthritis (RA). The patients were enrolled from 2008–2016 in a prospective registry at a single academic center.

Results. In women with RA, patient-reported disease activity was associated with preterm birth (odds ratio [OR] 5.9 [95% confidence interval (95% CI) 1.5, 23.9]) and gestational age in weeks (β = –1.5 [95% CI –2.6, –0.4]). The physician assessment of disease activity also predicted preterm (OR 2.1 [95% CI 1.2, 3.5]), small for gestational age births (OR 1.8 [95% CI 1.03, 3.1]), and gestational age in weeks (β = –0.6 [95% CI –0.9, –0.02]). Alternatively, in women with SLE, patient-reported disease activity measures, including the Health Assessment Questionnaire, pain, or global health measures, were not associated with adverse pregnancy outcomes. However, physician measures of SLE disease activity are associated with preterm birth (OR 2.9 [95% CI 1.3, 6.3]), cesarean delivery (OR 2.3 [95% CI 1.0, 5.3]), and preeclampsia (OR 2.8 [95% CI 1.3, 6.3]). The results did not appear to be driven by lupus nephritis or antiphospholipid syndrome.

Conclusion. For women with RA, patient-reported measures of disease activity were associated with adverse pregnancy outcomes, and thus may be useful adjuncts to physician-reported measures in identifying pregnancies at greater risk. In contrast, in SLE, while physician measures of disease activity helped predict several adverse pregnancy outcomes, no patient-reported measures were associated with adverse outcomes.

INTRODUCTION

Women with systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA) who are pregnant are at increased risk for preterm delivery, cesarean delivery, preeclampsia, and delivery of neonates who are small for gestational age (SGA) (1–9). The risk is particularly high among women with lupus, who are also at an increased risk for fetal loss (2). Among pregnancies in women with SLE, active disease at the time of conception and during pregnancy have been shown to increase the risk of preterm birth and complications such as intrauterine growth restriction and pregnancy-induced hypertension (10). In up to 60% of women, pregnancy can induce a lupus flare in those with previously quiescent disease, and 15–30% have moderate-to-severe disease activity during pregnancy (11–16). Women with SLE who are pregnant are also more likely to have complications including thrombosis, infection, and thrombocytopenia, as well as comorbidities (such as diabetes and hypertension) that increase the risk of adverse outcomes (5).

While pregnancy outcomes in women with well-controlled RA are similar to those in the general population (17), higher levels of disease activity are associated with preterm birth, reduced gestational age at birth, high rates of cesarean delivery, and SGA infants (6–9,18). While it has traditionally been taught that RA will remit in pregnancy (19), recent studies have found no improvement or even a worsening of disease in many patients during pregnancy (20,21).

Nathaniel Harris, MD, PhD, Amanda Eudy, PhD, Megan Clowse, MD, MPH: Duke University Medical Center, Durham, North Carolina.

Dr. Clowse has served as a paid consultant for UCB (more than $10,000).

Address correspondence to Megan Clowse, MD, MPH, Box 3535 Trent Drive, Durham, NC 27710. E-mail: megan.clowse@duke.edu.

Submitted for publication September 14, 2017; accepted in revised form June 12, 2018.

SIGNIFICANCE & INNOVATIONS

- Women with systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA) are at increased risk for preterm delivery, cesarean delivery, preeclampsia, and delivery of neonates who are small for gestational age.
- We hypothesized that patient-reported measures of disease activity might provide useful adjuncts to physician-reported measures in identifying pregnancies at greater risk for adverse pregnancy outcomes. The present study demonstrated an interesting contrast between the utility of existing patient-reported measures of disability in SLE when compared to RA.
- To our knowledge, we are the first to evaluate how patient-reported outcomes correlate with adverse pregnancy outcomes in patients with SLE. Prior studies have relied on physician-reported measures of disease activity to assess rheumatic disease activity during pregnancy.
- Almost half of the pregnancies in women with SLE were in African American women, providing rich racial diversity to this study.

Prior studies have relied on physician-reported measures to assess rheumatic disease activity during pregnancy. However, many obstetricians are inexperienced in the treatment of autoimmune disease in women, and most rheumatologists are inexperienced in the treatment of women who are pregnant. We hypothesized that patient-reported measures of disease might provide useful adjuncts to physician-reported measures in identifying pregnancies at greater risk for adverse pregnancy outcomes. This could help physicians who are inexperienced with this population select which patients require closer follow-up and give possible referrals to providers who specialize in autoimmune disease and pregnancy and maternal-fetal medicine. Patient-reported measures of RA activity have been well tested and validated outside of pregnancy and include the Health Assessment Questionnaire (HAQ), RAPID 3, visual analog scales of pain and global health, and the Rheumatoid Arthritis Disease Activity Index (22,23). Some studies suggest that these can be used to enhance doctor-patient communication and save time for the physician (24). Less is known about the utility of patient-reported measures in patients with SLE (25,26). In this study, we sought to understand the utility of several patient-reported measures in identifying women at increased risk for adverse pregnancy outcomes.

PATIENTS AND METHODS

Sequential pregnant women with SLE or RA were enrolled in the Duke Autoimmunity in Pregnancy Registry at their first rheumatology clinic appointment during pregnancy. The registry was approved by the Duke University Institutional Review Board (Pro00000756) and informed consent was obtained from all patients at the time of enrollment. Each pregnancy was followed prospectively through the postpartum period, with standard-of-care visits that included data collection of both patient- and physician-reported measures of activity for the registry. Patient-reported outcomes (PROs) were obtained at nearly every visit. The measures collected since the establishment of the registry include the HAQ, visual analog scales of pain and general health, the Short Form 36 Health Survey, and a depression screen. These measures were in wide use and validated at least outside of pregnancy at the time the registry was established. Measures such as the Systemic Lupus Activity Questionnaire and Lupus Quality of Life PRO were not in wide use at the time the registry was established in 2007 (27,28). All patients were assessed by a single rheumatologist, and the physician's global assessment (PhGA) was completed without knowledge of any patient survey measures.

Adverse outcomes. We assessed 6 outcomes (preterm birth, gestational age, cesarean delivery, SGA, preeclampsia, and pregnancy loss) by patient report in the weeks following delivery and, when available, by chart review. Preterm birth was defined as any live birth before 37 weeks. Gestational age in weeks describes the week of pregnancy at which the child was delivered, calculated either from the last menstrual period or an early ultrasound measurement. SGA was defined as being below the tenth percentile in weight for the week of gestation of delivery, by sex (29). Instances of preeclampsia were defined by an obstetric diagnosis of preeclampsia, because it was not possible to document blood pressures and urine protein levels for all patients at delivery.

Disease activity. *Patient-reported measures*. The principal measure of patient-reported disease activity was the HAQ, which was developed as a measure of physical function in a variety of rheumatic diseases (30,31). Because certain activities (such as bending over to pick up items) may become more difficult in pregnancy, it is clear that the HAQ score may be affected by pregnancy (20,32,33). Using the study by de Man et al (33) as an example, for our study we removed from the HAQ questions related to dressing oneself, tying shoes, getting in and out of bed, climbing up 5 steps, taking a bath, bending down and picking up clothing off the floor, and getting in and out of a car. We also utilized visual analog scales to measure pain in the last week and general health (both scales range 0–10, with 0 representing no disease activity). Our study included analyses using visual analog scales for pain and general health as patient-reported measures of disease activity.

Physician-reported measures. We compared the patient-reported measures of disease activity with physician-reported measures. The physician-reported measures served as benchmarks for the utility of the patient measures. The physician meas-

ure of disease activity used in our analyses was the PhGA (34). For SLE, the score ranges from 0–3, with 0 representing no disease activity, a 0.5 or 1 representing mild-to-moderate activity, and a 1.5 typically representing significant non-nephritis disease activity or mild nephritis. Anyone assigned a 2 or above has active internal organ disease, usually nephritis. The range of the PhGA in SLE differs from the PhGA in RA. In RA, patients are assessed on a scale of 0–10 with 0 representing no disease activity. We selected the maximum value of the HAQ and PhGA during the pregnancy. The maximum value for each PRO for each patient was chosen as the best proxy for whether a patient experienced high disease activity during pregnancy. Sensitivity analyses were performed using average PRO scores, as well as the HAQ score for each trimester; all results were similar. We selected the maximum value because we believed it would be unlikely to mask a disease flare (as might occur when using an average value), and would alleviate the risk of missing the importance of disease activity in any given trimester. For RA patients, we also assessed the 28-joint count measure, a summary of the number of tender and swollen joints. Joints that appeared to be swollen due to pregnancy alone were not included in the joint count assessment.

Confounding variables. We used a directed acyclic graph (DAG) to model the relationship between potential confounding variables to derive the minimally sufficient set of confounders to adjust for in our models that was reduced based on a 10% change in beta estimates (Figure 1) (35,36). We selected race, living arrangements, education level, and body mass index (BMI) as key confounders for pregnancy outcomes. Race was designated as either African American or other. Women were either living with a spouse/partner or not; this variable is used as a surrogate marker for socioeconomic status and social support. Education level is separated into those who have at least a college education and those who do not. BMI is a continuous variable based on the first trimester BMI.

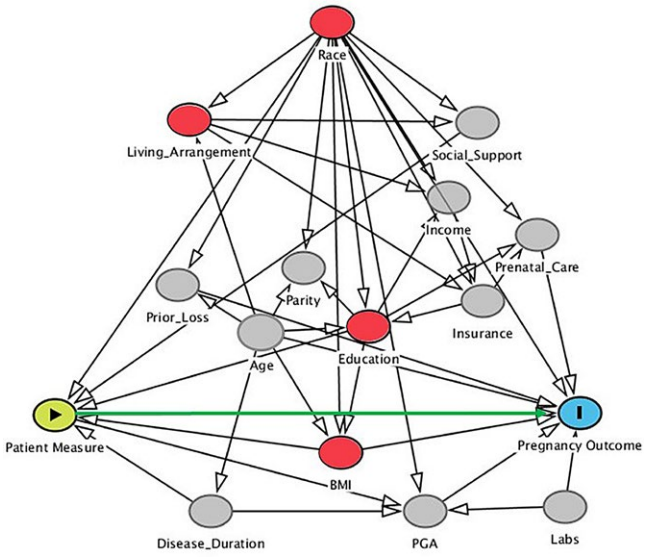


Figure 1. Directed acyclic graph (DAG; based on creation and evaluation of DAGs by Textor et al [52]) modeling the relationship between potential confounding variables. BMI = body mass index; PGA = physician's global assessment.

Statistical analysis. Univariate analyses of each disease measure and pregnancy outcome were conducted, followed by multivariable logistic and linear regression models, adjusting for the confounding variables of race, living arrangement, education, and BMI. Results from the logistic regression models are reported as odds ratios. Due to collinearity between the PhGA and the HAQ, pain level, and general health measures in RA patients, we did not include measures of patient- and physician-reported disease activity in the same model. Rather, separate analyses were conducted for the RA pregnancies.

Lupus nephritis and antiphospholipid syndrome (APS) are both predictors of poor pregnancy outcome in women with SLE, but we did not consider them as confounders based on the DAG model, because while they are expected to influence the PhGA and outcomes, they are not expected to directly impact measures of patient-reported disability such as the HAQ or pain level. Stratified multivariable analyses were not possible due to the small number of patients with lupus nephritis, APS, or antiphospholipid antibodies (aPLs). Instead, sensitivity analyses were conducted comparing the frequency of events in the pregnancies in women with a history of lupus nephritis, those with APS, and those with positive aPLs and those without. APS was defined based on the Sapporo criteria (37). All data analyses were performed using Stata statistical software (version 14.0).

RESULTS

Our patient panel included 145 women with SLE and 80 women with RA (including 17 patients with juvenile rheumatoid arthritis). There were substantial differences in the baseline characteristics of these 2 populations. Patients with RA were more likely to be older, white, living with a spouse or partner, and to have completed at least a 4-year college education (Table 1). Half of the lupus patients had a history of a positive Ro antibody, but only 17% had a history of lupus nephritis. Ten percent of women with SLE and 6% with RA had positive aPLs, 2 of whom met the criteria for APS. The average number of visits for patients with a diagnosis of SLE was 3.5. The average number of visits for patients with RA was 3.2.

Rates of adverse events were higher in pregnant patients with SLE. Preterm birth occurred in 30% of SLE pregnancies, cesarean delivery in 55% of SLE pregnancies, and pregnancy loss occurred in almost 10%. Preterm birth occurred in 14% of RA pregnancies, cesarean deliveries occurred in 37%, and pregnancy loss occurred in 4.5% of women with RA. In each analysis discussed below, the maximum value of the PhGA and PRO during the pregnancy was used. In all cases we performed

Table 1. Patient characteristics of women with systemic lupus erythematosus and rheumatoid arthritis*

Characteristics	SLE (n = 145)	RA (n = 80)
Maternal age at delivery, mean (range) years	30 (20–45)	33 (18–49)
Race/ethnicity		
White	63 (43)	63 (79)
African American	73 (50)	11 (14)
Asian	4 (3)	4 (5)
Other	5 (3)	2 (3)
Single		
Living with spouse/partner	98 (68)	68 (85)
Living alone	31 (21)	11 (14)
Missing	16 (11)	1 (1)
Education		
Completed college	75 (52)	60 (75)
Less than college	55 (38)	19 (24)
Missing	15 (10)	1 (1)
1st trimester BMI		
<25	41 (28)	29 (36)
25–30	37 (26)	29 (36)
>30	49 (34)	16 (20)
Missing	18 (12)	6 (8)
SSA/Ro+		
Positive	71 (49)	5 (6)
Negative	70 (48)	51 (63)
Missing	4 (3)	24 (30)
Nephritis		
History of nephritis	25 (17.2)	–
No history	120 (82.8)	–
aPL labs		
Positive	12 (8)	5 (6)
Negative	133 (86)	38 (48)
Missing	8 (6)	37 (46)

* Values are the number (%) of patients, unless indicated otherwise. SLE = systemic lupus erythematosus; RA = rheumatoid arthritis; BMI = body mass index; aPLs = antiphospholipid antibodies.

sensitivity analyses using average PRO scores, as well as the HAQ score by trimester; there was no substantive change in the results.

Lupus disease activity and preterm birth. Among lupus pregnancies, the preterm birth rate rose steadily with physician-assessed disease activity (Table 2). In lupus patients, 20% of pregnancies with a maximum PhGA score of 0 delivered preterm compared to 60–80% of pregnancies with a maximum PhGA score of 2 or 3 (the PhGA score in SLE ranges from 0–3, with 3 representing maximum disease activity). This tight correlation between physician-assessed lupus activity and preterm birth remained in multivariable analysis (Table 3), with each 1-unit increase in PhGA increasing the risk for preterm birth almost 3-fold, and shortening pregnancy by an estimated 4 weeks. Alternatively, the patient-reported measures reported by women with SLE, the HAQ, and the scores for pain and general health were not predictive of preterm birth. While the effect of HAQ on gestational age of babies born in patients with lupus approached statistical significance, the direction was unexpected. The effect estimate would suggest that a higher HAQ score, and thus, greater patient-reported disease activity, is associated with a longer pregnancy.

RA disease activity and preterm birth. In univariate and multivariable analyses, patient-reported disability and

Table 2. The rate of preterm birth based on the physician's global assessment and patient-reported disability with univariate logistic regression modeling*

Assessment	Preterm birth	OR (95% CI), *P*
SLE PhGA		3.0 (1.6–5.9), <0.01
0	20	
0.5	28	
1	33	
1.5	60	
2–3	86	
SLE HAQ		1.3 (0.6–2.5), 0.5
0–0.4	24	
0.5–0.9	33	
1–1.4	30	
1.5–1.9	60	
2–3	0	
RA PhGA		1.8 (1.2–2.7), <0.01
0–1.9	9	
2–3.9	15	
4–5.9	60	
6–7.9	50	
8–10	100	
RA HAQ		4.4 (1.4–14.0), 0.01
0	9	
0.5	0	
1	18	
1.5	50	
2–3	100	

* Values are the percent of preterm births for patients with systemic lupus erythematosus (SLE) and those with rheumatoid arthritis (RA), based on physician's global assessment (PhGA) and patient-reported disability (health assessment questionnaire [HAQ]). A PhGA score of 0 represents no disease activity and a HAQ score of 0 represents no disability. OR = odds ratio; 95% CI = 95% confidence interval.

Table 3. Association between disease activity by patient-report (HAQ) and physician-report (PhGA)*

	SLE HAQ‡	SLE PhGA‡	RA HAQ§	RA PhGA§
Gestational age, weeks†	1.40 (1.3, 6.3), 0.06	−2.9 (−5.3, −2.7), <0.01	−1.5 (−2.6, − 0.4), <0.01	−0.6 (−0.9, −0.2), <0.01
Preterm birth	0.99 (0.4, 2.4), 0.98	2.9 (1.3, 6.3), 0.01	5.9 (1.5, 23.9), 0.01	2.1 (1.2, 3.5), <0.01
Cesarean delivery	0.5 (0.2, 1.3), 0.15	2.3 (1.0, 5.3), 0.046	1.4 (0.5, 4.0), 0.51	1.3 (0.9, 1.8), 0.15
Preeclampsia	0.96 (04, 2.4), 0.9	2.8 (1.3, 6.3), 0.01	0.5 (0.1, 2.9), 0.5	0.7 (0.4, 1.4), 0.4
SGA	0.6 (0.2, 1.5), 0.3	0.7 (0.3, 1.7), 0.4	1.6 (0.4, 7.3), 0.5	1.8 (1.03, 3.1), 0.04
Fetal loss	0.2 (0.02, 2.1), 0.2	(0.1, 3.1), 0.6	0.1 (0.00, 13), 0.2	1.0 (0.5, 2.2), 0.9

* Values are the odds ratio (95% confidence interval [95% CI]), *P* value unless indicated otherwise. Multivariable regression analysis adjustments were made for race, living arrangements, education level, and body mass index. SLE = systemic lupus erythematosus; RA = rheumatoid arthritis; SGA = small for gestational age.
† Values are β (95% CI), *P*-value.
‡ Health Assessement Questionnaire (HAQ) and physician's global assessment (PhGA) were included in the same multivariate model.
§ HAQ and PhGA were assessed in separate multivariate models, due to collinearity.

pain, as well as physician-reported RA disease activity were statistically significant predictors of preterm birth. For each 1-unit increase in the HAQ, the odds of preterm birth increased almost 6-fold and shortened pregnancy by 1.5 weeks. Pain was associated with shorter pregnancies, but only approached statistical significance in preterm birth; in both cases the effect size is smaller than with the HAQ (see Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23621/abstract). The patient's assessment of general health was not associated with preterm birth or pregnancy length. Living without a partner or spouse was associated with an increase in the risk of preterm birth.

Physician-reported RA activity had a similar impact on pregnancy outcomes as patient-reported activity (Tables 2 and 3). A 1-unit increase in the PhGA, (which is reported here on a scale from 0–10, in contrast to SLE) was associated with a 2-fold increase in preterm birth and shortened pregnancy 0.5 weeks. For gestational age, the PhGA coefficient was smaller than the HAQ; however, the HAQ varies on a scale from 0–3 whereas the PhGA for RA patients varies from 0–10. The joint count was associated with preterm birth (OR 1.26 [95% confidence interval (95% CI) 1.1–1.5]), with a 1-unit increase in joint count corresponding to a pregnancy shortened by a day (see Supplementary Table 1, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.23621/abstract).

Cesarean delivery. As expected based on the impact of BMI on cesarean delivery rates in the general population, in both lupus and RA, a higher BMI was associated with increased odds of cesarean delivery (OR 1.2) (the range for first trimester was BMI 17–71 in our data set). The change in the rate of cesarean delivery as maternal weight increased from the normal to obese range was substantial (Table 4). In lupus patients, 7 out of 8 women with a BMI >40 delivered by cesarean section, a rate of 88%.

Physician-reported lupus activity was associated with increased odds of cesarean delivery, with a 1-unit increase in the PhGA associated with 2.3 greater odds of cesarean delivery (*P* = 0.046). Patient-reported measures in women with lupus were not associated with cesarean delivery. RA activity was not associated with cesarean delivery, whether physician- or patient-reported.

SGA. In the case of SGA children, the only statistically significant variable in any analysis was physician-reported disease activity in RA patients. A 1-unit increase in the PhGA was associated with 1.7 greater odds of having an SGA infant.

Preeclampsia. The only statistically significant variable in any analysis was physician-reported lupus activity. A 1-unit increase in PhGA was associated with 3 times greater odds of preeclampsia (*P* = 0.01). This effect remained similar and statistically significant in models with HAQ, pain, and general health as the patient-reported measures.

Pregnancy loss. None of the measures that we included in our models correlated with pregnancy loss. There was no difference in rates when comparing those with and without lupus nephritis or aPLs.

Multivariable analyses stratified by nephritis or aPLs/APS were not possible due to the small number of patients with these findings. The rates of preterm birth were higher among patients with a history of nephritis than in those without (48% versus 24%)

Table 4. Cesarean delivery rate by body mass index*

BMI category	SLE	RA
Normal weight (<25)	30	24
Overweight (25–30)	56	33
Obese (>30)	64	62

* Values are the percent of patients with systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA). BMI = body mass index.

as were rates of preeclampsia (32% versus 13%); however, rates of cesarean delivery, SGA births, and pregnancy loss were nearly identical. Positive aPLs were uncommon in this population, with only 17 pregnancies in women with positive aPLs (12 with SLE, 5 with RA). Only 2 women met the criteria for APS. Among lupus patients, rates of pregnancy loss were similar in those with and without aPLs (9% versus 8%). Rates of pregnancy loss in RA patients with aPLs were lower than those without such antibodies, but were not significantly different (0% versus 5%).

DISCUSSION

The present study demonstrated an interesting contrast between the utility of existing patient-reported measures in SLE when compared to those in RA. Among patients with lupus, only measures of physician-reported disease activity, and not measures of patient-reported disability, pain, or general health, correlated with adverse pregnancy outcomes. In contrast, in RA, both patient and physician-reported measures were useful predictors of adverse pregnancy events. The HAQ was adjusted for pregnancy; therefore, any changes in this measure were unlikely to be due to pregnancy alone.

As has been demonstrated previously, the physician measure of disease activity proved to be a reliable predictor of adverse pregnancy outcomes in lupus patients, including preterm birth, cesarean delivery, and preeclampsia (27,38). A higher physician score was associated with substantial decreases in gestational age, as well as increased odds of preterm birth, preeclampsia, and cesarean delivery. When compared to the patient-reported measures, an interesting pattern emerged. The HAQ, which measures disability, as well as measures of global health and pain, did not correlate with adverse pregnancy outcomes in lupus. Moreover, in the models for gestational age and cesarean delivery, the effect of patient-reported disability suggested that higher levels of pain and disability were actually associated with larger babies, longer pregnancies, and lower odds of preterm birth. However, these results were not statistically significant. It is likely that these measures do not correlate well with outcomes because the HAQ does not correlate with inflammation in lupus (and thus pregnancy outcomes). It is also interesting that outcomes in women with lupus did not correlate with their general level of pain or their sense of their general health. As much attention is placed on controlling lupus activity in pregnancy, a key component of care is identification and treatment of lupus activity. To our knowledge, we are the first to study how patient-reported outcomes correlate with pregnancy outcomes in patients with SLE. It appears, however, that the surveys about physical function, pain, and global health did not capture the types of lupus activity that impact pregnancy outcomes.

In patients with RA, both patient- and physician-reported measures appeared to be useful as prognosticators of adverse pregnancy outcomes and could assist physicians in identifying and treating pregnancies at higher risk for adverse outcomes. Greater patient-reported disability and pain were associated with shorter pregnancies, smaller babies, and an increase in the risk of preterm birth. There was no correlation with cesarean delivery or pregnancy loss.

An important strength of this study was the depth of data collected for each pregnancy, which allowed for all models to be adjusted for several factors known to impact pregnancy outcomes. Race, particularly in patients with lupus, contributes to increased rates of adverse pregnancy outcomes (39). African American and Hispanic women have higher rates of cesarean delivery, preterm labor, preeclampsia, and gestational hypertension (40). Almost half of the pregnancies in women with SLE were in African Americans, which provided rich racial diversity to the present study. Living with a partner or spouse reflects social support that likely impacts financial and physical access to prenatal and postpartum care; in the present study, this was associated with preterm birth in women with RA (41,42) in this study. Education reflects socioeconomic status and health literacy (43). It likely impacts adherence to medication, prenatal care, income, and other factors relevant to disease activity and pregnancy outcomes (44,45). Finally, BMI affects both perceptions of disability and disease activity, as well as birth outcomes (46,47).

In addition, in the present study the data were stratified by 2 lupus-related parameters that impact pregnancy outcomes. A history of lupus nephritis is associated with higher rates of maternal complications, such as preeclampsia (48,49) and lupus flares, as well as poor pregnancy and fetal outcomes (49,50). APS antibodies have been shown to increase the risk of pregnancy loss, preterm birth, and maternal hypertension (10,38,51). The generalizability of this study may be limited, as the study population was composed of volunteers at a tertiary referral center. Women more prone to pregnancy loss may have never presented for evaluation, and so our data may underrepresent the rates of pregnancy loss, hampering our ability to detect a relationship between disease activity and loss. In addition, we possessed incomplete data on prior pregnancies, including pregnancy outcomes such as prior cesarean deliveries. We were thus unable to account for the effect of prior pregnancies and adverse pregnancy outcomes in our analysis. It is possible that the measures we used were not appropriate for measuring the types of disease activity that influence pregnancy outcomes in women with lupus. The most important predictors of lupus outcomes are signs not typically felt by a patient, including proteinuria, aPLs, thrombocytopenia, and hypertension (5,12). It appears that the HAQ, pain, and general health measures provide a window into the symptoms experienced by many women with lupus, but do not measure the signs that are key to pregnancy success.

In conclusion, our study demonstrated a clear difference in the utility of patient-reported measures in the pregnancies of women with SLE and RA. Among women with SLE, patient-reported measures of physical functioning, pain, and global

health did not correlate with adverse pregnancy outcomes, but the physician's assessment of lupus activity did provide a useful tool in prediction of preterm birth, cesarean delivery, and preeclampsia. Our findings suggest that our patient-reported measures of lupus activity were inadequate for identifying pregnancies at highest risk for pregnancy complications. In contrast, for women with RA, both the patient- and physician-reported measures were useful predictors of adverse pregnancy outcomes, suggesting that patient-reported measures may be useful adjuncts to the physician assessment in identifying and treating pregnancies at higher risk for adverse events.

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. Harris 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. Harris, Clowse.
Acquisition of data. Harris, Clowse.
Analysis and interpretation of data. Harris, Eudy.

REFERENCES

1. Clowse ME, Wallace DJ, Weisman M, James A, Criscione-Schreiber LG, Pisetsky DS. Predictors of preterm birth in patients with mild systemic lupus erythematosus. Ann Rheum Dis 2013;72: 1536–9.
2. Clowse ME, Magder LS, Witter F, Petri M. The impact of increased lupus activity on obstetric outcomes. Arthritis Rheum 2005;52: 514–21.
3. Barnabe C, Faris PD, Quan H. Canadian pregnancy outcomes in rheumatoid arthritis and systemic lupus erythematosus. Int J Rheumatol 2011. E-pub ahead of print.
4. Madazli R, Yuksel MA, Oncul M, Imamoglu M, Yilmaz H. Obstetric outcomes and prognostic factors of lupus pregnancies. Arch Gynecol Obstet 2014;289:49–53.
5. Clowse ME, Jamison M, Myers E, James AH. A national study of the complications of lupus in pregnancy. Am J Obstet Gynecol 2008;199:127. e1–e6.
6. Bharti B, Lee SJ, Lindsay SP, Wingard DL, Jones KL, Lemus H, et al. Disease severity and pregnancy outcomes in women with rheumatoid arthritis: results from the organization of teratology information specialists autoimmune diseases in pregnancy project. J Rheumatol 2015;42:1376–82.
7. Bowden AP, Barrett JH, Fallow W, Silman AJ. Women with inflammatory polyarthritis have babies of lower birth weight. J Rheumatol 2001;28:355–9.
8. Nørgaard M, Larsson H, Pedersen L, Granath F, Askling J, Kieler H, et al. Rheumatoid arthritis and birth outcomes: a Danish and Swedish nationwide prevalence study. J Intern Med 2010;268: 329–37.
9. Reed SD, Vollan TA, Svec MA. Pregnancy outcomes in women with rheumatoid arthritis in Washington State. Matern Child Health J 2006;10:361–6.
10. Ko HS, Ahn HY, Jang DG, Choi SK, Park YG, Park IY, et al. Pregnancy outcomes and appropriate timing of pregnancy in 183 pregnancies in Korean patients with SLE. Int J Med Sci 2011;8:577.
11. Chakravarty EF, Colón I, Langen ES, Nix DA, El-Sayed YY, Genovese MC, et al. Factors that predict prematurity and preeclampsia in pregnancies that are complicated by systemic lupus erythematosus. Am J Obstet Gynecol 2005;192:1897–904.
12. Clowse ME. Lupus activity in pregnancy. Rheum Dis Clin North Am 2007;33:237–52.
13. Cortés-Hernández J, Ordi-Ros J, Paredes F, Casellas M, Castillo F, Vilardell-Tarres M. Clinical predictors of fetal and maternal outcome in systemic lupus erythematosus: a prospective study of 103 pregnancies. Rheumatology 2002;41:643–50.
14. Petri M, Howard D, Repke J. Frequency of lupus flare in pregnancy: the Hopkins Lupus Pregnancy Center experience. Arthritis Rheum 1991;34:1538–45.
15. Ruiz-Irastorza G, Lima F, Alves J, Khamashta M, Simpson J, Hughes G, et al. Increased rate of lupus flare during pregnancy and the puerperium: a prospective study of 78 pregnancies. Rheumatology 1996;35:133–8.
16. Wong KL, Chan FY, Lee CP. Outcome of pregnancy in patients with systemic lupus erythematosus: a prospective study. Arch Intern Med 1991;151:269–73.
17. De Man YA, Hazes JM, van der Heide H, Willemsen SP, de Groot CJ, Steegers EA, et al. Association of higher rheumatoid arthritis disease activity during pregnancy with lower birth weight: results of a national prospective study. Arthritis Rheum 2009;60:3196–206.
18. Wallenius M, Salvesen KÅ, Daltveit AK, Skomsvoll JF. Rheumatoid arthritis and outcomes in first and subsequent births based on data from a national birth registry. Acta Obstet Gynecol Scand 2014;93:302–7.
19. Hench P. The ameliorating effect of pregnancy on chronic atrophic (infectious rheumatoid) arthritis, fibrositis, and intermittent hydrarthrosis. Mayo Clin Proc 1938;13:7.
20. Barrett JH, Brennan P, Fiddler M, Silman AJ. Does rheumatoid arthritis remit during pregnancy and relapse postpartum? Results from a nationwide study in the United Kingdom performed prospectively from late pregnancy. Arthritis Rheum 1999;42:1219–27.
21. De Man YA, Dolhain RJ, van de Geijn FE, Willemsen SP, Hazes JM. Disease activity of rheumatoid arthritis during pregnancy: results from a nationwide prospective study. Arthritis Care Res (Hoboken) 2008;59:1241–8.
22. Stucki G, Liang MH, Stucki S, Brühlmann P, Michel BA. A self-administered rheumatoid arthritis disease activity index (RADAI) for epidemiologic research. Arthritis Rheum 1995;38:795–8.
23. Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. Arthritis Rheum 1993;36:729–40.
24. Pincus T, Castrejón I. Are patient self-report questionnaires as "scientific" as biomarkers in "treat-to-target" and prognosis in rheumatoid arthritis? Curr Pharm Des 2015;21:241–56.
25. Jolly M, Pickard AS, Block JA, Kumar RB, Mikolaitis RA, Wilke CT, et al. Disease-specific patient reported outcome tools for systemic lupus erythematosus. Semin Arthritis Rheum 2012;42:56–65.
26. Jolly M, Utset T. Can disease specific measures for systemic lupus erythematosus predict patients health related quality of life? Lupus 2004;13:924–6.
27. Buyon JP, Kalunian KC, Ramsey-Goldman R, Petri MA, Lockshin MD, Ruiz-Irastorza G, et al. Assessing disease activity in SLE patients during pregnancy. Lupus 1999;8:677–84.
28. Holloway L, Humphrey L, Heron L, Pilling C, Kitchen H, Højbjerre L, et al. Patient-reported outcome measures for systemic lupus erythematosus clinical trials: a review of content validity, face validity and psychometric performance. Health Qual Life Outcomes 2014;12:116.
29. Oken E, Kleinman KP, Rich-Edwards J, Gillman MW. A nearly continuous measure of birth weight for gestational age using a United States national reference. BMC Pediatr 2003;3:6.

30. Bruce B, Fries JF. The Stanford health assessment questionnaire: dimensions and practical applications. Health Qual Life Outcomes 2003;1:20.
31. Anderson J, Caplan L, Yazdany J, Robbins ML, Neogi T, Michaud K, et al. Rheumatoid arthritis disease activity measures: American College of Rheumatology recommendations for use in clinical practice. Arthritis Care Res (Hoboken) 2012;64:640–7.
32. Camacho E, Farragher T, Lunt M, Verstappen S, Bunn D, Symmons D. The relationship between post-onset pregnancy and functional outcome in women with recent onset inflammatory polyarthritis: results from the Norfolk Arthritis Register. Ann Rheum Dis 2010;69:1834–7.
33. De Man YA, Hazes JM, van de Geijn FE, Krommenhoek C, Dolhain RJ. Measuring disease activity and functionality during pregnancy in patients with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2007;57:716–22.
34. Petri M, Hellmann D, Hochberg M. Validity and reliability of lupus activity measures in the routine clinic setting. J Rheumatol 1992;19:53–9.
35. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology 1999:37–48.
36. Weng HY, Hsueh YH, Messam LL, Hertz-Picciotto I. Methods of covariate selection: directed acyclic graphs and the change-in-estimate procedure. Am J Epidemiol 2009;169:1182–90.
37. Wilson WA, Gharavi AE, Koike T, Lockshin MD, Branch DW, Piette JC, et al. International consensus statement on preliminary classification criteria for definite antiphospholipid syndrome: report of an international workshop. Arthritis Rheum 1999;42:1309–11.
38. Buyon JP, Kim MY, Guerra MM, Laskin CA, Petri M, Lockshin MD, et al. Predictors of pregnancy outcomes in patients with lupus: a cohort study. Ann Intern Med 2015;163:153–63.
39. Kaplowitz ET, Ferguson S, Guerra M, Laskin CA, Buyon JP, Petri M, et al. Contribution of socioeconomic status to racial/ethnic disparities in adverse pregnancy outcomes among women with systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2018;70:230–5.
40. Clowse ME, Grotegut C. Racial and ethnic disparities in the pregnancies of women with systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2016;68:1567–72.
41. Luo ZC, Wilkins R, Kramer MS, Fetal and Infant Health Study Group of the Canadian Perinatal System. Disparities in pregnancy outcomes according to marital and cohabitation status. Obstet Gynecol 2004;103:1300–7.
42. Hein HA, Burmeister LF, Papke KR. The relationship of unwed status to infant mortality. Obstet Gynecol 1990;76:763–8.
43. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011;155:97–107.
44. Blumenshine P, Egerter S, Barclay CJ, Cubbin C, Braveman PA. Socioeconomic disparities in adverse birth outcomes: a systematic review. Am J Prev Med 2010;39:263–72.
45. Parker JD, Schoendorf KC, Kiely JL. Associations between measures of socioeconomic status and low birth weight, small for gestational age, and premature delivery in the United States. Ann Epidemiol 1994;4:271–8.
46. Cnattingius S, Bergström R, Lipworth L, Kramer MS. Prepregnancy weight and the risk of adverse pregnancy outcomes. N Engl J Med 1998;338:147–52.
47. Sebire NJ, Jolly M, Harris J, Wadsworth J, Joffe M, Beard R, et al. Maternal obesity and pregnancy outcome: a study of 287,213 pregnancies in London. Int J Obes Relat Metab Disord 2001;25:1175.
48. Huong DLT, Wechsler B, Vauthier-Brouzes D, Beaufils H, Lefebvre G, Piette J. Pregnancy in past or present lupus nephritis: a study of 32 pregnancies from a single centre. Ann Rheum Dis 2001;60:599–604.
49. Wagner SJ, Craici I, Reed D, Norby S, Bailey K, Wiste HJ, et al. Maternal and fetal outcomes in pregnant patients with active lupus nephritis. Lupus 2009;18:342–7.
50. Rahman FZ, Rahman J, Al-Suleiman SA, Rahman MS. Pregnancy outcome in lupus nephropathy. Arch Gynecol Obstet 2005;271:222–6.
51. Julkunen H, Jouhikainen T, Kaaja R, Leirisalo-Repo M, Stephansson E, Palosuop T, et al. Fetal outcome in lupus pregnancy: a retrospective case-control study of 242 pregnancies in 112 patients. Lupus 1993;2:125–31.
52. Textor J, van der Zander B, Gilthorpe MS, Liskiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. Int J Epidemiol 2016;45:1887–94.

AMERICAN COLLEGE OF RHEUMATOLOGY
EDUCATION • TREATMENT • RESEARCH

Arthritis Care & Research
Vol. 71, No. 3, March 2019, pp 398–405
DOI 10.1002/acr.23599

Poverty, Neighborhoods, Persistent Stress, and Systemic Lupus Erythematosus Outcomes: A Qualitative Study of the Patients' Perspective

Edward Yelin,[1] Laura Trupin,[2] Jared Bunde,[2] and Jinoos Yazdany[2]

Objective. To obtain the perspective of individuals with systemic lupus erythematosus (SLE) regarding the role of poverty, neighborhood, and chronic stress in SLE outcomes.

Methods. Through annual structured interviews as part of the Lupus Outcomes Study, 723 persons with SLE were followed from 2003 to 2015 in order to establish the effect of combinations of poverty, persistent poverty, residence in an area of concentrated poverty, access to health care, and chronic stress on accumulated damage. We obtained a sample of 28 of the 723 individuals on the basis of household income, geography, and outcomes in their last interview, and administered qualitative interviews to explore their perspectives on the impact of these factors on SLE outcomes. The interviews were recorded, transcribed, and analyzed using a grounded theory approach.

Results. Persons in poverty frequently reported that poverty necessitated a choice to deal with food, medical care, and housing insecurity on a daily basis and to relegate their management of SLE to occurrences of disease flares. They also reported that exposure to crime in their neighborhoods was a stressor that triggered worse disease activity. Affluent participants reported that neighborhood neither helped nor hindered dealing with SLE, because they relied on networks not tied to neighborhoods to deal with SLE.

Conclusion. Mitigating poverty and reducing exposure to crime through moving to safer neighborhoods are factors identified by individuals with SLE as potentially critical in disease outcomes.

INTRODUCTION

The relationship between low income and poor disease outcomes has been established in numerous studies of the general population (1,2) and in those of individuals with specific clinical conditions, including systemic lupus erythematosus (SLE) (3–7). In a prior study (8), we used a national longitudinal sample of persons with SLE to establish that current poverty and persistent poverty at one point predict the extent of accumulated damage five years later, with the effect accentuated among the poor living in areas of concentrated poverty and lessened among those who permanently exit poverty. Additionally in our previous study, we demonstrated the extent to which individuals' perceived stress over the last month (9) accounted for a significant amount of the effect of poverty on disease damage.

However, there were several questions that emerged from the present study, including what was it about poverty that resulted in higher levels of damage (including but not limited to the effects of differential access to the health care system); what were the kinds of stressors that were particularly salient in the lives of those with SLE; what were the phenomena that intensified or reduced the effects of stress for poor and nonpoor alike, or that had a greater effect on the poor; and what were the detrimental and beneficial aspects of the neighborhood in helping individuals cope with SLE.

Because many of these questions had not previously been addressed systematically in the SLE literature, we embarked on an exploratory analysis using qualitative interviews with 28 persons with SLE who were systematically sampled from the larger study, in order to provide insight into the specific processes that may result in higher levels of disease damage.

Dr. Yelin's work was supported by an Investigator in Health Policy award from the Robert Wood Johnson Foundation and the NIH (grants P60-AR-053308 and 2R01-AR-056476 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases). Dr. Yazdany's work was supported by the NIH (grant K23-AR-060259 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases).

[1]Edward Yelin, PhD: University of California, San Francisco and University of California, Berkeley; [2]Laura Trupin, MPH, Jared Bunde, RN, MS, Jinoos Yazdany, MD, MPH: University of California, San Francisco.

Address correspondence to Edward Yelin, PhD, University of California, San Francisco, Box 0920, San Francisco, CA 94143-0920. E-mail: ed.yelin@ucsf.edu.

Submitted for publication November 15, 2017; accepted in revised form May 15, 2018.

SIGNIFICANCE & INNOVATIONS

- Prior studies have established that personal poverty and living in areas of concentrated poverty were associated with increased levels of disease damage in systemic lupus erythematosus (SLE) and that access to and quality of medical care were not the principal reasons for the increased damage.
- To our knowledge, this study is the first to document that attention to securing adequate food, medical care, and housing relegates focusing on disease to a secondary concern, except at times of disease flares.
- To our knowledge, this study is also the first to document the extent to which the experience and/or fear of crime is a stressor that affects persons in poverty who have SLE.

MATERIALS AND METHODS

Overview. In the previous study, broad themes relating poverty and living in an area of concentrated poverty to heightened damage were identified, but not the specific triggers. In such instances, qualitative interviews can be helpful in uncovering what individuals living with SLE perceive to be the reasons that low income or adverse neighborhood circumstances may lead to damage, along with the impact of differential experiences with the health care system. We developed a series of open-ended questions that delved into these issues and then used a grounded theory approach to extract meaning from the interview data (10).

Data source. The data source for the research was the Lupus Outcomes Study (LOS) (11). The LOS began in 2003, enrolling individuals with confirmed SLE diagnoses who had previously participated in genetics studies of lupus and had been recruited nationally from a combination of clinical and nonclinical sources. The principal data collection for the LOS was an annual structured telephone interview that covered the status of the SLE, enumeration of medications and health care, and standard demographic items. The content of the annual survey has been fully described in prior publications (7,11,12).

LOS participants reported on their household income in each annual survey, which, when combined with data on household size, enabled us to categorize each of them into those whose household income was at or below 125% of the federal poverty level versus above this level, which is the study definition of poverty (7). The extent of poverty in the participants' local neighborhoods was obtained by matching their geocoded addresses to information from the American Community Survey (13,14) at the level of the census block group, which encompassed between 600 and 3,000 individuals in the immediate neighborhood. We defined areas of concentrated poverty as those in which >30% of households were at or below the poverty level (15).

Sampling. Sampling for the qualitative interviews was based on a combination of specific criteria that were identified in the prior quantitative study (as outlined below) and purposive sampling based on an assessment by the study interviewers to determine which respondents to the structured annual survey would likely be informative respondents (16).

There were 723 respondents to the 2014–2015 LOS annual survey. From those 723 individuals, we limited the cohort of those targeted for qualitative interviews to those defined as poor (n = 116) or who reported household incomes of ≥$100,000 (n = 94) to highlight the impact of monetary resources on access to care and outcomes. We further stratified the individuals in poverty into those who did (n = 50) and did not (n = 66) live in areas meeting the study definition of concentrated poverty. For both the poverty and affluent groups, we included those who were at or above the highest quartile of disease damage versus below, as measured by the Brief Index of Lupus Damage (17,18). To ensure adequate geographic diversity, we also oversampled among respondents living outside of California. Finally, among those who were eligible on the basis of income, residence in areas of concentrated poverty, degree of disease damage accumulated, and residence inside and outside of California, we set a higher priority on respondents designated by the interviewers as likely being informative.

Because of the long history of participation in the LOS annual quantitative interviews, all patients we asked to participate in the qualitative interviews agreed to participate; we did not have to go beyond those targeted as being potentially informative respondents. Qualitative interviews were completed with 28 individuals, 11 from the affluent group, and 17 from the poverty group (among whom 11 lived in areas of concentrated poverty). Twelve of the 28 individuals who completed qualitative interviews lived outside of California (5 from among the affluent group and 7 from the poverty group).

After obtaining verbal consent from each participant, interviews were conducted by a single individual (JB) and were recorded and transcribed. The study was approved by the University of California, San Francisco Institutional Review Board.

Content and conduct of qualitative interview. The content of the interviews was guided by a list of major themes to be covered, in order to ensure that all participant responses accounted for the same material. The themes were drawn from the findings of the quantitative study about the relationship among income, neighborhood poverty, stress, and accumulated damage. There were 2 overarching themes. The first theme concerned health care for individuals with SLE, including gaining access to SLE care, maintaining access, and maneuvering through the health care system. The second theme concerned the stressors that affect the status of their SLE. Also, we explicitly asked about the negative and positive aspects of their neighborhoods in regard to access to SLE care or the status of their SLE.

We conducted a pretest of the interview protocol with study investigators and with several SLE clinic patients who were not LOS participants. Analysis of the pretest interviews resulted in the revised interview guide shown in Table 1. Interviews conducted with this guide took 45–60 minutes to complete.

Analyses. The broad themes of the interviews derived from the prior quantitative study, but in the analysis we used a grounded theory approach to capture and understand information on processes that differentiate the impact of low income, experience and meaning of stress, and beneficial and detrimental effects of neighborhood on SLE damage. In order to accomplish the goal of highlighting the processes that translate these characteristics into adverse disease outcomes, we systematically reviewed the interview content to ensure that we had captured the universe of responses about the processes (e.g., confirming that no other specific aspect of neighborhoods were being mentioned as having positive or negative effects). We then sought to see if the responses adhered to specific patterns, for example, certain responses about specific phenomena being more common among the poor might suggest a potential explanation for why the poor experienced heightened levels of damage (19).

To begin the process of combing the data, the interviewer summarized the responses of each respondent at the conclusion of the interview as they related to the study themes. One group representative (EY) listened to the recordings, reviewed the transcripts, and then coded the responses for the health care theme (whether the individual was able to gain access to health care and maintain it) and the stress theme. After review of the audio tapes and transcripts, respondents' comments about stressors were coded to those traditionally associated with the disease in qualitative studies, such as the uncertainty of when flares will occur, dealing with the impacts of organ manifestations and generalized fatigue and pain, and the effect on functioning in and outside the home (20) or "daily hassles," as one respondent described. However, given the study's focus on the impact of poverty, we also coded responses that attributed SLE status to socioeconomic status and financial needs (such as worries about paying for food, rent, or medical care) and to adverse neighborhood conditions. As a check on the interpretation of the interview content and coding, the interviewer's contemporaneous summary of the interview was compared to the systematic review of the transcript and audiotape for the gist of the information derived from the interview and resolution of discrepancies, if necessary.

RESULTS

Table 2 describes the characteristics of the 28 study participants and shows the diversity in demographics and extent of SLE. On average, respondents were slightly younger than 50 years old and had SLE for more than 2 decades. Most respondents were women, and more than two-thirds were members of racial and

Table 1. Interview guide for patients with systemic lupus erythematosus (SLE)

1. Health care theme: Gaining and maintaining access to health care in general and principal provider specifically for SLE
   a. How long between first symptoms and diagnosis?
   b. How long between diagnosis and reaching steady state of a treatment plan with health care providers?
      i. Regular source of care
      ii. Principal provider of care
   c. How did you get to the regular source of care and a principal provider of care?
      i. What were fits and starts?
      ii. How much was done yourself and how much by involving others?
         (1) Others who knew people who had SLE or other serious conditions?
         (2) Others who provided tangible and intangible support?
   d. Once in system, what has worked about the regular source of care and principal provider and what not?
      i. Insurance issues
      ii. Accessibility of care
      iii. Communication/coordination
         (1) Falling between cracks
      iv. Shared decision-making
   e. What about SLE makes it difficult to get the right doctor and stay with him or her?
2. Living with SLE theme: Description for someone newly diagnosed with SLE, what makes this illness difficult for you
   a. In your mind, what triggers a flare in your lupus?
      i. What are the stresses that you know will make your lupus worse?
      ii. What helps to keep it in check?
   b. How do you get help from your family and friends to deal with these issues of having SLE?
      i. What do they do to help?
      ii. What do you wish they could do for you but don't?
   c. Moving beyond your immediate family and friends, what in your neighborhood helps you deal with lupus and what gets in the way?
      i. Are there a lot of people with whom you can talk about your lupus?
      ii. Do these people help you deal with your lupus by doing things for you or talking things through with you?
      iii. What in the neighborhood got in the way of getting connected to the regular source of care and main lupus doctor?
      iv. What in the neighborhood helps you deal with your lupus day by day and what is a source of stress?
         (1) What in this neighborhood helps reduce the stress of having lupus or at least helps you deal with your lupus?
         (2) What role if any does your neighborhood play in triggering lupus flares?
3. Open-ended closing, to determine if there are any issues not covered by the health care and living with SLE themes
   a. What would be the ideal physician to help you manage your lupus? What would that physician do for you?
   b. What do you need to help you deal with your lupus?
   c. What would you say to people who ask what they or your friends and family can do to help you deal with your lupus?
   d. What would you say to people who ask what would be a good neighborhood for someone with lupus to live in?

Table 2. Characteristics of study respondents (n = 28)*

Age mean ± SD (range), years	49.1 ± 12.1 (22–70)
Disease duration mean ± SD (range), years	20.5 ± 7.6 (5–23)
Female	25
Race/ethnicity	
White (not Hispanic)	8
Hispanic of any race	4
African American	10
Asian American	4
Other races/ethnicities	2
Disease activity in worst quartile†	8
Disease damage in worst quartile‡	14
High level of depressive symptoms§	8

* Values are the number of patients, unless indicated otherwise.
† Systemic Lupus Activity Questionnaire score ≥15.
‡ Brief Index of Lupus Damage (17,18) score ≥5.
§ Center for Epidemiologic Studies Depression Scale score ≥24.

ethnic minorities (as a result of oversampling among the poor and those living in areas of concentrated poverty.

Access to and maintenance of SLE care. The results of the qualitative interviews after coding are shown in Table 3. With respect to health care access (according to income status), half of the 28 respondents reported obtaining a diagnosis of SLE and gaining access to SLE care within 6 months, and 20 reported maintenance of access and the ability to maneuver through the health care system to get the care needed after ultimately gaining initial access (although as shown below, not without some challenges).

There were no appreciable differences according to poverty status, those who do and do not live in areas of concentrated poverty, and in those reporting high versus low levels of damage in gaining access to SLE care within 6 months. However, it appears that those at or below poverty were slightly less likely than the affluent to maintain and successfully maneuver through the health system thereafter (10 of 17 of the poor versus 10 of 11 of the affluent).

Even though many study participants gained health care access quickly and maintained it over time, initial access and then maintenance of that access presented challenges, as one participant described. "I was seeing a specialist and then I went to not having insurance and not being able to go to the specialist. So, I started going to the community clinic, where it's income based…" (poor woman living in area of high concentrated poverty in rural California).

Another participant noted a loss of insurance before getting Medicaid. "When I was working at the hospital it was right before I was getting ready to be hired on full time. Then, they diagnosed me. Then, next thing I know I was put on SSI [Supplemental Security Income] and I got Medi-Cal [the California Medicaid Program] and food stamps for my children, then everything started going downhill and I started surgeries and getting sicker and sicker. In and out of the hospital. The kids thought I was going to die" *(*poor woman living in middle class area near Central Valley city in California).

Additional challenges were noted by another respondent. "And, not faulting the physicians, but that's the way our healthcare is structured right now. You can't describe your story, or even halfway give the details to put it all together to make some sense, or to start doing — for the doctors to take the time to think back to medical school, what they did and what they learned. Now, it's like, oh, okay, if I do this, we'll give you this and you're out the door. See you in two weeks" *(*affluent woman living in suburban Arizona).

Traditional and socioeconomic stressors. Table 3 also includes information on the frequency with which respondents reported traditional or specific socioeconomic stressors. The report of traditional stressors as triggers for SLE flares did not differ among the groups defined by poverty, residence in a high poverty area, or extent of damage; indeed, all 28 respondents were coded as indicating that both environmental triggers (such as sun exposure) or common daily hassles (such as family conflict, overwork, difficult commutes, etc.) affected their SLE.

Only one of the affluent respondents reported socioeconomic stressors, which in this case was difficulty in paying for medical care despite having insurance. In contrast, two-thirds of the poor respondents mentioned such stressors. When further dividing the poor into those living in and out of areas of concentrated poverty, it did not appear that the poor outside of such areas were able to escape socioeconomic stressors. Further analysis of the 2 groups and comparison to those in the higher income category indicates the reasoning behind why the poor living outside of concentrated poverty areas still experienced socioeconomic stressors. Although the poor inside areas of concentrated poverty lived in areas in which an average of 49% of the residents were poor, the poor outside those areas lived in areas in which an average of 18% were poor, and certainly many lived adjacent to areas with higher concentrations of the poor. Thus, the residential environments of the 2 groups of the poor were not as different as we had anticipated. Contrastingly, the affluent respondents lived in areas in which an average of only 8% of the residents were poor (data on concentrations of the poor not shown).

Several respondents spoke to the inherent stress of facing monetary uncertainty. "Money stresses make it [the SLE] worse. Definitely. Because I worry about money all the time and how my brother's paying my mortgage and how I wish it was me. How I wish I could be normal and work" (aforementioned poor woman living near Central Valley city).

Another participant similarly noted the inherent stress. "It's rough. I pay like $50 a month—whenever I get my check—$50 a

Table 3. Experiences of persons with systemic lupus erythematosus (SLE) in health care access and stressors that affect disease status, 2015*

	Health care issues			Stressors mentioned			
Group	No.	Gained SLE care access in <6 months	Maintained SLE care access*	No.	Traditional stressors	Socioeconomic stressors	Adverse neighborhood effects
Income							
High (household ≥$100,000)	11	6	10	11	11	1	0
In poverty	17	8	10	17	17	11	13
Residence (among those in poverty)							
Areas of concentrated poverty	11	6	5	11	11	6	8
Not in areas of concentrated poverty	6	2	5	6	6	5	5
Disease status							
Highest quartile of damage	14	6	10	14	14	3	3
Lowest three quartiles of damage	14	8	10	14	14	7	10

* Values are the number of participants. Access to SLE care required maneuverability in some instances.

month [for the motorized cart] is all I can afford to pay until it's paid. So, I'm going to go to a collection agency" (poor woman living in rural Indiana who relocated from more expensive city).

One respondent spoke of this uncertainty and the impact of money concerns on bandwidth. "No. No, it was a nightmare. Especially when you're dealing, you know, with cardiologists and pulmonologists and nephrologists and every other -ologist. Now and then, it gets a little bumpy…There's not a really good path" (poor woman living in rural Northwest, when asked about medical expenses). This participant, when asked about a better healthcare system, added, "Well, it certainly wouldn't cost as much… We ended up filing bankruptcy because of my medical bills."

Another study respondent further addressed monetary issues. "Well, I had to move in [with my family]— I had to leave my apartment, well, I lost everything" (poor woman living in rural Missouri who relocated from Chicago).

One participant stated that the cost of medical care may also affect the more affluent. "Well, right now I'm paying $215 every month for a hospitalization I had last June, because … I don't have $5,000 sitting around" (affluent woman living in exurban part of the San Francisco Bay area).

Many individuals with SLE are forced to delay dealing with anything other than with food, housing, and medical care insecurity due to monetary concerns. "This was all in June. Oh, there was one more — I can't think of it now, but I know there were five things on my plate and it kind of hit me spiraling down. So, I had to deal with them as they came" (a poor woman living in an area of concentrated poverty in a Bay Area central city). This particular respondent made a distinction between daily hassles and more chronic financial concerns. "I think pretty much any stress that will keep me worried about it a week later [is the] kind of stress. So, the long-term stress. Not the oh my God am I going to make it to work on time kind of stress…"

Another study participant suggested that financial concerns might translate into less than optimal monitoring of the SLE, even for someone of her means, envisioning a situation in which she didn't have to worry about bills. "It would be Dr. Moneybags, and he would pay my bills. And then I would go see my rheumatologist regularly. … For them to check on my lupus and stuff, those special tests cost a fortune. And I have to pay 40 or 60%, or whatever my fricking percentage is. It's a lot" (aforementioned affluent woman living in exurban San Francisco Bay Area location).

The starkest distinction between affluent and poor study participants, regardless of whether they resided in an area of concentrated poverty, had to do with the role of neighborhood in their lives. For the affluent, neighborhood played no role in dealing with SLE. Indeed, most relied on sources outside of their neighborhoods, such as relatives or professional connections, to find and maintain health care or deal with their disease. "Well, my wife's a nurse. So she knew that from her knowledge that Brigham and Women's Hospital in Boston had a lupus center…and my father is a doctor who recommended I get a second opinion there" (an affluent participant, New England). Professional connections and willingness to persevere were cited by one respondent. "Yes, I'm an advocate but I'm also a nurse and counselor. That's how I get to the care I need" (woman living in Arizona). With respect to neighborhood, another participant responded similarly to others in the affluent group in indicating that she knew that living in a good neighborhood could be helpful but that it wasn't something she relied on. "Well, I guess I'd have to say, it's a very quiet

neighborhood, so … I don't think the neighborhood contributes to bringing on any kind of lupus symptom" (woman living in an affluent Boston suburb).

However, the study participants who were poor repeatedly mentioned the experience of crime, the fear of crime, and being witness to illicit activities as primary factors affecting the status of their SLE. Participants noted stressors that made their SLE worse. "You know, there's been [three] instances where they's try to break in my house and that just threw me for a loop. I was highly agitated by that… the latest one was the one where they knocked on the door— they knocked on the screen door" *(*poor woman living in area of concentrated poverty in large Midwestern city). Another participant also described stressors that impacted her illness. "Oh, God, the sirens — I hear one now. It's constant, now. … I hear a lot of police. A lot of killings over here, robberies. So, when you do come in at night, it's the skids — like, 'OK, can I make it in here safely, without being mugged or robbed?'" (poor woman living in area of concentrated poverty in another Midwestern city).

Noting the difficult experience in her current neighborhood, another study participant described her vision of a better neighborhood. "Helps reduce the stress or helps me deal [with the SLE]? Here you hear the hustle and bustle. Whereas when I go to a neighboring city it's more quiet. It's just — you don't feel like you're surrounded by all that chaos" (woman from a San Francisco Bay Area central city).

The poor who lived in rural areas were not immune from the adverse effects of neighborhood on SLE, as one study respondent addressed. "I live in a trailer park…. We would be considered to be living in the best of that part of town — for a town of 6,000 people, I mean. Yeah. And do [I] feel safe where [I] live? I have a loaded .357 on my nightstand. Yes, I do" (woman from the rural Northwest).

Although the actual experience or fear of crime was mentioned repeatedly, separating the impact of adverse neighborhoods from socioeconomic stressors is difficult. Ironically, the woman living in an area of concentrated poverty in a Midwestern city who was quoted previously indicated that, at long last, her neighborhood was beginning to change for the better. However, she experienced the effects of gentrification as a mixed blessing, knowing that the experience of crime may be lessened, but also knowing that gentrification may not serve the current residents because she will have to leave the neighborhood for one that is more affordable. "Well, you know, from what I'm seeing, I see businesses moving back in. And, you know, I see a lot of things coming back in. But I don't think it's geared towards the population of the neighborhood."

DISCUSSION

In the approach of behavioral science, asking people about their experiences may yield less insight than observance of their behavior over time. The reason that the revealed preference of observing behavior may be more reliable than asking individuals about their behavior (21) is based on the notion that individuals may not be able to understand why they do things or may say things to legitimate their behavior. Our previous study (8) was conducted with a behavioral science approach and demonstrated that current poverty, duration of poverty, and exiting poverty could explain the extent of subsequent accumulated damage in SLE, that living in areas of concentrated poverty accentuated the effect of poverty on disease damage, and that higher levels of perceived stress accounted for much of the effect of poverty.

There were limits to the behavioral science approach, however. We knew from the quantitative study (8) that poorer access to SLE care, lower quality of care, and poorer interactions with health care providers and health systems mattered in determining SLE health outcomes, but not all that much. We also knew that differences in health behaviors between the poor and nonpoor, such as smoking, played a relatively small role. We knew that stress, as evaluated by standardized measures, had a substantial effect on SLE health outcomes, but not whether those were actually the kinds of stresses that individuals experienced as triggers for flares in SLE. Finally, the literature on the health effects of neighborhoods has posited that the presence of positive things like parks and accessible transportation and the absence of negative things like noise and crime may differentially affect the poor and nonpoor, with the nonpoor experiencing few of the deleterious impacts of neighborhoods and the poor experiencing few of the good ones (22).

The results of the qualitative interviews reinforced the observation from the prior study about the role of health care. Participants who were affluent and those who were poor reported similar experiences in getting to a steady state in SLE care, although the affluent were more likely to state that they were able to maintain that steady state and successfully maneuver through the system to obtain the care that they needed. Both groups held similar views regarding the impact of traditional stressors for persons with SLE, including environmental challenges such as sun exposure, and daily hassles like difficult commutes, overwork, and family strains. Although both groups cited these stressors, they stated that the impact of traditional stressors is as much because their providers or health education materials have told them to reduce such exposures because the individuals with SLE themselves are central in their own perceptions of what matters in their SLE.

However, when asked about stresses that affect the status of their SLE, the poor were much more likely to report that socioeconomic stressors such as housing, food, and medical care insecurity affect them (with the exception of one affluent woman with high out-of-pocket payments for medical care services). The respondents indicated how these concerns weighed on them, forcing them to focus all attention on securing these basic necessities which, in turn, led them to focus on SLE symptoms only when they must. Their insights are consistent with a burgeoning amount of health economics literature that focuses on the scarcity of time and energy in a manner analogous to the

traditional focus of economics on the scarcity of financial wherewithal (23) and with the sociologic literature that documents the stress associated with housing insecurity (24).

The observation from our previous study of the accentuating effect of living in areas of concentrated poverty, beyond the effect of personal poverty, on damage accumulation led us to ask respondents about the salutary and adverse aspects of their neighborhoods. In contrast to what we expected to hear, the results of the qualitative interviews suggested that, for the affluent neighborhoods, damage accumulation mattered very little. Many of the affluent had secured access to health providers through connections through work or family scattered across the country, and none mentioned positive aspects of living in good neighborhoods as a relief from SLE symptoms or daily hassles. For the study participants who were poor, none mentioned positive aspects of neighborhoods as a way of dealing with SLE care or stresses that affected their SLE status. But many cited the actual experience of crime or the fear of crime as an important stressor or an experience that affected their well-being. Of course, in retrospect, the differences between the affluent and poor were consistent with the observation made repeatedly over the last century by sociologists, that some of us live in networks not defined by space, but by social connections (referred to by classical sociologists as "gemeinschaft," or by society and contemporary analysts as "anywhere" people) while others do, for better or worse (what the traditional sociologists referred to as "gesellschaft," or community and contemporary analysts call "somewhere" people) (25–28). However, the observation that the local community did not provide a haven in any way, shape, or form is consistent with the contemporary understanding of the impact of communities of concentrated poverty with respect to social phenomena in general (29) and the health effects for persons with SLE in particular (7,8).

The exposure to crime combined with the chronic stress of food, shelter, and medical care insecurity would appear to be the major factors that differentiated the viewpoints of poor and affluent people about SLE. These phenomena are fundamentally different than either the forms of stress psychologists catalog in the laboratory, such as being asked to give a speech in public or daily hassles due to family conflicts, difficult commutes, or even periodic overwork. The fact that these factors were mentioned in response to the same series of questions asked of the affluent and poor participants indicates both that they were not mentioned as a result of "leading" the conversation and that they were truly salient to the poor.

There is increasing evidence in research (not specifically focused on persons with SLE) that providing the poor with the means of moving to areas with lower concentrations of the poor, to better neighborhoods, redounds to them in terms of educational achievement and, down the road, higher earnings (30). There are preliminary data indicating that there can also be health benefits (31,32). However, helping the poor to move to better neighborhoods through housing vouchers may have the paradoxical effect of rendering them more "house poor," unless the housing subsidies are sufficient to absorb the higher housing costs. If not, then reducing the stress associated with exposure to crime by providing housing vouchers may increase housing insecurity, another key source of stress (24).

We began this inquiry with the supposition that the poor experienced a higher level of daily hassles, fewer of the beneficial aspects of neighborhoods, and more of the harmful ones. In response to the guided interviews for the present study, it is clear that the poor view chronic stress as manifest in food, housing, and medical care insecurity, and exposure to crime as being crucial to their experience of SLE. Neither daily hassles of living with a chronic illness nor the absence of positive aspects of neighborhoods were identified as producing our hypothesized results.

The thoughts of the persons with SLE with whom we conducted these extensive qualitative interviews may also help to explain why medical care doesn't play a more central role in why the poor have more adverse outcomes (as respondents reported that they are largely able to initially access and then maintain that access to care), and why the stresses associated with sustained financial insecurity and exposure to neighborhoods with high rates of crime do (as they are not able to focus on the SLE until they cannot do otherwise).

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. Yelin 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. Yelin, Trupin, Bunde, Yazdany.
Acquisition of data. Yelin, Bunde.
Analysis and interpretation of data. Yelin, Trupin, Yazdany.

REFERENCES

1. Raphael D. Social determinants of health: an overview of key issues and themes. In: Raphael D, editor. Social Determinants of Health. Toronto, Canada: Canadian Scholars' Press; 2009. p. 2–19.
2. Glymour M, Avendano M, Kawachi I. Socioeconomic status and health. In: Berkman L, Kawachi I, Glymour M, editors. Social epidemiology. New York: Oxford University Press; 2014. p. 17–62.
3. Karlson EW, Daltroy LH, Lew RA, Wright EA, Partridge AJ, Roberts WN, et al. The independence and stability of socioeconomic predictors of morbidity in systemic lupus erythematosus. Arthritis Rheum 1995;38:267–73.
4. Cooper GS, Treadwell EL, St.Clair EW, Gilkeson GS, Dooley MA. Sociodemographic associations with early disease damage in patients with systemic lupus erythematosus. Arthritis Rheum 2007;57:993–9.
5. Alarcón GS. Lessons from LUMINA: a multiethnic US cohort. Lupus 2008;17:971–6.
6. Demas KL, Costenbader KH. Disparities in lupus care and outcomes. Curr Opin Rheumatol 2009;21:102–9.
7. Trupin L, Tonner MC, Yazdany J, Julian LJ, Criswell LA, Katz PP, et al. The role of neighborhood and individual socioeconomic status in outcomes of systemic lupus erythematosus. J Rheumatol 2008;35:1782–8.

8. Yelin E, Trupin L, Yazdany J. A prospective study of the impact of current poverty, history of poverty, and exiting poverty on accumulation of disease damage in systemic lupus erythematosus. Arthritis Rheumatol 2017;69:1612–22.
9. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983;24:385–96.
10. Foley G, Timonen V. Using grounded theory method to capture and analyze health care experiences. Health Serv Res 2015;50:1195–1210.
11. Yelin E, Trupin L, Katz P, Criswell L, Yazdany J, Gillis J, et al. Work dynamics among persons with systemic lupus erythematosus. Arthritis Rheum 2007;57:56–63.
12. Tonner C, Trupin L, Yazdany J, Criswell L, Katz P, Yelin E. Role of community and individual characteristics in physician visits for persons with systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2010;62:888–95.
13. US Census Bureau. Design and methodology: American Community Survey. URL: https://www.census.gov/content/dam/Census/library/publications/2010/acs/acs_design_methodology.pdf.
14. Bee C. The geographic concentration of high-income households: 2007–2011. US Census Bureau; 2013.
15. Bishaw A. Areas with concentrated poverty: 1999. Washington (DC): US Census Bureau; 2005. Census 2000 Special Reports No. CENSR–16.
16. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015;42:533–44.
17. Yazdany J, Trupin L, Gansky SA, et al. Brief index of lupus damage: a patient-reported measure of damage in SLE. Arthritis Care Res (Hoboken) 2011;63:1170–7.
18. Katz P, Trupin L, Rush S, Yazdany J. Longitudinal validation of the Brief Index of Lupus Damage. Arthritis Care Res (Hoboken) 2014;66:1057–62.
19. Hennink MM, Kaiser BN, Marconi VC. Code saturation versus meaning saturation: how many interviews are enough? Qual Health Research 2017;27:591–608.
20. Sutanto B, Singh-Grewal D, McNeil HP, et al. Experiences and perspectives of adults living with systemic lupus erythematosus: thematic synthesis of qualitative studies: living with SLE. Arthritis Care Res 2013;65:1752–65.
21. Samuelson PA. Consumption theory in terms of revealed preference. Economica. 1948;15:243–53.
22. Macintyre S, Ellaway A. Neighborhoods and health: an overview. In: Kawachi I, Berkman L, editors. Neighborhoods and Health. New York: Oxford University Press; 2003. p. 20–44.
23. Shah AK, Mullainathan S, Shafir E. Some consequences of having too little. Science 2012;338:682–5.
24. Desmond M. Evicted: poverty and profit in the American city. New York: Crown; 2016.
25. Tonnies F. Community and society. Mineola (NY): Dover; 2002. p. 1887.
26. Weber M. On economy and society. London: Routledge; 2010.
27. Webber MM. Order in diversity: community without propinquity. Baltimore: Johns Hopkins; 1963.
28. Goodhart D. The road to somewhere. New York: Oxford University Press; 2017.
29. Wilson W. The truly disadvantaged: the inner city, the underclass, and public policy. Chicago: University of Chicago Press; 1987.
30. Chetty R, Hendren N, Katz L. The effects of exposure to better neighborhoods on children: new evidence from the moving to opportunity experiment. Am Econ Rev 2016;106:855–902.
31. Ludwig J, Sanbonmatsu L, Gennetian L, et al. Neighborhoods, obesity, and diabetes: a randomized social experiment. N Engl J Med 2011;365:1509–9.
32. Ludwig J, Duncan GJ, Gennetian LA, et al. Neighborhood effects on the long-term well-being of low-income adults. Science 2012;337:1505–10.

AMERICAN COLLEGE OF RHEUMATOLOGY
EDUCATION • TREATMENT • RESEARCH

Arthritis Care & Research
Vol. 71, No. 3, March 2019, pp 406–412
DOI 10.1002/acr.23593

Sleep Disturbance and Depression Symptoms Mediate Relationship Between Pain and Cognitive Dysfunction in Lupus

Teresa A. Lillis,[1] Vanessa Tirone,[1] Nisarg Gandhi,[2] Stacy Weinberg,[1] Ailda Nika,[1] Winston Sequeira,[1] Stevan E. Hobfoll,[1] Joel A. Block,[1] and Meenakshi Jolly[1]

Objective. To determine whether sleep disturbance and symptoms of depression mediate the relationship between pain and cognitive dysfunction (CD) in a sample of 115 patients with systemic lupus erythematosus (SLE).

Methods. A total of 115 patients with SLE completed questionnaires regarding pain, perceived stress, depression, sleep, and CD. Relationships among pain, sleep, depression, and CD were assessed using bootstrap mediation models, controlling for race/ethnicity, fibromyalgia diagnosis, current corticosteroid use, disease activity and damage, and perceived stress.

Results. Mediation analyses indicated that the effect of pain on CD was mediated by sleep disturbance ($\beta = 0.30$) and depression symptoms ($\beta = 0.33$). These effects were maintained even after controlling for the aforementioned covariates, of which only disease activity ($\beta = 0.20$) and stress ($\beta = 0.22$) remained significantly linked to CD (overall model $R^2 = 0.53$; all $P < 0.05$).

Conclusion. After controlling for disease activity and perceived stress, the relationship between pain and CD was explained by sleep disturbance and depression symptoms. Although these relationships need validation in longitudinal studies with additional measurement modalities, our findings may indicate promising, nonpharmacologic intervention avenues for SLE patients with pain and CD. Specifically, cognitive behavioral therapies for depression and sleep are known to reduce distress and enhance functioning across various psychosocial domains. Given the symptom burden of SLE, interventions that maximize potential benefits without the use of additional pharmacologic treatments may be of particular utility.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a chronic autoimmune disease that disproportionately affects young women (ages 15–43 years), ethnic and racial minorities, and individuals of lower socioeconomic status (1). SLE is characterized by a number of biomedical and neuropsychiatric symptoms that have been associated with reduced quality of life (QoL) and increased work-related disability. Recent estimates suggest that 15–51% of patients with SLE terminate employment within 2–15 years after diagnosis, and more than half (60%) are unemployed 20 years after diagnosis (2). As such, estimates of annual direct costs per patient due to SLE range from $13,494 to $55,344, while indirect costs have been estimated to range from $1,252 to $20,046 (1,3). Individual indirect costs have superseded direct costs in some cases due to the impact of the associated symptom profile of SLE (e.g., pain, fatigue, depression, impaired cognition) on work-related disability (1,2). SLE patients report significantly reduced QoL across physical, mental, occupational, and social functioning domains, which often exceed reductions in QoL associated with other, more common chronic diseases, including hypertension, congestive heart failure, and major depression (4,5).

Given the cumulative symptom burden of SLE and associated challenges of traditional biomedical treatment approaches, clinicians and researchers have begun to approach SLE from a biopsychosocial framework and identify behavioral and psychological symptom targets for intervention-related strategies (6). Of the numerous behavioral and psychological symptoms associated

Supported by a grant from The Cornell Brewer Foundation.
[1]Teresa A. Lillis, PhD, Vanessa Tirone, PhD, Stacy Weinberg, MD, Ailda Nika, MD, Winston Sequeira, MD, Stevan E. Hobfoll, PhD, Joel A. Block, MD, Meenakshi Jolly, MD: Rush University Medical Center, Chicago, Illinois; [2]Nisarg Gandhi, MD: John H. Stroger Hospital, Chicago, Illinois.

Dr. Jolly is a coinventor of the LupusPRO tool. Rush University Medical Center and the University of Illinois at Chicago hold copyrights to LupusPRO.

Address correspondence to Teresa A. Lillis, PhD, Rush University Medical Center, Dept. of Behavioral Sciences, 1645 West Jackson Boulevard, Suite 400, Chicago, IL. E-mail: teresa_a_lillis@rush.edu.

Submitted for publication December 10, 2017; accepted in revised form April 24, 2018.

SIGNIFICANCE & INNOVATIONS

- Cognitive dysfunction (CD) and pain are common symptoms among patients with systemic lupus erythematosus (SLE) and adversely affect quality of life and productivity.
- Sleep issues and depression, also commonly seen in patients with SLE, may impact both pain and CD.
- Through the utilization of mediation models, we found that sleep disturbance and depression symptoms accounted for the relationship between pain and CD, even after controlling for disease activity and stress.
- Given the symptom burden of SLE, our findings may indicate promising, nonpharmacologic intervention avenues, including cognitive behavioral therapies for depression and sleep, for patients with SLE and concurrent pain and CD.

with SLE, approximately 20–80% of patients with SLE report experiencing cognitive dysfunction (CD) (7,8). The American College of Rheumatology (ACR) defines CD in the context of SLE as any significant deficit in complex attention, executive functioning, memory, visual-spatial processing, language, or processing speed (9). CD has been associated with a number of deleterious outcomes in patients with SLE, including reduced QoL and increased work-related disability (10). Currently, there is no consensus on the etiology of CD in SLE, and this has been cited as a contributing factor to the lack of effective interventions for CD in SLE populations (11). Although previous studies have assessed the influence of other common SLE-related symptoms (e.g., anxiety), treatments (e.g., corticosteroid use), and disease factors (e.g., activity and severity) on CD outcomes, thus far, the majority of these studies have failed to establish reliable associations with CD outcomes that would indicate promising targets for intervention (12).

To our knowledge, no previous study has concurrently investigated the roles of pain, sleep disturbance, and depression symptoms in SLE-related CD outcomes. Pain is one of the most commonly reported symptoms of patients with SLE, with 65% reporting pain as the most difficult aspect of their disease to manage (13,14). Pain is also a well-established correlate of impaired cognitive functioning in the general population and is thought to undermine normal cognitive functioning, via utilization of limited cognitive resources, alteration of neuroplasticity, and dysregulation of neurochemistry (15).

Further, 17–75% of patients with SLE report symptoms of depression (16,17) and depression has been linked with both pain and cognitive impairment. Specifically, along with the core features of a major depressive episode (i.e., sad mood and/or loss of interest or pleasure in normally enjoyed activities) (18), depression is also frequently characterized by changes in cognition, including distorted thinking patterns, poor concentration, impaired problem-solving, and reductions in working memory (19). Moreover, ~85% of patients with chronic pain report symptoms of depression (20,21), and there is evidence to suggest that depression plays a mediating role in the relationship between pain and cognitive impairment in rheumatoid arthritis (RA) populations (22). Additionally, 25–92% of patients with SLE report disrupted sleep (23–25). Often characterized by reports of difficulty falling and/or staying asleep, disrupted or fragmented sleep patterns have also been associated with difficulties in sustaining attention, poor memory consolidation, and impaired executive functioning (26,27), and with increased pain sensitivity in healthy individuals (28) and in those with chronic pain (29).

Taken together, these data indicate that there is a substantial overlap among pain, depression, sleep disturbance, and impaired cognition processes in the general population as well as in individuals with chronic diseases. Importantly, little is known about the interplay among these constructs in patients with SLE, and thus, mediation modeling is needed to determine how these variables overlap in SLE populations. Mediation is an empirical statistical method that integrates theory and data analysis, and is said to occur when the association between a predictor variable (i.e., pain) and a criterion variable (i.e., CD) is explained by one or more intervening or mediating variables (i.e., sleep disturbance and depression symptoms). Although other analyses adjust for potential confounders of a statistical relationship to remove nuisance variation, mediation is useful for considering whether one or more mediating variables participate in a cascade of influence from one variable to the next. Investigation of sleep disturbance and depression symptoms as potential pathways between pain and CD in SLE populations may provide insight into the underlying mechanisms of SLE-related CD and indicate new treatment targets for improvement of CD in patients with SLE.

Accordingly, mediation modeling was utilized in the present study in order to ascertain whether the association between pain and CD could be explained or mediated by sleep disturbance and depression symptoms in a sample of patients with SLE, even after controlling for a number of demographic, medical, and SLE-related treatment and disease covariates. Given the lack of longitudinal data on these constructs in SLE populations, and the lack of existing findings demonstrating that depression plays a mediating role in the relationship between pain and cognitive impairment in RA populations, we also evaluated a plausible alternative model in which positive relationships between sleep disturbance and CD would be accounted for by pain and depression symptoms.

PATIENTS AND METHODS

A cross-sectional sample of 115 consecutive patients with SLE was recruited from an outpatient rheumatology clinic. Adult

patients (ages >18 years) who had physician-confirmed SLE diagnoses (based on ACR criteria [9]) were eligible to participate. In order to collect information on demographics, pain, perceived stress, depression, sleep, and cognitive functioning, participants completed a series of psychosocial questionnaires. Information regarding physician-assessed SLE disease activity (Safety of Estrogens in Lupus Erythematosus National Assessment [SELENA] version of the SLE Disease Activity Index [SLEDAI]) (30), disease damage (Systemic Lupus International Collaborating Clinics/ACR Damage Index [SDI]) (31), concurrent diagnoses of fibromyalgia (FM), and current corticosteroid use were gathered through a review of the charts in participants' electronic medical records (EMR). All participants completed written, informed consent prior to study participation and this study was approved by the Institutional Review Board of Rush University Medical Center.

Primary study variables. *Cognitive dysfunction*. Two items from the Lupus Patient-Reported Outcome (LupusPRO) tool (version 1.8) (32) were used to assess CD. Respondents were asked to rate, on a scale from 0 (none of the time) to 4 (all of the time), the degree to which they had experienced poor memory and lack of concentration over the previous 4 weeks. Items were summed to form total scores, with higher scores indicating higher CD ($\alpha = 0.99$).

Pain. The Brief Pain Inventory (33) is a 4-item self-report measure that asks respondents to rate their pain at its worst over the previous week, at its best over the previous week, on average, and at the current moment on a scale from 0 (no pain) to 10 (pain as bad as you can imagine). An index pain score was then calculated by totaling the 4 ratings, with higher scores indicating more severe pain ($\alpha = 0.96$).

Sleep disturbance. Sleep disturbance was assessed using 3 items from the LupusPRO tool (version 1.8) (32). Respondents were asked to rate, on a scale of 0 (none of the time) to 4 (all of the time), the degree to which they had experienced unrefreshing sleep, poor sleep quality, and feelings of being worn out. Items were then summed to form total scores, with higher scores indicating higher sleep disturbance ($\alpha = 0.99$).

Depression. Although participants completed the Patient Health Questionnaire 9 (PHQ-9) (34), their responses to the first 2 items were used to assess for symptoms of depression. This decision was made due to the overlap of the other PHQ-9 items assessing for depression-related changes in sleep and depression-related changes in cognitive functioning, which would be statistically conflated with our dependent variable and sleep-related mediator variable. As such, participants functionally completed the PHQ-2 (35) which asked respondents to rate, on a scale of 1 (not at all) to 3 (nearly every day), the frequency with which they have experienced a depressed mood or loss of interest/pleasure in previously enjoyable activities over the previous 2 weeks. Higher scores indicated more severe symptoms of depression, and scores of ≥3 were suggestive of a potential major depressive episode ($\alpha = 0.79$).

Covariates. *Demographics and stress*. Race/ethnicity was assessed via self-report (white, Hispanic/Latino, African American, Asian, or other), and perceived stress was evaluated with the Perceived Stress Scale 4 (PSS-4) (36), a 4-item self-report measure that asks respondents to rate the degree to which, on a scale of 0 (never) to 4 (very often) scale, they have felt like things were not going their way, that they were not in control of important things in their life, that they were not confident in their ability to handle problems, and that they could not overcome difficulties in their lives. Total scores ranged from 0–16, with higher scores indicating higher perceived stress ($\alpha = 0.97$).

Disease activity and damage. SELENA-SLEDAI (30) was used to assess SLE disease activity. SELENA-SLEDAI is a cumulative and weighted index that evaluates disease activity across 24 different disease factors (range 0–105, where 0 = inactive disease). The index was performed by a rheumatologist, with scores >4 indicating active disease. SLICC-ACR/SDI was used by the rheumatologist to evaluate SLE-related damage (31). Scores can range from 0 to 47 (where 0 = no damage).

Concurrent fibromyalgia, current corticosteroid use. A diagnosis of concurrent FM (yes/no) was assessed via review of EMR for documentation of a diagnosis of FM (37) in physician notes for all participants. Current use of corticosteroid medication (yes/no) was assessed via chart review of EMRs for all participants.

Statistical analysis. SPSS version 23.0 software was used to calculate descriptive statistics of primary study variables and covariates and bivariate correlations in order to determine the strength and direction of relationships between primary study variables and covariates. Given the well-documented race-related difference in subjective pain outcomes (38), preliminary one-way analyses of variance were also conducted to determine whether any of our primary study variables differed as a function of race/ethnicity. These analyses indicated that the pain index scores, LupusPRO sleep disturbance total score, LupusPRO CD total score, and PHQ-2 total score differed significantly as a function of African American race. Accordingly, the race/ethnicity covariate was entered in subsequent analyses as a dichotomous variable (African American = 1, non–African American = 0). There were only a small amount of missing data among our primary variables and covariates (0–4.3%). Given that no one variable had >5% of its data missing, missing values were mean-replaced (39).

The PROCESS macro, a regression path analysis tool that can be used to estimate direct and indirect effects in single and multiple mediation models with SPSS software (40,41), was then used to evaluate our hypothesized relationships among pain, sleep, depression symptoms, and CD (primary study variables) with bootstrapped mediation models controlling for African American race,

FM status, current corticosteroid use, SLEDAI, SDI, and perceived stress (covariates). Per reported best practices for evaluating the significance of indirect effects (42), the PROCESS macro uses random resample bootstrapping to generate confidence intervals from which the significance of indirect effects can be evaluated. In addition to assessing our hypothesized model, we also evaluated a plausible alternative model where pain and depression symptoms functioned as mediators of the relationship between sleep disturbance and CD.

RESULTS

Sample characteristics. The cross-sectional sample was 90% female and had a mean ± SD age of 40 ± 13.28 years (range 19–72 years). Racial composition was 52% African American, 24% white, 11% Hispanic, 5% Asian, and 8% other. Of the sample, 55% percent scored above the clinical cutoff for active disease (mean ± SD SLEDAI score 4.92 ± 3.97; range 0–20). The mean ± SD SDI score was 0.67 ± 1.09 (range 0–6), and 15% of the sample had a concurrent diagnosis of FM. Approximately 61% of the sample was being treated with corticosteroids at the time of participation in the study. In terms of other medication use, 74% were being treated with hydroxychloroquine, 22% mycophenolate mofetil, 10% azathioprine, 6% methotrexate, and 3% cyclophosphamide. Descriptive statistics and bivariate correlations between the primary study variables and covariates are presented in Table 1. CD was significantly correlated with pain (r = 0.39), sleep disturbance (r = 0.56), and depression symptoms (r = 0.57; all $P < 0.05$). Twenty-six percent of the sample scored above the clinical cutoff on the PHQ-2 for a suggestive major depressive episode.

Mediation analyses. Mediation analyses of our hypothesized model (Figure 1) indicated that the effect of pain on CD was mediated by sleep disturbance (β = 0.30) and depression symptoms (β = 0.29). As noted previously, pain was significantly correlated with CD at the zero-order level ($P < 0.05$); however, after accounting for the effects of sleep disturbance and depression symptoms, the relationship between pain and CD was no longer significant. Thus, sleep disturbance and depression symptoms were significant independent mediators of the relationship between pain and CD. Further, these effects were maintained even after controlling for the aforementioned covariates, of which only SLEDAI (β = 0.20) and stress (β = 0.22) remained significantly associated with CD (overall model $R^2 = 0.53$; all $P < 0.05$).

As noted previously, given the lack of longitudinal data on the relationships among pain, sleep disturbance, depression symptoms, and CD in SLE populations and the lack of existing

Table 1. Descriptive statistics and bivariate correlations for primary study variables and covariates*

		1	2	3	4	5	6	7	8	9
Primary study variables										
Total pain index	13.15 ± 12.73 (0–40)	1								
Sleep disturbance (LupusPRO, sleep)	62.80 ± 29.21 (0–100)	0.65†	1							
Depression (PHQ-2)	1.71 ± 1.84 (0–6)	0.42†	0.52†	1						
CD (LupusPRO, cog.)	72.66 ± 28.05 (0–100)	0.39†	0.56†	0.57‡	1					
Covariates										
African American race, no. (%)	60 (52)	0.21‡	0.20‡	0.26†	0.22‡	1				
Comorbid fibromyalgia, no. (%)	17 (15)	0.25†	0.18	0.30†	0.30†	0.06	1			
Steroid use, no. (%)	70 (61)	0.24‡	0.04	0.15	−0.01	0.12	−0.18	1		
SLE activity (SLEDAI)	4.91 ± 4.05 (0–20)	0.20‡	0.01	0.03	−0.15	0.09	−0.09	0.42†	1	
SLE damage (SDI)	0.67 ± 1.09 (0–6)	−0.08	−0.12	−0.07	−0.13	0.13	−0.10	0.13	0.11	1
Perceived stress	6.21 ± 3.51 (0–14)	0.34†	0.37†	0.57†	0.49†	0.19‡	0.15	0.05	0.10	0.02

* Values are the mean ± SD (range) unless indicated otherwise. LupusPRO = Lupus Patient-Reported Outcome tool; PHQ-2 = Patient Health Questionnaire 2; CD = cognitive dysfunction; cog. = cognition; SLE = systemic lupus erythematosus; SLEDAI = SLE Disease Activity Index; SDI = Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index.
† $P < 0.01$
‡ $P < 0.05$

findings showing that depression plays a mediating role in the relationship between pain and cognitive impairment in RA populations (22), we also evaluated a plausible alternative model in which positive relationships between sleep disturbance and CD would be accounted for by pain and depression symptoms. Our alternative model (Figure 2) demonstrated that sleep disturbance and CD remained significantly linked even after accounting for the effects of pain and depression symptoms (β = 0.31). In addition, pain was not a significant mediator (β = 0.054; $P = 0.66$) in this alternative model, but depression symptoms were (β = 0.273). Finally, perceived stress (β = 0.227) was the only covariate that remained significantly associated with CD in the alternative model (all $P < 0.05$, unless otherwise noted).

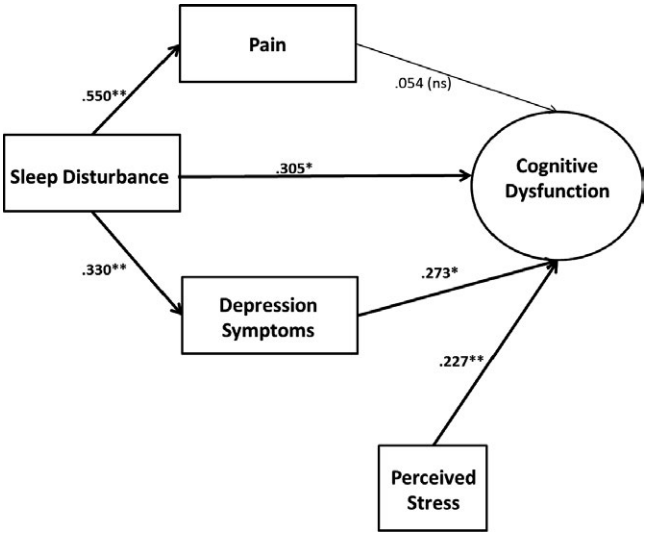


Figure 2. Alternative mediation model of sleep disturbance and cognitive dysfunction (CD). Standardized indirect effect of sleep disturbance on CD via pain = 0.03; SE 0.07 (95% confidence interval [95% CI] –0.13, 0.09). Standardized indirect effect of sleep disturbance on CD via depression symptoms = 0.09; SE 0.06 (95% CI 0.02, 0.20). Standardized regression coefficients presented in model; * = $P < 0.05$; ** = $P < 0.01$. Nonsignificant covariate paths (i.e., fibromyalgia diagnosis, current steroid use, race/ethnicity, Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index) not shown.

DISCUSSION

Consistent with previous studies that have assessed the symptom burden of SLE populations, symptoms of pain, sleep disturbance, depression, and CD were common among the sample of patients with SLE in the present study. Mediation analyses of our hypothesized model indicated that the relationship between pain and CD was explained by concurrent sleep disturbance and depression symptoms. These findings extend beyond prior research on CD in SLE by demonstrating the aforementioned effects even after controlling for the influences of a number of demographic, medical, and SLE-related symptoms, which are treatment and disease variables that have previously been linked with CD.

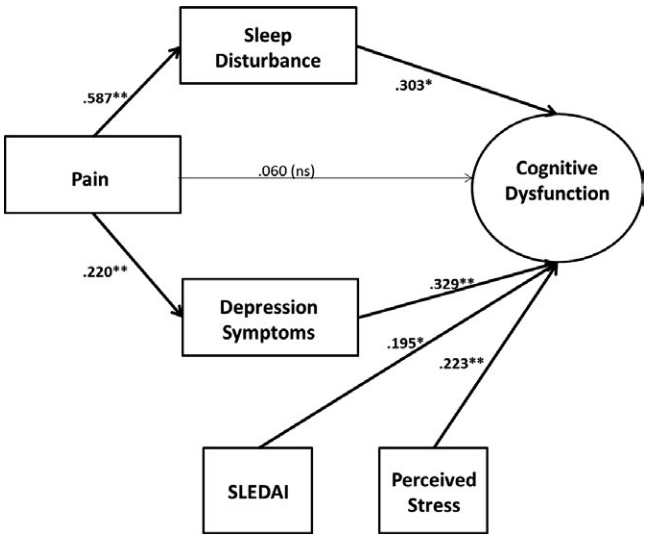


Figure 1. Hypothesized mediation model of pain and cognitive dysfunction (CD). Standardized indirect effect of pain on CD via sleep disturbance = 0.18; SE 0.08 (95% confidence interval [95% CI] 0.06, 0.32). Standardized indirect effect of pain on CD via depression symptoms = 0.07; SE 0.04 (95% CI 0.01, 0.17). Standardized regression coefficients presented in model; * = $P < 0.05$; ** = $P < 0.01$. Nonsignificant covariate paths (i.e., fibromyalgia diagnosis, current steroid use, race/ethnicity, Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index) not shown. SLEDAI = Systemic Lupus Erythematosus Disease Activity Index.

One explanation for these results is that disease-related flares may lead to increased pain, which could negatively impact overall sleep and mood. The cumulative degradation in mood and sleep may, in turn, deplete the cognitive resources needed to effectively manage pain. This explanation would also potentially account for the significant links between CD and disease activity (SLEDAI) and CD and perceived stress in the final model. In other words, a disease flare that elicits pain (and that would ultimately result in worsening CD) would likely be reflected in a positive association with SLEDAI and a positive association with perceived stress. It should be noted that these results may also reflect more general trends of the differential impact of depression and poor sleep processes among women in particular. Specifically, given that SLE is a chronic disease that predominantly impacts women and that women are more likely than men to report symptoms of depression and poor sleep (43,44), it would make sense that new or ongoing disease activity, stress, pain, and CD in a predominantly female sample of patients with SLE might be more strongly associated with concurrent decrements in mood and sleep. To that end, it is important to note that the cross-sectional nature of our data precludes the ability to draw causal inferences among our variables of interest. Although the testing of a plausible alternative model lends some additional support to the findings in our study, longitudinal models are needed to validate the directional

pathways we have posited and to determine whether sleep and mood degradation are indeed potential driving mechanisms of the relationship between pain and CD in SLE populations.

Despite the limitations of our study design, our findings may have a number of promising treatment-related implications. Currently, there are only 4 drugs that have been approved by the US Federal Drug Administration for the treatment of SLE, which include nonsteroidal antiinflammatory drugs, steroids, hydroxychloroquine, and belimumab (45,46). These biomedical treatment approaches and other immunosuppressive agents have generally been aimed at controlling disease activity and flares to avoid long-term damage and mortality. As noted previously, the larger symptom burden of SLE is significant (e.g., depression, anxiety, suicide, divorce, stress, fatigue, sleep problems, work-related disability, pain, etc.). Thus, treatment approaches that use a biopsychosocial framework may offer the best opportunity to optimize patient outcomes in SLE populations. To that end, our findings may indicate the utility of nonpharmacologic intervention avenues for patients with SLE with pain and CD. Specifically, cognitive behavioral therapies for depression and sleeping difficulties are known to reduce distress and enhance functioning across various psychosocial domains (47). Importantly, given the risk of disease-related damage in SLE (48) , optimization of pharmacologic interventions for patients with SLE is critical. However, nonpharmacologic interventions that maximize potential benefits without additional pharmacologic treatments may be of additional utility for SLE patient populations.

These results should be evaluated in the context of the limitations of the current study. Data were cross-sectional from a small, outpatient sample of patients with SLE. In addition, we cannot rule out the possibility that method variance may have accounted for some of our findings, given that CD, sleep disturbance, and depression symptoms were measured with brief, self-report screening tools (which may also have resulted in an over or underestimation of the presence and severity of these measured constructs). Although the LupusPRO tool (32) cognitive functioning items have been validated against other self-report cognitive screeners (e.g., the Multiple Sclerosis Neuropsychological Screener) (49,50), formal cognitive testing would strengthen the internal validity of our results. In addition, structured clinical interviews would be needed to affirm the presence or absence of a major depressive episode (51) and objective measurement of sleep (e.g., wrist actigraphy, polysomnography) (52) would be needed to corroborate the presence and severity of reported sleep disturbance. Although the findings revealed in our study may be of clinical relevance, they require validation in prospective, longitudinal studies with additional, non–self-report measurement modalities.

In conclusion, pain, sleep disturbance, and depression symptoms are frequently reported by patients with SLE and have known links to CD in the general population, but had not previously been concurrently investigated as potential explanatory factors in CD outcomes among patients with SLE. Along with factors previously linked to CD outcomes in SLE populations, including disease activity and perceived stress, the current study offers preliminary support for the roles of sleep disturbance and depression symptoms as potential mechanisms linking pain-related experiences to decrements in cognitive functioning in patients with SLE. These preliminary findings demonstrated the utility of further exploring a biopsychosocial treatment approach to SLE-related symptom management.

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. Lillis 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. Lillis, Tirone, Hobfoll, Block, Jolly.
Acquisition of data. Gandhi, Weinberg, Nika, Sequeira.
Analysis and interpretation of data. Lillis, Tirone, Hobfoll, Block, Jolly.

REFERENCES

1. Carter EE, Barr SG, Clarke AE. The global burden of SLE: prevalence, health disparities and socioeconomic impact. Nat Rev Rheumatol 2016;12:605–20.
2. Dhanhani AM, Gignac MA, Su J, Fortin PR. Work disability in systemic lupus erythematosus. Arthritis Rheum 2009;61:378–85.
3. Barber MR, Clarke AE. Socioeconomic consequences of systemic lupus erythematosus. Curr Opin Rheumatol 2017;29:480–5.
4. Jolly M. How does quality of life of patients with systemic lupus erythematosus compare with that of other common chronic illnesses? J Rheumatol 2005;32:1706–8.
5. Schmeding A, Schneider M. Fatigue, health-related quality of life and other patient-reported outcomes in systemic lupus erythematosus. Best Prac Res Clin Rheumatol 2013;27:363–75.
6. Havelka M, Lucanin JD, Lucanin D. Biopsychosocial model–the integrated approach to health and disease. Coll Antropol 2009;33:303–10.
7. Butt BA, Farman S, Khan SE, Saeed MA, Ahmad NM. Cognitive dysfunction in patients with Systemic Lupus Erythematosus. Pak J Med Sci 2017;33:59.
8. Huerta PT, Gibson EL, Rey C, Huerta TS. Integrative neuroscience approach to neuropsychiatric lupus. Immunol Res 2015;63:11–7.
9. ACR Ad Hoc Committee on Neuropsychiatric Lupus Nomenclature. The American College of Rheumatology nomenclature and case definitions for neuropsychiatric lupus syndromes. Arthritis Rheum 1999;42:599–608.
10. McElhone K, Abbott J, Teh LS. A review of health related quality of life in systemic lupus erythematosus. Lupus 2006;15:633–43.
11. Jeltsch-David H, Muller S. Neuropsychiatric systemic lupus erythematosus: pathogenesis and biomarkers. Nat Rev Neurol 2014;10:579–96.
12. Benedict RH, Shucard JL, Zivadinov R, Shucard DW. Neuropsychological impairment in systemic lupus erythematosus: a comparison with multiple sclerosis. Neuropsychol Rev 2008;18:149–66.
13. Pettersson S, Lövgren M, Eriksson LE, Moberg C, Svenungsson E, Gunnarsson I, Welin Henriksson E. An exploration of patient-reported symptoms in systemic lupus erythematosus and the relationship to health-related quality of life. Scand J Rheumatol 2012;41:383–90.
14. GfK Roper. Lupus Awareness Survey for the Lupus Foundation of America. Executive Summary Report. Washington (DC). GfK Roper Public Affairs & Corporate Communications; 2012 Oct.

15. Moriarty O, McGuire BE, Finn DP. The effect of pain on cognitive function: a review of clinical and preclinical research. Prog Neurobiol 2011;93:385–404.
16. Macêdo EA, Appenzeller S, Costallat LT. Depression in systemic lupus erythematosus: gender differences in the performance of the Beck Depression Inventory (BDI), Center for Epidemiologic Studies Depression Scale (CES-D), and Hospital Anxiety and Depression Scale (HADS). Lupus 2018;27:179–89.
17. Palagini L, Mosca M, Tani C, Gemignani A, Mauri M, Bombardieri S. Depression and systemic lupus erythematosus: a systematic review. Lupus 2013;22:409–16.
18. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 5th edition: DSM-5. Arlington (VA): American Psychiatric Publishing; 2013.
19. McIntyre RS, Xiao HX, Syeda K, Vinberg M, Carvalho AF, Mansur RB. The prevalence, measurement, and treatment of the cognitive dimension/domain in major depressive disorder. CNS Drugs 2015;29:577–89.
20. Williams LS, Jones WJ, Shen J, Robinson RL, Weinberger M, Kroenke K. Prevalence and impact of depression and pain in neurology outpatients. J Neurol Neurosurg Psychiatry 2003;74:1587–9.
21. Fishbain DA, Cutler R, Rosomoff HL, Rosomoff RS. Chronic pain-associated depression: antecedent or consequence of chronic pain? A review. Clin J Pain 1997;13:116–37.
22. Brown SC, Glass JM, Park DC. The relationship of pain and depression to cognitive function in rheumatoid arthritis patients. Pain 2002;96: 279–84.
23. Greenwood KM, Lederman L, Linder HD. Self reported sleep in systemic lupus erythematosus. Clin Rheumatol 2008;27:1147–51.
24. Costa DD, Bernatsky S, Drista M, Clarke AE, Dasgupta K, Keshani A, et al. Determinants of sleep quality in women with systemic lupus erythematosus. Arthritis Rheum 2005;53:272–8.
25. Chandrasekhara PK, Jayachandran NV, Rajasekhar L, Thomas J, Narsimulu G. The prevalence and associations of sleep disturbances in patients with systemic lupus erythematosus. Mod Rheumatol 2009;19:407–15.
26. McCoy JG, Strecker RE. The cognitive cost of sleep lost. Neurobiol Learn Mem 2011;96:564–82.
27. Krause AJ, Simon EB, Mander BA, Greer SM, Saletin JM, Goldstein-Piekarski AN, et al. The sleep-deprived human brain. Nat Rev Neurosci 2017;18:404–18.
28. Iacovides S, George K, Kamerman P, Baker FC. Sleep fragmentation hypersensitizes healthy young women to deep and superficial experimental pain. J Pain 2017;18:844–54.
29. Smith MT, Haythornthwaite JA. How do sleep disturbance and chronic pain inter-relate? Insights from the longitudinal and cognitive-behavioral clinical trials literature. Sleep Med Rev 2004;8:119–32.
30. Gladman DD, Ibañez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. J Rheumatol 2002;29:288–91.
31. Romero-Diaz J, Isenberg D, Ramsey-Goldman R. Measures of adult systemic lupus erythematosus: updated version of British Isles Lupus Assessment Group (BILAG 2004), European Consensus Lupus Activity Measurements (ECLAM), Systemic Lupus Activity Measure, Revised (SLAM-R), Systemic Lupus Activity Questionnaire for Population Studies (SLAQ), Systemic Lupus Erythematosus Disease Activity Index 2000 (SLEDAI-2K), and Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index (SDI). Arthritis Care Res (Hoboken) 2011;63 Supp 11:S37–46.
32. Jolly M, Pickard AS, Fogg L, Wilke CT, Lin HW, Mikolaitis RA, et al. US patient derived lupus-specific patient reported outcome measure (Lupus-PRO (c)): validity and reliability. Arthritis Rheum 2008;58, Suppl S805).
33. Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. Ann Acad Med Singapore 1994;23:129–38.
34. Kroenke K, Spitzer RL, Williams JB. The PHQ-9. J Intern Med 2001;16:606–13.
35. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, Falloon K, Hatcher S. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. Ann Fam Med 2010;8:348–53.
36. Lee EH. Review of the psychometric evidence of the perceived stress scale. Asian Nurs Res 2012;6:121–7.
37. Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL, et al. The American College of Rheumatology 1990 criteria for the classification of fibromyalgia: report of the multicenter criteria committee. Arthritis Rheum 1990;33:160–72.
38. Kim HJ, Yang GS, Greenspan JD, Downton KD, Griffith KA, Renn CL, Johantgen M, Dorsey SG. Racial and ethnic differences in experimental pain sensitivity: systematic review and meta-analysis. Pain 2017;158:194–211.
39. Little TD, Jorgensen TD, Lang KM, Moore EW. On the joys of missing data. J Pediatr Psychol 2013;39:151–62.
40. Hayes, AF. PROCESS: a versatile computational tool for observed variable mediation, moderation, and conditional process modeling. URL: http://www.afhayes.com/public/process2012.pdf.
41. Hayes AF. Introduction to mediation, moderation, and conditional process analysis: a regression-based approach. New York (NY): Guilford Publications; 2017.
42. Preacher KJ, Kelley K. Effect size measures for mediation models: quantitative strategies for communicating indirect effects. Psychol Methods 2011;16:93.
43. Kuehner C. Why is depression more common among women than among men? Lancet Psychiatry 2017;4:146–58.
44. Theorell-Haglöw J, Miller CB, Bartlett DJ, Yee BJ, Openshaw HD, Grunstein RR. Gender differences in obstructive sleep apnoea, insomnia and restless legs syndrome in adults: what do we know? A clinical update. Sleep Med Rev 2018;38:28–38.
45. Sugai DY, Gustafson CJ, De Luca JF, Davis SA, Jorizzo JL, O'Rourke KS, Feldman SR. Trends in the outpatient medication management of lupus erythematosus in the United States. J Drugs Dermatol 2014;13:545–52.
46. United States Federal Drug Administration. Lupus therapies continue to evolve. URL: https://www.fda.gov/ForConsumers/ConsumerUpdates/ucm398682.htm.
47. Carney CE, Edinger JD, Kuchibhatla M, Lachowski AM, Bogouslavsky O, Krystal AD, et al. Cognitive behavioral insomnia therapy for those with insomnia and depression: a randomized controlled clinical trial. Sleep 2017;40:zsx019.
48. De Achaval S, Suarez-Almazor ME. Treatment adherence to disease-modifying antirheumatic drugs in patients with rheumatoid arthritis and systemic lupus erythematosus. Int J Clin Rheumatol 2010;5:313.
49. Clowse, ME, Criscione-Schreiber LG, Jolly M, Keefe FJ, Somers TJ. Memory complaints in lupus patients: relationship to lupus activity, symptoms, quality of life, psychological distress, and coping [abstract]. Arthritis Rheum 2011;63 Suppl:1419.
50. Benedict RH, Cox D, Thompson LL, Foley F, Weinstock-Guttman B, Munschauer F. Reliable screening for neuropsychological impairment in multiple sclerosis. Mult Scler J 2004;10:675–8.
51. American Psychiatric Association, Steering Committee on Practice Guidelines. American Psychiatric Association practice guidelines for the treatment of psychiatric disorders: compendium 2006. Arlington (VA): American Psychiatric Association; 2006.
52. Valencia-Flores M, Resendiz M, Castano VA, Santiago V, Campos RM, Sandino S, Valencia X, Alcocer J, Ramos GG, Bliwise DL. Objective and subjective sleep disturbances in patients with systemic lupus erythematosus. Arthritis Rheum 1999;42:2189–93.

Arthritis Care & Research
Vol. 71, No. 3, March 2019, pp 413–418
DOI 10.1002/acr.23606

AMERICAN COLLEGE OF RHEUMATOLOGY
EDUCATION • TREATMENT • RESEARCH

BRIEF REPORT

Trends in Population-Based Incidence and Prevalence of Juvenile Idiopathic Arthritis in Manitoba, Canada

Natalie J. Shiff,[1] Kiem Oen,[2] Kristine Kroeker,[2] and Lisa M. Lix[2]

Objective. To estimate the incidence and prevalence of juvenile idiopathic arthritis (JIA) in children ages <16 years in the province of Manitoba, Canada, and to determine changes in estimates between 2000 and 2012.

Methods. JIA cases were ascertained from the administrative health data of Manitoba, using a validated case-finding algorithm. Annual incidence and prevalence rates were estimated using a generalized linear model with generalized estimating equations (GEEs), adjusting for sociodemographic characteristics. Changes in estimates were tested using piecewise regression models.

Results. A total of 455 cases of JIA prevalence met the inclusion criteria. Sex- and age-adjusted incidence estimates were 14.01 (95% confidence interval [95% CI] 13.52, 14.53) in 2000/2001 and 9.18 (95% CI 8.56, 9.85) in 2010/2011. Prevalence estimates were 65.33 (95% CI 63.87, 66.83) in 2000/2001 and 59.61 (95% CI 58.17, 61.08) in 2010/2011. A linear piecewise model provided the best fit to the data. There was a significant decrease in prevalence over the study period (–0.18 [95% CI –0.35, –0.02]; $P = 0.0292$) but no statistically significant change in incidence (–0.46 [95% CI –0.94, 0.01]; $P = 0.0571$). Sex-stratified models revealed a decrease for males in both prevalence (estimate –0.54 [95% CI –0.84, –10.25]; $P = 0.0003$) and incidence (estimate –1.02 [95% CI –2.02, –0.04]; $P = 0.0439$); there were no changes for females.

Conclusion. Few population-based longitudinal epidemiologic studies of JIA have been conducted. Our findings suggested a decrease in overall JIA prevalence, and in incidence and prevalence in men. Further research to validate these findings in other cohorts and to explore factors that contribute to this change will benefit future health care planning for JIA.

INTRODUCTION

Juvenile idiopathic arthritis (JIA) is the most common pediatric rheumatic disease, but reported incidence and prevalence estimates vary greatly (1). Study methodologies, differences in data sources, and geographic variations have been cited as potential explanations for the wide range of estimates (1). A recent systematic review (1) collated population estimates that were published from 1972 to 2011. Incidence estimates ranged from 1.6 to 23.0 per 100,000 individuals, and prevalence estimates from 8.8 to 400.0 per 100,000 (1). The review demonstrated that estimates may be influenced by changes in JIA classification.

Administrative health data are valuable for surveillance of chronic conditions like JIA. These data are collected whenever patients have encounters with the health care system. Although not created for research purposes, these data are relatively easy to access and enable investigation of trends in the prevalence and incidence of health conditions. Individuals with a condition of interest are identified using validated case-finding algorithms, in order to ascertain cases with high sensitivity and specificity. Few studies have used validated case-finding algorithms to produce population-level estimates of incidence and prevalence for chronic childhood arthritis. Recent work has validated a case-finding algorithm for JIA in

The views expressed herein are those of the authors and do not reflect the official policy of the Manitoba Centre for Health Policy, Manitoba Health, or other data providers. The data reported herein have been supplied by the Manitoba Population Research Data Repository housed at the Manitoba Centre for Health Policy, University of Manitoba and were derived from data provided by Manitoba.

Supported by a New Investigator Establishment grant from the Saskatchewan Health Research Foundation.

[1]Natalie J. Shiff, MD, MHSc: University of Florida, Gainesville and University of Saskatchewan, Saskatoon, Saskatchewan, Canada; [2]Kiem Oen, MD, Kristine Kroeker, MSc, Lisa M. Lix, PhD: University of Manitoba, Winnipeg, Manitoba, Canada.

Dr. Lix has received honoraria as Manitoba Research Chair (less than $10,000). No other disclosures relevant to this article were reported.

Address correspondence to Natalie J. Shiff, MD, MHSc, Department of Pediatrics, Division of Immunology, Rheumatology, and Allergy, Box 100296, University of Florida, 1600 SW Archer Road, Gainesville, FL 32610. E-mail: nshiff@ufl.edu.

Submitted for publication December 28, 2017; accepted in revised form May 22, 2018.

SIGNIFICANCE & INNOVATIONS

- This population-based longitudinal study contributes new evidence about the epidemiology of juvenile idiopathic arthritis (JIA).
- In the province of Manitoba, Canada, between 2000 and 2012, sex- and age-adjusted incidence estimates were 14.01 (95% confidence interval [95% CI] 13.52, 14.53) in 2000/2001 (highest value) and 9.18 (95% CI 8.56, 9.85) in 2010/2011, with a minimum estimate of 7.91 (95% CI 7.34, 8.53) in 2009/2010. Adjusted prevalence estimates by fiscal year (April 1 to March 31) were 65.33 (95% CI 63.87, 66.83) in 2000/2001 and 59.61 (95% CI 58.17, 61.08) in 2010/2011, with a minimum value of 52.86 (95% CI 51.56, 54.19) in 2009/2010.
- Overall JIA prevalence decreased significantly over the study period; incidence and prevalence of JIA in males also decreased significantly over the study period. Future research will investigate potential contributors to this changing epidemiology of JIA.

administrative health data for the pediatric population in the province of Manitoba, Canada, a region which has provincially administered universal health care and complete capture of the population in its administrative health databases (2). There is one pediatric rheumatology center in the province of Manitoba where all children with rheumatic diseases in the province are followed until they are transitioned to adult care at 18 years of age. A patient can be referred out-of-province only after he/she has been seen by physicians within the province and diagnostic or therapeutic problems cannot be resolved within the province. The aims of our study were to ascertain JIA incidence and prevalence in Manitoba in children ages <16 years, using administrative health data from 2000 to 2012, and to test for changes in these estimates over time.

PATIENTS AND METHODS

Data sources. The Population Research Data Repository, which is housed at the Manitoba Centre for Health Policy of the University of Manitoba contains records of health care contacts for all 1.2 million residents in this province. Diagnoses were coded according to either the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) or the ICD, Tenth Revision, Canada (ICD-10-CA) diagnosis codes (3). Databases accessed for the present study included the Manitoba Health Insurance Registry, which captures information about provincial health insurance dates of coverage, reasons for termination of coverage, and demographic characteristics (i.e., age, sex, residence location); the Hospital Discharge Abstracts Database (DAD), which includes diagnosis and procedure codes for patients at the point of discharge from Manitoba acute care facilities (ICD-9-CM codes were used up to March 31, 2004, with up to diagnosis 16 codes listed, and ICD-10-CA codes were used as of April 1, 2004, with up to 25 diagnosis codes listed); and the Medical Claims Database, which comprises the vast majority of physician billings for in-hospital and outpatient visits (each claim contains a single ICD-9-CM diagnosis code). Additionally, Statistics Canada census data from 2011 were used to define household income quintile, an area-level measure of socioeconomic status.

Study cohort. A previously validated case-finding algorithm (2) was used to retrieve cases from the DAD and Medical Claims Database between April 1, 1997 and March 31, 2012. The estimates of sensitivity, specificity, and positive predictive value (PPV) of the algorithm were 88.2% (95% CI 85.7, 90.7), 90.4% (95% CI 87.5, 93.3), and 93.9% (95% CI 92.0, 95.8), respectively, using data from 1980 to 2012; however, these accuracy statistics improved, resulting in a sensitivity of 95.1% (95% CI 91.3, 98.9), specificity of 94.9% (95% CI 90.0, 99.8), and PPV of 96.7% (95% CI 93.5, 99.9) when the case-finding algorithm was applied to data from 2005 to 2012 (2). In the present study, we chose to assess incidence and prevalence from 2000 to 2012, reflecting the time during which the most recent version of the International League of Associations for Rheumatology (ILAR) classification criteria for JIA (revised in 2001) (4) became standard nomenclature in Canada. Diagnosis codes used to define the case-finding algorithm were ICD-9-CM codes 714 (rheumatoid arthritis [RA]) and 720 (spondyloarthropathies), and ICD-10-CA codes M05 (seropositive RA), M06 (other RA), M08 (juvenile arthritis), and M45 (ankylosing spondylitis). The algorithm criteria included 1 day of health insurance coverage, and ≥1 relevant diagnosis in DAD records or ≥2 relevant diagnoses in physician billing claims (by any physician) that were at least 8 weeks apart within 2 years, prior to the 16th birthday of the case (2). The date of the first JIA-related diagnosis that satisfied the case-finding algorithm was considered the index date used to define year of diagnosis.

Statistical analysis. Annual crude incidence and point prevalence estimates were calculated for fiscal years (April 1 to March 31) 2000/2001 to 2011/2012. Members of the cohort who satisfied the case-finding algorithm for JIA were considered incident at the diagnosis index date. The washout period for ascertainment of incident cases was from April 1, 1997 to March 31, 2000. Individuals who met the criteria of the case-finding algorithm during this time period were excluded in incidence calculations in order to ensure that rates represented new cases and not existing cases with infrequent contact with the health care system (5,6); they were included in prevalence calculations for the 2000/2001 fiscal year, to ensure that cases were retained. The denominator for incidence and prevalence calculations was the number of insured Manitoba residents ages <16 years who had ≥1 day of health insurance coverage during the fiscal year, while the numerator for incidence estimates comprised the num-

ber of cohort members who met the criteria for the case-finding algorithm for the first time, with a diagnosis index date within the fiscal year. All ascertained cases remained prevalent cases until death or the 16th birthday, whichever occurred earlier.

Cases were described by age, sex, region of residence, and income quintile. Age at diagnosis was classified as 0–5 years, 6–10 years, and 11–15 years and region of residence as urban or rural. The Winnipeg and Brandon regions were considered urban. All other parts of the province were classified as rural.

Changes in incidence and prevalence rates were tested using generalized linear models with generalized estimating equations (GEEs), in order to account for correlation in the data. A Poisson distribution was selected to model JIA counts as the dependent variable, and the natural logarithm of the total population count was the model offset. Three different GEE models that contained the main effects of age group, sex, fiscal year, and region of residence were fit to the data. In the first model, fiscal year was a categorical variable; this model was used to produce age- and sex-adjusted estimates for each fiscal year. In the second model, fiscal year was a continuous variable; this model was used to produce the average annual rate of change, and also adjusted for age and sex. The third model was similar to the second, but included the 2-way interaction of fiscal year and region of residence in addition to main effects; this model was used to estimate the average annual rate of change, also adjusted for age and sex. The fit of each model was evaluated using the Quasi-Likelihood Information Criterion (QIC) statistic; comparisons of QIC statistics were made for unadjusted and adjusted models, with a smaller QIC statistic being indicative of a better fitting model (7).

Additionally, Poisson models were fit with a smoothing spline on fiscal year and covariates of sex, age group, and region of residence, in order to determine the nature of the trend (i.e., linear, curvilinear) in prevalence and incidence. Aikaike information criterion (AIC) was used to determine the appropriate number of degrees of freedom (i.e., fitted slopes) for the spline. The spline models had 1 to 5 degrees of freedom, which indicates increasing complexity (i.e., increasing departures from a linear trend). The spline models were first fit to the data for the entire population, after which sex-specific estimates were produced. A sensitivity analysis was conducted by excluding the first 2 years of data, as this was the timeframe in which the ILAR classification of JIA (4) was adopted in Canada. SAS software, version 9.4, was used for all modeling. Research was conducted in compliance with the Helsinki Declaration and was approved by Manitoba Health (HIPC# 2013/20114-19) and the Research Ethics Boards of the University of Manitoba Bannatyne Campus (H2013:246).

RESULTS

In total, 267 individuals were ascertained as incident JIA cases between April 1, 2000 and March 31, 2012. A total of 459 individuals ages <16 years had a diagnosis of JIA from April 1, 1997 to March 31, 2012; complete data were available for 455 prevalent cases. The demographic characteristics of the study cohort comprising all prevalent and incident cases at the time of meeting the criteria for the case-finding algorithm are shown in Table 1. Table 1 also contains the corresponding characteristics of the Manitoba population ages <16 years.

Using GEEs, the model with year as a categorical variable was the best fitting model to estimate changes in incidence over time (QIC 19.6 compared to QIC 44.7 for the null model with no covariates; QIC 29.9 and 31.0 for the 2 alternate models). Model estimates are shown in Table 2. There was no statistically significant difference in incidence estimates for residents of rural compared to urban areas (data not shown). All 3 of the adjusted models had better fit than the unadjusted model for prevalence (QIC –1634.4), with QICs of –7,870.8, –8,328.9, and –8,449.0. Prevalence estimates for the models using fiscal year as a continuous variable were also observed (see Supplementary Tables 1 and 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23606/abstract). There was no significant difference in the average JIA prevalence in urban and rural regions during the study period (see Supplementary Table 2, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.23606/abstract).

Table 1. Demographic characteristics at diagnosis index date for prevalent and incident cases of disease in the JIA cohort (2000–2012) and the Manitoba population ages <16 years in 2000/2001*

Characteristic	JIA cohort	Manitoba population
Sex		
Male	136 (29.9)	13,0733 (51.2)
Female	319 (70.1)	124,416 (48.8)
Age group, years		
0–5	134 (29.4)	89,403 (35.0)
6–10	119 (26.2)	83,358 (32.7)
11–15	202 (44.4)	82,388 (32.3)
Region of residence		
Rural	197 (43.3)	113,592 (44.5)
Urban†	258 (56.7)	141,557 (55.5)
Income quintile		
Q1 (lowest)	76 (16.7)	54,107 (21.2)
Q2	81 (17.8)	50,930 (20.0)
Q3	92 (20.2)	50,803 (19.9)
Q4	93 (20.4)	52,698 (20.6)
Q5 (highest)	113 (24.9)	46,611 (18.3)

* Values are the number (%) of individuals who met the case definition for juvenile idiopathic arthritis (JIA) in Manitoba between 2000 and 2012 and for whom income quintile data were available, and for the Manitoba population ages <16 years in the 2000/2001 fiscal year. Data shown are for 455 of the 459 cases.
† Individuals residing in Winnipeg and Brandon.

Table 2. Age- and sex-adjusted juvenile idiopathic arthritis incidence and point prevalence by fiscal year in the Manitoba population ages <16 years*

Fiscal year	Incidence (95% CI)	Prevalence (95% CI)
2000/2001	14.01 (13.52, 14.53)	65.33 (63.87, 66.83)
2001/2002	13.36 (12.86, 13.88)	70.36 (68.82, 71.94)
2002/2003	9.95 (9.54, 10.38)	70.48 (68.97, 72.04)
2003/2004	10.99 (10.40, 11.62)	75.19 (73.58, 76.84)
2004/2005	9.91 (9.32, 10.54)	70.89 (69.38, 72.44)
2005/2006	9.36 (8.86, 9.90)	67.10 (65.67, 68.56)
2006/2007	8.85 (8.42, 9.30)	64.28 (62.90, 65.70)
2007/2008	11.89 (11.32, 12.49)	61.57 (60.21, 62.97)
2008/2009	11.89 (11.30, 12.51)	61.49 (60.09, 62.92)
2009/2010	7.91 (7.34, 8.53)	58.02 (56.66, 59.42)
2010/2011	9.18 (8.56, 9.85)	59.61 (58.17, 61.08)
2011/2012	8.47 (7.93, 9.05)	52.86 (51.56, 54.19)

* Values are the incidence and point prevalence per 100,000 individuals. Ages by fiscal year (April 1 to March 31) are shown. 95% CI = 95% confidence interval.

During the 11-year study period, the sex and age-adjusted incidence estimates were 14.01 (95% CI 13.52, 14.53) in 2000/2001 (highest value) and 9.18 (95% CI 8.56, 9.85) in 2010/2011, with a minimum estimate of 7.91 (95% CI 7.34, 8.53) in 2009/2010. The adjusted prevalence estimates were 65.33 (95% CI 63.87, 66.83) in 2000/2001 and 59.61 (95% CI 58.17, 61.08) in 2010/2011, with a minimum value of 52.86 (95% CI 51.56, 54.19) in 2001/2012 and a maximum value of 75.19 in 2003/2004 (95% CI 73.58, 76.84) (Table 2).

Of the models with a smoothing spline for fiscal year, the model with the smallest AIC was the model with a single degree of freedom (i.e., linear trend model) for incidence and for prevalence among males (see Supplementary Table 3, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23606/abstract). The difference in AIC values for the models with 1 and 2 degrees of freedom for overall prevalence and prevalence in females was negligible, therefore the linear model was selected as the best fitting model. The overall incidence model had a nonsignificant slope estimate of –0.46 (P = 0.0571) (Table 3). Stratification by sex revealed a significant decrease in incidence for males (estimate –1.02, P = 0.0439) but no statistically significant decrease in incidence for females (estimate –0.27, P = 0.3390). There was a statistically significant decrease in prevalence over time, with an estimate of –0.18 (P = 0.0292). Prevalence models stratified by sex also showed a significant decrease in males (estimate –0.54, P = 0.0003) but not in females (estimate –0.01, P = 0.90) through the study period. The sensitivity analysis did not change the results for prevalence estimates (that is, prevalence estimates decreased); however, the decrease in incidence observed for males was not evident when the first 2 years of data were excluded.

Table 3. Estimates for JIA incidence and prevalence from 2000 to 2012*

Model	Estimate (95% CI)	*P*
Incidence		
Full cohort	–0.46 (–0.94, 0.01)	0.0571
Male	–1.02 (–2.02, –0.04)	0.0439
Female	–0.27 (–0.83, 0.28)	0.3390
Prevalence		
Full cohort	–0.18 (–0.35, –0.02)	0.0292
Male	–0.54 (–0.84, –0.25)	0.0003
Female	–0.01 (–0.21, 0.19)	0.8979

* Values are the linear spline estimates and 95% confidence intervals (95% CIs) for incidence and prevalence of juvenile idiopathic arthritis (JIA).

DISCUSSION

We calculated incidence and prevalence of JIA in the province of Manitoba from 2000 to 2012, using a case-finding algorithm for JIA that was previously validated in the administrative health data of Manitoba (3). As expected, there was a preponderance of female cases (female to male ratio 2.3:1). Although the 5-year age group with the largest number of prevalent cases was in the group ages 11–15 years, the majority of cases were ages <11 years, also as expected (8). The percentage of cases from urban and rural areas corresponded to the distribution of the Manitoba population ages <16 years.

We found age- and sex-adjusted incidence estimates in Manitoba ranging from 14.01 (95% CI 13.52, 14.53) to 7.91 (95% CI 7.34, 8.53), which are similar to those reported in similar studies in Minnesota (8.8 to 14.8), despite different methodologies (8). Studies in Olmsted County, Minnesota (9) and Manitoba (10) previously reported a cyclic incidence of juvenile arthritis, which was not seen in our data. We found age- and sex-adjusted point prevalence estimates that varied from 65.33 (95% CI 63.87, 66.83) to 58.02 (95% CI 56.66, 59.42). A previous study (11) found a higher crude cumulative prevalence estimate of 117 (95% CI 105, 130) cases per 100,000 individuals ages <19 years in Manitoba from 1995 to 2004, based only on codes for RA and juvenile arthritis. The limited diagnostic codes were expected to underestimate the true burden of disease, as JIA encompasses a number of different subtypes of pediatric inflammatory arthritis (4). The higher prevalence estimate could be explained by the inclusion of patients between 16 and 19 years of age with JIA or RA meeting case finding definitions, or high sensitivity of the ascertainment algorithm (2).

The present study showed similar incidence estimates but higher prevalence estimates than those reported in administrative data from Kaiser Permanente, a privately managed health care organization in northern California. In the latter, the incidence rate of JIA was 11.9 (95% CI 10.9, 12.9) per 100,000 person-years, and the 2009 point prevalence was 44.7 (95% CI 39.1, 50.2) per 100,000 persons among those ages ≤15 years, standardized to

the 2000 US census population (12). Ethnic differences may contribute to the difference between our findings and these rates, as Manitoba has relatively large proportions of First Nations peoples (American Indians in the US) who may be at particularly high risk of JIA (13), and individuals of European descent (although in this case one would also expect higher incidence rates in Manitoba) (14). We used a Poisson model to standardize by age and sex, while in a recent study by Harrold et al (12), age and sex were standardized by using a method that does not assume a specific distribution for the data. Both methods are valid choices; the difference in the estimates between the methods depends on the plausibility of the Poisson distribution assumption, which we showed to be reasonable using model fit statistics.

Our key finding was a decrease in overall prevalence and, for males, a decrease in both prevalence and incidence over time. Time trends for pediatric chronic arthritis have been previously reported in longitudinal population-based studies. In a study in Rochester, Minnesota (9), decreasing incidence and prevalence trend estimates for juvenile arthritis were noted for females between 1960 and 2013, with the decline occurring prior to 1993. In contrast, an increasing overall incidence of juvenile arthritis was noted in Finland between 1980 and 1995 (15). Of note, in the Minnesota study, the American College of Rheumatology classification for juvenile arthritis (16) was used, omitting enthesitis-related arthritis, which occurs more commonly in men. This difference in classification may explain the difference in results in regard to males compared to our study. An alternate possible explanation for our finding is the utilization of improved diagnostic accuracy for differentiating rheumatic from nonrheumatic musculoskeletal conditions in males, especially in regard to the diagnosis of enthesitis-related arthritis.

The advantages of the present study include the use of a validated case-finding algorithm, the availability of population-based provincial administrative databases, and a longitudinal analysis allowing examination for trends in the descriptive epidemiology of JIA over time. One known limitation of research using administrative health data is the potential misclassification of cases. We used a validated case-finding algorithm to address this limitation. We restricted our study to the years for which this algorithm performed best and when the current ILAR classification of JIA became standard nomenclature in Canada; however, our timeframe may not have been long enough to observe additional temporal patterns. Although case-finding algorithm accuracy performance statistics were consistent throughout the study period (2), and the fact that our results should not have been affected by the adoption of the ILAR classification of JIA in the early 2000s, sensitivity analysis did not confirm the decrease in incidence in males when the first 2 years of the observation period were excluded. The possibility therefore remains that the incidence rate in males did not change significantly over time.

This study provides population estimates of incidence and prevalence of JIA in Manitoba, Canada over time. To our knowledge, there are few longitudinal population-based studies that describe the epidemiology of JIA. Our findings support a possible temporal pattern of decreasing JIA prevalence overall, and decreasing prevalence and incidence in males over time. Our results require confirmation in other populations, but suggest the need to investigate factors potentially contributing to this changing epidemiology. We recommend that future studies assess temporal trends in JIA epidemiology, limit the population at risk to ages <16 years (the upper age limit for JIA according to ILAR), in order to allow comparisons among studies, use validated case-finding algorithms in order to provide reliable and accurate estimates, and use population-based datasets whenever possible. This methodology will advance knowledge regarding the disease burden of JIA with greater accuracy than relying solely on clinical records.

ACKNOWLEDGMENTS

The authors thank the Manitoba Centre for Health Policy for use of data contained in the Manitoba Population Research Data Repository under HIPC # 2013/2014-19.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Thierry S, Fautrel B, Lemelle I, Guillemin F. Prevalence and incidence of juvenile idiopathic arthritis: a systematic review. Joint Bone Spine 2014;81:112–7.
2. Shiff N, Oen K, Rabbani R, Lix L. Validation of administrative case ascertainment algorithms for chronic childhood arthritis in Manitoba, Canada. Rheumatol Int 2017;37:1575–84.
3. World Health Organization. International Classification of Diseases, Eleventh Revision. URL: http://www.who.int/classifications/icd/en/
4. Petty RE, Southwood TR, Manners P, Baum J, Glass DN, Goldenberg J, et al. International League of Associations for Rheumatology classification of juvenile idiopathic arthritis: second revision, Edmonton, 2001. J Rheumatol 2004;31:390–2.
5. Stringer E, Bernatsky S. Validity of juvenile idiopathic arthritis diagnoses using administrative health data. Rheumatol Int 2015;35:575–9.
6. Ehrmann Feldman D, Bernatsky S, Houde M. The incidence of juvenile rheumatoid arthritis in Quebec: a population data-based study. Pediatr Rheumatol Online J 2009;7:20.
7. Pan W. Akaike's information criterion in generalized estimating equations. Biometrics 2001;57:120–5.
8. Petty R, Laxer R, Lindsley C, Wedderburn L. Textbook of pediatric rheumatology. 7th ed. Philadelphia (PA): Elsevier; 2016. p. 188–237.
9. Krause ML, Crowson CS, Michet CJ, Mason T, Muskardin TW, Matteson EL. Juvenile idiopathic arthritis in Olmsted County, Minnesota, 1960–2013. Arthritis Rheumatol 2016;68:247–54.

10. Oen K, Fast M, Postl B. Epidemiology of juvenile rheumatoid arthritis in Manitoba, Canada, 1975–92: cycles in incidence. J Rheumatol 1995;22:745–50.
11. Shiff NJ, Lix LM, Oen K, Joseph L, Duffy C, Stringer E, et al. Chronic inflammatory arthritis prevalence estimates for children and adolescents in three Canadian provinces. Rheumatol Int 2015; 35:345–50.
12. Harrold LR, Salman C, Shoor S, Curtis JR, Asgari MM, Gelfand JM, et al. Incidence and prevalence of juvenile idiopathic arthritis among children in a managed care population, 1996–2009. J Rheumatol 2013;40:1218–25.
13. Saurenmann RK, Rose JB, Tyrrell P, Feldman BM, Laxer RM, Schneider R, et al. Epidemiology of juvenile idiopathic arthritis in a multiethnic cohort: ethnicity as a risk factor. Arthritis Rheum 2007;56:1974–84.
14. Government of Canada. Statistics Canada. Aboriginal Statistics at a Glance. Census Highlight Tables. Population Counts. http://www12.statcan.ca/census-recensement/2006/dp-pd/hlt/97-558/pages/page.cfm?Lang=E&Geo=PR&Code=01&Table=1&Data=Count&Sex=1&Age=1&StartRec=1&Sort=2&Display=Page.
15. Kaipiainen-Seppanen O, Savolainen A. Changes in the incidence of juvenile rheumatoid arthritis in Finland. Rheumatology (Oxford) 2001;40:928–32.
16. Brewer EJ Jr, Bass J, Baum J, Cassidy JT, Fink C, Jacobs J, et al. Current proposed revision of JRA criteria. Arthritis Rheum 1977;20 Suppl 2:195–9.

Arthritis Care & Research
Vol. 71, No. 3, March 2019, pp 419–426
DOI 10.1002/acr.23598

AMERICAN COLLEGE OF RHEUMATOLOGY
EDUCATION • TREATMENT • RESEARCH

Googling Gout: Exploring Perceptions About Gout Through a Linguistic Analysis of Online Search Activities

Kayla N. Jordan,[1] James W. Pennebaker,[1] Keith J. Petrie,[2] and Nicola Dalbeth[2]

Objective. To understand what terms people seeking information about gout use most frequently in online searches and to explore the psychological and emotional tone of these searches.

Methods. A large de-identified data set of search histories from major search engines was analyzed. Participants who searched for gout (n = 1,117), arthritis (arthritis search control group, age and sex-matched, n = 2,036), and a random set of age and sex-matched participants (general control group, n = 2,150) were included. Searches were analyzed using Meaning Extraction Helper and Linguistic Inquiry and Word Count.

Results. The most frequent unique searches in the gout search group included gout-related and food-related terms. Those who searched for gout were most likely to search for words related to eating or avoidance. In contrast, those who searched for arthritis were more likely to search for disease- or health-related words. Compared with the general control group, higher information seeking was observed for the gout and arthritis search groups. Compared with the general control group, both the gout and arthritis search groups searched for more food-related words and fewer leisure and sex-related words. The searches of both the gout and arthritis search groups were lower in positivity and higher in the frequency of sadness-related words.

Conclusion. The perception of gout as a condition managed by dietary strategies aligns with online information seeking about the disease and its management. In contrast, people searching for information about arthritis focus more on medical strategies. Linguistic analyses reflect greater disability in social and leisure activities and lower positive emotion for those searching for gout or arthritis.

INTRODUCTION

Gout is a common form of inflammatory arthritis, with increasing prevalence worldwide (1). The central cause of gout is chronic deposition of monosodium urate crystals (2). Although effective long-term management of gout is available, the results of many studies have shown low levels of prescription and adherence to urate-lowering therapy (3–5). In addition to dietary risk factors, there are many biologic factors that contribute to the development of gout, such as genetic risk factors, kidney disease, and medications (1,6). However, gout is perceived by patients, the general public, and many health care professionals as a self-inflicted condition that is caused primarily by dietary indiscretion (7–10). These perceptions have been present within popular culture for centuries (11,12) and are reflected in contemporary media depictions of the condition (13). These portrayals of the disease may contribute to stigma and a reduced willingness of patients to seek medical attention for gout management because of embarrassment about the condition (9).

Analysis of language has been used to understand and predict both physical and mental health. In the domain of mental health, the words people use have been found to be related to and predictive of outcomes such as depression (14,15), coping with trauma (16,17), and manic episodes (18). Analysis of online search behavior can be used to track and predict physical health issues such as pregnancy (19), breast cancer (20), and pancreatic cancer (21). Furthermore, the language people use and their online behavior can provide insights into how people perceive health issues. For example, analysis of language use in blogs and newspapers along with Wikipedia visits during a swine flu outbreak showed a rapid, short-term increase in information seeking on Wikipedia following the outbreak, an increase in anxiety

Supported by the National Science Foundation (IS-1344257) and the Templeton Foundation (48503).

[1]Kayla N. Jordan, MSc, James W. Pennebaker, PhD: University of Texas at Austin; [2]Keith J. Petrie, PhD, Nicola Dalbeth, MD, FRACP: University of Auckland, Auckland, New Zealand.

Dr. Dalbeth has received consulting and/or speaking fees or honoraria from Takeda, ArdeaBiosciences/AstraZeneca, Kowa, AbbVie, Janssen, and Horizon (less than $10,000 each).

Address correspondence to Nicola Dalbeth, MD, FRACP, Department of Medicine, Faculty of Medical and Health Sciences, University of Auckland, 85 Park Road, Grafton, Auckland, New Zealand. E-mail: n.dalbeth@auckland.ac.nz.

Submitted for publication November 21, 2017; accepted in revised form May 15, 2018.

SIGNIFICANCE & INNOVATIONS

- Linguistic analysis of online search behavior can be used to understand how people might perceive health conditions.
- People searching for knowledge about gout or arthritis engage in high levels of information seeking.
- The perception of gout as an illness that is primarily managed by dietary strategies aligns with online information seeking about the disease and its management.
- Linguistic analyses reflect greater disability in social and leisure activities and lower positive emotion for those searching for information about gout or arthritis than are found in controls.

and negative emotion in those talking about the outbreak, and greater concern with health and death (22). Collectively, previous research demonstrates the potential of language and online behavior to help researchers understand how people respond to health threats.

Building on this prior research and the high prevalence of seeking online health information (23), online search behavior can be explored to understand how people might perceive health conditions and the management of these conditions. The aim of this study was to understand what terms are used most frequently by people seeking information about gout in their online searches, and to explore the psychological and emotional tone of these searches using a linguistic analysis of search histories.

MATERIALS AND METHODS

In cooperation with Microsoft Research, a large de-identified data set of search histories from 200,000 consenting individuals was obtained from ComScore, a web analytics company, covering a 2-year period (2011–2013). In exchange for incentives such as cash and software, the company installed software on participants' computers to passively monitor online search activities. Time-stamped search terms were logged from major search engines (Google, Bing, Yahoo, Ask). Basic demographics (sex, age, income, and geographic region) were provided by the analytics company based on user self-reports. The data were de-identified by replacing names, addresses, number strings (e.g., phone numbers, zipcodes, and social security numbers), and URLs with anonymizing codes. For example, a zip code such as 55111 was replaced with "zipcode," or a first name such as Jane was replaced with "fname." All characters were translated into machine-readable characters. Common misspellings were corrected. Once the texts were cleaned, search terms were aggregated by participant over the entire period.

From the larger data set, 3 groups were identified: participants who searched for gout at least once (n = 1,117), participants who searched for arthritis (arthritis search control group, age- and sex-matched to the gout search group, n = 2,036), and a random set of participants matched for age and sex who had not searched for gout or arthritis (general control group, n = 2,150). The general control group size was chosen to approximately match the size of the arthritis search group. Additionally, the general control group was chosen to match for the age and sex composition of both the gout and arthritis search groups. For the analyzed data set with the gout, arthritis, and general control groups, the total number of search words was 25,000,811. All participants included in the analysis were age ≥18 years. The Office of Research Support at the University of Texas at Austin reviewed the study protocol and advised that because the study involved secondary use of a de-identified data set with no direct links to identifiers, institutional review board review and oversight were not required.

Meaning Extraction Helper, a word frequency software, was used to calculate search term frequencies from each participant's search history (http://meh.ryanb.cc, version 1.4.15). Search terms were also analyzed using Linguistic Inquiry and

Table 1. Demographics by search group*

	Gout (n = 1,117)	Arthritis (n = 2,036)	General control (n = 2,150)
Age, years			
18–24	18.35	21.91	16.98
25–34	16.03	18.66	18.37
35–44	15.67	15.37	20.70
45–54	20.68	16.55	15.16
≥55	22.56	20.92	28.79
Sex			
Unknown	0.45	0.44	0.00
Women	54.97	56.53	50.37
Men	44.58	43.03	49.63
Annual income, $US			
<15,000	11.28	9.38	8.93
15,000–24,999	6.45	5.75	6.00
25,000–39,999	10.03	9.58	8.65
40,000–59,999	16.03	19.94	17.12
60,000–74,999	27.39	22.30	24.51
75,000–99,999	12.35	12.43	12.05
≥100,000	8.59	10.76	10.37

* Values are the percentage.

Word Count (LIWC), a psychological text-analysis software from which psychological processes can be inferred based on the words people use (http://liwc.wpengine.com) (24). Search group membership was dummy-coded with the general control search group as the comparison group. Search group membership was correlated with individual word frequencies using Spearman correlations using the psych package for R (https://CRAN.R-project.org/package=psych, version = 1.7.8). One-way analysis of variance with Tukey's post hoc tests was used for LIWC categories to examine differences between the 3 search groups (gout search, arthritis search, and general control search) in terms of how the participants were searching for information, using SPSS software, version 22.0.

RESULTS

The demographic features of participants are shown in Table 1. Less than half of the participants were men, with a wide range of income and ages. Table 2 includes examples from the data set to show what people searched for when they searched for gout. An analysis of the 50 most frequent unique searches revealed gout-related terms (such as uric, purine, kidneys) and food-related terms (e.g., meats, atkins, mash) in the group searching for gout (Table 3). For the group searching for arthritis, there were no food-related terms in the 50 most frequent unique searches, but medical terms were commonly observed (e.g., fibromyalgia, inflammatory, psoriatic, Crohns, hospitalized).

Tables 4 and 5 show Spearman's correlations between search group membership and individual word frequencies, for the 10 most frequently correlated words. The majority of the words most often correlated with gout search group membership related to food or avoidance (e.g., food, diet, avoid, recipe, eat). In contrast, although some food words were correlated with the arthritis search group (e.g., food, diet, recipe, eat), stronger correlations were observed in the arthritis search group for disease or health-related words (e.g., rheumatoid, pain, disease, symptom). All of the words most often correlated with the arthritis search group membership were disease or health-related.

In the LIWC analysis, compared with the general control search group, the total word count was higher for the gout and arthritis search groups, indicating a higher level of information seeking in both groups (Figure 1 and Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23598/abstract). Insight words (e.g., know, learn, means) were also used more often in both the gout and arthritis search groups compared with the general control group. Although both the gout search group and the arthritis search group searched for more health-related words than the general control search group, the frequency of health-related words searched was lower for the gout search group than the arthritis search group. Compared with the general control search group, both the gout and arthritis search groups searched more frequently for ingestion-related words (mainly food-related terms like dish, eat, pizza).

Compared with the general control search group, the groups of participants searching for gout and arthritis searched for fewer leisure (e.g., cook, chat, movie) and sexual (e.g. love, sex) words (Figure 2 and Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23598/abstract). There were also fewer social words (e.g., family, talk, they) in the arthritis search group, with a similar trend for the gout search group (Tukey's $P = 0.052$ compared with the general control group). The searches of both the gout and arthritis search groups were less positive in emotional tone (lower in positivity and higher in the frequency of sadness-related words).

Overall, there were very few differences between the gout and arthritis search groups in the LIWC analysis, with the exception of a higher total word count and the use of health words by the arthritis search group, and a higher use of insight words (e.g., know, learn, means) by the gout search group (Figures 1 and 2, and Supplementary Table 1, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.23598/abstract).

Table 2. Examples of internet computer searches in the group of participants who searched for gout*

1	vw volkswagen beetle 2012 gout home treatment medications home remedies for gout pain home remedies for gout pain relief
2	youtube craigslist ebay black spider helicopter youube ebay amozon allopurinol side effects gout diet free gout diet foods that help lower uric acid omeprazole
3	horse trailer gout diet low purine diet xname bow reviews gun shows gun town
4	list of gout food cause urid acid in the body opusbank
5	female naked pics craigslist female naked pics female naked pics female naked pics gahout gaout gout symptoms and causes
6	what causes gout what populations are susceptible to gout anti reject drugs for organ transplants anti reject drugs for organ transplants gout gout
7	truth or fiction truth or fiction snopes truth or fiction snopes com truth or fiction snopes pledge of allegiance pledge of allegiance by obama and what causes hiccups what causes gout what causes fog what causes gurgling in sewer what causes gurgling in sewer lines

* Each numbered point shows a different individual's searches on a given day.

Table 3. Most frequent unique searches in the gout search group and the arthritis search group

Gout		Arthritis	
Word	% of people searching	Word	% of people searching
uric	6.27	fibromyalgia	6.53
sid	3.76	jwoww	4.52
meats	3.58	supposed	4.52
amusement	3.40	inflammatory	4.47
verdict	3.40	penney	4.42
wraps	3.40	psoriatic	4.42
reply	3.31	epilepsy	4.37
hoes	3.31	hoax	4.32
scams	3.31	brett	4.32
hates	3.31	asleep	4.27
vac	3.22	dvds	4.22
purine	3.22	controlled	4.17
mustard	3.13	feelings	4.17
hourly	3.13	equivalent	4.08
ache	3.13	outbreak	4.08
tsa	3.13	mitt	4.08
herman	3.13	weakness	4.03
erase	3.13	cds	4.03
renee	3.13	enjoy	4.03
rodman	3.13	ripa	3.98
atkins	3.13	kicked	3.98
comm	3.04	hugo	3.93
bananas	3.04	collision	3.93
hometown	3.04	andrea	3.93
directionsxnumber	3.04	mowry	3.88
airfare	3.04	addicted	3.88
herbs	3.04	crohn's	3.88
ego	2.95	postage	3.83
mash	2.95	began	3.83
arabia	2.95	patricia	3.83
bees	2.95	slams	3.83
buckle	2.95	sq	3.78
combs	2.95	menstrual	3.78
candidates	2.95	latifah	3.78
caroline	2.95	stains	3.78
jonah	2.95	shoppe	3.78
courteney	2.86	flexible	3.73
kidneys	2.86	popping	3.73
coop	2.86	hospitalized	3.73
dew	2.86	byxnumber	3.68
plow	2.86	transcript	3.68
keira	2.86	attempt	3.68
knightley	2.86	radius	3.68
coaches	2.86	criteria	3.68
martinez	2.86	expectancy	3.68

(continued)

Table 3 *(Cont'd)*

Gout		Arthritis	
Word	% of people searching	Word	% of people searching
blur	2.86	sightings	3.68
strawberries	2.86	organized	3.63
verification	2.86	miscarriage	3.63
jovovich	2.86	levi	3.63
gellar	2.86	jackman	3.63
pans	2.86	adjust	3.63
cinnamon	2.86	cries	3.63
doug	2.86	appears	3.63
–		attacked	3.63

DISCUSSION

This linguistic analysis of online search activities has shown that people searching for knowledge about gout or arthritis engage in high levels of information seeking. Consistent with prior research examining patient and community perceptions of disease (7,9,13,25), unique searches in the gout search group included more food-related words, and words related to food and dietary strategies were correlated with membership in the gout search group. Although similar strengths of correlation were observed between food-related words and the arthritis search group, much stronger correlations were observed between medical-related words and the arthritis search group. The perception of gout as an illness managed by dietary strategies aligns with online information seeking about the disease and its management. In contrast, people searching for information about arthritis are more focused on searching for information on medical strategies.

The LIWC analysis indicated that the emotional experiences of people searching for gout or arthritis were similar, with lower positive emotional tone and higher numbers of sadness-related words compared with searches by the general control participants. Our linguistic analyses reflect greater disability in social and leisure activities and lower positive emotion for those searching for medical conditions such as gout or arthritis, compared to people who do not search for these conditions. Although we cannot assume that all participants searching for gout had the condition, these findings align with prior studies that have shown higher rates of depression (26) and reduced social participation in people with gout (27).

Although gout occurs more frequently in men than women, notably less than half the participants searching for gout were men. These findings are consistent with the results of prior research, which has demonstrated that although men and women have equal access to the internet in the US, women

Table 4. Highest individual word frequencies correlated with the gout search group*

Word	Gout search group, r	Arthritis search group, r
food	0.16†	0.14†
diet	0.14†	0.11†
symptom	0.13†	0.24†
avoid	0.13†	0.04
acid	0.11†	0.10†
recipe	0.10†	0.14†
county	0.10†	0.11†
map	0.10	0.07
lawn	0.10	0.06
eat	0.10	0.14†

* In this analysis, different forms of the same word were combined (e.g., recipe, recipes). Spearman's correlations were calculated and the 10 words most correlated with the gout search group are shown.
† Significant at $P < 0.0001$.

Table 5. Highest individual word frequencies correlated with the arthritis search group*

Word	Arthritis search group, r	Gout search group, r
rheumatoid	0.40†	–0.12†
pain	0.29†	0.07
disease	0.25†	0.06
symptom	0.24†	0.13†
treatment	0.22†	0.08†
medical	0.21†	0.08†
knee	0.20†	0.02
blood	0.20†	0.09†
cancer	0.20†	0.06
syndrome	0.20†	0.05

* In this analysis, different forms of the same word were combined (e.g., symptom, symptoms). Spearman's correlations were calculated and the 10 words most correlated with the arthritis search group are shown.
† Significant at $P < 0.0001$.

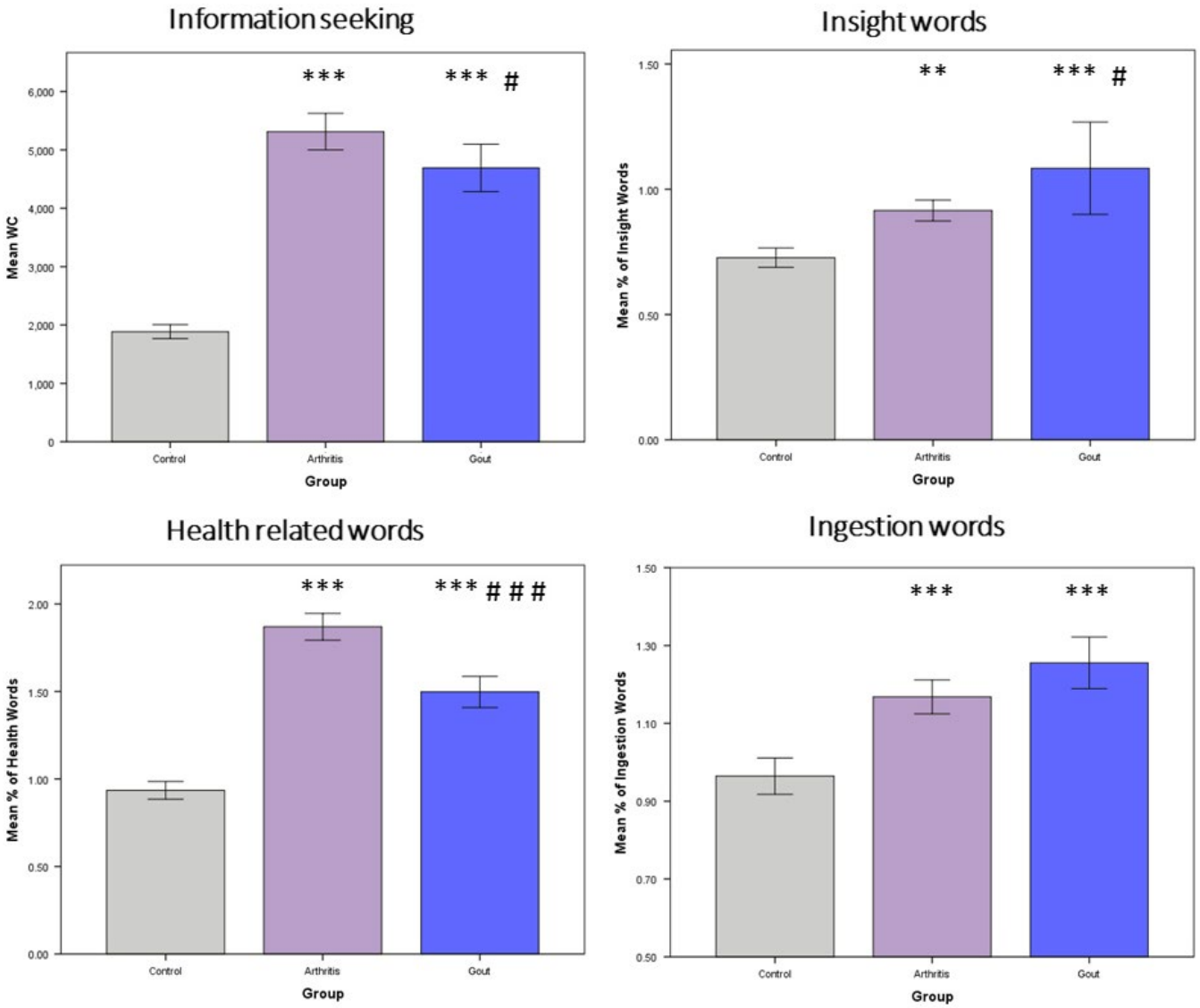


Figure 1. Linguistic Inquiry and Word Count analysis for information seeking, and insight, health-related, and ingestion words. Information seeking is a measure of word count (WC). All other categories represent the percentage of word count comprising words in that category. Data are shown as the mean (95% confidence interval) for the group who did not search for gout or arthritis (Control), the group of participants who searched for arthritis (Arthritis), and the group of participants who searched for gout (Gout). ** = $P < 0.01$; *** = $P < 0.001$ compared with general control group; # = $P < 0.05$ compared with the group of participants searching for arthritis; ### = $P < 0.001$ compared with the group of participants searching for arthritis, using Tukey's post hoc test. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.23598/abstract.

are more likely to gather health-related information online (23). Possibly women with gout searched for this condition more often because available education resources are of limited relevance to women with gout (28). However, women may also have searched for gout due to contact with a family member, friend, or acquaintance who was affected by gout, or for some unrelated reason.

Results of a previous qualitative study showed that gout is not seen as a form of arthritis, but may be viewed as a different kind of rheumatic disease (7). The observation in our study that people searching for arthritis search more often for medical words raises the possibility that including the word arthritis in a disease label, or in information provided about gout, may have an impact on understanding effective management strategies for gout. A recent study by our group has shown that supermarket shoppers view the condition "urate crystal arthritis" as a more chronic and serious condition that requires long-term medications, whereas the some condition labeled as "gout" is viewed as a condition caused by poor diet and over-consumption of alcohol, and requires dietary management (25).

The key limitation to this study is that very little personal information about participants was available, and we could not

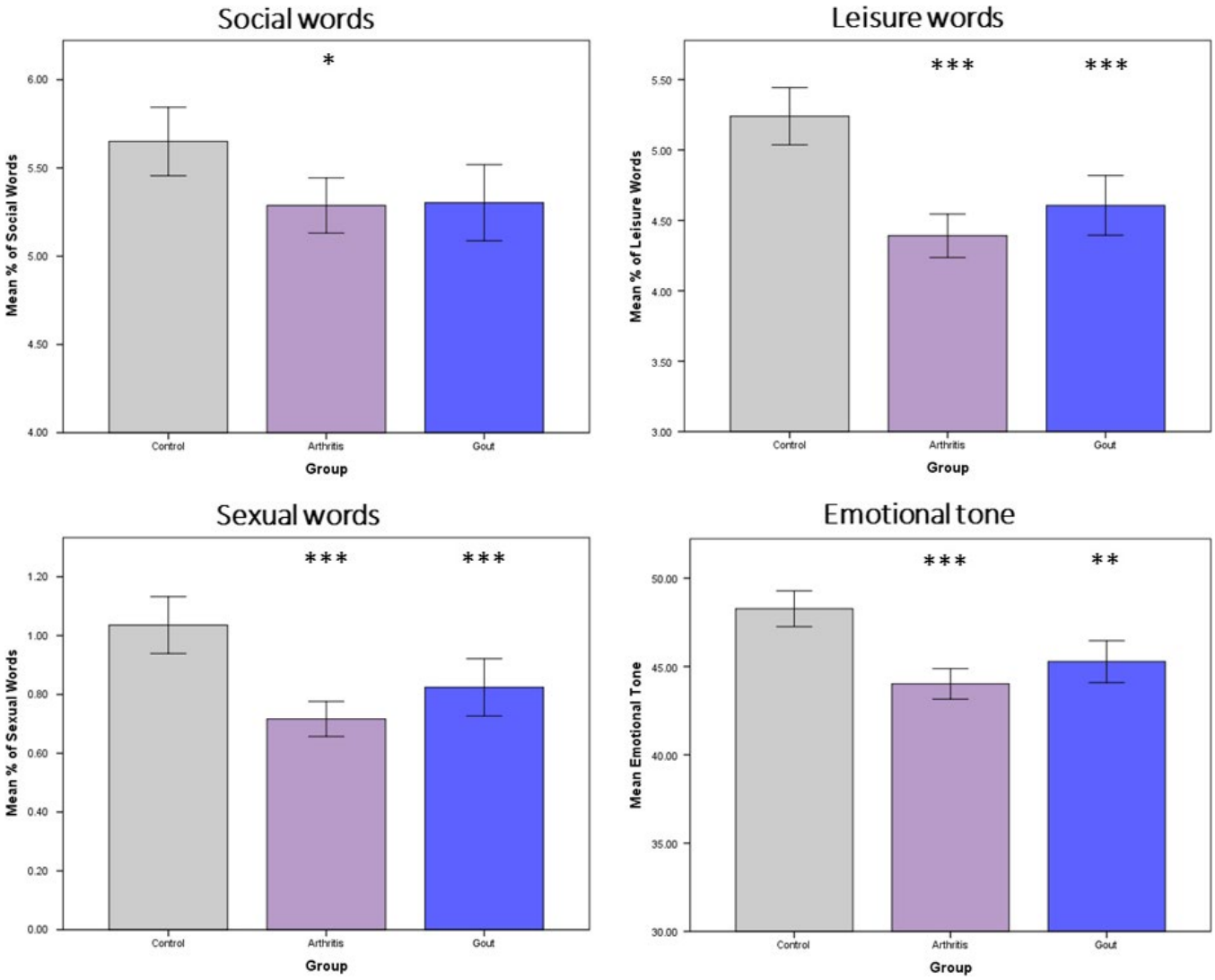


Figure 2. Linguistic Inquiry and Word Count analysis for social, leisure, and sexual words, and emotional tone. Tone is a standardized difference of positive and negative emotion words ranging from 0 to 100. All other categories represent the percentage of word count comprising words in that category. Data are shown as the mean (95% confidence interval) for the group who did not search for gout or arthritis (Control), the group of participants who searched for arthritis (Arthritis), and the group of participants who searched for gout (Gout). * = $P < 0.05$; ** = $P < 0.01$; *** = $P < 0.001$ compared with general control group, using Tukey's post hoc test. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.23598/abstract.

determine whether participants in the gout or arthritis search groups had the condition, or were searching for information about the condition due to concerns about a relative or friend, due to curiosity, or for information gathering as a student/health care professional. Furthermore, searching for a disease does not necessarily equate to a medical diagnosis. For example, results of prior research have shown that a relative Google search volume correlates with incidence and mortality rates of some, but not all cancers (29). Despite these limitations, the search patterns found in this study provide novel information about perceptions of gout and its management. The results are consistent with the popular perception that gout is a condition that is primarily managed by dietary restrictions.

ACKNOWLEDGMENTS

The authors thank Eric Horvitz and Ryen White for their assistance and work with data collection for this study.

AUTHOR CONTRIBUTIONS

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

Study conception and design. Jordan, Pennebaker, Petrie, Dalbeth.
Acquisition of data. Jordan, Pennebaker.
Analysis and interpretation of data. Jordan, Pennebaker, Petrie, Dalbeth.

REFERENCES

1. Kuo CF, Grainge MJ, Zhang W, Doherty M. Global epidemiology of gout: prevalence, incidence and risk factors. Nat Rev Rheumatol 2015;11:649–62.
2. Dalbeth N, Merriman TR, Stamp LK. Gout. Lancet 2016;388:2039–52.
3. Solomon DH, Avorn J, Levin R, Brookhart MA. Uric acid lowering therapy: prescribing patterns in a large cohort of older adults. Ann Rheum Dis 2008;67:609–13.
4. Singh JA, Akhras KS, Shiozawa A. Comparative effectiveness of urate lowering with febuxostat versus allopurinol in gout: analyses from large U.S. managed care cohort. Arthritis Res Ther 2015;17:120.
5. Harrold LR, Andrade SE, Briesacher BA, Raebel MA, Fouayzi H, Yood RA, et al. Adherence with urate-lowering therapies for the treatment of gout. Arthritis Res Ther 2009;11:R46.
6. Kottgen A, Albrecht E, Teumer A, Vitart V, Krumsiek J, Hundertmark C, et al. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. Nat Genet 2013;45:145–54.
7. Spencer K, Carr A, Doherty M. Patient and provider barriers to effective management of gout in general practice: a qualitative study. Ann Rheum Dis 2012;71:1490–5.
8. Humphrey C, Hulme R, Dalbeth N, Gow P, Arroll B, Lindsay K. A qualitative study to explore health professionals' experience of treating gout: understanding perceived barriers to effective gout management. J Prim Health Care 2016;8:149–56.
9. Lindsay K, Gow P, Vanderpyl J, Logo P, Dalbeth N. The experience and impact of living with gout: a study of men with chronic gout using a qualitative grounded theory approach. J Clin Rheumatol 2011;17:1–6.
10. Chandratre P, Mallen CD, Roddy E, Liddle J, Richardson J. "You want to get on with the rest of your life": a qualitative study of health-related quality of life in gout. Clin Rheumatol 2016;35:1197–205.
11. Nuki G, Simkin PA. A concise history of gout and hyperuricemia and their treatment. Arthritis Res Ther 2006;8 Suppl 1:S1.
12. Porter R. Gout: framing and fantasizing disease. Bull Hist Med 1994;68:1–28.
13. Duyck SD, Petrie KJ, Dalbeth N. "You don't have to be a drinker to get gout, but it helps": a content analysis of the depiction of gout in popular newspapers. Arthritis Care Res (Hoboken) 2016;68:1721–5.
14. Edwards T, Holtzman NS. A meta-analysis of correlations between depression and first person singular pronoun use. J Res Personality 2017;68 Suppl C:63–8.
15. Rude S, Gortner E, Pennebaker J. Language use of depressed and depression-vulnerable college students. Cogn Emotion 2004;18:1121–33.
16. Fernández I, Páez D, Pennebaker JW. Comparison of expressive writing after the terrorist attacks of September 11th and March 11th. Int J Clin Health Psychol 2009;9:89–103.
17. Pulverman CS, Lorenz TA, Meston CM. Linguistic changes in expressive writing predict psychological outcomes in women with history of childhood sexual abuse and adult sexual dysfunction. Psychol Trauma 2015;7:50–7.
18. Yom-Tov E, White RW, Horvitz E. Seeking insights about cycling mood disorders via anonymized search logs. J Med Internet Res 2014;16:e65.
19. Fourney A, White RW, Horvitz E. Exploring time-dependent concerns about pregnancy and childbirth from search logs. In: proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems; 2015; Seoul, Republic of Korea. p. 737–46.
20. Paul MJ, White RW, Horvitz E. Search and breast cancer: on episodic shifts of attention over life histories of an illness. ACM Trans Web 2016;10:1–27.
21. Paparrizos J, White RW, Horvitz E. Screening for pancreatic adenocarcinoma using signals from web search logs: feasibility study and results. J Oncol Pract 2016;12:737–44.
22. Tausczik Y, Faasse K, Pennebaker JW, Petrie KJ. Public anxiety and information seeking following the H1N1 outbreak: blogs, newspaper articles, and Wikipedia visits. Health Commun 2012;27:179–85.
23. Fox S, Duggan M. Health online 2013. Pew Research Center. URL: http://www.pewinternet.org/2013/01/15/health-online-2013/.
24. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The development and psychometric properties of LIWC2015. Austin (TX): University of Texas at Austin; 2015.
25. Petrie KJ, MacKrill K, Derksen C, Dalbeth N. An illness by any other name: the effect of renaming gout on illness and treatment perceptions. Health Psychol 2018;37:37–41.
26. Lin S, Zhang H, Ma A. Association of gout and depression: a systematic review and meta-analysis. Int J Geriatr Psychiatry 2018;33:441–8.
27. Aati O, Taylor WJ, Horne A, Dalbeth N. Toward development of a tophus impact questionnaire: a qualitative study exploring the experience of people with tophaceous gout. J Clin Rheumatol 2014;20:251–5.
28. Richardson JC, Liddle J, Mallen CD, Roddy E, Prinjha S, Ziebland S, et al. "Why me? I don't fit the mould … I am a freak of nature": a qualitative study of women's experience of gout. BMC Womens Health 2015;15:122.
29. Wehner MR, Nead KT, Linos E. Correlation among cancer incidence and mortality rates and internet searches in the United States. JAMA Dermatol 2017;153:911–4.

Arthritis Care & Research
Vol. 71, No. 3, March 2019, pp 427–434
DOI 10.1002/acr.23607

AMERICAN COLLEGE OF RHEUMATOLOGY
EDUCATION • TREATMENT • RESEARCH

Gout, Hyperuricemia, and Crystal-Associated Disease Network Consensus Statement Regarding Labels and Definitions for Disease Elements in Gout

David Bursill,[1] William J. Taylor,[2] Robert Terkeltaub,[3] Masanari Kuwabara,[4] Tony R. Merriman,[5] Rebecca Grainger,[2] Carlos Pineda,[6] Worawit Louthrenoo,[7] N. Lawrence Edwards,[8] Mariano Andrés,[9] Ana Beatriz Vargas-Santos,[10] Edward Roddy,[11] Tristan Pascart,[12] Ching-Tsai Lin,[13] Fernando Perez-Ruiz,[14] Sara K. Tedeschi,[15] Seoyoung C. Kim,[15] Leslie R. Harrold,[16] Geraldine McCarthy,[17] Nitin Kumar,[18] Peter T. Chapman,[19] Anne-Kathrin Tausche,[20] Janitzia Vazquez-Mellado,[21] Marwin Gutierrez,[22] Geraldo da Rocha Castelar-Pinheiro,[10] Pascal Richette,[23] Eliseo Pascual,[9] Mark C. Fisher,[24] Ruben Burgos-Vargas,[21] Philip C. Robinson,[25] Jasvinder A. Singh,[26] Tim L. Jansen,[27] Kenneth G. Saag,[28] Ole Slot,[29] Tillmann Uhlig,[30] Daniel H. Solomon,[15] Robert T. Keenan,[31] Carlo Alberto Scire,[32] Edyta Biernat-Kaluza,[33] Mats Dehlin,[34] George Nuki,[35] Naomi Schlesinger,[36] Matthijs Janssen,[27] Lisa K. Stamp,[37] Francisca Sivera,[38] Anthony M. Reginato,[39] Lennart Jacobsson,[34] Frédéric Lioté,[23] Hang-Korng Ea,[23] Ann Rosenthal,[40] Thomas Bardin,[23] Hyon K. Choi,[24] Michael S. Hershfield,[41] Christine Czegley,[42] Sung Jae Choi,[43] and Nicola Dalbeth[44]

Objective. The language currently used to describe gout lacks standardization. The aim of this project was to develop a consensus statement on the labels and definitions used to describe the basic disease elements of gout.

Methods. Experts in gout (n = 130) were invited to participate in a Delphi exercise and face-to-face consensus meeting to reach consensus on the labeling and definitions for the basic disease elements of gout. Disease elements and labels in current use were derived from a content analysis of the contemporary medical literature, and the results of this analysis were used for item selection in the Delphi exercise and face-to-face consensus meeting.

Results. There were 51 respondents to the Delphi exercise and 30 attendees at the face-to-face meeting. Consensus agreement (≥80%) was achieved for the labels of 8 disease elements through the Delphi exercise; the remaining 3 labels reached consensus agreement through the face-to-face consensus meeting. The agreed labels were monosodium urate crystals, urate, hyperuric(a)emia, tophus, subcutaneous tophus, gout flare, intercritical gout, chronic gouty arthritis, imaging evidence of monosodium urate crystal deposition, gouty bone erosion, and podagra. Participants at the face-to-face meeting achieved consensus agreement for the definitions of all 11 elements and a recommendation that the label "chronic gout" should not be used.

Conclusion. Consensus agreement was achieved for the labels and definitions of 11 elements representing the fundamental components of gout etiology, pathophysiology, and clinical presentation. The Gout, Hyperuricemia, and Crystal-Associated Disease Network recommends the use of these labels when describing the basic disease elements of gout.

INTRODUCTION

The language used to describe gout has substantial deficiencies in both accuracy and precision (1–3). A recent computational analysis of the medical literature found that the terminology used for gout, its treatment, and outcomes, was inconsistent, ambiguous, and frequently misleading (1). A widely accepted nomenclature for gout would have important

G-CAN was supported by Ironwood Pharmaceuticals, Horizon Pharmaceuticals, SOBI, Takeda, Teijin, Selecta, and CymaBay.

[1]David Bursill, MBBS: University of Auckland, Auckland, New Zealand, and Adelaide Medical School, University of Adelaide, South Australia, Australia; [2]William J. Taylor, PhD, MBChB, Rebecca Grainger, BMedSci, MBChB, PhD: University of Otago, Wellington, and Hutt Valley District Health Board, Lower Hutt, New Zealand; [3]Robert Terkeltaub, MD: Veterans Affairs Medical Center and University of California, San Diego; [4]Masanari Kuwabara, MD, PhD: Toranomon Hospital, Tokyo, Japan, and University of Colorado Denver, Aurora; [5]Tony R. Merriman, PhD: University of Otago, Dunedin, New Zealand; [6]Carlos Pineda, MD, PhD: Instituto Nacional Rehabilitación Luis Guillermo Ibarra Ibarra, Mexico City, Mexico; [7]Worawit Louthrenoo, MD: Chiang Mai University, Chiang Mai, Thailand; [8]N. Lawrence Edwards, MD, MPH: University of Florida College of Medicine, Gainesville; [9]Mariano Andrés, MD, Eliseo Pascual, MD, PhD: Hospital Universitario de Alicante and Universidad Miguel Hernández, Alicante, Spain; [10]Ana Beatriz Vargas-Santos,

SIGNIFICANCE & INNOVATIONS

- The terminology currently used for gout lacks standardization and has notable deficiencies in its precision, accuracy, and clarity.
- This international collaborative project provides a consensus statement on the labels and definitions given to the basic disease elements of gout.
- The adoption of this standardized nomenclature could benefit communications in gout-related research and management.

implications for standardization in research and clinical settings.

The Gout, Hyperuricemia, and Crystal-Associated Disease Network (G-CAN) is an international, multidisciplinary collaboration, dedicated to advancing the treatment of the crystal deposition–associated disorders. In recognition of the current deficits in gout terminology, G-CAN has supported the development of a Gout Nomenclature Project, aimed at establishing consensus on the labels and definitions used to describe this disease. The intended audience for this consensus statement includes health care professionals and nonphysician scientists in both clinical and research settings. Here we describe the methodology and results of the G-CAN Gout Nomenclature Project concerning the fundamental components of gout etiology, pathophysiology, and clinical presentation, referred to hereafter as the "basic disease elements."

MATERIALS AND METHODS

There were 3 stages to this project. A content analysis of the literature was undertaken to identify the basic disease elements and the labels currently used to denote them. The findings from this analysis provided a framework for the subsequent group consensus exercises. A Delphi exercise was used to achieve consensus agreement on the recommended labels for each disease element. A subsequent face-to-face consensus meeting had 2 aims: to reach agreement on any labels that had not achieved consensus through the Delphi exercise, and to reach consensus agreement on definitions for each of the disease elements. An overview of the project is shown in Figure 1.

Content analysis of the literature. The purpose of this analysis of the literature was to identify the range of elements that collectively represent the basic disease elements of gout. Furthermore, the analysis aimed to determine the number and frequency of labels used for each of these elements in the contemporary medical literature. This analysis has been described separately (2). Briefly, all journal articles containing reference to gout or hyperuricemia were examined within the 10 highest-ranked general rheumatology and the 5 highest-ranked general internal medicine journals (according to the 2015 Thomson-Reuters Journal Citation Reports) from January 1, 2012 to January 31, 2017, inclusive.

Delphi exercise. The principles of the Delphi technique have been described in detail elsewhere (4). In essence, it entails

MD, Geraldo da Rocha Castelar-Pinheiro, MD, PhD: State University of Rio de Janeiro, Rio de Janeiro, Brazil; [11]Edward Roddy, MD: Keele University, Keele, UK; [12]Tristan Pascart, MD, PhD: Lille Catholic University and Saint-Philibert Hospital, Lomme, France; [13]Ching-Tsai Lin, MD: Taichung Veterans General Hospital, Taichung, Taiwan; [14]Fernando Perez-Ruiz, MD, PhD: University of the Basque Country, Biscay, and Cruces University Hospital and Biocruces Health Research Institute, Baracaldo, Spain; [15]Sara K. Tedeschi, MD, MPH, Seoyoung C. Kim, MD, ScD, MSCE, Daniel H. Solomon, MD, MPH: Harvard Medical School, and Brigham and Women's Hospital, Boston, Massachusetts; [16]Leslie R. Harrold, MD, MPH: Corrona, LLC, Waltham, and University of Massachusetts Medical School, Worcester; [17]Geraldine McCarthy, MD: Mater Misericordiae University Hospital and University College, Dublin, Ireland; [18]Nitin Kumar, PhD: Henry Ford Hospital, Detroit, Michigan; [19]Peter T. Chapman, MD: Christchurch Hospital, Christchurch, New Zealand; [20]Anne-Kathrin Tausche, MD: University Hospital Carl Gustav Carus, Technical University Dresden, Dresden, Germany; [21]Janitzia Vazquez-Mellado, MD, PhD, Ruben Burgos-Vargas, MD: Hospital General de México and Universidad Nacional Autónoma de México, Mexico City, Mexico; [22]Marwin Gutierrez, MD: Instituto Nacional Rehabilitación, Mexico City, Mexico; [23]Pascal Richette, MD, PhD, Frédéric Lioté, MD, PhD, Hang-Korng Ea, MD, PhD, Thomas Bardin, MD: Hôpital Lariboisière, Assistance Publique-Hopitaux de Paris, and INSERM UMR-1132 and Université Paris Diderot, Paris, France; [24]Mark C. Fisher, MD, Hyon K. Choi, MD, DrPH: Harvard Medical School and Massachusetts General Hospital Boston; [25]Philip C. Robinson, MBChB, PhD: University of Queensland School of Medicine and Royal Brisbane and Women's Hospital, Herston, Queensland, Australia; [26]Jasvinder A. Singh, MD, MPH: Veterans Affairs Medical Center, Birmingham, and University of Alabama at Birmingham; [27]Tim L. Jansen, MD, PhD, Matthijs Janssen, MD, PhD: Viecuri Medical Centre, Venlo, The Netherlands; [28]Kenneth G. Saag, MD, MSc: University of Alabama at Birmingham; [29]Ole Slot, MD: Rigshospitalet Glostrup, Glostrup, Denmark; [30]Tillmann Uhlig, PhD: Diakonhjemmet Hospital, Oslo, Norway; [31]Robert T. Keenan, MD, MPH, MBA: Duke University School of Medicine, Durham, North Carolina; [32]Carlo Alberto Scire, MD, PhD: University of Ferrara, Ferrara, and Italian Society for Rheumatology, Milan, Italy; [33]Edyta Biernat-Kaluza, PhD: ORLIK, Warsaw, Poland; [34]Mats Dehlin, MD, PhD, Lennart Jacobsson, MD, PhD: Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; [35]George Nuki, MBBS: University of Edinburgh, Edinburgh, UK; [36]Naomi Schlesinger, MD: Rutgers–Robert Wood Johnson Medical School, Piscataway, New Jersey; [37]Lisa K. Stamp, MBChB, PhD: University of Otago, Christchurch, New Zealand; [38]Francisca Sivera, MD, PhD: Hospital General Universitario de Elda, Alicante, Spain; [39]Anthony M. Reginato, MD, PhD: Warren Alpert School of Medicine at Brown University, Providence, Rhode Island; [40]Ann Rosenthal, MD: Medical College of Wisconsin and the Clement J. Zablocki Veterans Affairs Medical Center, Milwaukee; [41]Michael S. Hershfield, MD: Duke University Medical Center, Durham, North Carolina; [42]Christine Czegley, MSc: Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany; [43]Sung Jae Choi, MD, PhD: University of California, San Diego, and Korea University Ansan Hospital, Ansan, South Korea; [44]Nicola Dalbeth, MD: University of Auckland, Auckland, New Zealand.

Dr. Taylor has received honoraria from Pfizer New Zealand (less than $10,000). Dr. Terkeltaub has received research support from Ardea/AstraZeneca and Ironwood, and consulting fees from SOBI, Kowa, Horizon, and Relburn (less than $10,000 each), and from Selecta (more than $10,000). Dr. Merriman has received research funding and consulting fees from Ardea Biosciences and consulting fees from Horizon (less than $10,000 each). Dr. Grainger has received speaking fees from AbbVie, Pfizer, and Janssen (less than $10,000 each) and a research grant from AbbVie. Dr. Louthrenoo has received speaking and/or consulting fees from Astellas Pharma and speaking fees from American Taiwan Biopharm (less than $10,000 each). Dr. Edwards has received consulting fees from Horizon, Ironwood, and Takeda (less than $10,000 each). Dr. Andrés has received consulting fees from AstraZeneca, Grünenthal, and Horizon (less than $10,000 each). Dr. Pascart has received consulting fees from Ipsen Pharma and Mayoly Spindler (less than $10,000 each). Dr. Perez-Ruiz has received speaking and/or consulting fees from Amgen, Ardea Biosciences, AstraZeneca, Faes Farma, Menarini, Horizon, and Pharmaceutical Laboratories ROVI (less than $10,000 each),

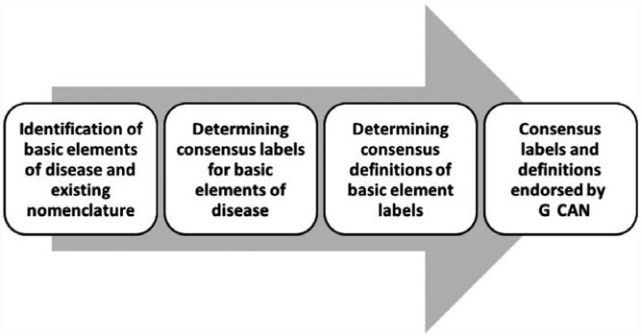


Figure 1. Schema for the Gout Nomenclature Project. G-CAN = Gout, Hyperuricemia, and Crystal-Associated Disease Network.

a series of rounds over which proposed options are sequentially refined by expert opinion until consensus is met. The Delphi exercise in this project consisted of 3 rounds, each conducted as an online survey using Qualtrics software. Experts in gout, identified by their membership in G-CAN, were invited by email to participate in the first survey (n = 130). Members were invited to participate in subsequent rounds if they had completed the preceding survey.

In each survey, the disease elements were represented using a variety of methods, descriptive text, clinical photographs, diagnostic images, and graphs, to accurately portray the element in question. Respondents were asked to select and rank their preferred labels for each element. The label options provided were derived from the content analysis of the literature and represented the most commonly identified labels. Labels were not included if they had been identified less than twice in the body of literature examined and no more than 10 options were given for each element. There were 2 exceptions to this format. The label "podagra" was unique in that there were essentially no other labels identified through the literature analysis that referred to involvement of a specific anatomic region. For this element, respondents were asked to vote whether "podagra" was an acceptable label to denote "an episode of acute inflammation of the first metatarsophalangeal joint" (with relevant images shown). Because "hyperuric(a)emia" was the only label used to describe "an elevated circulating level of the final enzymatic product generated by xanthine oxidase in purine metabolism," respondents were asked to vote on whether this word was an acceptable label for this element. The label options provided in the first Delphi survey are shown in Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23607/abstract.

Respondents could comment on labels they felt strongly for or against. In accordance with the principles of the Delphi process, the second and third rounds included group feedback consisting of previous results and thematic summaries of comments provided. The label options in the second and third rounds were progressively refined, depending on the achievement of consensus or other selection criteria. Consensus was defined as a situation in which at least 80% of respondents selected a label as their

and from Grünenthal (more than $10,000). Dr. Kim has received research grants from AstraZeneca, Roche, Pfizer, Bristol-Myers Squibb, and Merck. Dr. Harrold has stock options from Corrona, LLC, has received consulting fees from Pfizer, MedIQ, and Bristol-Myers Squibb (less than $10,000 each), and has received a research grant from Takeda. Dr. Tausche has received speaking and/or consulting fees from Berlin-Chemie, Menarini, Ipsen, SOBI, Novartis, and Ardea Biosciences/AstraZeneca (less than $10,000 each). Dr. Vazquez-Mellado has received speaking and/or consulting fees from Takeda (less than $10,000). Dr. Gutierrez has received speaking and/or consulting fees from Novartis, AbbVie, Union Chimique Belge, Esaote SpA, Janssen Pharmaceutica, Bristol-Myers Squibb, and Merck Sharp & Dohme (less than $10,000 each). Dr. Richette has received speaking fees from Ipsen, Menarini AZ, and Grünenthal (less than $10,000 each). Dr. Pascual has received speaking and/or consulting fees from Ipsen, Grünenthal, AstraZeneca and AMPEL Biosolutions LLC (less than $10,000 each). Dr. Robinson has received speaking fees from Menarini and speaking fees and research grants from AstraZeneca (less than $10,000 each). Dr. Singh has received research grants from Takeda and Savient, has received consulting fees from Savient, Takeda, Regeneron, Merz, Iroko, Bioiberica, Crealta/Horizon, Allergan, WebMD, and United Business Media (less than $10,000 each), has received research grants from Horizon, and is the editor and the Director of the UAB Cochrane Musculoskeletal Group Satellite Center on Network Meta-analysis. Dr. Jansen has received speaking and/or consulting fees from AbbVie, Ardea/AstraZeneca, Bristol-Myers Squibb, Grünenthal, Janssen, Menarini, Celgene, Eli Lilly, Novartis, Pfizer, Roche, Sandoz, and Union Chimique Belge (less than $10,000 each), and has received research grants from AbbVie and Ardea/AstraZeneca/Grünenthal. Dr. Saag has received consulting fees and/or research grants from Ironwood/AstraZeneca, Horizon, SOBI, and Takeda (less than $10,000 each). Dr. Solomon has received research grants from Pfizer, Amgen, Eli Lilly, AstraZeneca, Bristol-Myers Squibb, and Genentech. Dr. Keenan has received consulting fees from Horizon, Ironwood, and AstraZeneca (less than $10,000 each) and received research grants from SOBI. Dr. Scire has received consulting fees from AstraZeneca (less than $10,000). Dr. Dehlin has received consulting fees from Grünenthal (less than $10,000). Dr. Nuki has received consulting fees from Grünenthal, Menarini, Savient, and Ardea/AstraZeneca (less than $10,000 each). Dr. Schlesinger has received grants from AstraZeneca, and has received consulting fees from Celgene, AstraZeneca, Horizon, SOBI, and Bristol-Myers Squibb (less than $10,000 each). Dr. Janssen has received speaking fees from Grünenthal (less than $10,000). Dr. Stamp has received grants from Ardea Biosciences and has received speaking fees from Amgen (less than $10,000). Dr. Sivera has received speaking and/or consulting fees from Ardea, AstraZeneca, Grünenthal, Ipsen Pharma, Menarini, Novartis, Ipsen Pharma, and Mayoly-Spindler (less than $10,000 each), and has received research grants from Ardea, AstraZeneca, Grünenthal, Ipsen Pharma, Mayoly-Spindler, Menarini, Novartis, and SOBI. Dr. Jacobsson has received speaking and/or consulting fees from AbbVie, Novartis, and Eli Lilly (less than $10,000 each). Dr. Lioté has received consulting fees from Ardea, AstraZeneca, Grünenthal, Ipsen Pharma, Menarini, and Novartis (less than $10,000 each), and has received research grants from Ardea, AstraZeneca, Grünenthal, Ipsen Pharma, Mayoly-Spindler, Menarini, Novartis, and SOBI. Dr. Bardin has received research grants from AstraZeneca, Ipsen, and Menarini and has received speaking and/or consulting fees from Ampel BioSolutions, Astellas, AstraZeneca, Biomex, Grünenthal, Ipsen, Menarini, Novartis, Savient, and SOBI (less than $10,000 each). Dr. Choi has received consulting fees from Ardea/AstraZeneca and Takeda (less than $10,000 each). Dr. Hershfield receives royalties from Horizon as a co-inventor of pegloticase (Krystexxa). Dr. Dalbeth has received speaking and/or consulting fees or grants from Kowa, Horizon, Takeda, Menarini, Teijin, Amgen, Pfizer, Ardea Biosciences, AstraZeneca, and Cymabay (less than $10,000 each).

Address correspondence to Nicola Dalbeth, Bone and Joint Research Group, Department of Medicine, Faculty of Medical and Health Sciences, University of Auckland, 85 Park Road, Grafton, Auckland 1023, New Zealand. E-mail: n.dalbeth@auckland.ac.nz.

Submitted for publication December 10, 2017; accepted in revised form May 22, 2018.

first choice. If consensus agreement had not been achieved for a label following the first round, it would be included as an option in the second round if ranked within the top 3 by at least 20% of respondents. The 2 top-ranked labels from the second round were included in the third round in cases where consensus had not yet been reached.

Face-to-face consensus meeting and G-CAN endorsement. The face-to-face meeting was held on June 14, before the European League Against Rheumatism 2017 Congress in Madrid, Spain. Invitations to attend were extended to all G-CAN members. Participation in the Delphi process was not a prerequisite for attendance. The meeting was conducted as a facilitated discussion on a list of specific questions. First, the group discussed and voted on labels for elements that had not achieved consensus through the Delphi process. All attendees were given the opportunity to express their opinion on the remaining label options; however, the introduction of new labels was not permitted. Voting rounds, conducted by a show of hands, were held for each of these labels. Consensus was defined as at least 80% of participants in agreement for a proposed label. Attendees who abstained from voting were not included in calculating the percentage of those in agreement.

After consensus had been achieved for all labels, the definitions for each element were addressed using the same facilitated discussion approach. Prior to the meeting, draft definitions were circulated to attendees to provide a basis for these discussions. Attendees could express their views on each element with the aim of constructing an accurate, yet concise, definition for each label. Key concepts raised were documented and incorporated into sequentially modified definitions that were then put to a show-of-hands vote. This process continued until consensus agreement was achieved, defined as at least 80% of participants agreeing with the proposed definition. The results of the group consensus exercises and recommendations were submitted to the G-CAN Board of Directors, with endorsement following their review.

RESULTS

Results of the content analysis of the literature. Eleven basic disease elements of gout were identified; the text descriptions of these elements are shown in Table 1. Further details on the range and frequency of labels used to describe these elements have been described in detail elsewhere (2).

Results of the Delphi exercise and consensus. There were 51 G-CAN members who responded to the first survey (39% of all members). The respondents included 19 members from Europe (37%), 16 from North America (31%), 11 from the Asia-Pacific region (22%), and 5 from Latin America (10%). Respondents were predominantly rheumatologists (n = 46, 90%). Of the respondents who completed the first survey, 48 (94%) completed all 3 rounds.

Table 1. Text descriptions of the basic disease elements of gout identified in a content analysis of published literature

The pathogenic crystals in gout
The circulating form of the final enzymatic product generated by xanthine oxidase in purine metabolism in humans
An elevated circulating level of the final enzymatic product generated by xanthine oxidase in purine metabolism in humans
An episode of acute inflammation triggered by the presence of pathogenic crystals
The condition in which there is an absence of clinically evident inflammation after or between episodes of acute inflammation
Persistent inflammation induced by pathogenic crystals
A discrete collection of pathogenic crystals with associated host-response tissue
A discrete collection of pathogenic crystals with associated host-response tissue, detectable on physical examination
The presence of pathogenic crystal deposition on imaging
The presence of structural bone damage due to gout
An episode of acute inflammation of the 1st metatarsophalangeal joint

At the completion of the Delphi exercise, consensus had been achieved for labels to describe 8 of the 11 basic disease elements: monosodium urate crystals, urate, hyperuric(a)emia, intercritical gout, tophus, subcutaneous tophus, imaging evidence of monosodium urate crystal deposition, and podagra (Table 2).

Consensus was not achieved through the Delphi process for the labeling of 3 elements. The element described as "an episode of acute inflammation triggered by the presence of pathogenic crystals," narrowly failed to achieve consensus; 79% of respondents preferred the label "gout flare." For the element described as "persistent inflammation induced by pathogenic crystals," preferences were divided between 2 labels: chronic gout (69%) and chronic gouty arthritis (23%). Consensus was also not achieved for the element described as "the presence of structural bone damage due to gout"; 65% of participants preferred the label "bone erosion," while 33% favored "gouty erosion."

Results of the face-to-face meeting. The face-to-face consensus meeting was attended by 30 G-CAN members (23% of all members). Of these attendees, 24 (80%) had also participated in the Delphi exercise. The panel included 17 members from Europe (57%), 6 from the Asia-Pacific region (20%), 5 from North America (17%), and 2 from Latin America (7%). Most attendees were rheumatologists (n = 28, 93%). Some voting activities included only 29 attendees due to the late arrival of one member.

Table 2. Results of a Delphi exercise and face-to-face consensus meeting for agreement on the labels for the basic disease elements of gout*

	Delphi exercise		Face-to-face meeting		
Text description of element	Consensus (round)	Agreement, %	Consensus	Agreement, %	Agreed label
The pathogenic crystals in gout	Yes (1)	92	–	–	Monosodium urate crystals
The circulating form of the final enzymatic product generated by xanthine oxidase in purine metabolism in humans	Yes (3)	81	–	–	Urate
An elevated circulating level of the final enzymatic product generated by xanthine oxidase in purine metabolism in humans	Yes (1)	91	–	–	Hyperuric(a)emia
An episode of acute inflammation triggered by the presence of pathogenic crystals	No	–	Yes	83	Gout flare
The condition in which there is an absence of clinically evident inflammation after or between episodes of acute inflammation	Yes (3)	81	–	–	Intercritical gout
Persistent inflammation induced by pathogenic crystals [involving articular structures]†	No	–	Yes	83	Chronic gouty arthritis
A discrete collection of pathogenic crystals with associated host response tissue	Yes (1)	88	–	–	Tophus
A discrete collection of pathogenic crystals with associated host response tissue, detectable on physical examination	Yes (3)	83	–	–	Subcutaneous tophus
The presence of pathogenic crystal deposition on imaging	Yes (2)	86	–	–	Imaging evidence of MSU crystal deposition
The presence of structural bone damage due to gout	No	–	Yes	97	Gouty bone erosion
An episode of acute inflammation of the 1st metatarsophalangeal joint	Yes (1)	84	–	–	Podagra

* Consensus was defined as 80% or more agreement on the preferred label. MSU = monosodium urate.
† Text in brackets = modification of text description as presented in the face-to-face meeting.

Disease element labels. The face-to-face meeting was successful in achieving consensus agreement on the labeling of the 3 unresolved elements following the Delphi exercise. These consensus element labels were gout flare, chronic gouty arthritis, and gouty bone erosion (Table 2).

Consensus for the label "gout flare" was achieved without additional discussion, with 26 of 29 attendees (90%) voting in favor. Group discussion identified the fact that the text description "persistent inflammation induced by pathogenic crystals" used in the Delphi exercise was ambiguous; this ambiguity was thought to contribute to the failure to achieve consensus on the labeling of this element. Although gout-related chronic inflammation has multisystem effects, this disease element was intended to represent persistent inflammation in joint structures. Following this clarification, 24 of 29 (83%) voted in favor of "chronic gouty arthritis" as the preferred label. Furthermore, it was proposed that the label "chronic gout" should be not be used. The use of the word "chronic" in this context was felt to be redundant and potentially confusing when gout, by its very nature, is characterized by chronic crystal deposition. Consensus for the recommendation that "the label 'chronic gout' should be avoided" was achieved with 28 of 29 (97%) voting in favor of this recommendation.

When discussing the label for the element described as "the presence of structural bone damage due to gout," it was proposed that "bone erosion" and "gouty erosion" be combined to form "gouty bone erosion." The group considered that the use of "gouty" was desirable for disease specificity and to maintain consistency in this same approach used with other consensus labels. Inclusion of "bone erosion" was judged

to be in alignment with the accepted nomenclature of bone damage caused by other rheumatologic diseases. Consensus agreement for the label "gouty bone erosion" was achieved, with 28 of 29 (97%) voting in favor.

Disease element definitions. Consensus agreement was achieved for the definitions of all 11 basic disease elements of gout (Table 3). Further details on voting results are shown in Supplementary Table 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23607/abstract. Relevant issues arising from group discussions on the composition of these definitions are outlined here.

For "hyperuric(a)emia," it was considered beyond the scope of this nomenclature project to include a specific threshold of blood urate concentration in its definition. Apart from the complexities involving urate solubility, and the multifactorial relationship between circulating urate levels and the clinical presentation of gout, the point was made that a specific concentration may inadvertently promote a potentially inappropriate treatment target. Therefore, the final accepted definition was modified to read "an elevated blood urate concentration over the saturation threshold."

For "gout flare," it was agreed that the definition should include the qualifying description "a clinically evident episode" because it was deemed important to specify that such an episode should be detectable either by history or physical examination. When considering "intercritical gout," participants agreed to include the words "gout" and "crystal deposition" within the definition to reinforce the concept of gout as a chronic disease of crystal deposition, even in the absence of clinically evident acute inflammation. For the element labeled "chronic gouty arthritis," it was agreed the definition should refer specifically to "joint" involvement to maintain consistency with the use of "arthritis" within the label.

The definition for "imaging evidence of monosodium urate crystal deposition" was constructed to include the phrase "findings that are highly suggestive of monosodium urate crystals on an imaging test" as crystal deposits are not always directly visualized, depending on the imaging modality used. The definition of "gouty bone erosion" was worded to reflect the aforementioned changes in the element label. The definition explicitly refers to "a cortical break in bone" and the gout-specific radiologic findings of an "overhanging edge with sclerotic margins." Even with these characteristic findings, etiologic association is addressed as "suggestive of gout" in recognition that causality cannot be proven.

The group recognized that the consensus definition for "podagra" differs from the literal translation of this word (Greek: pod "foot," and agra "seizure/trap"). However, it was accepted by most participants that this term has evolved to refer specifically to the first metatarsophalangeal joint rather than the entire first toe or foot, or including the first interphalangeal joint.

DISCUSSION

We present the agreed labels and definitions for the basic disease elements of gout endorsed by G-CAN (Table 3). Our own research, consistent with that of others, has shown considerable shortfalls in the terminology used to represent gout in the scientific literature (1,2). The use of inconsistent, inaccurate, and misleading terms has the potential to obscure interpretation of research and clinical communication. A clear and agreed terminology of

Table 3. G-CAN endorsed labels and definitions of the basic disease elements of gout following a Delphi exercise and face-to-face consensus meeting*

Consensus label	Consensus definition
Monosodium urate crystals	The pathogenic crystals in gout (chemical formula $C_5H_4N_4NaO_3$)
Urate	The circulating form of the final enzymatic product generated by xanthine oxidase in purine metabolism in humans (chemical formula $C_5H_3N_4O_3^-$)
Hyperuric(a)emia	Elevated blood urate concentration over the saturation threshold
Gout flare	A clinically evident episode of acute inflammation induced by monosodium urate crystals
Intercritical gout	The asymptomatic period after or between gout flares, despite the persistence of monosodium urate crystals
Chronic gouty arthritis†	Persistent joint inflammation induced by monosodium urate crystals
Tophus	An ordered structure of monosodium urate crystals and the associated host tissue response
Subcutaneous tophus	A tophus that is detectable by physical examination
Imaging evidence of monosodium urate crystal deposition	Findings that are highly suggestive of monosodium urate crystals on an imaging test
Gouty bone erosion	Evidence of a cortical break in bone suggestive of gout (overhanging edge with sclerotic margin)
Podagra	A gout flare at the 1st metatarsophalangeal joint

* G-CAN = Gout, Hyperuricemia, and Crystal-Associated Disease Network.
† The G-CAN recommendation is that the label "chronic gout" should be avoided.

the basic disease elements, which promotes an improved conceptual understanding of the disease and its pathogenesis, could also have a positive impact on patient care. The objective of this project was to address these deficits by developing a complete nomenclature for these basic elements, based on consensus expert opinion.

One of the important outcomes of this project is the agreement reached on the labeling of the element "urate." The common use of "uric acid" for this element, often interchangeably with "urate," is a potential source of confusion to clinicians, researchers, and patients alike (2). The label "urate" is consistent with the biochemical characteristics of uric acid as a weak acid that is largely ionized at pH 7.4, thereby making it the predominant form in blood (5).

This consensus statement also makes important recommendations on the nomenclature of elements that refer to the clinical features of gout. A key consideration is the commonly used classification of gout as being either "acute" or "chronic." The label "acute gout" suggests that the deleterious effects of this disease are limited only to those periods of intense inflammatory response triggered by the presence of monosodium urate crystals. In the medical literature, the label "chronic gout" has been used by some authors to imply a distinct presentation of the disease, such as the presence of structural joint disease or extensive tophi (6,7). These 2 commonly used labels misrepresent the pathophysiology of gout, an inherently chronic disease, in which crystal deposition is present regardless of the inflammatory state. The labels recommended by G-CAN in this consensus statement reflect contemporary understanding of disease pathophysiology. Adoption of this nomenclature may help overcome the common misperception of gout as an episodic disease that requires only acute management (8).

The label "tophus" (Latin: stone) is inextricably linked to the clinical manifestations of monosodium urate crystal deposition (9,10). In this consensus statement, we recommend the label "tophus" when referring generally to monosodium urate deposition and its host tissue response. The label "subcutaneous tophus" should be used for tophi evident during physical examination; the use of "subcutaneous" is intended to add specificity while recognizing that tophi may in fact occur in other clinically evident anatomic locations (e.g., intradermal). When referring to crystal deposits detectable by imaging, the label "imaging evidence of monosodium urate crystal deposition" should be used. The label and definition for this element are applicable to any form of imaging; modality-specific nomenclature falls outside the scope of this project. For "gouty bone erosion," this label and its definition recognize the characteristic appearance of bone damage in gout.

Finally, this consensus statement supports the use of "podagra" to represent the element of a gout flare affecting the first metatarsophalangeal joint. This is a commonly used term that has been associated with gout for many centuries (9,10). Due to its historical significance and widespread acceptance, the use of this label has been deemed acceptable.

A limitation of this project is that it addresses only the basic disease elements of gout and not the name of the disease itself, nor does it deal with the nomenclature for disease staging, management, and outcome measures. However, establishing a nomenclature of the basic elements was prioritized due to the foundation it provides for the labeling of more complex concepts; these are intended to be addressed in future work by G-CAN. The response rate to the group consensus exercises was modest (39% for the Delphi, 23% for the face-to-face consensus meeting), and participants were predominantly male (73%). There was, however, international representation from experts in the field. A further potential limitation was the use of existing terminology identified by a content analysis of the medical literature, rather than devising novel terms. This decision was made to avoid adding to the already vast array of terms used to describe gout (2). Refinement of established terminology was considered the most effective way of expediting widespread acceptance among clinicians and nonphysician researchers.

In summary, this G-CAN consensus statement regarding the labels and definitions for basic disease elements in gout represents an important step in promoting standardized nomenclature for the fundamental elements of this disease. Consensus terminology for the basic disease elements is a fundamental requirement for further work to address more advanced concepts, including the disease states of gout. G-CAN recommends the use of these labels when describing the basic disease elements of gout in clinical settings and in the scientific literature, with the goal of improving the accuracy and consistency of the language used to describe gout.

ACKNOWLEDGMENT

The authors thank Pamela Love, G-CAN Executive Director and Board Secretary, for assisting in the dissemination of the Delphi surveys and organization of the face-to-face consensus meeting.

AUTHOR CONTRIBUTIONS

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

Study conception and design. Bursill, Taylor, Terkeltaub, Dalbeth.

Acquisition of data. Bursill, Taylor, Terkeltaub, Kuwabara, Merriman, Grainger, Pineda, Louthrenoo, Edwards, Andrés, Beatriz Vargas-Santos, Roddy, Pascart, Lin, Perez-Ruiz, Tedeschi, Kim, Harrold, McCarthy, Kumar, Chapman, Tausche, Vazquez-Mellado, Gutierrez, da Rocha Castelar-Pinheiro, Richette, Pascual, Fisher, Burgos-Vargas, Robinson, Singh, Jansen, Saag, Slot, Uhlig, Solomon, Keenan, Scire, Biernat-Kaluza, Dehlin, Nuki, Schlesinger, Janssen, Stamp, Sivera, Reginato, Jacobsson, Lioté, Ea, Rosenthal, Bardin, Choi, Hershfield, Czegley, Choi, Dalbeth.

Analysis and interpretation of data. Bursill, Taylor, Terkeltaub, Dalbeth.

ROLE OF THE STUDY SPONSOR

Ironwood Pharmaceuticals, Horizon Pharmaceuticals, SOBI, Takeda Pharmaceutical Company, Teijin, Selecta, and CymaBay 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 Ironwood Pharmaceuticals, Horizon Pharmaceuticals, SOBI, Takeda Pharmaceutical Company, Teijin, Selecta, or CymaBay.

REFERENCES

1. Edwards NL, Malouf R, Perez-Ruiz F, Richette P, Southam S, DiChiara M. Computational lexical analysis of the language commonly used to describe gout. Arthritis Care Res (Hoboken) 2016;68:763–8.
2. Bursill D, Taylor W, Terkeltaub R, Dalbeth, N. The nomenclature of the basic disease elements of gout: a content analysis of contemporary medical journals. Semin Arthritis Rheum 2018;48:456–61.
3. Simkin PA. Towards a coherent terminology of gout. Ann Rheum Dis 1993;52:693–4.
4. Jones J, Hunter D. Qualitative research: consensus methods for medical and health services research. BMJ 1995;311:376–80.
5. Hyndman D, Liu S, Miner JN. Urate handling in the human body. Curr Rheumatol Rep 2016;18:34.
6. Qaseem A, Harris RP, Forciea MA. Management of acute and recurrent gout: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2017;166:37–51.
7. Seth R, Kydd AS, Buchbinder R, Bombardier C, Edwards CJ. Allopurinol for chronic gout. Cochrane Database Syst Rev 2014;10:CD006077.
8. Spencer K, Carr A, Doherty M. Patient and provider barriers to effective management of gout in general practice: a qualitative study. Ann Rheum Dis 2012;71:1490–5.
9. MacKenzie C. Gout and hyperuricemia: an historical perspective. Curr Treatm Opt Rheumatol 2015;1:119–30.
10. Nuki G, Simkin PA. A concise history of gout and hyperuricemia and their treatment. Arthritis Res Ther 2006;8 Suppl 1:S1.

Arthritis Care & Research
Vol. 71, No. 3, March 2019, pp 435–447
DOI 10.1002/acr.23595

AMERICAN COLLEGE OF RHEUMATOLOGY
EDUCATION • TREATMENT • RESEARCH

Randomized Controlled Trial to Evaluate an Internet-Based Self-Management Program in Systemic Sclerosis

Dinesh Khanna,[1] Jennifer Serrano,[1] Veronica J. Berrocal,[1] Richard M. Silver,[2] Pedro Cuencas,[3] Sharon L. Newbill,[4] Josephine Battyany,[5] Cynthia Maxwell,[6] Mary Alore,[7] Laura Dyas,[8] Robert Riggs,[9] Kerri Connolly,[9] Saville Kellner,[10] Jody J. Fisher,[1] Erica Bush,[1] Anjali Sachdeva,[11] Luke Evnin,[12] Dennis W. Raisch,[13] and Janet L. Poole[13]

Objective. In a pilot study, our group showed that an internet-based self-management program improves self-efficacy in systemic sclerosis (SSc). The objective of the current study was to compare an internet-based self-management program to a patient-focused educational book developed to assess measures of self-efficacy and other patient-reported outcomes in patients with SSc.

Methods. We conducted a 16-week randomized, controlled trial.

Results. Of the 267 participants who completed baseline questionnaires and were randomized to the intervention (internet: www.selfmanagescleroderma.com) or control (book) group, 123 participants (93%) in the internet group and 124 participants (94%) in the control group completed the 16-week randomized controlled trial (RCT). The mean ± SD age of all participants was 53.7 ± 11.7 years, 91% were women, and 79.4% had some college or a higher degree. The mean ± SD disease duration after diagnosis of SSc was 8.97 ± 8.50 years. There were no statistical differences between the 2 groups for the primary outcome measure (Patient-Reported Outcomes Measurement Information System Self-Efficacy for Managing Symptoms: mean change of 0.35 in the internet group versus 0.94 in the control group; $P = 0.47$) and secondary outcome measures, except the EuroQol 5-domain instrument visual analog scale score ($P = 0.05$). Internet group participants agreed that the self-management modules were of importance to them, the information was presented clearly, and the website was easy to use and at an appropriate reading level.

Conclusion. Our RCT showed that the internet-based self-management website was not statistically superior to an educational patient-focused book in improving self-efficacy and other measures. The participants were enthusiastic about the content and presentation of the self-management website.

INTRODUCTION

Systemic sclerosis (SSc; scleroderma) is a rare autoimmune disease that universally affects the skin and is associated with aberrant vasculopathy and fibrosis of internal organs (1,2). Currently, there is no cure for SSc. In addition to having the highest mortality rate among the rheumatic diseases, SSc is characterized by disfigurement, hand contractures, fatigue, sleep disorders, low self-esteem, pain, and severe Raynaud's phenomenon, all of which are associated with significant functional and work disability and a decrement in quality of life. In addition, loss of productivity per person with the disease in the US is estimated to be $10,764 per year (3).

Clinicaltrials.gov identifier: NCT02494401.

The statements presented in this article are solely the responsibility of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute, its Board of Governors, or its Methodology Committee.

Supported by a Patient-Centered Outcomes Research Institute Award (CER-1310-08323). Dr. Khanna's work was supported by the NIH/National Institute of Arthritis and Musculoskeletal and Skin Diseases (K24-AR-063129).

[1]Dinesh Khanna, MD, MS, Jennifer Serrano, RN, Veronica J. Berrocal, PhD, Jody J. Fisher, BA, Erica Bush, BS: University of Michigan, Ann Arbor; [2]Richard M. Silver, MD: Medical University of South Carolina, Charleston; [3]Pedro Cuencas: Local Leverage Media, Dallas, Texas; [4]Sharon L. Newbill, PhD: Folkstone Evaluation Anthropology, Pace, Florida; [5]Josephine Battyany, MD: Scleroderma Foundation Southern California Chapter, Culver City, California; [6]Cynthia Maxwell, MBA: Johns Island, South Carolina; [7]Mary Alore, MBA: Shelby Township, Michigan; [8]Laura Dyas, MA, LSW, LPC: Scleroderma Foundation Michigan Chapter, Southfield; [9]Robert Riggs, Kerri Connolly, BS: National Scleroderma Foundation, Danvers, Massachusetts; [10]Saville Kellner: Lake Industries, Revenue Media Group, and JLS Financial Inc., Henderson, Nevada; [11]Anjali Sachdeva, MFA: University of Pittsburgh, Pittsburgh, Pennsylvania; [12]Luke Evnin, PhD: MPM Capital, Boston, Massachusetts, and Scleroderma Research Foundation, San Francisco, California; [13]Dennis W. Raisch, PhD, Janet L. Poole, PhD, OTR/L: University of New Mexico, Albuquerque.

Dr. Khanna has received consulting fees from Actelion, AstraZeneca, Bayer, Boehringer Ingelheim, Chemomab, Corbus, Covis, Cytori, Eicos (with ownership interest), EMD Serono, Genentech/Roche, Gilead, GlaxoSmithKline, Sanofi-Aventis, and UCB (less than $10,000 each), and has received grants from Bayer, Bristol-Myers Squibb, Boehringer Ingelheim, Genentech/Roche, Pfizer, and Sanofi-Aventis.

Address correspondence to Dinesh Khanna, MD, MS, Professor of Medicine, Scleroderma Program, Division of Rheumatology, Department of Internal Medicine, University of Michigan, 300 North Ingalls Building, Room 7C27, Ann Arbor, MI 48109-5422. E-mail: khannad@med.umich.edu.

Submitted for publication November 6, 2017; accepted in revised form May 1, 2018.

SIGNIFICANCE & INNOVATIONS

- Systemic sclerosis is a rare disease and many patients do not have access to educational programs.
- We performed a randomized, controlled trial comparing an internet-based self-management program to a patient-focused educational book in measures of self-efficacy and other patient-reported outcomes.
- The self-management website was not superior to a patient-focused educational book in improving self-efficacy and other measures.
- The participants were enthusiastic about the content and presentation of the self-management website and endorsed it for dissemination.

Because SSc is a rare disease, many patients with SSc do not have access to educational programs or support groups. To address the lack of educational programs, a self-management program consisting of a workbook and DVD was developed and then tested in a small sample of patients with SSc (4). Improvements in pain, depression, and fatigue, as well as positive feedback from the participants, led to the conversion of all the modules in the booklet and the DVD to an internet format. In a pilot study of the internet version of the self-management program, participants logged on to a website and proceeded through the modules and learning activities at their own pace over the course of 10 weeks (5). Participants were encouraged to log on to the discussion board, an interactive component of the website, and respond to discussion questions posted for each module. The pilot study showed significant and positive changes for self-efficacy, ability to manage care, health efficacy, fatigue, and depression (5).

Since the initial development of the self-management program, new therapies and recommendations for laboratory and diagnostic tests and pharmacologic treatments have emerged (1). Thus, the self-management program was revised and updated with input from patient partners and stakeholders (the Scleroderma Foundation and the Scleroderma Research Foundation) (6). In this article, we report our findings in a randomized controlled trial (RCT) conducted to evaluate the efficacy of the internet-based self-management program versus the patient book developed for patients with SSc for improving self-efficacy and other patient-reported outcome measures. We hypothesized that the internet-based self-management program was superior to the book in primary (self-efficacy) and secondary patient-reported outcome measures.

PATIENTS AND METHODS

Participants. Patients with SSc were recruited from the University of Michigan and the Medical University of South Carolina (identified by scleroderma clinics), and via websites and social media sources of the Scleroderma Foundation and the Scleroderma Research Foundation (self-identified SSc). Inclusion criteria were being residents of the US, having a diagnosis of SSc, being age ≥18 years, having basic computer literacy and access to a computer with internet and email capabilities, having communication skills in English, and being willing to complete the study protocol. This study was conducted in accordance with the Helsinki Declaration, and all participants provided informed consent. The study was approved by institutional review boards of the University of New Mexico, the University of Michigan, and the Medical University of South Carolina.

Outcome measures. Demographic information, including age, sex, type of scleroderma (diffuse, limited/sine, overlap disease) as reported by the participant, length of time since disease onset, self-rated health, education level, marital status, and ethnicity, was collected. Self-efficacy is the belief that one can carry out a behavior necessary to reach a desired goal, even when a situation contains unpredictable and stressful elements (7). Self-efficacy is a major determinant of behavior and behavioral change, and acts as a key mediator in attaining self-management skills in chronic diseases (8,9). To measure self-efficacy (10), we administered the PROMIS Self-Efficacy for Managing Chronic Conditions instrument, which comprises 5 domains: managing symptoms, daily activities, medications and treatments, emotions, and social interactions. Each domain consists of 8 items scored from 1 (not at all confident) to 5 (very confident), with higher scores indicative of greater self-efficacy. The scales were standardized to the US population so that the mean was 50 units and the SD was 10 units, and results were scored by uploading the data at http://www.healthmeasures.net/explore-measurement-systems/promis. We used the domain for managing symptoms as the primary outcome measure.

The PROMIS-29 Profile version 2.0 measure contains 29 items, 1 on pain intensity and 4 items in each of the following domains: physical function, anxiety, depression, fatigue, sleep disturbance, pain interference, and satisfaction with social roles (11). With the exception of physical function, which does not include a time frame, all item banks referenced the past 7 days. Items were scored from 1 (unable to do/never/not at all) to 5 (without any difficulty/always/very much). All scales, except the pain intensity item, were standardized to the US population so that the mean was 50 and SD was 10 units, and they were scored using the method described at http://www.healthmeasures.net/explore-measurement-systems/promis. The Patient Health Questionnaire 8 (PHQ-8) is an 8-item questionnaire that is commonly used to measure depressive symptoms (12). A score of ≥10 is consistent with depressed mood.

The Patient Activation Measure (PAM) is a 13-item measure that assesses patient knowledge, skill, and confidence for self-management (13). Each item is scored from 1 (strongly disagree) to 4 (strongly agree). Scores are then summed, yielding a total score that can range from 13.0 to 52.0. The summed score

is finally transformed into a 0–100 scale, with higher scores indicating more confidence and knowledge in patients managing their condition. PAM scores were categorized into 4 levels: level 1, the individual is disengaged and overwhelmed; level 2, the individual is aware but struggling; level 3, the individual is taking action; and level 4, the individual is maintaining behavior (https://www.insigniahealth.com/products/pam-survey). The PAM has been extensively used in different self-management courses (14).

The EuroQol 5-domain instrument (EQ-5D) and quality-adjusted life years (QALYs) provide a generic health-related quality of life assessment. The EQ-5D incorporates patient-reported outcomes across the domains of mobility, self-care, activity, pain, and anxiety. Using a conversion algorithm, patient responses are converted into a health utility measure, ranging from 0.0 (death) to 1.0 (full or optimal health). The Brief Satisfaction with Appearance Scale is a 6-item scale measuring body image concerns and social discomfort with body parts. It is scored from 0 to 36, with higher scores associated with greater dissatisfaction.

Participants in both groups completed questionnaires at baseline and post-intervention at 16 weeks. A program evaluation was performed by asking participants in the intervention group to complete a questionnaire to gauge the content and presentation of the modules and to provide other feedback to the investigators.

Sample size. Sample size calculation was based on an analysis of pre/post changes in the Chronic Disease Self-Efficacy Scale in our pilot internet study (5). Based on data from the pilot study, we expected that the effect size in the intervention group would be approximately 0.50 (medium effect size as suggested by Cohen) (15), and we anticipated a negligible effect size in the control group (effect size = 0.10). Using a significance level of 0.05, we estimated that recruiting 100 participants in each group would yield an 80% power for detecting this difference between the intervention group and the control group. Assuming a conservative attrition rate of 25% during the study, we planned to enroll 125 patients in each group.

Randomization. Participants who met the inclusion criteria were sent instructions to review an electronic consent form through a Qualtrics platform. Once signed consents were obtained, participants were invited to complete the baseline questionnaire. Participants who completed the consent form were randomized to either an intervention or control group. Randomization was performed using a 1:1 ratio and via computer-generated block randomization, with stratification based on the PHQ-8 score (<10 or ≥ 10) to ensure that subjects with depressive symptoms were equally distributed in the 2 groups. Stratification based on the PHQ-8 score was used because we hypothesized that participants who reported being depressed may have poor coping and self-management skills. Although the assignment to either group was random, to ensure that the proportion of patients with more or fewer depressive symptoms was approximately the same in both groups, after every 50 patients were recruited, the assignment of patients to each group up to that point was cross-tabulated with respect to PHQ-8 scores. In addition, block randomization of patients occurred in groups of 50. This process allowed us to divide the intervention groups into 5 waves of 25 participants, so that the discussion board groups were small enough to encourage participation.

Intervention. Patients randomized to the internet program received a link to the self-management website, as well as a password and user name. The site could be accessed only via a secured website, with 1 module focus made available per week. The 15 modules included a basic overview, coping and body image, exercise, self-advocacy, pain management, activities of daily living, fatigue and energy conservation, tips for families and caregivers, muscle and lung disease with a focus on African Americans, gastrointestinal tract, Raynaud's disease, sexuality and scleroderma, mouth and teeth care, clinical trials, and emergencies. Two investigators (SLN and JLP) posted weekly questions regarding the modules on the discussion board and moderated the online discussion as necessary (see Supplementary Appendix A, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23595/abstract). Participants were asked to log on to the discussion board at least once weekly.

Those allocated to the control group received a copy of The Scleroderma Book: A Guide for Patients and Families, by Dr. Maureen Mayes. This book is the authoritative, educational book most requested and used by patients with a diagnosis of scleroderma. To date, it is the only credible resource written for patients and includes sections on early diagnosis, symptoms, coping with the disease, and resources for patients. Participants randomized to the control group were sent the textbook and were given 16 weeks to read it. A variety of strategies were used to maintain participant engagement in both groups during the intervention, including phone calls or email contact at 4, 8, and 12 weeks, and an incentive of $150 in the form of gift cards during the course of the study.

Statistical analysis. Summary statistics of the baseline demographic variables were computed for all the patients enrolled in the study. For each of these variables, summary statistics were calculated for the group of patients as a whole and stratified by treatment group (intervention versus standard care). Group differences for these characteristics were tested using either *t*-tests, Wilcoxon's rank-sum test for dependent samples, a proportion test, or chi-square tests, depending on the type of data (continuous versus categorical, or normally distributed versus not normally distributed).

To compare group differences between the intervention and control groups post-intervention, we considered only subjects with both baseline and follow-up data available. For those subjects, we computed the change in the scores from baseline to follow-up for continuous variables. For categorical variables, such as the PAM levels, we generated contingency tables presenting the joint distribution of the categorical classes at both baseline and follow-up (e.g., what percentage of patients were

categorized as having PAM level 1 at baseline and PAM level 1 at follow-up, and so forth). For both continuous and categorical variables, we tested whether there was a significant difference between the 2 groups either in the change in the scores or in the joint distributions of the categorical variables. Specifically, for continuous variables, we assessed whether there was a significant difference in the change in the scores in the control and internet groups by performing either *t*-tests, if the change in score appeared to be continuous and normally distributed, or by using Wilcoxon's tests if a normal distribution did not seem appropriate. For categorical variables, we assessed whether there was a significant difference between the two groups in the joint distribution of the categorical variable at baseline and follow-up in the 2 groups using Fisher's exact test due to small counts in some of the contingency table cells. For each test, we used a significance level of 0.05, with no adjustment for multiple testing.

RESULTS

A total of 267 subjects agreed to participate in the study and were randomized to either the internet or control groups. Of these 267 participants who completed baseline questionnaires and were randomized to the intervention (internet) or control (book) group (Figure 1), 123 participants (93%) in the internet and 124 participants (94%) in the control groups completed the 16-week RCT. The 2 groups were similar at baseline with respect to the demographic variables (Table 1). Overall, the mean ± SD age was 53.7 ± 11.7 years, 91% were women, 82.8% were white, and 79.4% had some college or higher degree. The mean ± SD disease duration after diagnosis of SSc was 8.97 ± 8.50 years, with 44.9% classifying themselves as having limited/sine and 43.1% as diffuse SSc.

Regardless of the group, participants had similar mean scores on patient-reported outcome measures (PROs), except

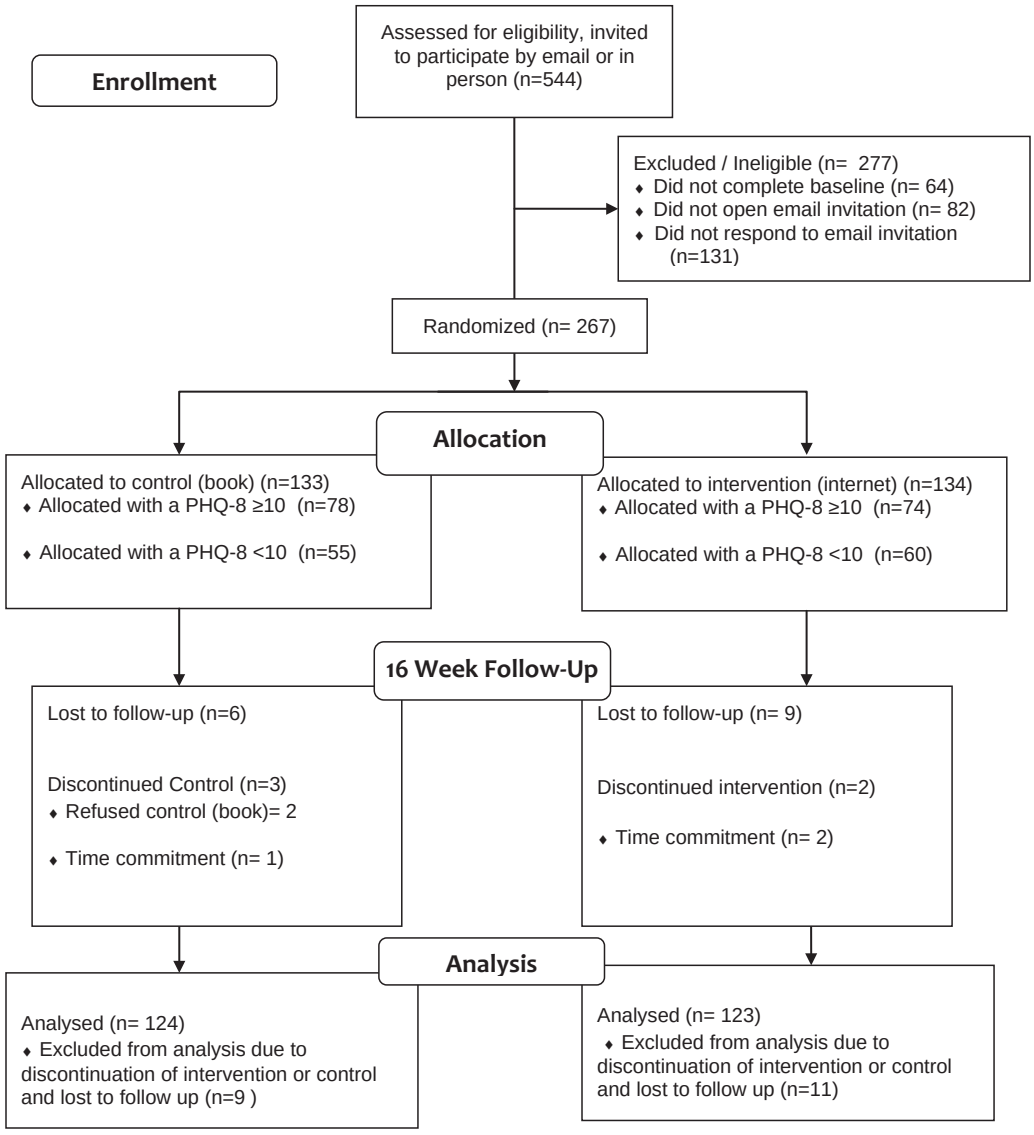


Figure 1. Flow diagram for participants in the trial, using the Consolidated Standards of Reporting Trials. PHQ-8 = Patient Health Questionnaire.

Table 1. Baseline characteristics of the 267 participants in the randomized clinical trial*

Characteristic	Values	Intervention (n = 134)	Control (n = 133)	*P*
Age, mean ± SD years	53.7 ± 11.7	54.3 ± 10.1	52.9 ± 13.1	0.33
Women	91 (243)	91.8 (123)	90.2 (120)	0.82
Race				
White	82.8 (221)	83.6 (112)	82.0 (109)	0.85
African American	7.5 (20)	5.2 (7)	9.8 (13)	0.24
Asian/Asian American	1.5 (4)	0.7 (1)	2.3 (3)	0.61
Native Hawaiian/other Pacific Islander	0.7 (2)	1.5 (2)	0	0.48
Other	1.5 (4)	0.7 (1)	2.25 (3)	1
Multiracial	6 (16)	8.2 (11)	3.8 (5)	0.2
Ethnicity				
Hispanic	4.1 (11)	5.2 (7)	3.0 (4)	0.55
Non-Hispanic	77.5 (207)	78.4 (105)	76.7 (102)	0.86
Other	15 (40)	14.9 (20)	15.0 (20)	1
Unknown	3.4 (9)	1.5 (2)	5.3 (7)	0.17
Education (years)				
High school (9–12)	20.6 (55)	20.1 (27)	21.1 (28)	0.98
College/university (13–16)	48.3 (129)	49.3 (66)	47.4 (63)	0.85
Graduate school (17–22)	27 (72)	26.9 (36)	27.1 (36)	1
Postgraduate school (≥23)	4.1 (11)	3.7 (5)	4.5 (6)	0.99
Marital status				
Single, never married	11.6 (31)	7.5 (10)	15.8 (21)	0.05
Married	63.7 (170)	70.9 (95)	56.4 (75)	0.02
Widowed	3.4 (9)	1.5 (2)	5.3 (7)	0.17
Divorced/separated	21.3 (57)	20.1 (27)	22.6 (30)	0.74
Employment status				
Working full time (≥20 hours/week)	35.6 (95)	35.8 (48)	35.3 (47)	1
Working part time (<20 hours/week)	6.7 (18)	7.5 (10)	6.0 (8)	0.82
On disability or sick leave	26.2 (70)	23.9 (32)	28.6 (38)	0.46
Retired	22.1 (59)	23.9 (32)	20.3 (27)	0.58
Not working but looking for work	2.3 (6)	1.5 (2)	3.0 (4)	0.67
Other	7.1 (19)	7.5 (10)	6.8 (9)	1
Self-defined scleroderma subtype				
Limited/sine	44.9 (120)	42.5 (57)	47.4 (63)	0.5
Diffuse	43.1 (115)	42.5 (57)	43.6 (58)	0.96
Overlap	11.6 (31)	14.1 (19)	9.0 (12)	0.26
Unknown	0.4 (1)	0.7 (1)	0	1
Patient-reported disease duration, mean ± SD years				
After first diagnosis from doctor	8.97 ± 8.50	8.72 ± 7.81	9.23 ± 9.17	0.63
After first scleroderma symptoms	11.91 ± 10.10	12.20 ± 9.33	11.62 ± 10.84	0.64
Overall health				
Excellent	1.1 (3)	0.7 (1)	1.5 (2)	1
Very Good	12.4 (33)	12.7 (17)	12.0 (16)	1
Good	42.7 (114)	44.8 (60)	40.6 (54)	0.57

(Continues)

Table 1. *(Cont'd)*

Characteristic	Values	Intervention (n = 134)	Control (n = 133)	*P*
Fair	37.4 (100)	34.3 (46)	40.6 (54)	0.35
Poor	6.4 (17)	7.5 (10)	5.3 (7)	0.63
US geographic region				
Midwest	50.2 (134)	54.5 (73)	45.9 (61)	0.2
Northeast	8.6 (23)	5.2 (7)	12.0 (16)	0.08
South	20.6 (55)	20.9 (28)	20.3 (27)	1
West	20.6 (55)	19.4 (26)	21.8 (29)	0.74

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

for the EQ-5D visual analog scale, which showed statistically higher scores in the internet group (Table 2). For the PROMIS self-efficacy and PROMIS-29 measures, the scores ranged from being similar between groups (PROMIS Self-Efficacy for Managing Medications and Treatment) to being 1.00 SD below the mean score in the US population (PROMIS-29 physical function scale). The mean ± SD PHQ-8 score was 8.67 ± 5.18, and 43.1% participants had depressed mood. Regarding the PAM scores, 18.7% and 59.6% of participants had PAM level 3 and PAM level 4, respectively.

Table 2 shows the mean change scores for the 2 groups between baseline and post-intervention at 16 weeks for all variables. There were no statistically significant differences between the 2 groups for the primary outcome measure (PROMIS Self-Efficacy for Managing Symptoms: mean change of 0.35 in the internet group versus 0.94 in the control group; $P = 0.47$) and other PROs, except for a significant difference between the internet and control groups for changes in the way the EQ-5D index changed from baseline to follow-up.

Because we recruited a group of participants who had a high level of patient activation (approximately 60% had PAM level 4), and long disease duration, we assessed the participants with early disease (<2 years and <5 years), PHQ-8 score <10, PHQ-8 score ≥10, and PAM levels 1 and 2 (Table 3). Again, there were no differences between the 2 groups, except for PROMIS Self-Efficacy for Managing Symptoms favoring the control group in early disease duration ($P = 0.03$) (Table 4), and EQ-5D self-care favoring the control group for those with PHQ-8 scores ≥10 ($P = 0.02$).

Discussion board evaluation. Of the 134 participants randomized to the internet group, 81 (61.4%) visited the discussion board, with 79 (59.8%) posting at least 1 comment over the 16-week RCT. An average of 8 comments were posted per user, with an average of 58.21 minutes reviewing each module. At the end of the 16-week RCT, 100 participants (74.6%) completed a course evaluation, in which they were asked to rate each module as helpful, slightly helpful, not helpful at all, or they did not review the module (see Supplementary Figure 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23595/abstract). An average of 75.4% of participants rated the modules as being helpful. Key modules (those that had more than 60 hours of time spent) included Scleroderma: A Basic Overview, Coping and Body Image/Appearance, Exercise, Self-Advocacy, and Dysphagia and the Digestive Tract. The course evaluation showed that 67.9% of participants agreed that the discussion board addressed important issues about scleroderma, with 44.5% agreeing the discussion board increased their understanding of scleroderma, and 63.0% agreeing the discussion board was a good way to learn from patients with scleroderma. When asked about their impression of the self-management course, an overwhelming 93.0% of participants agreed that the modules were of importance to them; 94.0% agreed that the information was presented clearly, with the website being easy to use, and at an appropriate reading level (Figure 2). We also provided access to the internet site for the participants who were randomized to the control group. In summary, 49 participants responded to the survey and 91.84% agreed that the information was presented clearly, and 93.75% agreed that the website was easy to use.

DISCUSSION

Using input from US Scleroderma Foundations and patient partners, we refined a previously developed internet program and tested it in the current RCT. Although we could not show any difference in the primary and secondary outcome measures compared to using the book, participants from the intervention showed overwhelming support and enthusiasm for the content and presentation on the website (see Supplementary Figure 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23595/abstract).

Based on input from the patient and stakeholder partners, we stratified the randomization with respect to PHQ-8 scores as <10 versus ≥10, because we hypothesized that participants who have depressed mood may exhibit poor coping skills. Although participants with PHQ-8 scores of ≥10 had lower scores on self-efficacy and PROMIS-29 scores (Table 3), there was no benefit in the inter-

Table 2. Mean patient-reported outcome measures at baseline, 16 weeks, and changes score over 16 weeks*

Scales	Baseline			16 weeks			Change		
	Internet (n = 134)	Control (n = 133)	*P*	Internet (n = 123)	Control (n = 124)	*P*	Internet (n = 123)	Control (n = 124)	*P*
PROMIS self-efficacy									
Managing emotions	46.78 ± 9.29†	46.34 ± 8.75	0.69	46.6 ± 9.20	47.20 ± 9.33	0.62	0.09 ± 6.28‡	0.73 ± 5.34	0.29
Managing symptoms	47.41 ± 9.15	47.58 ± 7.81	0.87	47.53 ± 8.50	48.61 ± 8.70	0.32	0.35 ± 6.12	0.94 ± 6.79	0.47
Managing daily activities	44.83 ± 7.60	44.71 ± 6.77	0.89	45.18 ± 7.84	45.37 ± 7.70	0.85	0.17 ± 5.11	0.79 ± 4.52	0.32
Managing social interactions	46.57 ± 9.74	46.66 ± 8.96	0.93	46.91 ± 9.63	47.43 ± 9.50	0.67	0.43 ± 6.94	0.99 ± 7.42	0.54
Managing medications and treatment	49.15 ± 9.06	50.3 ± 8.29	0.28	49.0 ± 8.99	50.85 ± 9.28	0.05	−0.54 ± 7.59	0.70± 6.29	0.16
PROMIS-29									
Physical function	40.63 ± 6.95	40.17 ± 6.15	0.59	40.81 ± 8.14	40.86 ± 7.30	0.96	0.05 ± 5.71	0.75 ± 4.83	0.25
Social role	43.74 ± 8.63	44.39 ± 7.68	0.51	45.82 ± 9.46	46.68 ± 8.86	0.46	2.05 ± 17.19	2.24 ± =15.58	0.93
Anxiety	53.85 ± 9.72	54.11 ± 10.31	0.83	54.14 ± 10.25	53.06 ± 10.11	0.41	0.16 ± 6.93	−0.86 ± 7.64	0.27
Depression	51.04 ± 10.03	51.59 ± 9.69	0.64	51.29 ± 9.53	51.22 ± 9.70§	0.96	0.08 ± 6.40	−0.13 ± 6.76	0.8
Fatigue	58.47 ± 10.52	58.91 ± 10.31	0.73	58.06 ± 10.59	59.24 ± 10.96§	0.67	−0.005 ± 6.90	0.20 ± 7.19	0.82
Pain interference	57.99 ± 9.37	57.92 ± 9.22	0.95	57.09 ± 9.20	57.37 ± 9.50	1	−0.65 ± 6.96	−0.80 ± 6.95	0.87
VAS pain intensity	4.08 ± 2.21	4.25 ± 2.26	0.65	4.13 ± 2.29	4.14 ± 2.26	0.98	0.11 ± 1.78	−0.19 ± 1.78	0.19
Sleep disturbance	55.22 ± 6.96	56.18 ± 7.84	0.29	52.08 ± 7.28	52.80 ± 7.91	0.46	−2.95 ± 6.22	−3.40 ± 6.62	0.58
PHQ-8	8.61 ± 5.39	8.72 ± 4.98	0.87	7.44 ± 5.56	7.40 ± 5.65	0.96	−1.19 ± 5.02	−1.27 ± 4.99	0.89
EQ-5D									
Mobility	1.53 ± 0.52	1.61 ± 0.49	0.18	1.54 ± 0.50	1.63 ± 0.49	0.14	0.0 ± 0.48	0.01 ± 0.47	0.89
Self-care	1.28 ± 0.48	1.35 ± 0.49	0.19	1.33 ± 0.51	1.32 ± 0.49	0.93	0.04 ± 0.41	−0.02 ± 0.37	0.13
Usual activities	1.78 ± 0.50	1.79 ± 0.43	0.82	1.78 ± 0.54	1.77 ± 0.46	0.94	0.01 ± 0.47	−0.02 ± 0.38	0.76
Pain discomfort	1.93 ± 0.48	2.02 ± 0.43	0.11	1.93 ± 0.43	1.96 ± 0.45	0.57	0.03 ± 0.44	−0.06 ± 0.44	0.08
Anxiety	1.56 ± 0.61	1.63 ± 0.62	0.34	1.57 ± 0.57	1.57 ± 0.63	0.87	0.02 ± 0.47	−0.05 ± 0.51	0.19
EQ-5D VAS	67.47 ± 18.01	63.72 ± 17.33	0.05	68.28 ± 18.61	64.92 ± 19.13	0.14	0.37 ± 16.39	1.40 ± 16.57	0.62
EQ-5D index	0.71 ± 0.18	0.69 ± 0.16	0.08	0.72 ± 0.17	0.71 ± 0.17	0.69	−0.002 ± 0.14	0.02 ± 0.14	0.05
SWAP	17.1 ± 9.53	16.81 ± 8.13	0.96	16.47 ± 9.47	16.76 ± 9.08	0.81	−0.82 ± 10.58	−0.31 ± 9.56	0.69
PAM level, no. (%)									
Level 1	14 (10.45)	14 (10.53)	1	10 (8.13)	12 (9.68)	0.82	NA	NA	
Level 2	15 (11.19)	15 (11.28)	1	13 (10.57)	13 (10.48)	1	NA	NA	
Level 3	23 (17.16)	27 (20.30)	0.53	26 (21.14)	22 (17.74)	0.52	NA	NA	
Level 4	82 (61.19)	77 (57.89)	0.62	74 (60.16)	77 (62.10)	0.79	NA	NA	

* Values are the mean ± SD unless indicated otherwise. PROMIS = Patient-Reported Outcomes Measurement Information System; VAS = visual analog scale; PHQ-8 = Patient Health Questionnaire 8; EQ-5D = EuroQol 5-domain instrument; SWAP = Brief Satisfaction with Appearance Scale; PAM = Patient Activation Measure; NA = not applicable.
† N = 133.
‡ N = 122.
§ N = 123.

Table 3. Mean patient-reported outcome measures disease duration, PHQ-8, and PAM levels at baseline*

	Disease <2 years			Disease <5 years			PHQ-8 <10			PHQ-8 ≥10			PAM level 1–2		
Scales	Treatment baseline (n = 8)	Control baseline (n = 12)	*P*	Treatment baseline (n = 42)	Control baseline (n = 40)	*P*	Treatment baseline (n = 73)	Control baseline (n = 79)	*P*	Treatment baseline (n = 55)	Control baseline (n = 50)	*P*	Treatment baseline (n = 26)	Control baseline (n = 28)	*P*
PROMIS self efficacy															
Managing emotions	46.80 ± 9.25	44.06 ± 10.24	0.54	49.26 ± 10.33†	46.14 ± 8.37	0.14	49.47 ± 8.29‡	47.59 ± 8.58	0.17	43.45 ± 9.76	44.33 ± 9.00	0.63	39.55 ± 7.31	40.53 ± 6.74	0.61
Managing symptoms	49.96 ± 10.72	47.51 ± 5.51	0.56	50.32 ± 9.25	47.15 ± 7.05	0.08	49.81 ± 8.44	48.04 ± 7.89	0.18	44.28 ± 9.05	46.86 ± 7.97	0.12	39.06 ± 6.09	42.25 ± 6.08	0.06
Managing daily activities	46.11 ± 7.84	43.16 ± 4.48	0.36	46.53 ± 7.77	44.52 ± 6.04	0.19	46.59 ± 7.93	45.87 ± 7.37	0.56	42.77 ± 6.40	42.79 ± 5.38	0.99	40.05 ± 5.74	42.22 ± 6.41	0.20
Managing social interactions	47.31 ± 7.28	47.18 ± 8.54	0.97	49.10 ± 9.90	47.01 ± 8.32	0.3	49.33 ± 9.38	46.61 ± 9.22	0.07	43.13 ± 9.30	45.93 ± 8.34	0.11	38.76 ± 7.26	39.99 ± 5.33	0.48
Managing medications and treatment	53.62 ± 5.08	50.58 ± 7.48	0.29	50.23 ± 9.45	49.53 ± 8.41	0.72	50.92 ± 8.64	50.94 ± 7.97	0.99	47.07 ± 9.30	48.88 ± 8.76	0.31	42.01 ± 6.52	45.29 ± 8.50	0.12
PROMIS-29															
Physical function	43.79 ± 8.95	38.92 ± 5.06	0.19	43.26 ± 8.05	39.98 ± 5.96	0.04	42.04 ± 8.17	41.28 ± 6.28	0.52	38.95 ± 6.57	38.4 ± 5.64	0.65	36.18 ± 5.38	37.0 ± 6.43	0.61
Social role	41.56 ± 9.64	46.78 ± 6.53	0.21	41.40 ± 8.00	44.44 ± 8.22	0.09	41.52 ± 7.80	43.19 ± 7.28	0.18	46.51 ± 8.88	46.46 ± 8.03	0.98	50.63 ± 8.70	47.84 ± 9.68	0.27
Anxiety	56.18 ± 9.54	58.38 ± 9.10	0.61	52.86 ± 10.2	53.98 ± 9.14	0.6	50.53 ± 8.16	51.39 ± 8.86	0.53	57.93 ± 10.23	58.52 ± 11.17	0.78	59.72 ± 9.07	61.68 ± 9.03	0.43
Depression	52.34 ± 10.14	54.95 ± 11.60	0.60	49.87 ± 9.36	50.76 ± 10.21	0.68	47.76 ± 7.79	48.8 ± 8.01	0.42	55.14 ± 11.21	55.98 ± 10.73	0.69	58.88 ± 10.31	58.23 ± 7.87	0.8
Fatigue	57.89 ± 12.33	61.91 ± 9.46	0.45	54.82 ± 10.53	60.33 ± 8.81	0.01	55.22 ± 10.27	56.77 ± 9.51	0.34	62.11 ± 9.67	62.04 ± 10.74	0.97	63.21 ± 7.79	64.31 ± 9.06	0.63
Pain interference	59.69 ± 12.28	59.55 ± 7.92	0.98	55.62 ± 10.06	57.47 ± 9.33	0.39	55.85 ± 8.59	56.51 ± 9.34	0.65	60.55 ± 9.72	60.40 ± 8.62	0.93	62.41 ± 8.61	62.45 ± 8.65	1.0
VAS pain intensity	4.63 ± 2.72	4.83 ± 2.29	0.86	4.05 ± 2.12	4.35 ± 1.89	0.5	3.67 ± 2.08	3.91 ± 2.23	0.49	3.78 ± 1.93	4.1 ± 2.28	0.44	4.27 ± 1.61	4.21 ± 1.99	0.91
Sleep disturbance	56.21 ± 6.49	57.04 ± 8.33	0.81	54.53 ± 7.47	56.02 ± 8.41	0.4	53.63 ± 7.12	55.08 ± 7.51	0.22	56.95 ± 6.20	57.59 ± 7.90	0.65	58.09 ± 6.25	59.89 ± 6.56	0.31
PHQ-8	11.62 ± 3.62	10.42 ± 5.26	0.55	9.26 ± 5.43	8.80 ± 5.40	0.7	4.49 ± 2.74	5.35 ± 2.55	0.05	13.76 ± 3.23	13.62 ± 3.33	0.82	11.54 ± 6.33	11.57 ± 5.55	0.98

(Continues)

Table 3. *(Cont'd)*

Scales	Disease <2 years			Disease <5 years			PHQ-8 <10			PHQ-8 ≥10			PAM level 1–2		
	Treatment baseline (n = 8)	Control baseline (n = 12)	*P*	Treatment baseline (n = 42)	Control baseline (n = 40)	*P*	Treatment baseline (n = 73)	Control baseline (n = 79)	*P*	Treatment baseline (n = 55)	Control baseline (n = 50)	*P*	Treatment baseline (n = 26)	Control baseline (n = 28)	*P*
EQ-5D															
Mobility	1.5 ± 0.53	1.42 ± 0.51	0.75	1.45 ± 0.50	1.45 ± 0.50	0.99	1.48 ± 0.50	1.54 ± 0.50	0.43	1.62 ± 0.53	1.72 ± 0.45	0.26	1.89 ± 0.33	1.75 ± 0.44	0.21
Self-care	1.38 ± 0.52	1.25 ± 0.45	0.59	1.31 ± 0.52	1.35 ± 0.48	0.59	1.22 ± 0.45	1.30 ± 0.49	0.23	1.36 ± 0.52	1.42 ± 0.50	0.49	1.65 ± 0.63	1.5 ± 0.51	0.41
Usual activities	1.75 ± 0.89	1.83 ± 0.39	0.63	1.76 ± 0.58	1.83 ± 0.45	0.5	1.75 ± 0.50	1.76 ± 0.43	0.86	1.82 ± 0.51	1.84 ± 0.42	0.76	2.0 ± 0.4	1.93 ± 0.38	0.51
Pain discomfort	1.88 ± 0.64	2.08 ± 0.51	0.44	1.79 ± 0.52	2 ± 0.39	0.04	1.88 ± 0.47	1.99 ± 0.38	0.10	1.96 ± 0.47	2.08 ± 0.49	0.22	2.08 ± 0.27	2.14 ± 0.45	0.49
Anxiety	1.75 ± 0.70	2 ± 0.60	0.40	1.52 ± 0.67	1.63 ± 0.59	0.33	1.43 ± 0.52	1.49 ± 0.55	0.45	1.71 ± 0.69	1.84 ± 0.68	0.31	1.89 ± 0.65	2.0 ± 0.61	0.50
EQ-5D VAS	62.0 ± 20.99	58.25 ± 21.83	0.91	70.88 ± 17.85	63.85 ± 18.67	0.09	71.6 ± 16.95	71.6 ± 15.63	0.10	62.89 ± 18.90	56.66 ± 18.06	0.09	57.31 ± 16.12	57.5 ± 18.88	0.97
EQ-5D index	0.68 ± 0.26	0.64 ± 0.18	0.51	0.74 ± 0.20	0.70 ± 0.14	0.09	0.75 ± 0.15	0.72 ± 0.13	0.06	0.67 ± 0.20	0.63 ± 0.19	0.30	0.59 ± 0.16	0.60 ± 0.19	0.83
PAM, mean scores															
Raw score	43.88 ± 4.85	40.92 ± 5.96	0.28	45.0 ± 5.22	41.58 ± 5.99	0.007	44.77 ± 5.16	43.72 ± 5.54	0.23	42.25 ± 6.75	42.50 ± 5.98	0.84	34.88 ± 3.15	34.82 ± 2.68	0.94
Activation score	69.89 ± 13.92	62.96 ± 17.14	0.28	73.95 ± 15.92	64.14 ± 16.57	0.008	73.21 ± 15.24	70.36 ± 15.70	0.26	66.88 ± 19.04	66.56 ± 16.70	0.93	46.07 ± 5.96	45.6 ± 4.91	0.76
PAM level, no. (%)															
1	0 (0)	2 (16.67)	0.49	2 (4.76)	8 (20)	0.05	3 (4.11)	7 (8.86)	0.33	10 (18.18)	7 (14.0)	0.61	12 (46.15)	14 (50.0)	0.79
2	2 (25.0)	2 (16.67)	1.0	4 (9.52)	5 (12.5)	0.73	8 (10.96)	6 (7.59)	0.58	6 (10.91)	8 (16.0)	0.57	14 (53.85)	14 (50.0)	0.79
3	1 (12.5)	4 (33.33)	0.6	6 (14.29)	10 (25.0)	0.27	11 (15.07)	18 (22.78)	0.30	10 (18.18)	8 (16.0)	0.8	0	0	1.0
4	5 (62.5)	4 (33.33)	0.36	30 (71.43)	17 (42.5)	0.01	51 (69.86)	8 (60.76)	0.31	29 (52.73)	27 (54.0)	1.0	0	0	1.0
SWAP	13.0 ± 10.20	18.17 ± 10.36	0.33	16.38 ± 9.92	17.42 ± 8.21	0.6	16.58 ± 8.80	16.59 ± 8.12	0.48	17.76 ± 10.44	17.56 ± 7.90	0.91	19.38 ± 11.54	17.43 ± 8.04	0.48

* Values are the mean ± SD unless indicated otherwise. PHQ-8 = 8-item Patient Health Questionnaire; PAM = Patient Activation Measure; PROMIS = Patient-Reported Outcomes Measurement Information System; VAS = visual analog scale; EQ-5D = EuroQol 5-domain instrument; SWAP = Brief Satisfaction with Appearance Scale.
† N = 41.
‡ N = 72.

Table 4. Mean patient-reported outcome measures disease duration, PHQ-8, and PAM levels at 16-week follow-up*

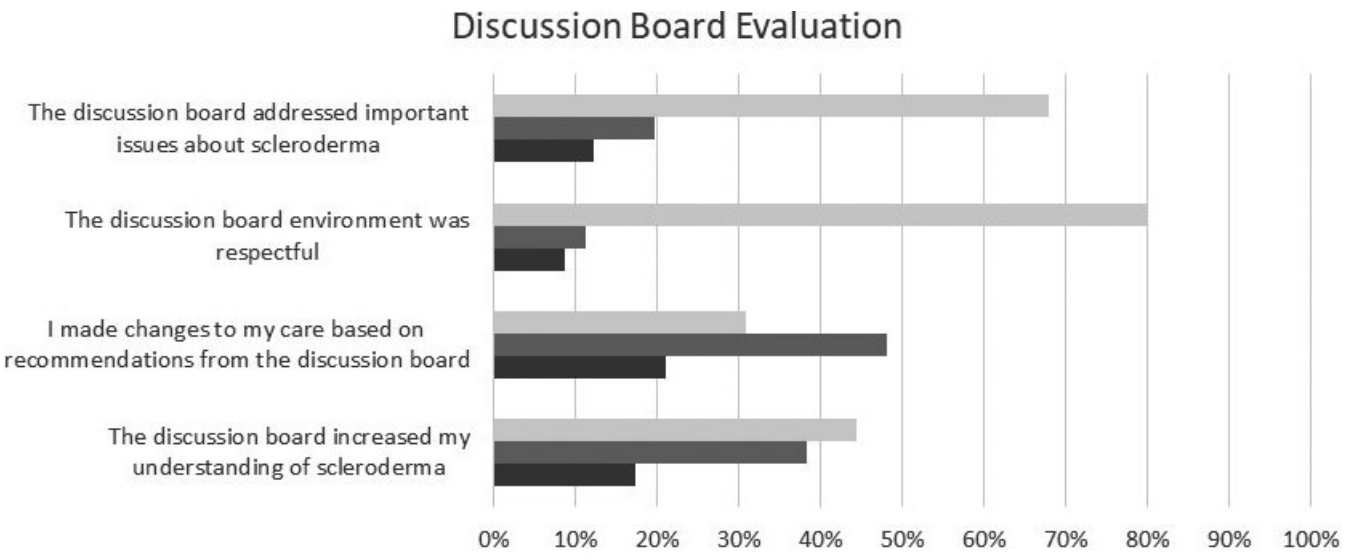
	Disease <2 years			Disease <5 years			PHQ-8 <10			PHQ-8 ≥10			PAM level 1–2		
Scales	Treatment change (n = 8)	Control change (n = 9)	*P*	Treatment change (n = 42)	Control change (n = 40)	*P*	Treatment change (n = 68)	Control change (n = 74)	*P*	Treatment change (n = 55)	Control change (n = 50)	*P*	Treatment change	Control change	*P*
PROMIS self-efficacy															
Managing emotions	−1.06 ± 7.04	0.11 ± 7.52	0.74	−1.51 ± 7.93†	0.93 ± 6.28	0.13	−0.24 ± 6.91‡	0.91 ± 6.06	0.30	0.10 ± 5.47	0.47 ± 5.80	0.76	1.37 ± 5.35	1.74 ± 5.27	0.80
Managing symptoms	−4.03 ± 6.92	0.38 ± 7.50	0.23	−1.28 ± 6.50	1.85 ± 5.98	0.03	0.15 ± 6.79	1.16 ± 6.93	0.39	0.59 ± 5.22	0.63 ± 6.65	0.97	2.79 ± 5.91	1.81 ± 7.45	0.59
Managing daily activities	−0.98 ± 2.62	1.96 ± 5.67	0.19	1.25 ± 6.27	0.35 ± 4.53	0.46	−0.01 ± 6.13	0.52 ± 4.63	0.57	0.4 ± 3.51	1.19 ± 4.37	0.31	0.48 ± 2.98	1.05 ± 4.04	0.55
Managing social interactions	−1.66 ± 11.19	0.97 ± 9.79	0.62	−0.20 ± 6.77	0.53 ± 6.94	0.63	−0.55 ± 7.53	0.76 ± 7.52	0.30	1.63 ± 5.98	1.33 ± 7.33	0.82	1.50 ± 8.65	1.85 ± 7.20	0.87
Managing medications and treatment	−2.15 ± 5.77	2.94 ± 7.42	0.13	0.54 ± 7.07	1.85 ± 6.90	0.40	−1.14 ± 6.84	0.86 ± 6.48	0.08	0.20 ± 8.44	0.47 ± 5.05	0.85	1.22 ± 6.91	−0.77 ± 5.53	0.25
PROMIS-29															
Physical function	−1.53 ± 3.89	−0.72 ± 8.73	0.81	0.56 ± 5.74	0.12 ± 5.64	0.73	0.04 ± 5.44	0.72 ± 4.72	0.43	0.05 ± 3.66	0.80 ± 5.05	0.39	0.33 ± 2.71	0.94 ± 4.82	0.57
Social role	4.95 ± 18.94	−0.74 ± 15.97	0.52	7.20 ± 16.12	2.66 ± 15.27	0.19	6.44 ± 15.91	5.13 ± 15.42	0.62	−3.37 ± 17.31	−2.03 ± 14.96	0.67	−10.38 ± 15.97	−4.25 ± 17.97	0.19
Anxiety	0.03 ± 8.85	−0.52 ± 9.77	0.91	0.24 ± 6.89	−2.23 ± 8.40	0.15	−0.35 ± 6.03	−0.46 ± 7.72	0.93	0.80 ± 7.91	−1.46 ± 7.58	0.14	0.44 ± 9.46	−3.0 ± 6.35	0.13
Depression	2.30 ± 7.69	−1.11 ± 9.85	0.44	1.17 ± 6.32	−0.60 ± 6.68	0.22	0.75 ± 6.09	0.77 ± 7.30	0.99	−0.74 ± 6.74	−1.45 ± 5.68	0.56	−1.42 ± 7.30	−2.44 ± 6.04	0.58
Fatigue	−3.06 ± 9.81	−1.13 ± 8.26	0.67	0.46 ± 7.33	−1.82 ± 5.90	0.12	0.18 ± 7.43	0.30 ± 6.61	0.92	−0.24 ± 6.25	0.06 ± 8.04	0.84	0.38 ± 6.74	−0.22 ± 5.17	0.72
Pain interference	−0.33 ± 6.07	−0.81 ± 4.56	0.86	−0.51 ± 7.52	−0.24 ± 6.38	0.86	−1.05 ± 7.34	−1.05 ± 7.04	1.0	−0.16 ± 6.48	−0.43 ± 6.86	0.84	−1.65 ± 6.34	0.08 ± 6.15	0.32
VAS pain intensity	−0.13 ± 1.13	0 ± 1.32	0.84	0.12 ± 1.77	−0.25 ± 1.79	0.35	0.12 ± 1.86	−0.16 ± 1.94	0.38	0.10 ± 1.69	−0.22 ± 1.53	0.30	−0.04 ± 1.43	0.07 ± 1.12	0.76
Sleep disturbance	−4.29 ± 7.23	−2.18 ± 7.05	0.55	−3.70 ± 6.47	−2.36 ± 5.78	0.32	−2.96 ± 6.28	3.48 ± 6.75	0.64	2.94 ± 6.20	−3.28 ± 6.60	0.78	−2.66 ± 5.74	−4.26 ± 7.46	0.38
PHQ-8	−3.63 ± 8.33	0.33 ± 8.22	0.34	−2.19 ± 5.77	−1.33 ± 5.01	0.47	0.85 ± 3.39	0.84 ± 3.98	0.98	−3.71 ± 5.57	−4.40 ± 4.72	0.49	−1.08 ± 4.94	−1.68 ± 4.74	0.65

(Continues)

Table 4. *(Cont'd)*

	Disease <2 years			Disease <5 years			PHQ-8 <10			PHQ-8 ≥10			PAM level 1–2		
Scales	Treatment change (n = 8)	Control change (n = 9)	*P*	Treatment change (n = 42)	Control change (n = 40)	*P*	Treatment change (n = 68)	Control change (n = 74)	*P*	Treatment change (n = 55)	Control change (n = 50)	*P*	Treatment change	Control change	*P*
EQ-5D															
Mobility	0 ± 0.53	0.22 ± 0.44	0.40	−0.05 ± 0.54	0.08 ± 0.42	0.26	−0.07 ± 0.43	0.03 ± 0.50	0.21	0.09 ± 0.52	−0.02 ± 0.43	0.23	−0.08 ± 0.39	−0.04 ± 0.33	0.68
Self-care	0 ± 0	0 ± 0	NA	−0.05 ± 0.44	−0.05 ± 0.45	0.79	0.01 ± 0.37	0.01 ± 0.39	0.99	0.07 ± 0.47	−0.08 ± 0.34	0.02	0 ± 0.49	−0.04 ± 0.43	0.52
Usual activities	0 ± 0.53	0.11 ± 0.33	0.66	0 ± 0.44	−0.05 ± =0.45	0.61	−0.03 ± 0.52	−0.04 ± 0.35	0.94	0.05 ± 0.40	0.02 ± 0.43	0.68	0.08 ± 0.48	0 ± 0.27	0.64
Pain discomfort	0.13 ± 0.35	0 ± 0.5	0.61	0.02 ± 0.47	0 ± 0.39	0.8	0.06 ± 0.45	−0.05 ± 0.43	0.13	0 ± 0.43	−0.08 ± 0.44	0.35	0.04 ± 0.34	−0.04 ± 0.51	0.54
Anxiety	−0.13 ± 0.35	−0.22 ± 0.44	0.66	0.05 ± 0.49	−0.08 ± =0.53	0.28	−0.01 ± 0.44	−0.05 ± 0.49	0.61	0.07 ± 0.50	−0.04 ± 0.53	0.18	0.04 ± 0.34	−0.14 ± 0.65	0.09
EQ-5D VAS	3.88 ± 12.57	−1.67 ± 15.26	0.74	−1.24 ± 15.28	−0.08 ± 15.54	0.73	0.59 ± 15.27	−1.11 ± 14.95	0.51	0.09 ± 17.82	5.10 ± 18.25	0.16	2.0 ± 16.76	0.29 ± 15.25	0.70
EQ-5D index	0.03 ± 0.13	0.02 ± 0.17	1.0	0.01 ± 0.15	0.009 ± 0.15	0.66	−0.003 ± 0.11	0.01 ± 0.13	0.22	−0.002 ± 0.16	0.03 ± 0.17	0.10	−0.007 ± 0.15	0.03 ± 0.21	0.35
PAM															
Mean raw score	0.75 ± 2.66	4.56 ± 7.55	0.19	31.13 ± 14.33	28.70 ± 13.52	0.43	1.25 ± 4.76	1.23 ± 5.37	0.98	25.07 ± 15.17	26.16 ± 15.61	0.72	23.76 ± 15.69	26.01 ± 17.41	0.61
Mean activation score	3.30 ± 7.86	11.81 ± 23.54	0.33	2.18 ± 12.09	6.13 ± 14.29	0.18	4.01 ± 14.65	3.64 ± 16.50	0.89	0.44 ± 15.18	2.10 ± 15.14	0.58	12.58 ± 15.21	15.24 ± 17.29	0.55
SWAP	1.0 ± 7.80	−6.0 ± 9.25	0.11	−2.24 ± 9.87	−2.50 ± 10.13	0.91	−1.22 ± 10.64	−0.66 ± 9.25	0.74	−0.33 ± 10.57	0.20 ± 10.08	0.79	0.73 ± 13.96	1.82 ± 8.81	0.74

* Values are the mean ± SD unless indicated otherwise. PHQ-8 = 8-item Patient Health Questionnaire; PAM = Patient Activation Measure; PROMIS = Patient-Reported Outcomes Measurement Information System; VAS = visual analog scale; EQ-5D = EuroQol 5-domain instrument; SWAP = Brief Satisfaction with Appearance Scale.
† N = 41.
‡ N = 67.



	The discussion board increased my understanding of scleroderma	I made changes to my care based on recommendations from the discussion board	The discussion board environment was respectful	The discussion board addressed important issues about scleroderma
Agree	44.45%	30.87%	80.00%	67.90%
Neither Agree nor Disagree	38.27%	48.15%	11.25%	19.75%
Disagree	17.28%	20.99%	8.75%	12.34%

Figure 2. Discussion board evaluation, showing an illustration of the responses received by internet participants after completion of the 16-week randomized, controlled trial.

net group compared to the control group. Our baseline data suggest that we recruited a group of highly motivated (approximately 60% of patients had PAM level 4), highly educated participants (80% had attended at least some college), who have been dealing with their disease for a long time (the mean time since diagnosis was 9 years). When we focused only on participants with early disease (<2 years and <5 years), PHQ-8 score ≥10, and PAM levels 1 and 2, we found no difference between the internet group versus the control group, although the sample sizes in these subgroups were very small and may be related to a Type II error.

Patients with chronic diseases such as SSc self-manage their illnesses on a daily basis. A central concept in self-management education is self-efficacy (16), which is a major determinant of behavior and behavioral change and acts as a key mediator of the attainment of self-management skills in patients with chronic diseases (8,9). Published work suggests that self-management skills are associated with improved clinical outcomes and reduce costs associated with arthritis (16). Because SSc is a rare disease (designated as an orphan disease by the Food and Drug Administration), Scleroderma Foundation Chapters and/or support groups do not exist in every state in the US. Many patients with SSc have not met anyone else with the disease (17,18). Patients living outside major metropolitan areas may not have access to health care providers with a specialized knowledge of SSc. Thus, SSc patients feel isolated from sources of support and education programs. The only educational programs specifically focused on scleroderma are offered via written materials, webinars, and annual conferences through the Scleroderma Foundation, and by state and/or local chapters of the Scleroderma Foundation and the Scleroderma Research Foundation. These offerings are credible sources of information, but patients may need to search through a website or wait for the next conference, meeting, or webinar.

Having an internet program that contains all the information and resources on self-management in 1 site and 1 format that can be quickly updated may be very useful to meet the needs of patients with scleroderma and their families and/or caregivers. Creators of the Arthritis Self-Management Program and Chronic Disease Self-Management Program developed internet versions of their successful programs, with outcomes similar to those achieved with the group format (19,20). The advantages of internet programs are that they are easily accessible, can be shared with family members, caregivers, and/or health professionals, and can be viewed as many times as needed for reinforcement or as symptoms change with disease progression. However, the existing self-management programs for patients with arthritis or other chronic illness do not address the specific needs of scleroderma patients related to body image changes, skin and wound management, gastrointestinal involvement, lung involvement, Raynaud's phe-

nomenon and ulcerations, and disability. This gap was exemplified by a recent study showing that information available on the internet is not meeting the health care needs of systemic scleroderma patients (21).

Our study also provides insight into the design of the next trial. First, it highlights the fact that the majority of the patients with SSc using the internet materials were well-educated, classified themselves as white, and were well-versed in management of their disease (approximately 80% had attended college or higher education, 83% were white, and 60% had PAM level 4). Future studies should focus on recruiting participants with lower PAM levels (likely to be nonwhite and less-educated participants) who have lower self-efficacy scores (10), and who would likely benefit from self-management courses. Second, patients with earlier disease may benefit, because published data suggest that patients' adjustment to a chronic disease improves with time (22).

Our RCT has many strengths. We recruited and retained >90% of participants over a period of 16 weeks. In addition, we collaborated with patient partners and stakeholders and recruited participants from both academic and nonacademic settings, providing generalizability for our results. Last, this is one of the largest studies evaluating a self-management or behavioral intervention in patients with SSc.

In conclusion, our RCT showed that the internet-based self-management website was not superior to the patient-focused textbook in improving self-efficacy and other measures. High patient activation scores and near-normal self-efficacy scores may have contributed to this result. However, participants were overwhelmingly enthusiastic, indicating a need for an internet program that is credible and easily accessible.

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. Khanna 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. Khanna, Serrano, Silver, Cuencas, Newbill, Battyany, Maxwell, Alore, Dyas, Riggs, Connolly, Kellner, Sachdeva, Evnin, Raisch, Poole.

Acquisition of data. Khanna, Battyany, Maxwell, Alore, Dyas, Connolly, Kellner, Poole.

Analysis and interpretation of data. Khanna, Berrocal, Fisher, Bush.

REFERENCES

1. Denton CP, Khanna D. Systemic sclerosis. Lancet 2017;390:1685–99.
2. Shreiner AB, Murray C, Denton C, Khanna D. Gastrointestinal manifestations of systemic sclerosis. J Scleroderma Relat Disord 2016;1:247–56.
3. Singh MK, Clements PJ, Furst DE, Maranian P, Khanna D. Work productivity in scleroderma: analysis from the University of California, Los Angeles scleroderma quality of life study. Arthritis Care Res (Hoboken) 2012;64:176–83.
4. Poole JL, Skipper B, Mendelson C. Evaluation of a mail-delivered, print-format, self-management program for persons with systemic sclerosis. Clin Rheumatol 2013;32:1393–8.
5. Poole JL, Mendelson C, Skipper B, Khanna D. Taking charge of systemic sclerosis: a pilot study to assess the effectiveness of an internet self-management program. Arthritis Care Res (Hoboken) 2014;66:778–82.
6. Newbill S, Khanna D, Serrano J, Battyany J, Rosson D, Maxwell C, et al. Use of focus groups and patient partners to revise an internet self-management program [abstract]. Arthritis Rheum 2015;67 Suppl 10.
7. Ashford S, Edmunds J, French DP. What is the best way to change self-efficacy to promote lifestyle and recreational physical activity? A systematic review with meta-analysis. Br J Health Psychol 2010;15:265–88.
8. Holman HR, Lorig K. Perceived self-efficacy in self-management of chronic disease, in self-efficacy: thought control of action. New York: Hemispheres Publications; 1992.
9. Lorig KR, Sobel DS, Ritter PL, Laurent D, Hobbs M. Effect of a self-management program on patients with chronic disease. Eff Clin Pract 2001;4:256–62.
10. Gruber-Baldini AL, Velozo C, Romero S, Shulman LM. Validation of the PROMIS(R) measures of self-efficacy for managing chronic conditions. Qual Life Res 2017;26:1915–24.
11. Kwakkenbos L, Thombs BD, Khanna D, Carrier ME, Baron M, Furst DE, et al. Performance of the Patient-Reported Outcomes Measurement Information System-29 in scleroderma: a scleroderma patient-centered intervention network cohort study. Rheumatology (Oxford) 2017;56:1302–11.
12. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009;114:163–73.
13. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. Health Serv Res 2005;40:1918–30.
14. Hibbard JH, Mahoney ER, Stock R, Tusler M. Do increases in patient activation result in improved self-management behaviors? Health Serv Res 2007;42:1443–63.
15. Cohen J. Statistical power analysis for the behavioral sciences. 1st ed. Cambridge (MA): Academic Press; 1977.
16. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. JAMA 2002; 288:2469–75.
17. Joachim G, Acorn S. Life with a rare chronic disease: the scleroderma experience. J Adv Nurs 2003;42:598–606.
18. Mendelson C, Poole JL. Become your own advocate: advice from women living with scleroderma. Disabil Rehabil 2007;29:1492–501.
19. Fries JF, Carey C, McShane DJ. Patient education in arthritis: randomized controlled trial of a mail-delivered program. J Rheumatolgy 1997;24:1378–83.
20. Lorig KR, Ritter PL, Laurent DD, Plant K. The internet-based arthritis self-management program: a one-year randomized trial for patients with arthritis or fibromyalgia. Arthritis Rheum 2008;59:1009–17.
21. Van der Vaart R, Repping-Wuts H, Drossaert CH, Taal E, Knaapen-Hans HK, van de Laar MA. Need for online information and support of patients with systemic sclerosis. Arthritis Care Res (Hoboken) 2013;65:594–600.
22. Raymakers AJ, Tsao NW, Marra CA, Clements PJ, Khanna D. Health state utilities and disease duration in systemic sclerosis: is there an association? J Rheumatol 2016;43:1832–7.

ARP Announcements

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

ACR/ARP Annual Meeting

November 8–13, 2019, Atlanta

Applications Invited for *Arthritis & Rheumatology* Editor-in-Chief (2020–2025 Term)

The American College of Rheumatology Committee on Journal Publications announces the search for the position of Editor, *Arthritis & Rheumatology*. The official term of the next *Arthritis & Rheumatology* editorship is July 1, 2020–June 30, 2025; however, some of the duties of the new Editor will begin during a transition period starting April 1, 2020. ACR members who are considering applying should submit a nonbinding letter of intent by May 1, 2019 to the Managing Editor, Jane Diamond, at jdiamond@rheumatology.org, and are also encouraged to contact the current Editor-in-Chief, Dr. Richard Bucala, to discuss details; initial contact should be made via e-mail to richard.bucala@yale.edu. Applications will be due by June 21, 2019 and will be reviewed during the summer of 2019. Application materials are available on the ACR web site at https://www.rheumatology.org/Learning-Center/Publications-Communications/Journals/A-R.

New Division Name

Rheumatology is truly a people specialty: We often develop lifelong relationships with our patients as well as our colleagues. We increasingly recognize that providing the best rheumatologic care requires a team effort. The collegial nature of our specialty is reflected in the ACR's mission statement: To empower rheumatology professionals to excel in their specialty.

In keeping with this mission, we are pleased to announce that our health professionals' membership division is changing its name to Association of Rheumatology Professionals (ARP). This name change highlights the dedication of the ACR to serve the entire rheumatology community. It also reflects our broadened base of interprofessional members (administrators, advanced practice nurses, health educators, nurses, occupational therapists, pharmacists, physical therapists, physician assistants, research teams, and more).

The name is new, but our commitment and promise remain the same: We are here for you, so you can be there for your patients.

ARP Membership

The Association of Rheumatology Professionals (ARP), a division of the American College of Rheumatology, appreciates your continued membership and looks forward to serving you another year. Membership costs range from $30 to $140. ARP welcomes nurse practitioners, nurses, physician assistants, office staff, researchers, physical therapists, occupational therapists, assistants, and students. Student membership is complimentary; the Annual Meeting registration fee is waived for students who submit the required student verification letter. For information, go to www.rheumatology.org and select "Membership" or call 404-633-3777 and ask for an ARP staff member.

New for 2019: Education for Rheumatology Professionals

Whether you are new to a rheumatology practice or just need a rheumatology refresher, kick off 2019 with high-quality education for the entire interprofessional team. All 19 Advanced Rheumatology Course activities have been updated with all-new interactive content, including mini-quizzes. You can also register for 11 brand new Advanced eBytes, which are complimentary to ARP members. For information on pricing, credits hours, and registration go to www.rheumatology.org, click the drop down box "I AM A" next to the Membership tab and select "Health Professional Education."