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. 12, December 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 • December 2019 • NO. 12

Special Articles

Rheumatoid Arthritis

Pediatrics

Osteoarthritis

Systemic Lupus Erythematosus

Cover image: The figure on the cover (from Cheung et al, page 1640) is magnetic imaging data from a healthy control showing the calculated transverse relaxation time map.

Arthritis Care & Research
Vol. 71, No. 12, December 2019, pp 1531–1539
DOI 10.1002/acr.24040

American College of Rheumatology
Empowering Rheumatology Professionals

SPECIAL ARTICLE

2019 American College of Rheumatology Recommended Patient-Reported Functional Status Assessment Measures in Rheumatoid Arthritis

Claire E. H. Barber,[1] JoAnn Zell,[2] Jinoos Yazdany,[3] Aileen M. Davis,[4] Laura Cappelli,[5] Linda Ehrlich-Jones,[6] Donna Everix,[7] J. Carter Thorne,[8] Victoria Bohm,[1] Lisa Suter,[9] Alex Limanni,[10] and Kaleb Michaud[11]

Objective. To develop American College of Rheumatology (ACR) recommendations for patient-reported Functional Status Assessment Measures (FSAMs) for use in routine clinical practice in patients with rheumatoid arthritis (RA).

Methods. We convened a workgroup to conduct a systematic review of published literature through March 16, 2017 and abstract FSAM properties. Based upon initial search results and clinical input, we focused on the following FSAMs appropriate for routine clinical use: the Health Assessment Questionnaire (HAQ) and derived measures and the Patient-Reported Outcomes Measurement Information System (PROMIS) tool. We used the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) 4-point scoring method to evaluate each FSAM, allowing for overall level of evidence assessment. We identified FSAMs fulfilling a predefined minimum standard and, through a modified Delphi process, selected preferred FSAMs for regular use in most clinic settings.

Results. The search identified 11,835 articles, of which 56 were included in the review. Descriptions of the measures, properties, study quality, level of evidence, and feasibility were abstracted and scored. Following a modified Delphi process, 7 measures fulfilled the minimum standard for regular use in most clinic settings, and 3 measures were recommended: the PROMIS physical function 10-item short form (PROMIS PF10a), the HAQ-II, and the Multidimensional HAQ.

Conclusion. This work establishes ACR recommendations for preferred RA FSAMs for regular use in most clinic settings. These results will inform clinical practice and can support future ACR quality measure development as well as highlight ongoing research needs.

INTRODUCTION

Functional status is an important outcome in rheumatology and relates to measures of functioning that capture the interaction between a person's health condition and their ability to participate in activities (1). Poor functional status is associated with work disability (2), poor quality of life (3), and is one of the strongest predictors of mortality in rheumatoid arthritis (RA) (2,4–7). Functional

Supported by the American College of Rheumatology. Dr. Barber's work was supported by the Canadian Institutes of Health Research and the University of Calgary Department of Medicine. Dr. Davis' work was supported by the CIHR. Dr. Cappelli's work was supported by the Jerome Greene Foundation and Bloomberg Philanthropies. Dr. Ehrlich-Jones' work was supported by the National Institute on Disability, Independent Living, and Rehabilitation Research, the Craig H. Neilsen Foundation, the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases grant 5R01-AR071091 and National Heart, Lung, and Blood Institute grant 1R01-HL132978-01A1), and the DOD. Dr. Thorne's work was supported by Mount Sinai and the University Health Network. Dr. Suter's work was supported by the VA and Yale New Haven Health System Center for Outcomes Research & Evaluation through a contract to the Centers for Medicare and Medicaid Services. Dr. Michaud's work was supported by the Rheumatology Research Foundation.

[1]Claire E. H. Barber, MD, PhD, FRCPC, Victoria Bohm, MSc, MPH: University of Calgary, Calgary, Alberta, Canada; [2]JoAnn Zell, MD: Denver Health, Denver, Colorado; [3]Jinoos Yazdany, MD, MPH: University of California, San Francisco; [4]Aileen M. Davis, PhD: Krembil Research Institute, University Health Network, and University of Toronto, Toronto, Ontario, Canada; [5]Laura Cappelli, MD: Johns Hopkins University, Baltimore, Maryland; [6]Linda Ehrlich-Jones, PhD, RN: Shirley Ryan AbilityLab, Chicago, Illinois; [7]Donna Everix, MPA, BS, PT: Mills Peninsula Health Services, Burlingame, California, and OnMyCare Home Health, Fremont, California; [8]Carter Thorne, MD, FRCPC: University of Toronto, Toronto, Ontario, Canada; [9]Lisa Suter, MD: Yale University, New Haven, Connecticut, and Veterans Affairs Medical Center, West Haven, Connecticut; [10]Alex Limanni, MD: Arthritis Centers of Texas, Dallas; [11]Kaleb Michaud, PhD: University of Nebraska Medical Center, Omaha, and FORWARD, the National Databank for Rheumatic Diseases, Wichita, Kansas.

Dr. Zell has received research support from Novartis. Dr. Yazdany has received consulting fees and/or speaking fees from Astra Zeneca and PRIME Education (less than $10,000 each) and research support from Pfizer and Astra Zeneca. Dr. Cappelli has received consulting fees from Regeneron/Sanofi (less than $10,000) and research support from Bristol-Myers Squibb. Dr. Ehrlich-Jones has received consulting fees from Zimmer Biomet (less than $10,000). Dr. Thorne has received consulting fees and/or speaking fees from AbbVie, Celgene, Novartis, Pfizer, Sandoz, Sanofi Genzyme, Serene, Amgen, and Medexus (less than $10,000 each). Dr. Limanni has received research support from GlaxoSmithKline, Janssen, Novartis, Gilead, and Pfizer. Dr. Michaud has received research support from Pfizer (ASPIRE award). No other disclosures relevant to this article were reported.

Address correspondence to Kaleb Michaud, PhD, Division of Rheumatology & Immunology, University of Nebraska Medical Center, 986270 Nebraska Medical Center, Omaha, NE 68198. E-mail: kmichaud@unmc.edu.

Submitted for publication April 17, 2019; accepted in revised form August 8, 2019.

status assessment measures (FSAMs) may be used in assessment of prognosis and aid in RA treatment decisions. Because of its importance, functional status assessment is included in guidelines for rheumatologic care for a number of conditions including RA (8). Assessment of functional status is captured by an American College of Rheumatology (ACR) RA quality measure (9) and is included in the Merit-Based Incentive Payment System, 1 of 2 payment tracks under the Quality Payment Program in the US emphasizing a value-based payment model (10).

In 2012 the ACR published recommendations on 6 RA disease activity measures (11). While no formalized document for ACR FSAM recommendations was developed, current ACR guidelines list collection of a standardized, validated FSAM as a key principle of RA treatment (8) and cite examples of commonly used FSAMs, including the Health Assessment Questionnaire (HAQ) disability index (DI), HAQ-II, Multidimensional HAQ (MDHAQ), and Patient-Reported Outcomes Measurement Information System (PROMIS) FSAMs, but do not make *specific* recommendations about their use in clinical practice. This work to provide initial recommendations on RA FSAMs was performed in parallel to an ACR workgroup updating the ACR's prior RA disease activity instrument recommendations.

The objectives of the RA FSAM workgroup were to provide RA patient-reported FSAMs meeting a minimum standard for regular use and preferred RA patient-reported FSAMs for regular use. These objectives reflect the fact that feasibility and clinical efficiency are important considerations in functional status assessment, supplementing minimum instrument performance standards.

METHODS

Study design. The ACR convened a workgroup of rheumatology professionals and rheumatologists to evaluate and recommend RA FSAMs. The workgroup developed a protocol and presented the process and preliminary findings at the 2017 ACR Annual Scientific Meeting in San Diego, California and obtained public comment following that presentation.

Search strategy. We conducted a systematic literature review, adhering to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis checklist (12). We searched Medline, Embase, Cochrane Library, and the Cumulative Index of Nursing and Allied Health databases, from study inception to March 16, 2017. We devised search terms according to a published search strategy for finding studies on measurement properties of patient-reported outcome instruments (13) from the Consensus-Based Standards for the selection of Health Measurement Instruments (COSMIN) group (URL: http://www.cosmin.nl/). This strategy uses MeSH terms and keywords across 3 themes: construct search (for assessment of functional status), population search (RA), and instrument search (including terms for instruments of interest, e.g., questionnaires, etc.). The Boolean search operator "AND" was used to combine the 3 search themes (see Supplementary Appendix A, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24040/abstract). We manually searched the reference lists of included articles to identify potentially relevant studies. Additionally, we contacted content experts to ensure search completeness. We reviewed reference lists of relevant published reviews. Included articles were hand-searched for any additional relevant publications.

Eligibility criteria and article selection. We included studies with the primary objective of developing, validating, or establishing psychometric properties of patient-reported FSAMs in RA. We applied the following exclusion criteria: non-English publications, studies validating FSAMs in non-RA populations, performance-based measures (e.g., grip strength, walk tests, etc.), FSAMs that assessed a single extremity or body part, and studies using FSAMs to validate another instrument (e.g., assessing validity of joint ultrasound using FSAMs). We excluded health-related quality of life measures or multidimensional measures including function as a single construct among many (e.g., Short Form 36 [SF-36]) and studies only evaluating the cross-cultural validity of FSAMs.

Two reviewers (CEHB and JZ) first independently screened titles and abstracts to determine eligible studies for full text review and then conducted a full text review of eligible studies independently in duplicate. Disagreements between reviewers were resolved by discussion between reviewers or with a third reviewer (KM) when necessary.

Data abstraction and study quality assessment. Two of the 3 independent reviewers (CEHB, JZ, or VB) conducted data abstraction in duplicate for 15% of included articles to obtain consistent abstraction. A single reviewer (CEHB) abstracted the remaining studies with additional spot-checking of data abstraction performed by a second reviewer (VB). All measure characteristics, including details on measure items, administration time, scoring, and interpretation were abstracted. FSAMs with limited publications in RA (≤3) and/or not commonly used in the US (as evidenced in the ACR's Rheumatology Informatics System for Effectiveness [RISE] registry [14]) were not further evaluated for methodologic quality using COSMIN as it was unlikely such measures would be recommended for use due to feasibility concerns.

We rated the methodologic quality of included studies using COSMIN checklists (15). Briefly, COSMIN is a standardized tool for assessing study properties including internal consistency, reliability, measurement error, content validity, structural validity, hypothesis testing, cross-cultural validity, responsiveness, and interpretability. For each measurement property, a checklist of 5–18 items is completed and rated on a 4-point scale (poor, fair, good, or excellent) based on predefined criteria. An overall score for each property is based on the lowest score for each checklist. To assess the study psychometric result quality, we

employed a rating scheme using criteria proposed by Terwee et al (16) as modified by Dobson et al (17).

Although not rated using the 4-point scale, COSMIN reporting also includes standardized abstraction of items relating to the interpretability of the measurement property (including percentage of missing items and handling of missing items, adequate sample size, floor and ceiling effects, and minimum important change or minimum important difference) and the generalizability of the study (including population characteristics and study setting) (16).

Level of evidence. We provided the level of evidence for each individual FSAM psychometric property, considering all studies evaluating each property and their result using criteria proposed by Hendrikx et al (18) (Table 1). Each RA FSAM psychometric property received a level of evidence rating of strong (+++ or – – –), moderate (++ or – –), limited (+ or –), conflicting (±), or unknown (?) (Table 2). Three authors (CEHB, JZ, and VB) defined the level of evidence, with disagreements settled by a fourth author (KM).

Feasibility. Although evaluating the administration feasibility of FSAMs is not part of COSMIN, the workgroup agreed it is integral to making a recommendation for routine clinical use. An overall feasibility assessment for each FSAM was based on the following criteria: number of questions, whether computer-based administration was required, and associated costs or use licenses. The overall feasibility was scored as very feasible = +++, moderately feasible = ++, feasible = +, and not feasible = –.

Selection process. Ten workgroup members identified and selected by the ACR Quality Measures Subcommittee Chairs, including clinicians and researchers with expertise in functional status measurement and an ACR Quality Measures Subcommittee Liaison (see Supplementary Appendix A, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24040/abstract) participated in a modified Delphi process to provide recommendations for the routine use of each FSAM. Only FSAMs with an overall assessment of adequate psychometric properties and feasibility (a rating of at least + on both) were reviewed. Members were given the study protocol and systematic review, including all COSMIN ratings and overall assessments. Prior to proceeding, members rated their comfort level with the study protocol and transparency, including the proposed modified Delphi process. During each of 3 rounds of the modified Delphi process, members rated each FSAM for ACR recommendation on a scale of 1 to 9 (where 1 = not recommended and 9 = essential to have). Following each round, members reviewed the results prior to re-rating. Following Round 2, workgroup members participated in a conference call to review and discuss the voting results, followed by a final round of voting. FSAMs were recommended if >80% of members (all but 1 member) rated the FSAM in the 7–9 range and excluded if >80% of ratings were in the 1–3 range, following best practices (19). FSAMs not achieving recommendation for inclusion or exclusion were deemed inconclusive. FSAMs deemed inconclusive at the end of voting remained on the list of measures fulfilling the minimum standard. The ACR Quality Measures Subcommittee reviewed these recommendations in parallel with the recommendations on functional status assessment, modifying as necessary based upon the goal of identifying preferred tools for regular use in most clinic settings, before voting. The ACR Quality of Care Committee and ACR Board of Directors reviewed and approved this article prior to publication.

Table 1. Rating the levels of evidence for the Functional Status Assessment Measures*

Level	Rating	Criteria
Strong	+++ or – – –	Consistent findings in multiple studies of good (methodologic) quality OR in one study of excellent quality
Moderate	++ or – –	Consistent findings in multiple studies of fair methodologic quality OR in one study of good methodologic quality
Limited	+ or –	One study of fair methodologic quality
Conflicting	±	Conflicting findings
Unknown	?	Only studies of poor methodologic quality
No evidence	0	No studies

* Positive result = +; negative result = −. Based on ref. 18.

RESULTS

A total of 11,835 articles underwent title and abstract screening; of those, 649 were eligible for full text review during which 571 articles were excluded (Figure 1). We identified 3 additional articles through hand searches, resulting in 81 included articles. After excluding 25 articles that were not based on the HAQ or PROMIS, 56 were subjected to COSMIN review, including 48 on HAQ-derived and 8 on PROMIS-derived instruments.

Patient-reported FSAMs. FSAMs ranged from simple visual analog scales to questionnaires with over 100 items (see Supplementary Appendix A, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24040/abstract). We excluded 19 FSAMs that had ≤3 RA-relevant publications and/or were rarely used in the US. The HAQ DI, 3 additional HAQ-derived measures (the modified HAQ [M-HAQ], MDHAQ, and HAQ-II), two PROMIS static forms (the physical function 10-item and 20-item [PF10a and PF20a]), and the PROMIS physical function Computer Adaptative Test (PF CAT) underwent COSMIN evaluation. Characteristics of included studies are shown in Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24040/abstract.

Table 2. Overall assessment of the psychometric properties of the evaluated Functional Status Assessment Measures in rheumatoid arthritis*

	HAQ				PROMIS		
Psychometric properties	HAQ DI	M-HAQ	MDHAQ	HAQ-II	PF10a	PF20a	PF CAT
Internal consistency	++	++	++	++	0	0	++
Reliability							
Retest	++	?	?	0	0	+	+
Interrater	?	0	0	0	0	0	0
Measurement error	?	++	0	0	0	0	++
Validity							
Structural	+++	++	–	+	0	0	0
Criterion	N/A	++	0	+	0	0	N/A
Hypothesis testing	++	++	++	+	++	++	++
Content	+	0	0	0	0	+++†	0
Responsiveness‡	++	++	0	+	++	++	++
Interpretability	+/–	–	+	++	++	++	++
Overall assessment§	+	+	+	++	++	++	++

* HAQ = Health Assessment Questionnaire; HAQ DI = HAQ disability index; M-HAQ = modified HAQ; MDHAQ = Multidimensional HAQ; PROMIS = Patient Reported Outcomes Measurement Information system; PF10a = PROMIS physical function 10-item form; PF20a = PROMIS physical function 20-item form; PF CAT = PROMIS physical function Computer Adaptive Test.
† This study also examined content validity of the entire PROMIS item bank.
‡ Due to substantial heterogeneity in the evaluation of responsiveness, due to a lack of a functional status gold standard, only the quality of the studies was considered, not the result.
§ Overall assessment: + was assigned if the measures demonstrated adequate psychometric qualities (i.e., the measure is valid for use in routine clinical practice and captures functional status and can be reliably followed over time), ++ was assigned if, in addition, the measure had evidence of superior development methodology resulting in a more robust measure with improved floor/ceiling effects, and +++ was assigned if there was an abundance of evidence supporting a superiorly developed measure. Ratings of – were reserved for measures without any evidence of basic validity for use in routine clinical practice.

Internal consistency. There was moderate evidence for all HAQ-derived measures and the PROMIS PF CAT, which were the instruments with available internal consistency data (Table 2 and Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24040/abstract). Cronbach's alpha was the most commonly reported internal consistency assessment and was always acceptable (α = 0.70–0.95) when reported.

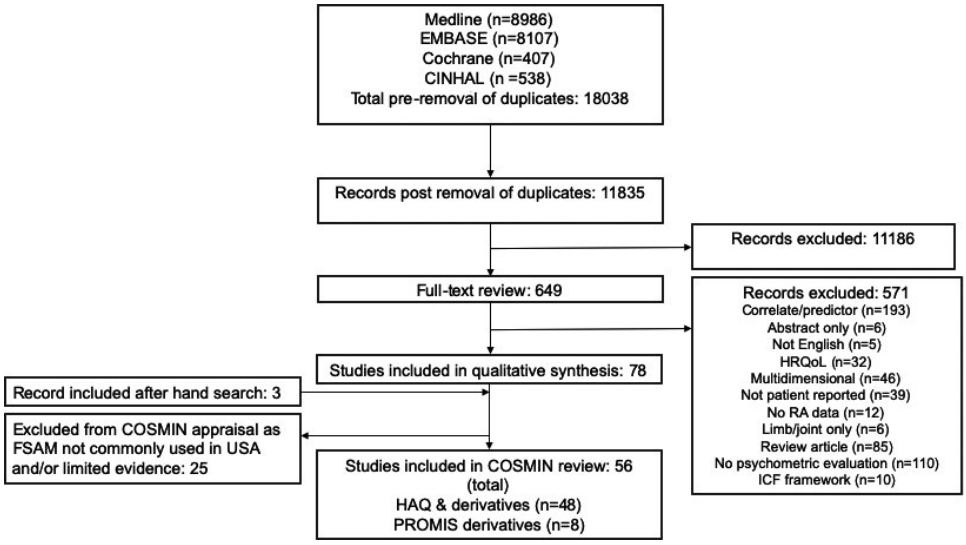

Figure 1. Flow diagram depicting manuscript selection for systematic review of functional status measures. COSMIN = Consensus-Based Standards for the Selection of Health Measurement Instruments; FSAM = Functional Status Assessment Measure; HAQ = Health Assessment Questionnaire; PROMIS = Patient-Reported Outcomes Measurement; HRQoL = health-related quality of life; RA = rheumatoid arthritis; ICF = International Classification of Functioning.

Reliability. The most common type of reliability testing, test–retest reliability, was usually assessed by interclass correlation coefficient (ICC). Reported ICCs were >0.7 for most domains (see Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24040/abstract). The HAQ DI reached a moderate reliability due to a single good COSMIN-rated study. Both the M-HAQ and MDHAQ had indeterminate reliability ratings because of only poor-quality studies. PROMIS measures had very limited reliability data and achieved a limited reliability rating for one FSAM.

Measurement error. According to COSMIN, the preferred measurement error statistics for classical test theory (CTT)–based studies are, in order of preference, standard error of measurement, limits of agreement, and smallest detectable change. Measurement error was only reported for the HAQ DI, M-HAQ, and PROMIS PF CAT, and each used a different method, which made comparisons challenging (see Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24040/abstract). The HAQ DI had only poor-quality studies, leading to an indeterminate assessment. The M-HAQ had a single fair-quality study that only provided 95% confidence intervals, supporting greater precision with an Item Response Theory (IRT)–based FSAM combining the SF-36 and M-HAQ than a non-IRT based measure (20). IRT-based measures use an item bank with specific questions related to a domain of health (21,22) that are evaluated for their correlation with a latent trait, in this case physical function (23). For the PROMIS PF CAT, study methods precluded COSMIN rating (24). However, results of the single study showed the PROMIS PF CAT had higher precision than the HAQ DI, based on root mean square errors. No study reported minimum important change, which should be greater than measurement error (16).

Content validity. The COSMIN content validity checklist assesses whether the authors appropriately judge item relevance and comprehensiveness. Very few articles explicitly evaluated RA FSAM content validity (see Supplementary Appendix A, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24040/abstract). A single, fair-quality article on the HAQ DI (25) yielded a limited rating. A study by Oude Voshaar et al (24) compared the PROMIS PF20, the PROMIS physical function item bank, the HAQ DI and the SF-36 physical function scale to the International Classification of Functioning, Disability, and Health (ICF) core set (26,27) for RA. Their high-quality study demonstrated that the PROMIS physical function item bank more comprehensively reflected all areas of RA-related physical function according to the ICF core set.

Structural validity. COSMIN structural validity reflects the "degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured" (i.e., functional status) (15). Factor analysis is the preferred CTT method, while IRT methods may also check item dimensionality. For good FSAM structural validity, factors should explain at least 50% of the variance (17). We identified 10 studies evaluating structural validity for the HAQ DI, M-HAQ, MDHAQ, and HAQ-II (see Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24040/abstract). Not all reported the percentage of variance explained by the models, because many used IRT-based methods, making comparisons challenging. In IRT, the model fit is examined to ensure the model reflects the true relationship between the underlying construct and the item response (28). Fit (or conversely misfit) of items describes the relationship between predicted and observed responses (28). One excellent study on the HAQ DI (29) yielded an overall strong weighting for structural validity despite lower-quality studies suggesting some misfitting HAQ items. We found 3 studies on M-HAQ (1 excellent, 1 fair, and 1 poor quality). However, the results of the methodologically strongest M-HAQ study concluded that an IRT-based scale combining the M-HAQ and SF-36 physical function scale had improved model fit versus the M-HAQ alone (20). The fair- and poor-quality studies identified misfitting M-HAQ items (2,30). A single, fair-quality HAQ-II study (2) demonstrated excellent structural validity compared to the HAQ DI, M-HAQ, and MDHAQ; however, limited evidence led to an overall low rating. The MDHAQ received a limited negative overall rating based upon 1 poor- (30) and 1 fair-quality study (2), which concluded the MDHAQ had 3 misfitting items. No study reported structural validity for the PROMIS-related measures in RA populations.

Criterion validity. Criterion validity assesses the degree to which instrument scores adequately reflect a gold standard. While there is no gold standard for RA FSAMs, in the case of HAQ-derived measures, the HAQ DI is considered the gold standard. Criterion validity evidence was assessed for the M-HAQ and HAQ-II (see Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24040/abstract). Given the fact that there were multiple studies of fair quality (2,31–33), the M-HAQ was assigned a moderate level of evidence. The HAQ-II received a limited evidence level based on a single fair-quality study (2).

Convergent validity. We found many instruments and variables assessing convergent validity between FSAMs, leading to heterogeneous results (see Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24040/abstract). Evidence of convergent validity was found for all instruments. However, the quality and number of studies varied, yielding a moderate level of evidence for all FSAMs except for HAQ-II. With only 1 fair-quality study, the HAQ-II received a limited rating (2).

Responsiveness. Responsiveness reflects an instrument's ability to detect change over time when true change has occurred. We identified responsiveness evidence for all FSAMs except the MDHAQ (see Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24040/abstract). COSMIN stipulates that hypotheses about expected change scores or correlations between instrument change scores and changes in other variables should be expressed. Hypotheses about expected effect size or similar measures including standardized response means can also be used when explicit hypotheses are made. Heterogeneity in approach across studies made comparisons using our selected approach difficult. Furthermore, FSAM responsiveness testing used disparate comparator outcomes (e.g., patient's perception of change, pain, disease activity, etc.). Based only on study quality (and not the results due to significant reporting heterogeneity), we found moderate evidence for the HAQ DI, HAQ-II, M-HAQ, and all PROMIS measures.

Floor and ceiling effects. According to the results of a study by Terwee et al (16), fewer than 15% of respondents achieve the highest or lowest possible scores in good quality instruments. Where evaluated, the M-HAQ had high percentages of patients with the lowest scores leading to an unfavorable overall rating. There was mixed information about the HAQ. The HAQ-II, MDHAQ, and PROMIS measures achieved moderate ratings (see Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24040/abstract).

Results of feasibility. While the HAQ DI, M-HAQ, MDHAQ, HAQ-II, and the PROMIS measures are all feasible because they are in current use in clinical practice, the shorter FSAMs (the M-HAQ, MDHAQ, HAQ-II, and PROMIS PF10a) received higher feasibility ratings (Table 3). The PROMIS PF CAT received a lower rating due to computer and proprietary software requirements.

Delphi selection of recommended measures. The results from the modified Delphi process are shown in Table 4. The PROMIS PF10a and HAQ-II reached consensus for recommended use and no FSAMs reached consensus for exclusion. Among FSAMs without consensus, the M-HAQ had the lowest mean panelist score and the MDHAQ had the highest mean score (3.1 and 6.6, respectively).

The ACR Quality Measures Subcommittee approved these 2 recommendations with only 1 modification, which was the additional recommendation of the MDHAQ. The MDHAQ was included in the measures for preferred use based upon Delphi rating, feasibility, current use, and strength of its inclusion in the prior (11) and concurrent (34) ACR RA disease activity measure recommendations within the Routine Assessment of Patient Index Data 3 (RAPID3), considerations beyond this current work that focused solely on function.

DISCUSSION

This work represents the first ACR recommendations on FSAMs for use in routine clinical practice in RA. It provides a systematic literature review and synthesis of the psychometric properties of widely used FSAMs as well as a modified Delphi expert panel process to assess the feasibility of routine clinical use. Only 3 FSAMs were recommended: the PROMIS PF10a, HAQ-II, and MDHAQ. Consensus for recommendation was not reached for an additional 4 measures (the HAQ DI, M-HAQ, PROMIS PF20a, and PROMIS PF CAT). These 4 additional FSAMs will be monitored for inclusion in future recommendations along with any new instruments. Importantly, an inconclusive recommendation when applied to the 4 measures in this article should not necessarily prevent these measures from being used. Rather, it highlights the fact that more information is necessary before recommending widespread use of these 4 measures over other measures.

The HAQ DI (35) is one of the oldest and most widely-used patient-reported FSAMs in rheumatology. A variety of adaptations of the HAQ DI were later developed to shorten the

Table 3. Feasibility of the Functional Status Assessment Measures reviewed*

	HAQ				PROMIS		
Feasibility properties	HAQ DI	M-HAQ	MDHAQ	HAQ-II	PF10a	PF20a	PF CAT
No. of questions	20†	8	10	10	10	20	Variable (~5)
Requires computer	No	No	No	No	Assessment center scoring preferred‡	Assessment center scoring preferred‡	Yes§
Proprietary license for use	No	No	No	No	No	No	Yes
Overall feasibility assessment	++	+++	+++	+++	+++	++	+

* HAQ = Health Assessment Questionnaire; HAQ DI = HAQ disability index; M-HAQ = modified HAQ; MDHAQ = Multidimensional HAQ; PROMIS = Patient Reported Outcomes Measurement Information System; PF10a = PROMIS physical function 10-item form; PF20a = PROMIS physical function 20-item form; PF CAT = PROMIS physical function Computer Adaptive Test; +++ = very feasible; ++ = moderately feasible; + = feasible; – = not feasible.
† Requires assessment of the use of 13 assistive devices or help from others with 8 activities, and examined content validity of the entire PROMIS item bank.
‡ Score conversion tables available.
§ Assessment center pricing is available at URL: http://www.healthmeasures.net/resource-center/about-us/pricing-for-services.

Table 4. Results from 3-round modified Delphi process for functional status assessment measures*

	HAQ				PROMIS		
	HAQ DI	M-HAQ	MDHAQ	HAQ-II	PF10a	PF20a	PF CAT
Round 1							
Mean	6.4	5.3	5.1	6.9	7.1	6.5	5.6
Ratings†	0/6/4	3/3/4	3/4/3	1/1/8	1/0/9	1/2/7	1/5/3
Round 2							
Mean	6.4	3.6	4.4	7.1	N/A	6.6	5.3
Ratings†	1/3/6	6/3/1	5/1/4	1/0/9	N/A	1/1/8	2/6/2
Round 3							
Mean	6.2	3.1	6.6	N/A	N/A	6.5	5.7
Ratings†	1/4/5	6/4/0	0/3/7	N/A	N/A	1/2/7	3/1/6
Final recommendation	Inconclusive	Inconclusive	Recommended‡	Recommended	Recommended	Inconclusive	Inconclusive

* HAQ = Health Assessment Questionnaire; HAQ DI = HAQ disability index; M-HAQ = modified HAQ; MDHAQ = Multidimensional HAQ; PROMIS = Patient Reported Outcomes Measurement Information system; PF10a = PROMIS physical function 10-item form; PF20a = PROMIS physical function 20-item form; PF CAT = PROMIS physical function Computer Adaptive Test; N/A = not applicable because measure included based on previous rounds of voting.
† Ratings were reported by the number of participant votes on a 1–9 Likert scale (1–3/4–6/7–9) where 1–3 = not recommended, 4–6 = sometimes recommended, 7–9 = essential to have; and >80% agreement required for recommendation.
‡ During review by the American College of Rheumatology (ACR) Quality Measures Subcommittee, the additional final recommendation of the MDHAQ for preferred use was based upon Delphi rating, feasibility, current use, and strength of its inclusion in the prior and concurrent ACR Rheumatoid Arthritis disease activity measure recommendations within the Routine Assessment of Patient Index Data 3 measure.

scale while maintaining or improving its original psychometric properties. The most commonly used adaptations include the M-HAQ (32), MDHAQ (36), and the HAQ-II (2). More recently, PROMIS measures have been developed and are widely used (URL: http://www.nihpromis.org). PROMIS is a National Institutes of Health initiative that aims to create a more efficient and precise resource for patient outcome measurement when compared to existing legacy instruments for use in a wide variety of chronic disease conditions (21). PROMIS measures evaluate physical, mental, and social health across different chronic conditions (37) and general population health (21). Although most FSAMs were developed using CTT, the PROMIS measures were developed using modern IRT methods. PROMIS measures are available in static short forms with a fixed number of questions and also as computer adaptive tests, which adapt to the ability level of the respondent. The results of all PROMIS measures are normalized to the US population and reported with a T score (mean ± SD 50 ± 10).

The PROMIS physical function measures evaluated in our study included the 10- and 20-item static forms (the PF10a and PF20a) and the PROMIS PF CAT. However, only the PROMIS PF10a was recommended by our panelists. While the PROMIS physical function measures were developed using rigorous methods and tested extensively in the general population and populations with chronic disease (22,38,39), there were few studies specific to patients with RA (24,40–45), impacting panelist ratings. Panelists concluded that the shorter 10-item instrument was likely more feasible for routine use in the clinic than the 20-item survey. While the adaptive PROMIS PF CAT usually requires the fewest items, the computer and proprietary software requirements reduced its feasibility.

The HAQ-II is a 10-item questionnaire developed using Rasch analysis and IRT-based methodology. Instrument development was aimed at addressing 4 main issues identified with the original HAQ DI and its derivatives: removing misfitting items, maximizing scale length, eliminating items with overlapping difficulties, and eliminating gaps in measurement along the continuum of functional status assessment (2). The resulting instrument includes 5 items from the original HAQ DI plus 5 new items. When compared to the M-HAQ, MDHAQ, and HAQ DI, the HAQ-II better captures the disability continuum. Gaps in the measurement of disability were found in all scales evaluated except the HAQ-II, indicating that the HAQ-II has the most favorable psychometric properties of the HAQ-derived instruments. The HAQ-II also has the least floor effect among the evaluated HAQ-derived measures.

Although the HAQ DI is the legacy FSAM, and has been extensively tested and used worldwide, its psychometric properties when compared to the HAQ-II and the newer PROMIS measures were felt to be less favorable. Additionally, the length and relatively complex scoring of the HAQ DI led to lower panelist ratings.

The MDHAQ was designed as a shorter version of the HAQ DI and includes 10 items (all items from the M-HAQ plus 2 additional items) (32). While the MDHAQ has greater feasibility than the original HAQ DI and more favorable psychometric properties compared to the M-HAQ (36), it performs less well when compared to the HAQ-II (2) or the PROMIS measures (44). A limitation in our assessment of the MDHAQ is that we did not evaluate the literature on the RAPID3 measure (46). The RAPID3 is a patient-reported disease activity tool that includes the MDHAQ, a measure of pain, and a patient global score (46). The psychometric and clinometric properties of the RAPID3 have been reviewed by the ACR RA Disease Activity Workgroup, which recommended the RAPID3 as an effective measure of RA disease activity. RAPID3 is also the most commonly collected disease activity measure in the RISE registry

(14). Given this, we additionally recommend the MDHAQ as a preferred FSAM.

The 8-item M-HAQ is derived from the HAQ DI (using 1 question from each domain) and is the shortest measure evaluated (22). Although the M-HAQ is highly correlated to the HAQ DI (32), the M-HAQ has significant floor effects and may not be as sensitive to clinical changes as longer scales (2). The panel did not reach consensus for excluding the M-HAQ; however, it received the lowest scores of all the FSAMs evaluated.

Our study had a number of strengths, including the rigorous and transparent methodologic assessment of the measures combined with expert opinion; however, there are some limitations. We did not subject all FSAMs to COSMIN assessment and consideration by our expert panel because it was felt unlikely that measures not already commonly used in the US would be included in our final recommendations. Therefore, it is possible that measures with highly favorable psychometric properties were not considered in generating our recommendations. Additionally, our review was conducted while only considering RA-specific data and English-language publications, and it is possible this limited the evidence on which our recommendations were based. After our systematic review was completed, the COSMIN group updated their checklist (47), and the study ratings could be different if the updated checklist was used. Given that the overall panelist ratings on the FSAMs weighed not only the psychometric properties as evaluated by COSMIN but also measured feasibility, it is less likely that the overall outcome of the process would have varied greatly from our present results by using the updated checklist. Patients were not involved in the panel, given the significant methodologic expertise required for the project; however, this work will inform ongoing measure development work, which includes patient partners. Lastly, given the paucity of psychometric data on some measures, further research in this area is warranted and it is possible that some of the recommendations may change in the future as a result of new findings.

In conclusion, we have presented the first ACR recommendations on FSAMs for routine use in clinical practice to be used for the assessment of functional status in RA, based on a rigorous systematic review and expert panel process. Although we only recommend 3 FSAMs, this work should not preclude the use of other identified measures but rather encourage the use of measures with the most favorable psychometric properties while highlighting the need for ongoing research in this area.

ACKNOWLEDGMENTS

We thank American College of Rheumatology staff members Amy Turner and Regina Parker for their support and assistance through the recommendation process.

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. Michaud 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. Barber, Zell, Yazdany, Davis, Cappelli, Suter, Limanni, Michaud.

Acquisition of data. Barber, Zell, Bohm.

Analysis and interpretation of data. Barber, Zell, Davis, Cappelli, Ehrlich-Jones, Everix, Thorne, Suter, Michaud.

REFERENCES

1. World Health Organization. International classification of functioning, disability and health (ICF). May 2001. URL: https://www.who.int/classifications/icf/en/.
2. Wolfe F, Michaud K, Pincus T. Development and validation of the Health Assessment Questionnaire II: a revised version of the health assessment questionnaire. Arthritis Rheum 2004;50:3296–305.
3. Cohen JD, Dougados M, Goupille P, Cantagrel A, Meyer O, Sibilia J, et al. Health assessment questionnaire score is the best predictor of 5-year quality of life in early rheumatoid arthritis. J Rheumatol 2006;33:1936–41.
4. Pincus T, Brooks RH, Callahan LF. Prediction of long-term mortality in patients with rheumatoid arthritis according to simple questionnaire and joint count measures. Ann Intern Med 1994;120:26–34.
5. Michaud K, Vera-Llonch M, Oster G. Mortality risk by functional status and health-related quality of life in patients with rheumatoid arthritis. J Rheumatol 2012;39:54–9.
6. Sokka T, Pincus T. Poor physical function, pain and limited exercise: risk factors for premature mortality in the range of smoking or hypertension, identified on a simple patient self-report questionnaire for usual care. BMJ Open 2011;1:e000070.
7. Yelin E, Trupin L, Wong B, Rush S. The impact of functional status and change in functional status on mortality over 18 years among persons with rheumatoid arthritis. J Rheumatol 2002;29:1851–7.
8. Singh JA, Saag KG, Bridges SL Jr, Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. Arthritis Rheumatol 2016;68:1–26.
9. Yazdany J, Robbins M, Schmajuk G, Desai S, Lacaille D, Neogi T, et al. Development of the American College of Rheumatology's rheumatoid arthritis electronic clinical quality measures. Arthritis Care Res (Hoboken) 2016;68:1579–90.
10. Department of Health & Human Services. Quality payment program: quality measures requirements. URL: https://qpp.cms.gov/mips/quality-measures.
11. 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.
12. PRISMA. Transparent reporting of systematic reviews and meta-analyses. URL: http://www.prisma-statement.org/.
13. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res 2009;18:1115–23.
14. Yazdany J, Bansback N, Clowse M, Collier D, Law K, Liao KP, et al. Rheumatology Informatics System for Effectiveness: a national informatics-enabled registry for quality improvement. Arthritis Care Res (Hoboken) 2016;68:1866–73.
15. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of

studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012;21:651–7.
16. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34–42.
17. Dobson F, Hinman RS, Hall M, Terwee CB, Roos EM, Bennell KL. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. Osteoarthritis Cartilage 2012;20:1548–62.
18. Hendrikx J, de Jonge MJ, Fransen J, Kievit W, van Riel PL. Systematic review of patient-reported outcome measures (PROMs) for assessing disease activity in rheumatoid arthritis. RMD Open 2016;2:e000202.
19. Diamond IR, Grant RC, Feldman BM, Pencharz PB, Ling SC, Moore AM, et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. J Clin Epidemiol 2014;67:401–9.
20. Martin M, Kosinski M, Bjorner JB, Ware JE Jr, MacLean R, Li T. Item response theory methods can improve the measurement of physical function by combining the modified health assessment questionnaire and the SF-36 physical function scale. Qual Life Res 2007;16:647–60.
21. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. J Clin Epidemiol 2010;63:1179–94.
22. Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware JE Jr. The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. J Clin Epidemiol 2014;67:516–26.
23. Fries JF, Witter J, Rose M, Cella D, Khanna D, Morgan-DeWitt E. Item response theory, computerized adaptive testing, and PROMIS: assessment of physical function. J Rheumatol 2014;41:153–8.
24. Oude Voshaar MA, ten Klooster PM, Glas CA, Vonkeman HE, Taal E, Krishnan E, et al. Validity and measurement precision of the PROMIS physical function item bank and a content validity-driven 20-item short form in rheumatoid arthritis compared with traditional measures. Rheumatology (Oxford) 2015;54:2221–9.
25. Goeppinger J, Doyle MA, Charlton SL, Lorig K. A nursing perspective on the assessment of function in persons with arthritis. Res Nurs Health 1988;11:321–31.
26. Cieza A, Brockow T, Ewert T, Amman E, Kollerits B, Chatterji S, et al. Linking health-status measurements to the international classification of functioning, disability and health. J Rehabil Med 2002;34:205–10.
27. Cieza A, Geyh S, Chatterji S, Kostanjsek N, Ustun B, Stucki G. ICF linking rules: an update based on lessons learned. J Rehabil Med 2005;37:212–8.
28. Siemons L, Krishnan E. A short tutorial on item response theory in rheumatology. Clin Exp Rheumatol 2014;32:581–6.
29. Cole JC, Motivala SJ, Khanna D, Lee JY, Paulus HE, Irwin MR. Validation of single-factor structure and scoring protocol for the Health Assessment Questionnaire-Disability Index. Arthritis Rheum 2005;53:536–42.
30. Pincus T, Sokka T, Kautiainen H. Further development of a physical function scale on a Multidimensional Health Assessment Questionnaire for standard care of patients with rheumatic diseases. J Rheumatol 2005;32:1432–9.
31. Uhlig T, Haavardsholm EA, Kvien TK. Comparison of the Health Assessment Questionnaire (HAQ) and the modified HAQ (M-HAQ) in patients with rheumatoid arthritis. Rheumatology (Oxford) 2006;45:454–8.
32. Pincus T, Summey JA, Soraci SA Jr, Wallston KA, Hummon NP. Assessment of patient satisfaction in activities of daily living using a modified Stanford Health Assessment questionnaire. Arthritis Rheum 1983;26:1346–53.
33. Nagasawa H, Kameda H, Sekiguchi N, Amano K, Takeuchi T. Differences between the Health Assessment Questionnaire Disability Index (HAQ DI) and the modified HAQ (M-HAQ) score before and after infliximab treatment in patients with rheumatoid arthritis. Mod Rheumatol 2010;20:337–42.
34. England BR, Tiong BK, Bergman MJ, Curtis JR, Kazi S, Mikuls TR, et al. 2019 update of the American College of Rheumatology Recommended Rheumatoid Arthritis Disease Activity Measures. Arthritis Care Res [Hoboken] 2019;71:1540–55.
35. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137–45.
36. Pincus T, Swearingen C, Wolfe F. Toward a multidimensional Health Assessment Questionnaire (MDHAQ): assessment of advanced activities of daily living and psychological status in the patient-friendly health assessment questionnaire format. Arthritis Rheum 1999;42:2220–30.
37. Cook KF, Jensen SE, Schalet BD, Beaumont JL, Amtmann D, Czajkowski S, et al. PROMIS measures of pain, fatigue, negative affect, physical function, and social function demonstrated clinical validity across a range of chronic conditions. J Clin Epidemiol 2016;73:89–102.
38. Hays RD, Spritzer KL, Amtmann D, Lai JS, Dewitt EM, Rothrock N, et al. Upper-extremity and mobility subdomains from the Patient-Reported Outcomes Measurement Information System (PROMIS) adult physical functioning item bank. Arch Phys Med Rehabil 2013;94:2291–6.
39. Rothrock NE, Hays RD, Spritzer K, Yount SE, Riley W, Cella D. Relative to the general US population, chronic diseases are associated with poorer health-related quality of life as measured by the Patient-Reported Outcomes Measurement Information System (PROMIS). J Clin Epidemiol 2010;63:1195–204.
40. Fries J, Rose M, Krishnan E. The PROMIS of better outcome assessment: responsiveness, floor and ceiling effects, and Internet administration. J Rheumatol 2011;38:1759–64.
41. Fries JF, Krishnan E, Rose M, Lingala B, Bruce B. Improved responsiveness and reduced sample size requirements of PROMIS physical function scales with item response theory. Arthritis Res Ther 2011;13:R147.
42. Hays RD, Spritzer KL, Fries JF, Krishnan E. Responsiveness and minimally important difference for the patient-reported outcomes measurement information system (PROMIS) 20-item physical functioning short form in a prospective observational study of rheumatoid arthritis. Ann Rheum Dis 2015;74:104–7.
43. Oude Voshaar MA, ten Klooster PM, Glas CA, Vonkeman HE, Krishnan E, van de Laar MA. Relative performance of commonly used physical function questionnaires in rheumatoid arthritis and a patient-reported outcomes measurement information system computerized adaptive test. Arthritis Rheumatol 2014;66:2900–8.
44. Bartlett SJ, Orbai AM, Duncan T, DeLeon E, Ruffing V, Clegg-Smith K, et al. Reliability and validity of selected PROMIS measures in people with rheumatoid arthritis. PLoS ONE 2015;10:e0138543.
45. Wahl E, Gross A, Chernitskiy V, Trupin L, Gensler L, Chaganti K, et al. Validity and responsiveness of a 10-item patient-reported measure of physical function in a rheumatoid arthritis clinic population. Arthritis Care Res (Hoboken) 2017;69:338–46.
46. Pincus T, Swearingen CJ, Bergman M, Yazici Y. RAPID3 (Routine Assessment of Patient Index Data 3), a rheumatoid arthritis index without formal joint counts for routine care: proposed severity categories compared to disease activity score and clinical disease activity index categories. J Rheumatol 2008;35:2136–47.
47. Mokkink LB, de Vet HC, Prinsen CA, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Qual Life Res 2018;27:1171–9.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 71, No. 12, December 2019, pp 1540–1555
DOI 10.1002/acr.24042

SPECIAL ARTICLE

2019 Update of the American College of Rheumatology Recommended Rheumatoid Arthritis Disease Activity Measures

Bryant R. England,[1] Benedict K. Tiong,[2] Martin J. Bergman,[3] Jeffrey R. Curtis,[4] Salahuddin Kazi,[5] Ted R. Mikuls,[1] James R. O'Dell,[1] Veena K. Ranganath,[2] Alex Limanni,[6] Lisa G. Suter,[7] and Kaleb Michaud[8]

Objective. To provide updated American College of Rheumatology (ACR) recommendations on rheumatoid arthritis (RA) disease activity measurements to facilitate a treat-to-target approach in routine clinical care.

Methods. A working group conducted a systematic literature review from the time of the prior ACR recommendations literature search. Properties of disease activity measures were abstracted, and study quality was assessed using the Consensus-Based Standards for the selection of Health Measurement Instruments 4-point scoring method, allowing for overall level of evidence assessment. Measures that fulfilled a minimum standard were identified, and through a modified Delphi process preferred measures were selected for regular use in most clinic settings.

Results. The search identified 5,199 articles, of which 110 were included in the review. This search identified 46 RA disease activity measures that contained patient, provider, laboratory, and/or imaging data. Descriptions of the measures, properties, study quality, level of evidence, and feasibility were abstracted and scored. Following a modified Delphi process, 11 measures fulfilled a minimum standard for regular use in most clinic settings, and 5 measures were recommended: the Disease Activity Score in 28 Joints with Erythrocyte Sedimentation Rate or C-Reactive Protein Level, Clinical Disease Activity Index, Simplified Disease Activity Index, Routine Assessment of Patient Index Data 3, and Patient Activity Scale-II.

Conclusion. We have updated prior ACR recommendations for preferred RA disease activity measures, identifying 11 measures that met a minimum standard for regular use and 5 measures that were preferred for regular use in most clinic settings.

Supported by the American College of Rheumatology. Dr. England's work was supported by the University of Nebraska Medical Center (Internal Medicine Scientist Development award, Mentored Scholars Program award, and Physician-Scientist Training Program award), the Rheumatology Research Foundation (Scientist Development award), and the NIH (National Institute of General Medical Sciences grant 1U54-GM-115458). Dr. Curtis' work was supported by the Patient-Centered Outcomes Research Institute, and the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases grant AR-064172). Dr. Kazi's work was supported by the Dartmouth Institute. Dr. Mikuls' work was supported by the VA, the Rheumatology Research Foundation, and the NIH (National Institute of General Medical Sciences grant 1U54-GM-115458). Dr. O'Dell's work was supported by the VA. Dr. Ranganath's work was supported by the Rheumatology Research Foundation. Dr. Suter receives salary support from the VA and Yale New Haven Health System, Center for Outcomes Research & Evaluation through a contract to the Centers for Medicare and Medicaid Services. Dr. Michaud's work was supported by the Rheumatology Research Foundation.

[1]Bryant R. England, MD, Ted R. Mikuls, MD, MSPH, James R. O'Dell, MD: VA Nebraska-Western Iowa Health Care System and University of Nebraska Medical Center, Omaha; [2]Benedict K. Tiong, MD, Veena K. Ranganath, MD, MS, RhMSUS: University of California, Los Angeles; [3]Martin J. Bergman, MD: Drexel University, Philadelphia, Pennsylvania; [4]Jeffrey R. Curtis, MD, MS, MPH: University of Alabama at Birmingham; [5]Salahuddin Kazi, MD: University of Texas Southwestern, Dallas; [6]Alex Limanni, MD: Arthritis Centers of Texas, Dallas; [7]Lisa G. Suter, MD: Yale University, New Haven, Connecticut, and VA Medical Center, West Haven, Connecticut; [8]Kaleb Michaud, PhD: University of Nebraska Medical Center, Omaha, and FORWARD, the National Databank for Rheumatic Diseases, Wichita, Kansas.

Dr. Bergman has received consulting fees, speaking fees, and/or honoraria from Boehringer Ingelheim, Horizon, Novartis, Pfizer, Sanofi, Sandoz, AbbVie, Celgene, Merck, Amgen, Genentech, and Regeneron (more than 10,000 each) and owns stock or stock options in Johnson & Johnson. Dr. Curtis has received consulting fees from Bristol-Myers Squibb, Roche, UCB, AbbVie, Janssen, Radius, Regeneron, Amgen, Eli Lilly and Company, Pfizer, the Corrona registry, and Myriad (less than $10,000 each) and research support from Amgen, Pfizer, and Radius. Dr. Mikuls has received consulting fees from Pfizer (less than 10,000) and research support from Bristol-Myers Squibb and Horizon. Dr. Ranganath has received consulting fees from Amgen and Bristol-Myers Squibb (less than $10,000 each) and research support from Pfizer, Genentech, Bristol-Myers Squibb, and Mallinckrodt. Dr. Limanni has received research support from GlaxoSmithKline, Janssen, Novartis, Gilead Sciences, and Pfizer. Dr. Michaud has received research support from Pfizer. No other disclosures relevant to this article were reported.

Address correspondence to Kaleb Michaud, PhD, Division of Rheumatology & Immunology, University of Nebraska Medical Center, 986270 Nebraska Medical Center, Omaha, NE 68198. E-mail: kmichaud@unmc.edu.

Submitted for publication April 17, 2019; accepted in revised form August 13, 2019.

SIGNIFICANCE & INNOVATIONS

- This is the first update to the American College of Rheumatology's recommended rheumatoid arthritis disease activity measures for regular clinical use.
- We used a systematic approach to identify and evaluate measures meeting a minimum standard for regular use that can be repeated in future updates and provide a path for research on existing or new measures.

INTRODUCTION

A treat-to-target strategy in rheumatoid arthritis (RA) was recommended in the 2015 American College of Rheumatology (ACR) RA Treatment Guidelines (1). In order to adhere to these recommendations, regular RA disease activity assessments must be made during routine care. Although the severity of chronic diseases such as diabetes mellitus or hypertension can be directly measured, no equivalent measurement exists in RA. Numerous RA disease activity measures have been proposed for this purpose, most incorporating data gathered from a combination of sources, including patient-reported measures, provider assessments, laboratory values, and/or imaging modalities. These measures may vary in terms of their performance (e.g., validity, reliability, responsiveness) and feasibility for regular use.

Recognizing the challenge that clinicians face selecting a disease activity measure due to multiple options and varying performance, the ACR convened a working group in 2008 to review the literature and provide recommendations on which RA disease activity measures were best suited for regular use (2). RA disease activity measures were identified through a literature review (3), which were then narrowed by an expert advisory panel. Recommendations were drafted after psychometric properties of the measures were compiled and practicing rheumatologists were surveyed. This process resulted in the recommendation of 6 RA disease activity measures: the Clinical Disease Activity Index (CDAI), Disease Activity Score in 28 joints (DAS28), Patient Activity Scale (PAS), Patient Activity Scale II (PAS-II), Routine Assessment of Patient Index Data 3 (RAPID3), and Simplified Disease Activity Index (SDAI) (2).

Since these original recommendations, additional RA disease activity measures have been reported, further studies characterizing the performance of these and other novel measures have been conducted, and imaging modalities have been developed for assessment of disease activity. Therefore, an update to the prior recommendations for selecting an RA disease activity measure was needed, including a critical evaluation of more recent literature. The ACR convened a working group to update these prior recommendations in conjunction with a separate effort to provide initial recommendations on functional status assessment in RA. The objectives of this RA disease activity measures working group were to provide recommendations for RA disease activity measures meeting a minimum standard for regular use, and preferred RA disease activity measures for regular use. The former objective was added since many measures may be valid, feasibility varies across different practices and healthcare systems, and providers may have experience with and be comfortable using certain disease activity measures.

METHODS

Study design. A working group composed of rheumatologists and rheumatology professionals, including one rheumatology professional diagnosed with RA, was convened by the ACR to update the recommended RA disease activity measures. A protocol was developed and agreed on by the working group for providing updated RA disease activity measure recommendations. The recommendation process and preliminary findings were presented in a special session at the 2017 ACR Annual Scientific Meeting held in San Diego, California and were then opened for public comment (from patients, providers, and other stakeholders) following that presentation.

Updated systematic literature review. In conjunction with the assistance of a medical librarian, we updated the prior literature review by searching Ovid Medline, Embase, and Cochrane databases from January 1, 2009 to January 25, 2017 for published original articles on RA disease activity measures using combinations of MeSH terms and keywords for rheumatoid arthritis, disease activity measures, and psychometric properties. We did not review components of composite measures individually as prior recommendations selected the composite measures over their individual components (2). A full description of the systematic literature review is shown in Supplementary Appendix 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24042/abstract. Systematic review inclusion criteria were published articles in the English language reporting on a psychometric property of an RA disease activity measure. The exclusion criteria were reports limited to diseases other than RA; reports assessing only cross-cultural validity, radiographic damage, or a single joint area; and measures not providing numerical values. Titles and abstracts were screened in duplicate by 2 authors (BRE and BKT) for relevance, followed by full text review in duplicate by 2 authors (BRE and BKT) to assess eligibility. Discordance after full text review was settled by a third-party reviewer (KM). Publications retrieved were reviewed to identify additional articles eligible for inclusion. RefWorks (ProQuest) was utilized for management of literature search results.

Data abstraction and study quality assessment. Study details and psychometric properties were abstracted and study quality was assessed from included studies, using the Consensus-Based Standards for the Selection of Health Measurement Instru-

ments (COSMIN) 4-point scoring as the template (4). An abstraction tool was developed and was piloted iteratively for data collection, then applied to the studies by an abstractor (BRE or BKT). To ensure abstraction consistency and quality, regular meetings occurred between the abstractors during the abstraction process.

Items abstracted from studies included those pertaining to the publication (author, year, journal), study (patient characteristics, sample size, setting, patient selection), disease activity measures (measures included, score distributions), and psychometric properties. Psychometric properties abstracted were internal consistency, reliability, measurement error, content validity, structural validity, hypotheses testing, and responsiveness (COSMIN properties [4,5]). Criterion validity was not abstracted because considering a distinct RA disease activity measure a "gold-standard" would bias recommendations. Rather, studies reporting criterion validity were abstracted as hypothesis testing (i.e., convergent validity).

Study quality assessment for each psychometric property was assessed using the COSMIN checklist with a 4-point scale (4). Using this method, each psychometric property reported in each study received a quality rating of excellent, good, fair, or poor. The score assigned to each property in each study represented the lowest score of all the criteria for that property.

Level of evidence. Abstracted data on psychometric properties and study quality were synthesized as others have previously reported (6,7). The psychometric properties for each RA disease activity measure received a level of evidence of strong (rating of +++ or - - -), moderate (rating of ++ or - -), limited (rating of + or -), conflicting (rating of ±), or unknown (rating of ?). See Supplementary Appendix 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24042/abstract, for details concerning the level of evidence grading system. Assessments of level of evidence were performed in duplicate (BRE and BKT), and discordance was settled by a third-party reviewer (KM).

Consideration of prior literature. A literature review was previously performed in conjuction with the 2012 ACR RA Disease Activity Recommendations (3). The psychometric properties of RA disease activity measures identified in the prior review were extracted according to the COSMIN groupings utilized in the current systematic review. Additionally, we searched for psychometric properties of studies not previously included in the prior literature review that were published before our search date. As study quality assessment was not part of the prior review, these results were not incorporated into the level of evidence grading with those from the current systematic review. Instead, these prior performance metrics were provided to the working group members for review during the selection (i.e., voting) process.

Feasibility. Validated scoring systems for the feasibility of RA disease activity measures do not currently exist. We scored feasibility on a scale of 0 to 4 (i.e., - to ++++) with scores ≥1 (+ to ++++) denoting measures feasible for regular use and scores of 4 (++++) representing the most feasible measures. The number of items included in the measure, time to complete, need for provider joint counts, need for laboratory testing, commercial availability, and need for advanced imaging were evaluated as part of the grading. All measures not commercially available or requiring advanced imaging (due to additional equipment, training, or consultation being required) were graded as 0 (i.e., - [not feasible for regular use]). Requirement of provider joint counts or laboratory testing both reduced the maximum score by 1 each. Consideration of number of items and completion time served as final modifiers of the feasibility grade. (The score was reduced by 1 if not feasibile in a routine clinic visit or by 2 if not feasible on the same day as the clinic visit.)

Selection process. The RA disease activity measures working group reviewed the literature search, abstracted data, level of evidence for each identified measure, prior literature for each measure, and feasibility scoring, as well as their own experience with these measures, to provide 2 recommendations on RA disease activity measures feasible for regular use in rheumatology clinics. First, the group identified RA disease activity measures meeting a minimum standard for regular use and second, the group selected measures with the most favorable psychometric properties and feasibility for preferred use.

Fulfilling the minimum standard for an RA disease activity measure in regular use was established by measures 1) providing a numerical value, 2) categorizing to ≥3 disease states that separate low, moderate, and high disease activity, 3) being feasible for regular measurement in the clinic, and 4) possessing adequate psychometric properties. Items were considered to meet the minimum standard for feasibility in regular use if the previously mentioned feasibility score was ≥1. Psychometrics were considered adequate if the level of evidence suggested at least moderate positive results in the COSMIN area of hypothesis testing plus 1 of the following: level of evidence suggesting at least moderate positive results in at least 1 other COSMIN area, level of evidence suggesting at least limited positive results in at least 2 COSMIN areas (one of which must be responsiveness), or a defined minimum important difference/minimum clinically important difference.

A modified Delphi process was utilized to generate the recommendations on RA disease activity measures for preferred use (8). Working group members and an ACR Quality Measures Subcommitee liason rated each measure that fulfilled the minimum standard on a scale of 1 to 9, where 9 = essential this measure be recommended for use. Ratings of 7 to 9 constituted a recommended measure for inclusion, while ratings of 4 to 6 were inconclusive and ratings of 1 to 3 were recommended measures for exclusion. Measures were *recommended* if >80% of members (all but 1) rated the measure in the 7 to 9 range and

Table 1. Characteristics of rheumatoid arthritis disease activity measures*

Measure†	Components	Items, no.	Formula	Range	Rem	Low DA	Moderate DA	High DA	Method of administration	MID, MCID
Clinical Disease Activity Index (CDAI)	28TJC (0–28); 28SJC (0–28); Pt Global VAS (0–10); Pr Global VAS (0–10)	4	28SJC + 28TJC + PtGA + PrGA	0–76	≤2.8	>2.8 to 10	>10 to 22	>22	Patient item, provider assessment	MCID 12 (12 for high DA, 6 for moderate DA, 1 for low DA)
Modified CDAI (Baker)	28SJC; Pr Global	2	28SJC + 2xPrGA	0–48	-	-	-	-	Provider assessment	-
Patient Derived CDAI	Pt 28TJC; Pt 28SJC; Pt Global, Pr Global	4	28SJC + 28TJC + PtGA + PrGA	0–76	≤2.8	>2.8 to 10	>10 to 22	>22	Patient items, provider assessment	-
Disease Activity Score (DAS)	RAI; SJC44; ESR; Pt Global	4	0.53938xSqrt(RAI) + 0.06465xSJC44 + 0.33ln(ESR) + 0.00722(PtGA)	0–10	<1.6	1.6 to <2.4	2.5 to <3.7	≥3.7	Patient item, provider assessment, lab	1.2
Disease Activity Score 28 Joints (DAS28)	28TJC (0–28); 28SJC (0–28); Pt Global VAS (0–10); ESR or CRP	3 or 4	0.56xSqrt(28TJC) + 0.28xSqrt(28SJC) + 0.70xln(ESR) + 0.014xPtGA OR 0.56xSqrt(28TJC)+ 0.28xSqrt(28SJC) + 0.36xln(CRP + 1) + 0.014xPtGA + 0.96	0–9.4	<2.6	2.6 to <3.2	3.2 to ≤5.1	>5.1	Patient item, provider assessment, lab	MID 1.2; MCID 1.2 (DAS28-ESR), 1.0 (DAS28-CRP)
Modified DAS28 (Baker)	28SJC; Pr Global; acute,-phase reactant (ESR or CRP)	3	0.40xln(ESR) + 0.17xSJC28 + 0.26xPrGA -OR- 0.49xln(CRP) + 0.15x28SJC + 0.22PrGA + 1	-	-	-	-	-	Provider assessment, lab	-
Modified DAS28 (no acute-phase reactants) (Bentley)	28TJC; 28SJC; mHAQ; Pain; Pr Global; Pt Global	6	0.53xsqrt(28TJC) + 0.31xsqrt(28SJC) + 0.25xmHAQ + 0.001xPain + 0.005xPrGA + 0.014xPtGA + 1.694	-	-	-	-	-	Patient items, provider assessment	-
Patient Derived DAS28	Pt 28TJC; Pt 28SJC; Pt Global; ESR or CRP	4	0.56xSqrt(28TJC) + 0.28xSqrt(28SJC) + 0.7xln(ESR) + 0.014xPtGA OR 0.56xSqrt(28TJC) + 0.28xSqrt(28SJC) + 0.36xln(CRP + 1) + 0.014xPtGA + 0.96	0–9.4	<2.6	2.6 to <3.2	3.2 to ≤5.1	>5.1	Patient items, lab	-

(Continued)

Table 1. *(Cont'd)*

Measure†	Components	Items, no.	Formula	Range	Rem	Low DA	Moderate DA	High DA	Method of administration	MID, MCID
Ultrasound Derived DAS28	Pt 28TJC; Pr 28TJC; US 28SJC; Pt Global; ESR or CRP	4	0.56xSqrt(28TJC) + 0.28xSqrt(28SJC) + 0.7xln(ESR) + 0.014xPtGA OR 0.56xSqrt(28TJC) + 0.28xSqrt(28SJC) + 0.36xln(CRP + 1) + 0.014xPtGA + 0.96	0–9.4	<2.6	2.6 to <3.2	3.2 to ≤5.1	>5.1	Patient items, provider assessment, lab, imaging modality	-
Global Arthritis Score (GAS)	Pain VAS (0–10); mHAQ (0–24); Pt reported 28TJC (0–28)	3	Pain + mHAQ + Pt 28TJC	0–62	-	-	-	-	Patient items	-
Hospital Universitario La Princesa Index (HUPI)	28TJC; 28SJC; Pt Global; acute-phase reactant (ESR or CRP)	4	Each component scored 0–3 based on cutoff values	0–12	-	≤2	>2 to ≤5	>5	Patient items, provider assessment, lab	4 (~DAS 1.2)
Individualized Ultrasound Score	Selects up to 7 or 12 of most affected joints for MSUS	Up to 7 or 12	Sum of individual joint scores (sum of US subscores divided by maximum score at the joint - GS synovial hypertrophy, PD vascularity, tenosynovitis (GS & PD)	-	-	-	-	-	Imaging modality	-
Individualized Composite Ultrasound Score	Selects up to 7 or 12 of most affected joints for MSUS and clinical examination	Up to 7 or 12	Sum of individual joint scores (sum of US subscores and joint subscores divided by maximum score at the joint)	-	-	-	-	-	Provider assessment, imaging modality	-
Kappa/Lambda Hybrid Antibody	Antibody measurement	1	Values generated referent to a standard curve	-	-	-	-	-	Lab	-
Mean Overall Index for Rheumatoid Arthritis (MOI-RA)	28SJC; 28TJC; Pt Global; Pr Global; Pain; HAQ; ESR	7	Mean of standardized values of individual components (each 0-100)	0–100	-	-	-	-	Patient item, provider assessment, lab	-
Multi-Biomarker Disease Activity Score (MBDA)	CRP; EGF; IL-6; Leptin; MMP-1; MMP-3; Resistin; SAA; TNFRI; VCAM-1; VEGF-A; YKL-40	12	(0.56sqrt(TJC) + 0.28SJC + 0.14PtGA + 0.36log(CRP+1) + 0.96) x 10.53 + 1 (Biomarker scores to predict above components)	1–100	≤25	26–29	30–44	>44	Lab	-
Optical Spectral Transmission (OST)	Bilateral PIP (1–5); MCP (1–5); wrist	22	Not reported	-	-	-	-	-	Imaging modality	-
Patient Activity Scale (PAS)	HAQ (0–3); Pain VAS (0–10); Pt Global VAS (0–10)	3	(HAQx3.33 + Pain VAS + PtGA VAS)/3	0–10	≤0.25	>0.26 to 3.70	3.71 to <8.0	≥8.0	Patient items	-

(Continued)

Table 1. *(Cont'd)*

Measure†	Components	Items, no.	Formula	Range	Rem	Low DA	Moderate DA	High DA	Method of administration	MID, MCID
Patient Activity Scale-II (PAS-II)	HAQ-II (0–3); Pain VAS (0–10); Pt Global VAS (0–10)	3	(HAQ-IIx3.33 + Pain VAS + PtGA VAS)/3	0–10	≤0.25	>0.26 to 3.70	3.71 to <8.0	≥8.0	Patient items	-
Patient Based Disease Activity Score (PDAS1)	Pt Global; Pt 50TJC; HAQ; ESR	4	0.19xPtGA + 0.842ln(ESR+2) + 0.432xln(PtTJC +2) + 0.271xHAQ	-	<3.5	3.5-4.5	4.5-4.8	>4.8	Patient items, lab	0.8 good response
Patient Based Disease Activity Score (PDAS2)	Pt Global; Pt 28SJC; HAQ; morning stiffness duration	4	2.667 + 0.021xPtGA + 0.483xHAQ + 0.033xPtSJC + 0.002xAM stiffness	-	<3.8	3.8–4.6	4.6–5.0	>5.0	Patient items	1.2 good response
Patient Reported Clinical Arthritis Activity (PRO-CLARA)	Recent Onset Arthritis Disability questionnaire (ROAD 0–10); Pt 16TJC (0–10); Pt Global (0–10)	3	ROAD + PtTJC + PtGA/3	0–10	-	-	-	-	Patient items	-
Rheumatoid Arthritis Disease Activity Index (RADAI)	Pt Global (0–10); Current swollen/tender joints (0–10); Pain (0–10); Duration morning stiffness (0–10 transformed); Tender joint list: (0–10 transformed)	5	(PtGA + swollen/tender + pain + AM stiffness + TJC)/5	0–10	-	<2.2	≥2.2 to ≤4.9	>4.9	Patient items	1-1.4
Rheumatoid Arthritis Disease Activity Index 5 (RADAI-5)	Pt Global 6 months (0–10); Pt active joint swollen/tender today (0–10); Pain (0–10); Pt general health (0–10); AM stiffness (0–10)	5	(PtGA + Pt swollen/ tender joints + Pain + Pt GH + AM stiffness) / 5	0–10	≤1.4	1.6–3.0	3.2–5.4	≥5.6	Patient items	-
Rheumatoid Arthritis MRI Scoring (RAMRIS)	Synovitis (7 areas scored 0–3); Osteitis/bone edema (23 areas scored 0–3), Erosion (23 areas scored 0–10)	3	Subcomponents or total score (sum)	-	-	-	-	-	Imaging modality	-
Routine Assessment of Patient Index Data 3 (RAPID3)	MDHAQ (0–10); Pain VAS (0–10); Pt Global VAS (0–10)	3	MDHAQ + Pain VAS + PtGA VAS	0–30	≤3	4–6	7–12	≥13	Patient items	MID 3.2-3.6
Routine Assessment of Patient Index Data 4 (RAPID4)	MDHAQ (0–10); Pain VAS (0–10); Pt Global VAS (0–10); RADAI Tender Joint List (0–10)	4	MDHAQ + Pain + Pt Global + RADAI-tender joint list	0–40	≤4	5–8	9–16	≥17	Patient items	-
Routine Assessment of Patient Index Data 5 (RAPID5)	MDHAQ (0–10); Pain VAS (0–10); Pt Global VAS (0–10); RADAI Tender Joint List (0–10); Pr Global (0–10)	5	MDHAQ + Pain + Pt Global + RADAI-Tender Joint List + Pr Global	0–50	≤5	6-10	11–20	≥21	Patient items, provider assessment	-
Simplified Disease Activity Index (SDAI)	28TJC (0–28); 28SJC (0–28); Pt Global VAS (0–10); Pr Global VAS (0–10); CRP (0–10)	5	28SJC + 28TJC + PtGA + PrGA + CRP	0–86	≤3.3	>3.3 to ≤11.0	>11.0 to ≤26	>26	Patient item, provider assessment, lab	16 ~ DAS MID1.2; MCID 13

(Continued)

Table 1. *(Cont'd)*

Measure†	Components	Items, no.	Formula	Range	Rem	Low DA	Moderate DA	High DA	Method of administration	MID, MCID
Modified SDAI (Baker)	28SJC; Pr Global; CRP	3	CRP + 28SJC + PrGA	-	-	-	-	-	Provider assessment, lab	-
Patient Derived SDAI	Pt 28TJC; Pt 28SJC; Pt Global VAS; Pr Global VAS; CRP	5	28SJC + 28TJC + PtGA + PrGA + CRP	0–86	≤3.3	>3.3 to ≤11.0	>11.0 to ≤26	>26	Patient items, provider assessment, lab	-
Ultrasound Derived SDAI	28TJC; US 28SJC; Pt Global VAS; Pr Global VAS; CRP	5	28SJC + 28TJC + PtGA + PrGA + CRP	0–86	≤3.3	>3.3 to ≤11.0	>11.0 to ≤26	>26	Patient item, provider assessment, lab, imaging modality	-
Simplified RA MRI Score (SAMIS)	Synovitis (7 areas scored 0–2); Osteitis/bone edema (15 areas scored 0–1); Erosion (15 areas scored 0–10)	3	Subcomponents or total score (sum)	-	-	-	-	-	Imaging modality	-
Swiss Sonography in Arthritis and Rheumatism Score (SONAR)	Bilateral elbow and wrist; MCP2–5, PIP2–5; Knee; B mode (0–3) and PD (0–3) for each joint	22	Sum of individual scores (PD and B mode)	0–66 B-mode, 0–66 PD	-	-	-	-	Imaging modality	-
Ultrasound 6 Joint (Perricone)	Bilateral wrist; MCP2; Knee; synovial effusion (0–3); Synovial proliferation (0–3); PD (0–3)	6	Sum of individual scores	0–54	-	-	-	-	Imaging modality	-
Ultrasound 6 Joint (Rosa)	Bilateral wrist; MCP2; MCP3 (0–2)	6	Sum of individual scores	0–12	-	-	-	-	Imaging modality	-
Ultrasound 6 Joint (Kawashiri)	Bilateral wrist; MCP2; MCP3 (0–3)	6	Sum of individual scores	0–18	-	-	-	-	Imaging modality	-
Ultrasound 7 Joint (Backhaus)	Unilateral (dominant side) wrist; MCP2; MCP3; PIP2; PIP3; MTP2; MTP5; Synovitis PDUS (0–3); synovitis GSUS (0–3); tenosynovitis GSUS (0–1); tenosynovitis PDUS (0–3); erosion.	7	Sum of individual scores	S-GSUS 0–27, S-PDUS 0–39, TS-GSUS 0–7, TS-PDUS 0–21, E 0–14	-	-	-	-	Imaging modality	-
Ultrasound 8 Joint (Yoshimi)	Bilateral wrist; MCP2; MCP3; knees (0–3 PDUS)	8	Sum of individual scores	0–24	-	-	-	-	Imaging modality	-
Ultrasound 12 Joint (Naredo)	Bilateral elbow; wrist; MCP2; MCP3; knee, ankle (synovitis 0–3; PD 0–3)	12	Sum of individual scores, *alternative: sum of # joints	-	-	-	-	-	Imaging modality	-
Ultrasound 14 Joint (Dale)	Bilateral wrist; MCP2; MCP3; PIP2; PIP3; MTP; MTP5; GSUS 0–3; PDUS 0–3	14	-	-	-	-	-	-	Imaging modality	-
Ultrasound 20 Joint (Dougados)	Bilateral MCP1–5; MTP1–5	20	0–3 each joint for B-mode and PD	-	-	-	-	-	Imaging modality	-

(Continued)

Table 1. *(Cont'd)*

Measure†	Components	Items, no.	Formula	Range	Rem	Low DA	Moderate DA	High DA	Method of administration	MID, MCID
Ultrasound 28 Joint (Dougados)	Bilateral shoulders, elbows, wrists; MCPs; PIPs; knees	28	0–3 each joint for B-mode and PD	-	-	-	-	-	Imaging modality	-
Ultrasound 38 Joint (Dougados)	Bilateral shoulders, elbows, wrists; MCPs; PIPs; knees; bilateral MTPs	38	0–3 each joint for B-mode and PD	-	-	-	-	-	Imaging modality	-
Ultrasound 78 Joint (Hammer)	B-mode and PD: bilateral PIP1–5, MCP1–5, CMC1–5, wrist (radiocarpal, intercarpal, radioulnar), elbow (anterior, posterior), shoulder (GHu, AC), hip, knee, ankle, foot (talonavicular, subtalar, calcaneocuboid, cuneonavicular), TMT1–5, MTP1–5, First IP	78	0–3 each joint for B-mode and PD	B-Mode (0–234); PD (0–234)	-	-	-	-	Imaging modality	-
Ultrasound Score A, B (Aga)	Bilateral GSUS and PDUS of (A): MCP1; MCP2; PIP3; radiocarpal; elbow; MTP1; MTP2; TP tendon and ECU tendon. (B): A joints + MCP5; MTP5. Each location graded 0–3.	A 18; B 22	Sum of individual scores	A 0–54; B 0–66	-	-	-	-	Imaging modality	-

* Rem = remission; DA = disease activity; MID = minimum important difference; MCID = minimum clinically important difference; TJC = tender joint count; SJC = swollen joint count; Pt = patient; VAS = visual analog score; Pr = provider; GA = global assessment; RAI = Ritchie Articular Index; ESR = erythrocyte sedimentation rate; sqrt = square root; ln = natural logarithm; CRP = C-reactive protein level; mHAQ = modified Health Assessment Questionnaire; MSUS = musculoskeletal ultrasound; US = ultrasound; GS = gray scale; PD = power Doppler; HAQ = Health Assessment Questionnaire; EGF = epidermal growth factor; IL-6 = interleukin-6; MMP-1 = matrix metalloproteinase 1; SAA = serum amyloid A; TNFRI = tumor necrosis factor receptor type I; VCAM-1 = vascular cell adhesion molecule 1; VEGF-A = vascular endothelial growth factor A; PIP = proximal interphalangeal joint; MCP = metacarpophalangeal joint; MRI = magnetic resonance imaging; MDHAQ = Multidimensional HAQ; MTP = metatarsophalangeal joint; CMC = carpometacarpal joint; GHu = glenohumeral joint; AC = acromioclavicular joint; TMT = tarsometatarsal joint; TP: = tibialis posterior; ECU = extensor carpi ulnaris.
† Study references are listed in Supplementary Appendix 8, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24042/abstract.

excluded if >80% of ratings were in the 1 to 3 range, following best practices (9). The voting process continued iteratively to a maximum of 3 voting cycles, with discussion of RA disease activity measures not fulfilling agreement held between voting cycles. Measures not achieving recommendation for inclusion or exclusion were deemed inconclusive. Measures deemed inconclusive remained on the list fulfilling the minimum standard.

The ACR Quality Measures Subcommittee reviewed these recommendations in parallel with the recommendations on functional status assessment, modifying as necessary based upon the goal of identifying preferred tools for regular use in most clinic settings, before voting. The ACR Quality of Care Committee and ACR Board of Directors reviewed and approved this article prior to publication.

RESULTS

Systematic literature review and identified disease activity measures. Our systematic literature review identified 5,199 articles (see Supplementary Appendix 3, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24042/abstract). After screening titles, abstracts, and full texts, 104 articles met criteria for inclusion in the study. A review of the retrieved publications identified an additional 6 articles fulfilling eligibility criteria, resulting in a total of 110 included studies. There was 98.2% agreement between the reviewers for study inclusion.

Characteristics of the individual studies are provided in Supplementary Appendix 4, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24042/abstract. The majority of studies had predominantly female participants, with a mean age in the 6th decade. Sample sizes, mean DAS28 score, location, design, and selection varied between studies.

Our search identified 47 RA disease activity measures. The components, number of items, scoring method, score range, disease activity category cutoffs, method of administration, and minimum important difference/minimum clinically important difference of each RA disease activity measure are

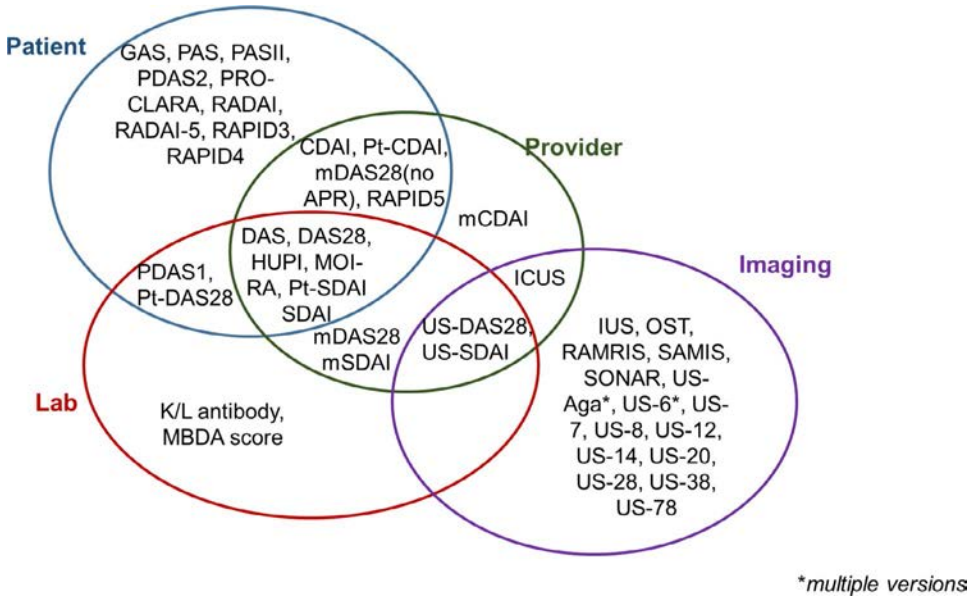

Figure 1. Venn diagram depicting the major domains of data (patient reported, provider assessment, laboratory, and imaging) included in rheumatoid arthritis (RA) disease activity measures, which are listed in the areas from which they are derived. GAS = Global Arthritis Score; PAS = Patient Activity Scale; PAS-II = Patient Activity Scale-II; PDAS2 = Patient Based Disease Activity Score 2; PRO-CLARA = Patient Reported Clinical Arthritis Activity; RADAI = RA Disease Activity Index; RADAI-5 = RA Disease Activity Index-5; RAPID3 = Routine Assessment of Patient Index Data 3; RAPID4 = Routine Assessment of Patient Index Data 4; CDAI = Clinical Disease Activity Index; Pt-CDAI = Patient Derived Clinical Disease Activity Index; mDAS28 = Modified Disease Activity Score in 28 joints; (no APR) = mDAS28 no acute-phase reactants; RAPID5 = Routine Assessment of Patient Index Data 5; mCDAI = Modified Clinical Disease Activity Index; DAS = Disease Activity Score; HUPI = Hospital Universitario La Princesa Index; MOI-RA = Mean Overall Index for RA; PDAS1 = Patient Based Disease Activity Score 1; Pt-DAS28 = Patient Derived DAS 28-Joint DAS; SDAI = Simplified Disease Activity Index; Pt-SDAI = Patient Derived SDAI; mSDAI = Modified SDAI; US-DAS28 = ultrasound-derived DAS28; US-SDAI = ultrasound-derived SDAI; ICUS = Individualized Composite Ultrasound Score; IUS = Individualized Ultrasound Score; OST = Optical Spectral Transmission; RAMRIS = RA Magnetic Resonance Imaging Scoring; SAMIS = Simplified RA Magnetic Resonance Imaging Score; SONAR = Swiss Sonography in Arthritis and Rheumatism Score; US-Aga = ultrasound sound score A & B proposed by Aga et al; US-6 = ultrasound 6 joint; US-7 = ultrasound 7 joint; US-8 = ultrasound 8 joint; US-12 = ultrasound 12 joint; US-14 = ultrasound 14 joint; US-20 = ultrasound 20 joint; US-28 = ultrasound 28 joint; US-38 = ultrasound 38 joint; US-78 = ultrasound 78 joint; K/L antibody = kappa/lambda hybrid antibody; MBDA score = Multibiomarker Disease Activity score.

Table 2. Level of evidence relevant to the psychometric properties of RA disease activity measures*

Measure†	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypotheses testing	Responsiveness
Clinical Disease Activity Index (CDAI)	+	++	+		++	+++	+++
Modified CDAI (Baker)						++	
Patient Derived CDAI		+	+			+	
Disease Activity Score (DAS)						++	
Disease Activity Score 28 Joints (DAS28)	++	++	++	- - -	++	+++	+++
Modified DAS28 (Baker)						++	
Modified DAS28 (Bentley)	?					++	++
Patient Derived DAS28		++	+			++	+
Ultrasound Derived DAS28		+				++	+
Global Arthritis Score (GAS)						++	++
Hospital Universitario La Princesa Index (HUPI)	?					++	++
Individualized Ultrasound Score							?
Individualized Composite Ultrasound Score							?
Kappa/Lambda Hybrid Antibody						+	
Mean Overall Index for RA (MOI-RA)						+	+
Multi-Biomarker Disease Activity Score (MBDA)		?		+++	+++	++	++
Optical Spectral Transmission (OST)						++	
Patient Activity Scale (PAS)						+	+
Patient Activity Scale-II (PAS-II)							
Patient Based Disease Activity Score (PDAS1)		+				+	
Patient Based Disease Activity Score (PDAS2)		+				+	
Patient Reported Clinical Arthritis Activity (PRO-CLARA)	+			?		++	+
Rheumatoid Arthritis Disease Activity Index (RADAI)		?	?			++	++
Rheumatoid Arthritis Disease Activity Index 5 (RADAI-5)	?				++	++	
Rheumatoid Arthritis MRI Scoring (RAMRIS)						- -	- -
Routine Assessment of Patient Index Data 3 (RAPID3)	?	?	?	+++		+++	+++
Routine Assessment of Patient Index Data 4 (RAPID4)				++		+	
Routine Assessment of Patient Index Data 5 (RAPID5)						++	++
Simplified Disease Activity Index (SDAI)	+	++	+			+++	+++
Modified SDAI (Baker)						++	
Patient Derived SDAI		+	+			++	
Ultrasound Derived SDAI		+				++	+
Simplified RA MRI Score (SAMIS)		+				?	
Swiss Sonography in Arthritis and Rheumatism (SONAR) Score		?	?			++	++
Ultrasound 6 Joint (Perricone)				+++		+	+
Ultrasound 6 Joint (Rosa)						+	
Ultrasound 6 Joint (Kawashiri)						?	
Ultrasound 7 Joint (Backhaus)		++		+++		++	++
Ultrasound 8 Joint (Yoshimi)				+++		++	
Ultrasound 12 Joint (Naredo)		+		+++		+	+
Ultrasound 14 Joint (Dale)						?	
Ultrasound 20 Joint (Dougados)	?	+					++
Ultrasound 28 Joint (Dougados)	?	+					++
Ultrasound 38 Joint (Dougados)	?	+					++
Ultrasound 78 Joint (Hammer)		?				?	?
Ultrasound Score A, B (Aga)				+++	+		+

* Grading scale: +++ or - - - = strong (consistent findings in multiple studies of good methodologic quality OR in one study of excellent methodologic quality); ++ or - - = moderate (consistent findings in multiple studies of fair methodologic quality OR in one study of good methodologic quality); + or - = limited (one study of fair methodologic quality); ± = conflicting (conflicting findings); ? = unknown (studies only of poor methodologic quality). RA = rheumatoid arthritis; MRI = magnetic resonance imaging.
† Study references are listed in Supplementary Appendix 8, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24042/abstract.

listed in Table 1. A Venn diagram illustrating the components (e.g., patient reported, provider assessment, laboratory values, and imaging modalities) of the identified RA disease activity measures is shown in Figure 1.

Properties of RA disease activity measures. The individual performance of RA disease activity measures in each study is provided in Supplementary Appendix 5 available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24042/abstract. The study quality assessment using the COSMIN checklist with 4-point scale is provided in Supplementary Appendix 6, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24042/abstract. Based on both the measure performance and study quality, an overall level of evidence was generated for each psychometric property for each RA disease activity measure (Table 2). This process was completed in duplicate with 96.6% agreement between raters in assessing the overall level of evidence for RA disease activity measures.

Hypothesis testing (testing hypotheses regarding relationships to other instruments measuring similar constructs, i.e., content validity) was the most frequently assessed psychometric property. Reliability and responsiveness were also frequently assessed for RA disease activity measures. The CDAI, DAS28, Multibiomarker Disease Activity (MBDA) score, RAPID3, and SDAI were the most frequently studied RA disease activity measures. Although negative content validity was reported for the DAS28, it should be noted this was based on one study of excellent quality that showed underestimation of radiographic progression in the feet, i.e., joints not included in the 28-joint count (10).

Properties of RA disease activity measures from before the current search period were collected from the prior review (3) and from hand searches for measures not previously included (see Supplementary Appendix 7, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24042/abstract). A full reference list of all articles identified and abstracted in the systematic literature review, as well as searches for earlier time periods, is shown in Supplementary Appendix 8, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24042/abstract.

Feasibility of RA disease activity measures. Feasibility scoring of the RA disease activity measures is shown in Table 3. Twenty-five measures were scored to be feasible for regular use in most clinics. Of these measures, 11 (44%) received a score of 4 (++++), 6 (25%) a score of 3 (+++), 5 (20%) a score of 2 (++), and 3 (12%) a score of 1 (+).

Recommended RA disease activity measures. Eleven measures fulfilled the minimum standard defined for RA disease activity measures for regular use (Table 4). Four measures (the CDAI, DAS28 using the erythrocyte sedimentation rate or C-reactive protein level [DAS28-ESR/CRP], RAPID3, and SDAI) were part of the prior ACR RA disease activity measure recommendations (2). Of the 7 measures not listed in the original recommendations, the Disease Activity Score (DAS) was a predecessor to the DAS28, the patient-derived DAS28 was derived from the DAS28, and the Routine Assessment of Patient Index Data 5 (RAPID5) was related to the RAPID3. The remaining measures were the Hospital Universitario La Princesa Index (HUPI), MBDA score, Rheumatoid Arthritis Disease Activity Index (RADAI), and RADAI-5. Of the 36 measures not fulfilling the minimum standard, 27 (75%) did not categorize into disease activity states, 28 (78%) did not have adequate psychometrics, and 22 (61%) were not scored as feasible for regular use (Table 4).

Results of the modified Delphi voting process are shown in Table 5. Four measures achieved consensus for preferred use: the CDAI, DAS28, RAPID3, and SDAI. The CDAI (mean score 8.8) and SDAI (mean score 7.6) achieved consensus during the first round of voting, the RAPID3 (mean score 7.6) during the second round of voting, and the DAS28 (mean score 7.6) during the third round of voting. The remaining 7 RA disease activity measures (mean score range 2.6–5.6) did not achieve consensus after the third round of voting and were deemed "inconclusive" for preferred use.

The ACR Quality Measures Subcommittee approved the previously mentioned recommendations with a single modification, which was the additional recommendation of PAS-II. This recommendation was based upon PAS-II feasibility, current use, strength of its inclusion in prior ACR recommendations that included evidence not captured in this current work, and alignment with the concurrent functional status assessment project (2).

DISCUSSION

Patient outcomes in RA, including physical function, quality of life, and achieving remission/low disease activity, have improved as a result of treatment advances, including the early initiation of treatment, treating to target, and novel therapeutics (11,12). Critical to adhering to a treat-to-target approach is the regular integration of disease activity measurement as part of routine care, a practice included in ACR RA treatment guidelines (1) and selected as a quality measure by the Centers for Medicare and Medicaid Services (Quality ID #177: Rheumatoid Arthritis: Periodic Assessment of Disease Activity). In this study, we have updated the initial ACR 2012 recommendations for RA disease activity measures (2) through an updated systematic literature review, RA disease activity measure performance assessment, study quality assessment, level of evidence synthesis, and a modified Delphi voting process. Five preferred RA disease activity measures for regular clinical use were selected: the CDAI, DAS28-ESR/CRP, PAS-II, RAPID3,

Table 3. Feasibility of RA disease activity measures*

Measure†	Items, no.	Time	Provider joint count	Lab testing required	Advanced imaging	Feasibility‡
Clinical Disease Activity Index (CDAI)	3	2–5 mins	Yes	No	No	+++
Modified CDAI (Baker)	2	5 mins	Yes	No	No	+++
Patient Derived CDAI	4	5 mins	No	No	No	++++
Disease Activity Score (DAS)	4	10 mins	Yes	Yes	No	+
Disease Activity Score 28 Joints (DAS28-ESR/CRP)	3 or 4	5 mins + lab	Yes	Yes	No	++
Modified DAS28 (Baker)	3	5 mins + lab	Yes	Yes	No	++
Modified DAS28 (no acute-phase reactants, Bentley)	6	5 mins	Yes	No	No	+++
Patient Derived DAS28	4	5 mins + lab	No	Yes	No	+++
Ultrasound Derived DAS28	4	N/R	No	Yes	Yes	-
Global Arthritis Score (GAS)	3	5 mins	No	No	No	++++
Hospital Universitario La Princesa Index (HUPI)	4	5 mins + lab	Yes	Yes	No	++
Individualized Ultrasound Score	Up to 7 or 12	N/R	No	No	Yes	-
Individualized Composite Ultrasound Score	Up to 7 or 12	N/R	No	No	Yes	-
Kappa/Lambda Hybrid Antibody	1	Not commercially available	No	Yes	No	-
Mean Overall Index for RA (MOI-RA)	7	10–20 mins + lab	Yes	Yes	No	+
Multi-Biomarker Disease Activity Score (MBDA, VECTRA)	12	Days	No	Yes	No	+
Optical Spectral Transmission (OST)	22	Not commercially available	No	No	Yes	-
Patient Activity Scale (PAS)	3	5 mins	No	No	No	++++
Patient Activity Scale-II (PAS-II)	3	2 mins	No	No	No	++++
Patient Based Disease Activity Score (PDAS1)	4	5–10 mins + lab	No	Yes	No	+++
Patient Based Disease Activity Score (PDAS2)	4	5–10 mins	No	No	No	++++
Patient Reported Clinical Arthritis Activity (PRO-CLARA)	3	5 mins	No	No	No	++++
Rheumatoid Arthritis Disease Activity Index (RADAI)	5	5 mins	No	No	No	++++
Rheumatoid Arthritis Disease Activity Index 5 (RADAI-5)	5	30 sec to 2 mins	No	No	No	++++
Rheumatoid Arthritis MRI Scoring (RAMRIS)	3	N/R	No	No	Yes	-
Routine Assessment of Patient Index Data 3 (RAPID3)	3	30 sec to 2 mins	No	No	No	++++
Routine Assessment of Patient Index Data 4 (RAPID4)	4	5–10 mins	No	No	No	++++
Routine Assessment of Patient Index Data 5 (RAPID5)	5	5–10 mins	No	No	No	++++
Simplified Disease Activity Index (SDAI)	5	2–5 mins + lab	Yes	Yes	No	++
Modified SDAI (Baker)	3	5 mins + lab	Yes	Yes	No	++
Patient Derived SDAI	5	5 mins + lab	No	Yes	No	+++
Ultrasound Derived SDAI	5	N/R	No	Yes	Yes	-
Simplified RA MRI Score (SAMIS)	3	N/R	No	No	Yes	-
Swiss Sonography in Arthritis and Rheumatism (SONAR) Score	22	20–30 mins	No	No	Yes	-
Ultrasound 6 Joint (Perricone)	6	14 mins	No	No	Yes	-
Ultrasound 6 Joint (Rosa)	6	5–12 mins	No	No	Yes	-
Ultrasound 6 Joint (Kawashiri)	6	N/R	No	No	Yes	-
Ultrasound 7 Joint (Backhaus)	7	10–20 mins	No	No	Yes	-
Ultrasound 8 Joint (Yoshimi)	8	N/R	No	No	Yes	-
Ultrasound 12 Joint (Naredo)	12	20–25	No	No	Yes	-
Ultrasound 14 Joint (Dale)	14	N/R	No	No	Yes	-
Ultrasound 20 Joint (Dougados)	20	N/R	No	No	Yes	-
Ultrasound 28 Joint (Dougados)	28	N/R	No	No	Yes	-
Ultrasound 38 Joint (Dougados)	38	N/R	No	No	Yes	-
Ultrasound 78 Joint (Hammer)	78	N/R	No	No	Yes	-
Ultrasound Score A, B (Aga)	A =18, B = 22	N/R	No	No	Yes	-

* RA = rheumatoid arthritis; Lab = laboratory; mins = minutes; N/R = not reported; sec = seconds; MRI = magnetic resonance imaging.
† Study references are listed in Supplementary Appendix 8, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24042/abstract.
‡ Feasibility was assessed by the number of items, time to complete, and the need for provider joint counts, laboratory testing, and advanced imaging. Feasibility was graded - to ++++ with + to ++++ meeting minimum feasibility for regular use. Scoring was as follows: measures started with a score of ++++; any measure not commercially available or requiring advanced imaging was graded -; requiring a provider joint count reduced feasibility by +; requiring a laboratory test reduced feasibility by +; number of items and time to completion were considered and score was reduced by + if not feasibile in a routine clinic visit or by ++ if not feasible on the same day as the clinic visit.

Table 4. RA disease activity measures assessment of minimum standard for regular use*

Measure†	Numeric	Categorizes 3–4 states	Feasible‡	Adequate psychometrics§	Meet minimum standard
Fulfilled minimum standard					
Clinical Disease Activity Index (CDAI)	+	+	+	+	+
Disease Activity Score (DAS)	+	+	+	+	+
Disease Activity Score 28 Joints (DAS28-ESR/CRP)	+	+	+	+	+
Patient Derived DAS28	+	+	+	+	+
Hospital Universitario La Princesa Index (HUPI)	+	+	+	+	+
Multi-Biomarker Disease Activity Score (MBDA score, VECTRA DA)	+	+	+	+	+
Rheumatoid Arthritis Disease Activity Index (RADAI)	+	+	+	+	+
Rheumatoid Arthritis Disease Activity Index 5 (RADAI-5)	+	+	+	+	+
Routine Assessment of Patient Index Data 3 (RAPID3)	+	+	+	+	+
Routine Assessment of Patient Index Data 5 (RAPID5)	+	+	+	+	+
Simplified Disease Activity Index (SDAI)	+	+	+	+	+
Did not fulfill minimum standard					
Modified CDAI (Baker)	+	-	+	-	-
Patient Derived CDAI	+	+	+	-	-
Modified DAS28 (Baker)	+	-	+	-	-
Modified DAS28 (Bentley)	+	-	+	+	-
Ultrasound Derived DAS28	+	+	-	+	-
Global Arthritis Score (GAS)	+	-	+	+	-
Individualized Ultrasound Score	+	-	-	-	-
Individualized Composite Ultrasound Score	+	-	-	-	-
Kappa/Lambda Hybrid Antibody	+	-	-	-	-
Mean Overall Index for RA (MOI-RA)	+	-	+	-	-
Optical Spectral Transmission (OST)	+	-	-	-	-
Patient Activity Scale (PAS)	+	+	+	-	-
Patient Activity Scale-II (PAS-II)	+	+	+	-	-
Patient Based Disease Activity Score (PDAS1)	+	+	+	-	-
Patient Based Disease Activity Score (PDAS2)	+	+	+	-	-
Patient Reported Clinical Arthritis Activity (PRO-CLARA)	+	-	+	+	-
Rheumatoid Arthritis MRI Scoring (RAMRIS)	+	-	-	-	-
Routine Assessment of Patient Index Data 4 (RAPID4)	+	+	+	-	-
Modified SDAI (Baker)	+	-	+	-	-
Patient Derived SDAI	+	+	+	-	-
Ultrasound Derived SDAI	+	+	-	+	-
Simplified RA MRI Score (SAMIS)	+	-	-	-	-
Swiss Sonography in Arthritis and Rheumatism (SONAR) Score	+	-	-	+	-
Ultrasound 6 Joint (Perricone)	+	-	-	-	-
Ultrasound 6 Joint (Rosa)	+	-	-	-	-
Ultrasound 6 Joint (Kawashiri)	+	-	-	-	-
Ultrasound 7 Joint (Backhaus)	+	-	-	+	-
Ultrasound 8 Joint (Yoshimi)	+	-	-	+	-
Ultrasound 12 Joint (Naredo)	+	-	-	-	-
Ultrasound 14 Joint (Dale)	+	-	-	-	-
Ultrasound 20 Joint (Dougados)	+	-	-	-	-
Ultrasound 28 Joint (Dougados)	+	-	-	-	-
Ultrasound 38 Joint (Dougados)	+	-	-	-	-
Ultrasound 78 Joint (Hammer)	+	-	-	-	-
Ultrasound Score A, B (Aga)	+	-	-	-	-

* RA = rheumatoid arthritis; ESR = erythrocyte sedimentation rate; CRP = C-reactive protein level; MRI = magnetic resonance imaging.
† Study references are listed in Supplementary Appendix 8, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.24042/abstract.
‡ Measures deemed feasible if feasibility scoring was ≥1 as shown in Table 3.
§ Measures were considered to have adequate psychometrics if the level of evidence suggested at least moderate positive results in the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) area of hypothesis testing plus had ≥1 of the following: level of evidence suggesting at least moderate positive results in another COSMIN area, level of evidence suggesting at least limited positive results in ≥2 COSMIN areas (one of which must be responsiveness), or a defined minimum important difference/minimum clinically important difference.

and SDAI. Seven additional RA disease activity measures that met a minimum standard for regular use were identified: the DAS, patient-derived DAS28, HUPI, MBDA score, RADAI, RADAI-5, and RAPID5. Preferred measures represent those with the most support for their performance and feasibility as assessed by the working group, while those fulfilling the minimum standard have adequate performance and feasibility for regular use. Clinicians can utilize these recommendations when selecting an RA disease

Table 5. Summary of 3-round Delphi method with recommendations for rheumatoid arthritis disease activity measures*

Measure	Round 1		Round 2†		Round 3		Final recommendation‡
	Mean	Rating 1–3/4–6/7–9‡	Mean	No. 1–3/4–6/7–9‡	Mean	Rating 1–3/4–6/7–9‡	
Clinical Disease Activity Index (CDAI)	8.8	0/0/10	N/A	N/A	N/A	N/A	Recommended
Simplified Disease Activity Index (SDAI)	7.6	0/1/9	N/A	N/A	N/A	N/A	Recommended
Routine Assessment of Patient Index Data 3 (RAPID3)	7.4	0/3/7	7.6	0/1/7	N/A	N/A	Recommended
28-Joint Disease Activity Score (DAS28)	7.6	0/2/8	7.1	0/2/6	7.6	1/0/9	Recommended
Rheumatoid Arthritis Disease Activity Index-5 (RADAI-5)	6.1	4/2/4	5.3	2/4/2	5.6§	2/4/3§	Inconclusive
Disease Activity Score (DAS)	5.0	3/4/3	3.8	5/2/1	4.2	4/5/1	Inconclusive
Patient Derived-DAS28	4.9	4/2/4	4.5	2/6/0	4.2	4/6/0	Inconclusive
Rheumatoid Arthritis Disease Activity Index (RADAI)	5.1	4/3/3	4.2	5/2/1	4.4	4/5/1	Inconclusive
Routine Assessment of Patient Index Data 5 (RAPID5)	5.2	4/1/5	4.5	2/5/1	3.8§	5/3/1§	Inconclusive
Multibiomarker Disease Activity (MBDA) score	4.2	7/1/2	3.5	5/2/1	3.2§	7/1/1§	Inconclusive
Hospital Universitario La Princesa Index (HUPI)	4.0	6/1/3	3.5	5/3/0	2.6	8/2/0	Inconclusive

* N/A = not applicable; ESR = erythrocyte sedimentation rate; CRP = C-reactive protein.
† Eight voters participated in round 2 voting.
‡ Ratings were on a 1–9 Likert scale, where 1–3 = not recommended, 4–6 = sometimes recommended, 7–9 = essential to have, and >80% agreement is required for recommendation.
§ There was one missing vote for this score.

activity measure for integration into their care for RA patients, and any of the 11 measures shown in Table 4 that meet the minimum standard reasonably satisfy quality measures for assessing RA disease activity.

The purpose of these recommendations was to assist clinicians in the care of RA patients by identifying RA disease activity measures and evaluating their performance and feasibility for regular use. These recommendations are not meant to dictate the specific RA disease activity measure a clinician utilizes. The working group recognizes that feasibility varies based on practice and provider. Furthermore, providers may have experience with and be comfortable with specific RA disease activity measures. Therefore, we aimed to identify not only preferred RA disease activity measures, but also RA disease activity measures that met a minimum standard by categorizing into disease activity states, possessing adequate psychometric properties, and being feasible for regular clinical use. For providers adopting an RA disease activity measure or aiming to integrate disease activity measurement into care through a standardized fashion (i.e., integration into the electronic health record), we recommend selecting a preferred RA disease activity measure (CDAI, DAS28-ESR/CRP, PAS-II, RAPID3, or SDAI).

In addition to not precluding the use of other RA disease activity measures, these recommendations importantly do not provide recommendations on disease activity measures in special circumstances. An example might include the use of musculoskeletal ultrasound or magnetic resonance imaging in a patient with a difficult or equivocal joint examination who is being considered for treatment escalation or withdrawal. There are certainly specific circumstances or patient populations where alternative disease activity assessments may be clinically indicated. Additionally, there are certain RA subpopulations where the validity of RA disease activity measures may vary. Disease activity scores including patient-reported measures are higher in patients with comorbid fibromyalgia (13), and disease activity scores including inflammatory markers are higher in obese patients (14). Providing recommendations for disease activity assessment in these specific situations or patient populations was beyond the scope of these recommendations and are left to the judgement of the treating clinician.

The preferred RA disease activity measures are largely unchanged from those previously recommended (2), with the difference being that the PAS was not recommended for preferred use in these updated recommendations. Both the PAS and PAS-II were infrequently studied since the time of the prior recommendations and subsequently did not satisfy the requirement of having demonstrated adequate psychometrics during this period. It is important to note that the PAS and PAS-II differ from the RAPID3 only by the functional status component of each composite measure. The PAS-II contains the Health Assessment Questionnaire II (HAQ-II) (15), while PAS contains the HAQ (16) and RAPID3 contains the Multidimensional Health Assessment Questionnaire (MDHAQ) (17). Assessment and recommendation of functional status measures in RA has been conducted in parallel, with recommendations for the use of Patient-Reported Outcomes Measurement Information System Physical Function 10, MDHAQ, and HAQ-II. Given the overlap between PAS-II and RAPID3 as well as the results from the parallel functional status assessment project, the Quality Measures Subcommittee additionally recom-

mended the PAS-II as a preferred measure. The consistency in the selection of preferred disease activity measures between the prior and current recommendations provides further support for these measures.

There are limitations to this effort. We conducted a systematic literature review from the time of the prior review. Therefore, generation of overall level of evidence from measure performance and study quality assessment was only able to be completed for studies since the initial review. Properties assessed early in measure development may not have been routinely re-assessed in later literature. Although not included into level of evidence, we synthesized data from the prior literature review as well as additional searches from before our current search period and provided these to working group members to inform the selection process. In contrast to the parallel functional status assessment recommendations, which were limited to patient-reported measures, we assessed RA disease activity measures with several different components: patient reported, provider assessment, laboratory, and imaging. The broad nature of these components makes selecting adequate measure performance and study quality assessment tools challenging. We selected the COSMIN checklist with 4-point scoring system to adapt for our study because it was designed to facilitate selection of health instruments in systematic reviews (18) and could be applied to both the RA disease activity and functional status assessment projects. While COSMIN was designed primarily for patient reported outcomes measures, it has been adapted beyond health-related patient-reported instruments (19,20). An updated COSMIN tool was developed after study inception that penalizes studies less for having smaller sample sizes and not reporting handling of missing data, which may affect the level of evidence grading (21). Finally, because there are no validated feasibility scoring systems for RA disease activity measures, we developed a scoring system to be used for this effort. Feasibility is inherently subjective based on varying viewpoints of different providers and practice types; therefore, we focused our feasibility scoring on identifying measures that could be regularly used by the majority of providers and practice types. As adoption of, and training in, the advanced imaging modalities continues to increase, the feasibility will need to be re-assessed in future efforts (22). While advanced imaging modalities were all deemed not feasible for regular use, all measures solely based on advanced imaging also did not fulfill the minimum standard by the absence of categorizing into 3 to 4 disease activity states.

There are several strengths to this effort. The working group was composed of content experts and practicing rheumatologists. The process and preliminary results were presented at the 2017 ACR Annual Scientific Meeting and underwent public comment. A systematic literature review with duplicate screening of articles for inclusion and standardized data abstraction was performed. Study quality was assessed using a standardized approach with a widely accepted tool and combined with the performance of RA disease activity measures to generate an overall level of evidence. A modified Delphi process was used to obtain final recommendations and incorporated the prior literature search as well as additional hand searches over the period before the current literature review.

In conclusion, we updated prior ACR recommendations for RA disease activity measures, providing recommendations for both measures that meet a minimum standard for regular use and preferred measures for regular use, specificially the CDAI, DAS28-ESR/CRP, PAS-II, RAPID3, and SDAI. These recommendations can assist clinicians with adhering to a treat-to-target approach for the management of RA but should not be interpreted as dictating the "proper" measure to be used in individual circumstances or clinical practices. As additional measures are developed and performance of measures is further characterized, these recommendations should again be evaluated.

ACKNOWLEDGMENTS

We thank Cynthia Schmidt, MD, MLS (University of Nebraska Medical Center McGoogan Library of Medicine) for her assistance performing the systematic literature review searches. We thank the American College of Rheumatology staff members Amy Turner and Regina Parker for their support and assistance through the recommendation process.

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. Michaud 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. England, Tiong, Curtis, Kazi, O'Dell, Limanni, Suter, Michaud.

Acquisition of data. England, Tiong.

Analysis and interpretation of data. England, Tiong, Bergman, Curtis, Mikuls, Ranganath, Suter, Michaud.

REFERENCES

1. Singh JA, Saag KG, Bridges SL Jr, Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. Arthritis Rheumatol 2016;68: 1–26.
2. 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.
3. Anderson JK, Zimmerman L, Caplan L, Michaud K. Measures of rheumatoid arthritis disease activity: Patient (PtGA) and Provider (PrGA) Global Assessment of Disease Activity, Disease Activity Score (DAS) and Disease Activity Score with 28-Joint Counts (DAS28), Simplified Disease Activity Index (SDAI), Clinical Disease Activity Index (CDAI), Patient Activity Score (PAS) and Patient Activity Score-II (PASII), Routine Assessment of Patient Index Data (RAPID), Rheumatoid Arthritis Disease Activity Index (RADAI) and Rheumatoid Arthritis Disease Activity Index-5 (RADAI-5), Chronic Arthritis

Systemic Index (CASI), Patient-Based Disease Activity Score With ESR (PDAS1) and Patient-Based Disease Activity Score without ESR (PDAS2), and Mean Overall Index for Rheumatoid Arthritis (MOI-RA). Arthritis Care Res (Hoboken) 2011;63 Suppl 11:S14–36.

4. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012;21:651–7.
5. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010;63:737–45.
6. Hendrikx J, de Jonge MJ, Fransen J, Kievit W, van Riel PL. Systematic review of patient-reported outcome measures (PROMs) for assessing disease activity in rheumatoid arthritis. RMD Open 2016;2:e000202.
7. Lee J, Kim SH, Moon SH, Lee EH. Measurement properties of rheumatoid arthritis-specific quality-of-life questionnaires: systematic review of the literature. Qual Life Res 2014;23:2779–91.
8. Dalkey N, Helmer O. An experimental application of the DELPHI method to the use of experts. Management Science 1963;9:458–67.
9. Diamond IR, Grant RC, Feldman BM, Pencharz PB, Ling SC, Moore AM, et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. J Clin Epidemiol 2014;67:401–9.
10. Bakker MF, Jacobs JW, Kruize AA, van der Veen MJ, van Booma-Frankfort C, Vreugdenhil SA, et al. Misclassification of disease activity when assessing individual patients with early rheumatoid arthritis using disease activity indices that do not include joints of feet. Ann Rheum Dis 2012;71:830–5.
11. Aga AB, Lie E, Uhlig T, Olsen IC, Wierod A, Kalstad S, et al. Time trends in disease activity, response and remission rates in rheumatoid arthritis during the past decade: results from the NOR-DMARD study 2000–2010. Ann Rheum Dis 2015;74:381–8.
12. Overman CL, Jurgens MS, Bossema ER, Jacobs JW, Bijlsma JW, Geenen R. Change of psychological distress and physical disability in patients with rheumatoid arthritis over the last two decades. Arthritis Care Res (Hoboken) 2014;66:671–8.
13. Ranzolin A, Brenol JC, Bredemeier M, Guarienti J, Rizzatti M, Feldman D, et al. Association of concomitant fibromyalgia with worse disease activity score in 28 joints, health assessment questionnaire, and short form 36 scores in patients with rheumatoid arthritis. Arthritis Rheum 2009;61:794–800.
14. George MD, Giles JT, Katz PP, England BR, Mikuls TR, Michaud K, et al. Impact of obesity and adiposity on inflammatory markers in patients with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2017;69:1789–98.
15. Wolfe F, Michaud K, Pincus T. Development and validation of the Health Assessment Questionnaire II: a revised version of the Health Assessment Questionnaire. Arthritis Rheum 2004;50:3296–305.
16. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137–45.
17. Pincus T, Sokka T, Kautiainen H. Further development of a physical function scale on a MDHAQ [corrected] for standard care of patients with rheumatic diseases. J Rheumatol 2005;32:1432–9.
18. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010;19:539–49.
19. Larsen CM, Juul-Kristensen B, Lund H, Sogaard K. Measurement properties of existing clinical assessment methods evaluating scapular positioning and function: a systematic review. Physiother Theory Pract 2014;30:453–82.
20. Vrijman C, Linthorst Homan MW, Limpens J, van der Veen W, Wolkerstorfer A, Terwee CB, et al. Measurement properties of outcome measures for vitiligo: a systematic review. Arch Dermatol 2012;148:1302–9.
21. Mokkink LB, de Vet HC, Prinsen CA, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Qual Life Res 2018;27:1171–9.
22. Torralba KD, Cannella AC, Kissin EY, Bolster MB, Salto LM, Higgs J, et al. Musculoskeletal ultrasound instruction in adult rheumatology fellowship programs. Arthritis Care Res (Hoboken) 2017. E-pub ahead of print.

Arthritis Care & Research
Vol. 71, No. 12, December 2019, pp 1556–1565
DOI 10.1002/acr.23799

American College of Rheumatology
Empowering Rheumatology Professionals

International Consortium for Health Outcome Measurement Set of Outcomes That Matter to People Living With Inflammatory Arthritis: Consensus From an International Working Group

Martijn A. H. Oude Voshaar,[1] Zofia Das Gupta,[2] Johannes W. J. Bijlsma,[3] Annelies Boonen,[4] Jeffrey Chau,[5] Delphine S. Courvoisier,[6] Jeffrey R. Curtis,[7] Benjamin Ellis,[8] Sofia Ernestam,[9] Laure Gossec,[10] Christine Hale,[11] Jennifer Hornjeff,[12] Katy Y. Y. Leung,[13] Merav Lidar,[14] Phillip Mease,[15] Kaleb Michaud,[16] Girish M. Mody,[17] Mwidimi Ndosi,[18] Christina H. Opava,[9] Geraldo R. C. Pinheiro,[19] Matthew Salt,[2] Enrique R. Soriano,[20] William J. Taylor,[21] Maria J. H. Voshaar,[1] Angelique E. A. M. Weel,[22] Maarten de Wit,[23] Nico Wulffraat,[24] Mart A. F. J. van de Laar,[25] and Harald E. Vonkeman[25]

Objective. The implementation of value-based health care in inflammatory arthritis requires a standardized set of modifiable outcomes and risk-adjustment variables that is feasible to implement worldwide.

Methods. The International Consortium for Health Outcomes Measurement (ICHOM) assembled a multidisciplinary working group that consisted of 24 experts from 6 continents, including 6 patient representatives, to develop a standard set of outcomes for inflammatory arthritis. The process followed a structured approach, using a modified Delphi process to reach consensus on the following decision areas: conditions covered by the set, outcome domains, outcome measures, and risk-adjustment variables. Consensus in areas 2 to 4 were supported by systematic literature reviews and consultation of experts.

Results. The ICHOM Inflammatory Arthritis Standard Set covers patients with rheumatoid arthritis (RA), axial spondyloarthritis, psoriatic arthritis, and juvenile idiopathic arthritis (JIA). We recommend that outcomes regarding pain, fatigue, activity limitations, overall physical and mental health impact, work/school/housework ability and productivity, disease activity, and serious adverse events be collected at least annually. Validated measures for patient-reported outcomes were endorsed and linked to common reporting metrics. Age, sex at birth, education level, smoking status, comorbidities, time since diagnosis, and rheumatoid factor and anti-citrullinated protein antibody lab testing for RA and JIA should be collected as risk-adjustment variables.

Conclusion. We present the ICHOM inflammatory arthritis Standard Set of outcomes, which enables health care providers to implement the value-based health care framework and compare outcomes that are important to patients with inflammatory arthritis.

Supported by Arthritis Research UK, the Karolinska Institutet, Maasstad Hospital, Santeon, Sheba Academic Medical Center Hospital, and Transparency in Healthcare.

[1]Martijn A. H. Oude Voshaar, PhD, Maria J. H. Voshaar, MSc: University of Twente, Enschede, The Netherlands; [2]Zofia Das Gupta, PhD, Matthew Salt, BSc: International Consortium for Health Outcomes Measurement, London, UK; [3]Johannes W. J. Bijlsma, PhD, MD: University Medical Center Utrecht, Utrecht, The Netherlands; [4]Annelies Boonen, PhD, MD: Maastricht University Medical Centre, Maastricht, The Netherlands; [5]Jeffrey Chau, BA: Hong Kong Psoriatic Arthritis Association, Hong Kong, China; [6]Delphine S. Courvoisier, PhD: University Hospitals of Geneva and University of Geneva, Geneva, Switzerland; [7]Jeffrey R. Curtis, PhD, MD: University of Alabama at Birmingham; [8]Benjamin Ellis, MSc, MD: Imperial College Healthcare NHS Trust, London, UK; [9]Sofia Ernestam, PhD, MD, Christina H. Opava, PhD: Karolinska Institutet, Stockholm, Sweden; [10]Laure Gossec, PhD, MD: Sorbonne Université and Pitié Salpêtrière Hospital, AP-HP, Paris, France; [11]Christine Hale, MD: Lockton Dunning Benefits, Dallas, Texas; [12]Jennifer Hornjeff, PhD: Columbia University Medical Centre, New York, New York; [13]Katy Y. Y. Leung, PhD, MD: Singapore General Hospital, Duke-NUS Medical School, Singapore; [14]Merav Lidar, PhD, MD: Sheba Medical Centre, Tel-Hashomer, Israel; [15]Phillip Mease, PhD, MD: Providence St. Joseph Health System, University of Washington, Seattle; [16]Kaleb Michaud, PhD, MD: University of Nebraska Medical Center Omaha, and the National Databank for Rheumatic Diseases, Wichita, Kansas; [17]Girish M. Mody, PhD, MD: University of KwaZulu-Natal, Durban, South Africa; [18]Mwidimi Ndosi, PhD: University of the West of England, Bristol, UK; [19]Geraldo R. C. Pinheiro, PhD, MD: Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil; [20]Enrique R. Soriano, PhD, MD: Instituto Universitario Hospital Italiano de Buenos Aires, Buenos Aires, Argentina; [21]William J. Taylor, PhD, MD: University of Otago, Wellington, New Zealand; [22]Angelique E. A. M. Weel, PhD, MD: Maasstad Hospital, Rotterdam, The Netherlands; [23]Maarten de Wit, PhD: VU University Medical Centre, Amsterdam Public Health, Amsterdam, The Netherlands; [24]Nico Wulffraat, PhD, MD: Wilhelmina Children's Hospital, Utrecht, The Netherlands; [25]Mart A. F. J. van de Laar, PhD, MD, Harald E. Vonkeman, PhD, MD: University of Twente, Enschede, The Netherlands, and International Consortium for Health Outcomes Measurement, London, UK.

Dr. Bijlsma has received consulting fees from AbbVie, Bristol-Myers Squibb, MSD, Pfizer, Roche, SUN, and UCB (less than $10,000 each). Dr. Boonen

SIGNIFICANCE & INNOVATIONS

- Standards for measuring and comparing treatment outcomes that matter to patients with inflammatory arthritis that are globally implementable are currently lacking.
- We used a modified Delphi procedure and systematic reviews of the literature to develop a standard set of outcomes that matter to patient with inflammatory arthritis.
- The patient-reported outcome measures we recommend for measuring pain, activity limitations, fatigue and assessment of overall emotional and physical health impact were linked to a common item response theory–based common metric, so that users of the set can select their preferred instrument for measuring these outcomes.

INTRODUCTION

The inflammatory arthritides are a group of systemic, immune-mediated rheumatic conditions, characterized by synovitis or inflammation of periarticular tissues and joint damage. The lifetime risk of adult-onset inflammatory arthritis has been estimated to be ~6% of the US population (1). The availability of strategies to diagnose the diseases earlier, and the availability of biologic and targeted small molecule therapies, in combination with early, tightly controlled treatment strategies have led to relevant improvements in outcomes for many patients over the last decades (2,3). However, these improvements have also resulted in an increased financial burden on health care systems (4,5).

The prevalence and case recognition of inflammatory arthritis is expected to increase further over the next decade, particularly in less economically developed countries (6). Hence, it will be increasingly important to optimize care and allocate available resources efficiently to improve or maintain quality of care. Value Based Healthcare (VBHC) is a framework for improving the quality and efficiency of health care, in which improving value for the patient is the central concept (7). Value is defined as the patient outcomes achieved, relative to financial costs for obtaining those outcomes. Within this framework, value can be increased by improving patient outcomes or by delivering the same outcomes at a lower cost. Public reporting of patient outcomes by health care providers is proposed as a mechanism that will accelerate identification and adoption of high value care, through shared learning and promoting benchmarking of outcomes that matter to patients.

In order for outcomes to be comparable between different health care providers, exact definitions for each relevant outcome are required. The outcomes need to be feasible to be collected in a variety of health care systems, and a set of relevant risk-adjustment variables should be included to ensure risk-adjusted comparisons of outcomes between providers that serve different patient populations. The International Consortium for Health Outcomes Measurement (ICHOM) initiative is working toward the global implementation of VBHC by developing standard sets of patient outcomes for a range of medical conditions (8). These standards are intended to be implemented in routine clinical practice and therefore complement earlier core sets and reporting standards intended for clinical research, including the work of the Outcome Measures in Rheumatology group (9).

The ICHOM process is grounded in a conceptual framework which distinguishes 3 hierarchically ordered tiers of outcome, including health status achieved/retained, the process of recovery, and the sustainability of health (10). In order to select the most relevant outcomes, outcome measures, and risk adjustment variables for particular conditions, various stakeholders including patients, physicians, policymakers and outcome experts are engaged in a consensus-building process that is supported by a systematic evaluation of the available evidence base, including critical evaluation of available instruments and evidence supporting their measurement properties. In order to further encourage the adoption and implementation of VBHC in rheumatology, the

has received honoraria from UCB, Pfizer, Eli Lilly, and Novartis (less than $10,000 each) and research support from AbbVie and Friesland Campina. Dr. Courvoisier has received consulting fees from Bristol-Myers Squibb (less than $10,000). Dr. Curtis has received honoraria from Amgen, Bristol-Myers Squibb, Corrona, Crescendo, Janssen, Pfizer, AbbVie, UCB, and Celgene (less than $10,000 each) and research support from Amgen, Bristol-Myers Squibb, Corrona, Crescendo, Janssen, Pfizer, AbbVie, UCB, and Celgene. Dr. Gossec has received consulting fees, speaking fees, and/or honoraria from Abbott, AbbVie, Celegne, Chugai, Janssen, MSD, Novartis, Pfizer, Roche, and UCB (less than $10,000 each). Dr. Leung has received consulting fees, speaking fees, and/or honoraria from Pfizer and Boehringer (less than $10,000 each). Dr. Mease has received consulting fees, speaking fees, and/or honoraria from Bristol-Myers Squibb, UCB, Celegne, Genentech, SUN, Corrona (less than $10,000 each), AbbVie, Amgen, Janssen, Eli Lilly, Novartis, and Pfizer (more than $10,000 each) and research support from Bristol-Myers Squibb, UCB, Corrona, Abbvie, Amgen, Janssen, Novartis, and Pfizer. Dr. Michaud has received research support from Pfizer. Dr. Mody has received consulting fees and/or honoraria from GlaxoSmithKline and AstraZeneca (less than $10,000 each). Dr. Ndosi has received speaking fees and/or honoraria from Pfizer (less than $10,000) and research support from Bristol-Myers Squibb. Dr. Pinheiro has received consulting fees and/or honoraria from AbbVie, AstraZeneca, Bristol-Myers Squibb, Janssen, Pfizer, Roche, and Sanofi Aventis (less than $10,000 each). Dr. Soriano has received consulting fees, speaking fees, and/or honoraria from AbbVie, Bristol-Myers Squibb, Pfizer, Janssen, UCB, Roche, and Novartis (less than $10,000 each) and research support from AbbVie, Bristol-Myers Squibb, Pfizer, Janssen, UCB, Roche, and Novartis. Dr. Weel has received research support from AbbVie, Pfizer, and UCB. Dr. Wulffraat has received speaking fees from Novartis (less than $10,000) and research support from Pfizer, AbbVie, and Sanofi. Dr. van de Laar has received consulting fees, speaking fees, and/or honoraria from Sanofi Genzyme, Pfizer, MSD, Bristol-Myers Squibb, Janssen-Cilag, Eli Lilly, Grunenthal (less than $10,000 each) and research support from AbbVie, Pfizer, MSD, Janssen-Cilag, Eli Lilly, Grunenthal, and Sobi. Dr. Vonkeman has received consulting fees, speaking fees, and/or honoraria from Roche, AbbVie, Celltrion, Sanofi Genzyme, Novartis, Pfizer, MSD, Bristol-Myers Squibb, and Janssen-Cilag (less than $10,000 each) and research support from Roche. No other disclosures relevant to this article were reported.

Address correspondence to Martijn A. H. Oude Voshaar, PhD, Department of Psychology, Health and Technology, University of Twente, PO Box 50 000, 7500 KA Enschede, The Netherlands. E-mail: A.H.Oudevoshaar@utwente.nl.

Submitted for publication May 1, 2018; accepted in revised form October 16, 2018.

aim of our study was to develop a globally applicable set of outcome measures that reflect outcomes that matter to patients with inflammatory arthritis, for providers to track in their clinical practice.

PATIENTS AND METHODS

Working group. A working group of outcome experts (n = 24) was convened by ICHOM. Working group members were carefully selected to ensure representation of relevant professional disciplines, different geographic areas, and the patient perspective. The working group included patient representatives (n = 6), registry leaders, and members with a professional background in adult and pediatric rheumatology, nursing, epidemiology, psychology, rehabilitation medicine, physiotherapy, and psychometrics. Working group members of all 6 inhabited continents were included. The efforts of the working group were guided and facilitated by a core project team.

Working group process. A modified Delphi approach that has been developed by ICHOM and was previously applied by ICHOM to develop standards for a number of other conditions (9,11–26) was used. The process involved reaching consensus in 4 major decision areas: determination of which inflammatory arthritis conditions and treatments to include in the Standard Set, a minimally sufficient set of outcomes relevant for each of the conditions, standardized definitions and time points for assessing these outcomes, and standardized definitions for risk-adjustment variables to ensure fair comparisons between health care providers who wish to implement the set. Each decision area was discussed during a dedicated video conference. A list of potentially relevant items (i.e., conditions/domains/time points/risk-adjustment variables) to be included in the Standard Set, along with supporting evidence (see below), was prepared by the project team preceding each meeting. These items were identified in a series of systematic literature reviews and/or consultation with external experts on the topic under consideration. A summary of the preparatory work was provided to working group members preceding each video conference. During each meeting, the items were discussed and expanded on or revised based on the input of the working group. Following each video conference, the project team circulated a summary of the discussions and a survey that asked working group members to provide feedback and vote on each item considered for inclusion in the Standard Set. For voting during the final survey, a 9-point Likert scale ranging from 1 ("not recommended") to 9 ("essential to have") was used. Items were included in the Standard Set if they were rated ≥7 by at least 70% of the working group or excluded if rated ≤3 by at least 70% of the working group. In other cases, the result was considered inconclusive and the item was discussed again during the following video conference. The conduct and reporting of the Delphi process followed reporting guidelines for core outcome set development using the Delphi approach (see Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23799/abstract) (27).

Preselection of relevant patient outcome domains. Preceding the video conference on the selection of outcome domains, 2 separate systematic literature reviews were performed in December 2016 using the PubMed database, in order to identify outcomes relevant to patients for the included conditions that were modifiable in principle and feasible to implement. In the first search, we used the "Cochrane Highly Sensitive Search Strategy for identifying randomized trials" (28) to identify 25 recent randomized trials in each of the conditions included in the Standard Set. For each randomized trial, we checked the relevant online repositories and conducted a second PubMed search using the name of each trial and trial registration number, in order to find additional publications on the same study. All outcomes assessed in each trial were extracted. Reports on randomized trials in languages other than English were excluded. In the second search, we identified reports on qualitative studies in which patients with 1 of the relevant conditions were asked about the most important outcomes of their disease. We included only papers in which an open-ended question format was used, in order to prevent investigator-imposed biases on the list of patient-generated outcome domains. All outcome domains considered important by patients were extracted from each paper by 2 reviewers independently. Disagreements were resolved during a consensus meeting with a third reviewer present. Previously published core set recommendations for outcome measurement in randomized trials were also consulted, as was the European League Against Rheumatism standardized data set for observational research (29–33). Finally, 2 patient advisory group sessions with inflammatory arthritis patients from the Netherlands and the US were organized by the project team to serve as a check on the comprehensiveness of the list of identified patient outcomes. The patient advisory group protocol was exempt from ethical review by the Chesapeake Institutional Review Board.

Preselection of outcome measures. All outcome measures used in any randomized trial identified in the initial systematic review, recommended for inclusion by working group members or previously endorsed by relevant consensus statements, were considered for inclusion. The instruments were reviewed with respect to outcome domains, evidence supporting psychometric properties, feasibility, licensing fees, and degree to which they were established in the field. In order to support this process, a systematic literature review was performed in May 2017 to identify papers that had reported on the measurement properties of 26 potentially relevant patient-reported outcome measures (PROMs). The methodologic quality of the 159 identified papers was assessed using the Consensus-Based Standards for the Selection of Health Status Measurement Instruments (COSMIN) checklist (34). The studies that were of high methodologic quality were

then used to rate the measurement properties of the 26 PROMs, using quality criteria proposed by Terwee et al and the International Society for Quality of Life Research (35,36). Comprehensibility, cost, and time needed for completion were all assessed to determine the feasibility of implementing specific PROMs. The Flesch-Kincaid grade level was calculated for the English language version of each PROM (37), information about fees payable for use of the instrument was retrieved from the copyright owner's website when applicable, and information on time to complete was retrieved from a previous series of reviews (38).

Preselection of risk-adjustment variables. A preliminary list of risk-adjustment variables was extracted from published reviews on risk factors and validated risk models. Previously published ICHOM Standard Sets were reviewed for definitions of demographic and socioeconomic variables.

External validation by health professionals and patient experts. After proceeding through all of the Delphi rounds, the preliminary Standard Set was made available and sent to various stakeholders for review (www.ichom.org). A patient survey was distributed through national patient organizations and the networks of the project team and working group members. Patients were asked to rate the importance of each outcome using a 9-point Likert scale ranging from "1 = not at all relevant" to "9 = essential," and were given the opportunity to suggest additional outcomes. Health professionals were recruited via the professional networks of the working group members and project team. Health professionals were asked to rate the relevance of each domain, provide feedback on feasibility of implementation of the Standard Set in clinical practice, and rate the appropriateness of the risk-adjustment variables and time points for assessment.

RESULTS

Scope. At the start, it was recognized by the working group that the same treatment goals and longitudinal outcomes (pain, physical function, fatigue) are relevant to most types of inflammatory arthritis. The ICHOM Inflammatory Arthritis Standard Set was therefore designed to evaluate treatment outcomes of patients with rheumatoid arthritis (RA), axial spondyloarthritis (SpA) and psoriatic arthritis (PsA), as well as juvenile idiopathic arthritis (JIA), and applies to all treatments for these conditions, including medication, surgery, and physical and occupational therapies. All working group members voted to include RA, axial SpA, and PsA, and 82% voted to include JIA. The inclusion of gout (59% voted to include initially), connective tissue diseases (41%) and vasculitis (36%) was also considered, especially as few outcome recommendations are available for the latter 2 conditions. However, after revisiting this topic at a subsequent meeting, it was decided that their inclusion might lead to a generic set of outcomes, which might insufficiently capture the outcomes that matter to patients with individual conditions, due to the variety of disease manifestations and disease courses.

Table 1. Final list of outcomes considered for inclusion in ICHOM IA set*

Outcome	Qualitative studies in which outcome was reported as important disease outcome	No. of IA conditions for which outcome is included in clinical trial core set	No. of IA conditions for which outcome was measured in ≥1 clinical trial, but not in core set	Working group members voted for inclusion (%)
Pain	27 (96)	3	1	100
Physical function	26 (93)	4	NA	100
Adverse events	11 (39)	0	4	95
Fatigue	23 (82)	2	2	90
Work/school ability and productivity	14 (50)	0	3	90
Overall physical and mental health impact	NA	4	NA	86
Inflammatory disease activity	6 (21)	4	NA	84
Participation restrictions	18 (64)	0	0	55
Joint damage	7 (25)	2	1	45
Mortality	1 (4)	0	0	40
Psychological well-being	22 (79)	0	0	35
Sleep	10 (36)	0	2	35
Coping & self-management	16 (57)	0	0	33
Financial impact	3 (11)	0	0	25
Joint stiffness	15 (54)	0	4	20
Joint range of motion	0 (0)	1	0	15
Physical appearance	7 (25)	0	0	10
Sexual functioning	8 (29)	0	0	10
Cognitive functioning	5 (18)	0	0	5
Fever	0 (0)	0	1	0

* Values are the number (%) unless indicated otherwise. ICHOM = The International Consortium for Health Outcome Measurement; IA = inflammatory arthritis; NA = not applicable.

Outcome domains. Twenty-four outcome domains were initially identified in the 130 randomized trial reports and 28 qualitative studies that were identified in the systematic literature reviews (references available upon request from the corresponding author). This list was expanded upon and refined several times based on group discussions with working group members. The final consolidated list of outcomes is presented in Table 1, together with a summary of both systematic reviews and the final voting results. The list of outcome domains assessed in randomized trials and their rank ordering reflects a preference in trials for clinical measures of disease manifestations and patient-reported outcomes of symptoms and their direct impact on functioning. The list and rank ordering of patient-generated outcome domains, on the other hand, somewhat deemphasized the importance of outcome measures that reflect the pathophysiology of the specific disease and included a wider variety of outcomes that reflect the different ways arthritis impacts the daily lives of patients. Besides PROMs of symptoms and basic functioning, the patient-generated list also included more generic outcomes, such as overall psychological well-being and participation restrictions. To characterize the core symptoms and their direct effect on functioning from the patient perspective, the working group recommends that providers assess pain, fatigue, and activity limitations (i.e., physical function). These were the most frequently used outcome domains in randomized trials and were reported as important by patients in almost all of the reviewed qualitative studies. These outcome domains were also the most frequently endorsed domains in the individual core sets for randomized trials of the respective conditions. In order to assess the impact of inflammatory arthritis on the daily lives of patients more broadly, the working group recommends an assessment of overall emotional and physical health impact, and work/school/housework productivity. Participation restrictions other than work or school productivity were also considered important. However, this domain was eventually excluded because of significant overlap with other included domains, and because experience with available measurement instruments is currently limited.

Assessments of inflammatory disease activity and therapeutic response are further recommended as measures of disease control, because the absence of signs and symptoms of disease is the primary treatment target for inflammatory arthritis and disease activity was considered the main determinant of overall impact of disease. Finally, adverse events should be recorded as a measure of disutility of care.

Outcome measures. The list of recommended PROMs is shown in Table 2. Supplementary Appendix 1 provides an overview of characteristics and ratings assigned to the measurement properties of these PROMs and includes an overview of the criteria used for assigning ratings (available online on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23799/abstract).

The working group agreed that a key property was feasibility in different settings globally. Therefore, instruments with a large number of items, or that could not be hand-scored were avoided. All endorsed PROMs are available in multiple languages and for each outcome domain, at least 1 PROM is recommended that was judged to have sufficient evidence supporting its measurement properties, based on the results of the systematic review. On the other hand, some instruments were included that do not (yet) meet psychometric standards of the COSMIN checklist. Several Patient-Reported Outcomes Measurement Information System (PROMIS) measures were included so that experience with innovative Item Response Theory (IRT)–based measures could accumulate. The RA and PsA impact of disease scores and the Assessment of SpondyloArthritis international Society (ASAS Health index) were recommended because these are patient/International Classification of Functioning, Disability, and Health–derived composite scores that assess the important domains of impact of RA and PSA, and axial SpA, respectively. An overview of the various measures that are recommended by the clinical guidelines issued by various national and international rheumatology societies is provided (see Supplementary Table 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23799/abstract. For each of the ICHOM outcome domains, the endorsed outcome measures are congruent with the various guidelines. Users of the Standard Set may select preferred instruments to assess each outcome from the list of endorsed PROMs presented in Table 2. The shortest recommended combination of PROMs to assess all outcome domains is the numerical rating scale/visual analog scale (VAS) to measure fatigue, overall emotional and physical health impact, and pain; the Work Productivity and Activity Impairment questionnaire to measure work/school/housework ability and productivity; and the Health Assessment Questionnaire II to measure activity limitations. This combination of PROMs is free to use for all users and totals 19 questions, which most patients should be able to fill out in 5 minutes. The endorsed PROMs of pain, fatigue, overall well-being, and activity limitations have been linked to a common reporting metric, such that outcomes can be compared between providers that have used different instruments (39). Linked scores for benchmarking outcomes using any of the recommended instruments of the Standard Set can be obtained from www.tihealthcare.nl/en/expertise/common-metrics.

In order to track disease activity and therapeutic response, it is proposed that major evidence-based guidelines are followed (40,41). Patients and health care providers should specify target disease activity levels for individual patients (preferably remission; if not, feasible low disease activity) and assess at each visit whether this target was achieved. Disease activity should be monitored using a validated and internationally recognized clinical composite score.

Table 2. Overview of measures endorsed for assessing patient-reported outcomes included in ICHOM IA Standard Set*

Outcome, endorsed instruments	Construct validity	Reliability	Responsiveness	Flesch-Kincaid grade level†
Pain				
NRS/VAS	>1	>1	>1	7
SF-36 bodily pain	>1	>1	>1	8
PROMIS Short Form v1.0– Pain Interference 8a	0	0	1	6
PedsQL aches and pain‡	0	1	0	0
Fatigue				
BRAF-MD	1	>1	1	5
FACIT-F	1	>1	>1	3
NRS/VAS	>1	1	>1	8
PROMIS Short Form v1.0– Fatigue 8a	0	0	1§	5
PedsQL 4.0 fatigue‡	0	1	0	0
Activity limitations				
HAQ DI	>1	>1	>1	3
HAQ-II	1	>1	0	4
MHAQ	1	>1	>1	3
PROMIS Short Form v2.0 – Physical Function 10a	1	0	¶	4
MDHAQ score	1	>1	1	3
BASFI	0	>1	>1	7
C-HAQ	1	>1	>1	4
JAMAR	0	0	0	2
Health impact				
PROMIS global health	0	0	0	8
EQ-5D#	>1**	>1	>1	6
SF-6D	1	>1	>1	9
RAID (for RA)	0	>1	1	10
PSAID (for PsA)	0	>1	1	12
ASAS Health Index	1	>1	0	6
Patient/parent global assessment (NRS or VAS)	0	>1	>1	7
Work/school/ housework ability and productivity, WPAI	1	1	1	8

***** Values are the number of studies of good methodologic quality that noted favorable properties, unless indicated otherwise. ICHOM = International Consortium for Health Outcome Measurement; IA = inflammatory arthritis; NRS = numerical rating scale; VAS = visual analog scale; SF-36 = Short Form 36 Health Survey; PROMIS = Patient-Reported Outcome Measurement Information System; PedsQL = Pediatric Quality of Life Inventory; BRAF-MD = Bristol Rheumatoid Arthritis Fatigue Multidimensional Questionnaire; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; HAQ DI; Health Assessment Questionnaire Disability Index; MHAQ = Modified Health Assessment Questionnaire Disability Index; MDHAQ = Multidimensional Health Assessment Questionnaire; BASFI = Bath Ankylosing Spondylitis Functional Activity Index; C-HAQ = Childhood Health Assessment Questionnaire; JAMAR = The Juvenile Arthritis Multidimensional Assessment Report; EQ-5D = EuroQol 5 dimensions; SF-6D = Medical Outcome Study Short Form 6D; RAID = Rheumatoid Arthritis Impact of Disease; RA = rheumatoid arthritis; PSAID = Psoriatic Arthritis Impact of Disease; PsA = psoriatic arthritis; ASAS = Assessment of SpondyloArthritis international Society; WPAI =Work Productivity and Activity Impairment Questionnaire.
† Estimated using the Flesch-Kincaid grade level statistic.
‡ Instrument is intended for pediatric populations.
§ Unfavorable properties according to 1 study with good methodologic quality.
¶ Mixed findings in studies of good methodologic quality.
# Including EQ-5D-Y for pediatric patients.
** Unfavorable properties according >1 study of good methodologic quality.

Risk-adjustment variables. Most risk-adjustment variables included in the set (Table 3) apply to all patients, and care was taken to include risk-adjustment variables that are relevant and applicable in a variety of health care systems. Year of birth and sex were included as demographic variables. Education level was ultimately chosen as the only indirect measure of social economic status (SES). Other SES-related variables were considered important, but difficult to collect, due to restrictions on recording race/ethnicity in some countries, area-based measure of SES possibly being unavailable for each country, and patients potentially feeling reluctant to report on their income/wealth. For baseline status indicators, we included smoking status, comorbidities, diagnosis, time since diagnosis, and rheumatoid factor and anti–cyclic citrullinated protein antibody for RA and JIA.

Table 3. Case-mix variables*

Variable	Definition (response options)	Timing†	Data source
Age	Year of birth	Baseline	Patient
Sex	Sex at birth (Female/male)	Baseline	Patient
Smoking status	Never/former /current	Baseline	Patient
Education level	Highest attained education ISED classification (none/primary/secondary/tertiary)	Baseline	Patient
Comorbidities	Present/absent/unknown: chronic lung disease, myocardial infarction, other heart disease, stroke, hypertension, diabetes mellitus, fracture, depression, ulcer or stomach problem, cancer, central sensitization to pain, obesity (i.e., BMI ≥30)	Baseline	Clinical
Diagnosis	Physician reported diagnosis (RA/SpA/PsA/JIA)	Baseline and annually	Clinical
Disease duration	Year of diagnosis	Baseline	Clinical
Immunologic‡	Rheumatoid factor and ACPA positivity (yes/no)	Baseline	Clinical

* ISED = Institute for the Study of Education and Human Development; BMI = body mass index; RA = rheumatoid arthritis; SpA = spondyloarthritis; PsA = psoriatic arthritis; JIA = juvenile idiopathic arthritis; ACPA = anti-citrullinated protein antibody.
† Baseline defined as first measurement for patient.
‡ Only for RA and JIA.

HLA–B27 was excluded since it is not routinely collected in the health care system. Comorbidities should be assessed using the Rheumatic Disease Comorbidities Index (42), modified to include central sensitization to pain (e.g., fibromyalgia) and obesity. In order to avoid misclassification of early symptoms that may or may not reflect those specific to the inflammatory arthritis diagnosis of interest, we elected to include time since diagnosis rather than time since symptom onset.

Data collection time points. To allow meaningful outcomes comparisons between health care providers, we recommend that all risk-adjustment variables to be collected at the first assessment. All PROMs and clinical measures should also be collected at the first assessment and annually thereafter. In instances of active disease, we recommend that the patient's disease activity status be recorded at least every 6 months, but likely more frequently, at the discretion of the patient and their health care provider. Adverse events should be collected at each assessment point after baseline. A reference guide with detailed instructions for implementation and exact definitions for all of the data elements can be downloaded (www.ICHOM.org). Finally, we stress that these recommendations are intended only for quality improvement purposes and should not be understood as more than minimally acceptable clinical guidelines in patients with established disease. Especially in patients with early disease, more frequent monitoring may be required.

Open review. Eighty-three health care professionals, the majority of which (95%) were clinician/researchers, and 630 people living with inflammatory arthritis from the US, France, Argentina, The Netherlands, and Brazil reviewed the Standard Set. All outcomes included in the ICHOM Inflammatory Arthritis Set were considered very relevant by patients and health care professionals (Figure 1). Similar to the results of the systematic reviews that were used for identifying outcome domains, patients considered clinical measures slightly less relevant compared with the patient-reported outcomes. A large majority of patients (81.3%) felt that the set comprehensively covers all the relevant outcome domains of their disease. The health care professionals predominantly shared this view. Only 3 outcomes were suggested to be missing by >1 reviewer: financial impact (n = 2), joint damage (n = 2), and patient satisfaction (n = 3). Psychological well-being (12.6%) and participation restrictions (5.4%) were the only outcomes that were reported as missing from the set by >2% of patient reviewers (see Supplementary Appendix 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23799/abstract). The included risk-adjustment variables were rated very relevant by 91.8% of health professionals.

DISCUSSION

We present a standard set of outcomes for inflammatory arthritis that health care providers worldwide can use in routine clinical care to help quantify the value provided for patients in different centers, countries, and health care systems. This Standard Set was developed through consensus of an international working group with expertise across a range of disciplines relevant to outcome assessment and care for patients with inflammatory arthritis. We used a multiple methods approach, in which extensive patient input as well as published qualitative and quantitative data were used to develop a minimally sufficient set of outcomes that we believe represents outcomes that matter to patients with inflammatory arthritis. We also proposed time points for data collection and relevant risk-adjustment variables to enable comparisons between providers with different patient populations. Feasibility of implementation in different health care systems was a central priority. Therefore, we included PROMs that are not only widely accepted measures of the respective domains but that are also available in multiple translations, and for each outcome there is at least 1 PROM free to use. However, for the use of several

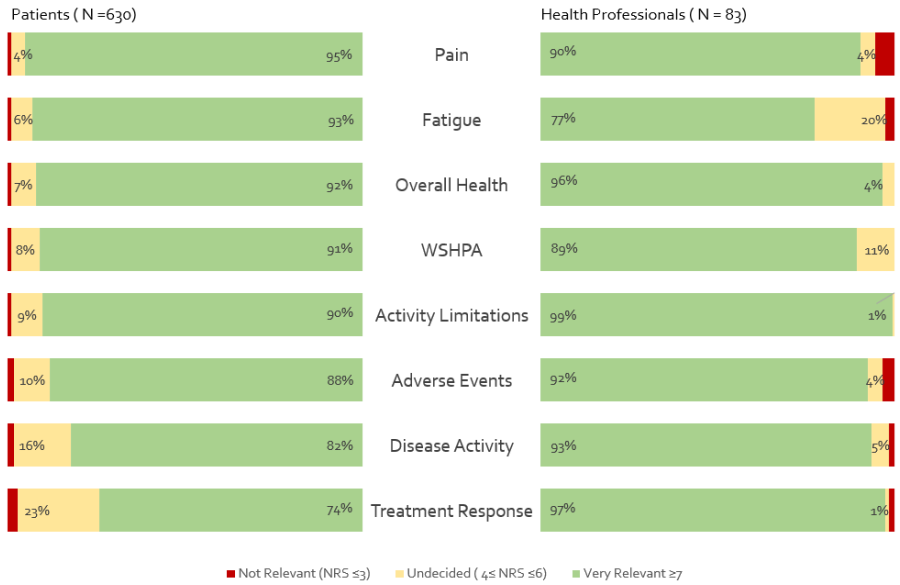

Figure 1. Relevance of outcomes included in The International Consortium for Health Outcome Measurement inflammatory arthritis set according to patient and health professional open review. WSPHA = Work/School/Household productivity and ability; NRS = numerical rating scale.

instruments, including the Pediatric Quality of Life Inventory, the Medical Outcome Study Short Form 6D, Functional Assessment of Chronic Illness Therapy (FACIT), and EuroQol 5-domain license fees may apply. Patient-Reported Outcomes Measurement Information System (PROMIS) has also introduced financial charges for the use of some of their products, including their computerized adaptive tests. Using the PROMIS Assessment Centre platform will incur a $5,000 USD charge per study per year.

One of the challenges faced with international standardization of patient outcomes data collection is that a variety of well validated and frequently used PROMs are typically available to assess the different patient outcomes. The working group for the ICHOM Depression & Anxiety and Chronic Kidney Disease Standard Sets previously responded to this challenge by endorsing PROMs that can be mapped to the PROMIS metric, using resources provided by the PROMIS PROsetta project (43,44). This way, users of these sets use 1 PROM for each domain, but results can be scored on the PROMIS metric. In the work on the ICHOM inflammatory arthritis Standard Set presented here, this is taken one step further, by linking multiple PROMs to an IRT-based common reporting metric (39). This makes it easier for new or ongoing data collection initiatives to contribute their data, since it allows users of the ICHOM inflammatory arthritis set to choose 1 instrument from a number of alternatives for each domain. Provided that 1 of the endorsed instruments (Table 2) is collected, outcomes can be compared with those from other health care providers who use the ICHOM Standard Set. For example, outcomes assessed using the VAS scale for fatigue can be directly compared with outcomes of a different group of patients assessed using the FACIT–Fatigue subscale. In principle, PROMs could be added to and removed from the list of endorsed instruments, without affecting comparability of the outcomes. The ICHOM list of recommended PROMs overlaps significantly with current clinical guidelines. Moreover, the results of 2 systematic reviews of various national RA patients' registries show that the majority of the PROMs that are currently collected in the reviewed registries are also included in the ICHOM Inflammatory Arthritis Standard Set. The IRT approach also allows each of the ICHOM inflammatory arthritis outcomes to be assessed using computerized adaptive tests, which would help achieve optimally precise scores with minimal numbers of questions (45,46).

Since the ICHOM Inflammatory Arthritis Standard Set is intended to reflect outcomes that are important to patients, the extensive input from patients is a strength of this work. We included 6 patient representatives in the working group, derived the list of outcomes from published qualitative studies in which patients reported outcomes that matter to them, organized 2 patient advisory group sessions with patients that were not included in the working group to review the final list of outcomes to be voted on by working group members, and the final version of the Standard Set was reviewed by 630 patients from various countries. The inclusion of working group members with diverse geographic and professional backgrounds is also a strength.

We do, however, acknowledge that different results might have been obtained had other working group members been selected. We also realize that it may prove challenging to collect all the requested information for all health care providers at all time points. In particular, inflammatory disease activity may prove logistically challenging to track in some health care systems, as it requires clinical assessment of joint involvement and, in some cases, laboratory assessments. In such situa-

tions, we would encourage users of the set to at least monitor the PROMs. All patient-reported outcomes can be collected using a minimum of 20 items, which could be further reduced using computerized adaptive testing or targeted short forms. Finally, we acknowledge that the value of the ICHOM Inflammatory Arthritis Standard Set has not yet been proven in practice. ICHOM aims to partner with several interested institutions to pilot test the Standard Set. Furthermore, a steering committee has been established that will periodically review the Standard Set, including lessons learned from the pilot phase and other applications of the set. This will include, but will not be limited to, reviewing PROMs that are endorsed in the Inflammatory Arthritis Standard Set, the ease in accessing and monitoring these PROMs, and the outcomes related to personal goals that individual patients identify.

In summary, we propose a standard set of outcomes for patients with inflammatory arthritis that providers of care for patients with inflammatory arthritis can track to facilitate the global reporting of outcome data and shared learning. A detailed reference guide is available (www.ichom.org).

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. Oude Voshaar 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. Oude Voshaar, Das Gupta, van de Laar, Vonkeman.

Acquisition of data. Oude Voshaar, Das Gupta, Bijlsma, Boonen, Chau, Courvoisier, Curtis, Ellis, Ernestam, Gossec, Hale, Hornjeff, Leung, Lidar, Mease, Michaud, Mody, Ndosi, Opava, Pinheiro, Salt, Soriano, Taylor, Voshaar, Weel, de Wit, Wulffraat, van de Laar, Vonkeman.

Analysis and interpretation of data. Oude Voshaar, Das Gupta, Bijlsma, Boonen, Chau, Courvoisier, Curtis, Ellis, Ernestam, Gossec, Hale, Hornjeff, Leung, Lidar, Mease, Michaud, Mody, Ndosi, Opava, Pinheiro, Salt, Soriano, Taylor, Voshaar, Weel, de Wit, Wulffraat, van de Laar, Vonkeman.

REFERENCES

1. Crowson CS, Matteson EL, Myasoedova E, Michet CJ, Ernste FC, Warrington KJ, et al. The lifetime risk of adult-onset rheumatoid arthritis and other inflammatory autoimmune rheumatic diseases. Arthritis Rheum 2011;63:633–9.
2. Stoffer MA, Schoels MM, Smolen JS, Aletaha D, Breedveld FC, Burmester G, et al. Evidence for treating rheumatoid arthritis to target: results of a systematic literature search update. Ann Rheum Dis 2016;75:16–22.
3. Schoels MM, Braun J, Dougados M, Emery P, Fitzgerald O, Kavanaugh A, et al. Treating axial and peripheral spondyloarthritis, including psoriatic arthritis, to target: results of a systematic literature search to support an international treat-to-target recommendation in spondyloarthritis. Ann Rheum Dis 2014;73:238–42.
4. Mody GM, Cardiel MH. Challenges in the management of rheumatoid arthritis in developing countries. Best Pract Res Clin Rheumatol 2008;22:621–41.
5. Kumar B. Global health inequities in rheumatology. Rheumatology (Oxford) 2017;56:4–5.
6. Cross M, Smith E, Hoy D, Carmona L, Wolfe F, Vos T, et al. The global burden of rheumatoid arthritis: estimates from the global burden of disease 2010 study. Ann Rheum Dis 2014;73:1316–22.
7. Porter M, Teisberg E. Redefining health care: creating value-based competition on results. Brighton (MA): Harvard Business Review Press; 2006.
8. Porter ME, Larsson S, Lee TH. Standardizing patient outcomes measurement. N Engl J Med 2016;374:504–6.
9. Werneke M. A proposed set of metrics for standardized outcome reporting in the management of low back pain. Acta Orthop 2016;87:88–90.
10. Porter ME. Measuring health outcomes: the outcomes hierarchy. N Engl J Med 2010;363:2477–81.
11. Martin NE, Massey L, Stowell C, Bangma C, Briganti A, Bill-Axelson A, et al. Defining a standard set of patient-centered outcomes for men with localized prostate cancer. Eur Urol 2015;67:460–7.
12. Mahmud I, Kelley T, Stowell C, Haripriya A, Boman A, Kossler I, et al. A proposed minimum standard set of outcome measures for cataract surgery. JAMA Ophthalmol 2015;133:1247–52.
13. Allori AC, Kelley T, Meara JG, Albert A, Bonanthaya K, Chapman K, et al. A standard set of outcome measures for the comprehensive appraisal of cleft care. Cleft Palate Craniofac J 2017;54:540–54.
14. Rodrigues IA, Sprinkhuizen SM, Barthelmes D, Blumenkranz M, Cheung G, Haller J, et al. Defining a minimum set of standardized patient-centered outcome measures for macular degeneration. Am J Ophthalmol 2016;168:1–12.
15. Rolfson O, Wissig S, van Maasakkers L, Stowell C, Ackerman I, Ayers D, et al. Defining an international standard set of outcome measures for patients with hip or knee osteoarthritis: consensus of the International Consortium for Health Outcomes Measurement Hip and Knee Osteoarthritis Working Group. Arthritis Care Res (Hoboken) 2016;68:1631–9.
16. Mak KS, van Bommel AC, Stowell C, Abrahm JL, Baker M, Baldotto CS, et al. Defining a standard set of patient-centred outcomes for lung cancer. Eur Respir J 2016;48:852–60.
17. Ong WL, Schouwenburg MG, van Bommel AC, Stowell C, Allison KH, Benn KE, et al. A standard set of value-based patient-centered outcomes for breast cancer: the International Consortium for Health Outcomes Measurement (ICHOM) Initiative. JAMA Oncol 2017;3:677–85.
18. De Roos P, Bloem BR, Kelley TA, Antonini A, Dodel R, Hagell P, et al. A consensus set of outcomes for Parkinson's disease from the International Consortium for Health Outcomes Measurement. J Parkinsons Dis 2017;7:533–43.
19. Obbarius A, van Maasakkers L, Baer L, Clark DM, Crocker AG, de Beurs E, et al. Standardization of health outcomes assessment for depression and anxiety: recommendations from the ICHOM Depression and Anxiety Working Group. Qual Life Res 2017;26:3211–25.
20. Morgans AK, van Bommel AC, Stowell C, Abrahm JL, Basch E, Bekelman JE, et al. Development of a standardized set of patient-centered outcomes for advanced prostate cancer: an international effort for a unified approach. Eur Urol 2015;68:891–8.
21. Foust-Wright C, Wissig S, Stowell C, Olson E, Anderson A, Anger J, et al. Development of a core set of outcome measures for OAB treatment. Int Urogynecol J 2017;28:1785–93.
22. Akpan A, Roberts C, Bandeen-Roche K, Batty B, Bausewein C, Bell D, et al. Standard set of health outcome measures for older persons. BMC Geriatr 2018;18:36.
23. Salinas J, Sprinkhuizen SM, Ackerson T, Bernhardt J, Davie C, George MG, et al. An international standard set of patient-centered outcome measures after stroke. Stroke 2016;47:180–6.

24. Kim AH, Roberts C, Feagan BG, Banerjee R, Bemelman W, Bodger K, et al. Developing a standard set of patient-centred outcomes for inflammatory bowel disease: an international, cross-disciplinary consensus. J Crohns Colitis 2018;12:408–18.
25. Zerillo JA, Schouwenburg MG, van Bommel AC, Stowell C, Lippa J, Bauer D, et al. An international collaborative standardizing a comprehensive patient-centered outcomes measurement set for colorectal cancer. JAMA Oncol 2017;3:686–94.
26. McNamara RL, Spatz ES, Kelley TA, Stowell CJ, Beltrame J, Heidenreich P, et al. Standardized outcome measurement for patients with coronary artery disease: consensus from the International Consortium for Health Outcomes Measurement (ICHOM). J Am Heart Assoc 2015;4:e001767.
27. Sinha IP, Smyth RL, Williamson PR. Using the Delphi Technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. PLoS Med 2011;8:e1000393.
28. Higgins J, Green S. Cochrane handbook for systematic reviews of interventions version 5.1. The Cochrane Collaboration; 2011. URL: www.cochrane-handbook.org.
29. Chatzidionysiou K, Nikiphorou E, Gossec L, Hyrich K, Filip C, van Eijk-Hustings Y, et al, on behalf of the EULAR Task Force for Standardising Minimum Data Collection in Rheumatoid Arthritis Observational Research. The Eular Task Force for Standardising Minimum Data Collection in Rheumatoid Arthritis Observational Research: results of a hierarchical literature [abstract]. Ann Rheum Dis 2016;75:166.
30. 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.
31. Giannini EH, Ruperto N, Ravelli A, Lovell DJ, Felson DT, Martini A. Preliminary definition of improvement in juvenile arthritis. Arthritis Rheum 1997;40:1202–9.
32. Orbai AM, de Wit M, Mease P, Shea JA, Gossec L, Leung YY, et al. International patient and physician consensus on a psoriatic arthritis core outcome set for clinical trials. Ann Rheum Dis 2017;76:673–80.
33. Van der Heijde D, Bellamy N, Calin A, Dougados M, Khan MA, van der Linden S. Preliminary core sets for endpoints in ankylosing spondylitis: assessments in Ankylosing Spondylitis Working group. J Rheumatol 1997;24:2225–9.
34. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010;19: 539–49.
35. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34–42.
36. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. Qual Life Res 2013;22:1889–905.
37. Kincaid J, Fishburne RP Jr, Rogers RL, Chissom BS. Derivation of new readability formulas (automated readability index, fog count and Flesch reading ease formula) for navy enlisted personnel. Millington (TN): Chief of Naval Technical Training Naval Air Station Memphis; 1975.
38. Katz PP. Introduction to special issue: patient outcomes in rheumatology, 2011 [editorial]. Arthritis Care Res (Hoboken) 2011;63 Suppl 11:S1–3.
39. Oude Voshaar MA, Vonkeman HE, Courvoisier D, Finkckh A, Gossec L, Leung YY, et al. Towards standardized patient reported physical function outcome reporting: linking ten commonly used questionnaires to a common metric. Qual Life Res 2019;28: 187–97.
40. Singh JA, Saag KG, Bridges SL Jr, Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. Arthritis Rheumatol 2016;68: 1–26.
41. Smolen JS, Aletaha D, Bijlsma JW, Breedveld FC, Boumpas D, Burmester G, et al. Treating rheumatoid arthritis to target: recommendations of an international task force. Ann Rheum Dis 2010;69: 631–7.
42. England BR, Sayles H, Mikuls TR, Johnson DS, Michaud K. Validation of the Rheumatic Disease Comorbidity Index. Arthritis Care Res (Hoboken) 2015;67:865–72.
43. Obbarius A, van Maasakkers L, Baer L, Clark DM, Crocker AG, de Beurs E, et al. Standardization of health outcomes assessment for depression and anxiety: recommendations from the ICHOM Depression and Anxiety Working Group. Qual Life Res 2017;26:3211–25.
44. Verberne W, Das-Gupta Z, Allegretti A, Bart HA, van Biesen W, García-García G, et al. Development of an international standard set of value-based outcome measures for patients with chronic kidney disease: a report of the International Consortium for Health Outcomes Measurement (ICHOM) CKD working group. Am J Kidney Dis 2019;73:372–84.
45. Van der Linden W, Glas C. Elements of adaptive testing (statistics for social and behavioral sciences). New York: Springer; 2010.
46. Van der Linden W. Linear models for optimal test design. New York: Springer; 2005.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 71, No. 12, December 2019, pp 1566–1575
DOI 10.1002/acr.23900

Patient-Reported Outcome Data From an Early Rheumatoid Arthritis Trial: Opportunities for Broadening the Scope of Treating to Target

Kristien Van der Elst,[1] Patrick Verschueren,[1] Veerle Stouten,[2] Sofia Pazmino,[2] An De Groef,[2] Diederik De Cock,[2] Johan Joly,[3] Philip Moons,[4] and René Westhovens[1]

Objective. Treating early, intensively, and to target leads to rapid disease control, preventing joint damage and loss of function in early rheumatoid arthritis (RA). We report the effect of such an approach on patient-reported outcomes and explore the contribution of rapid and persistent disease control to well-being after 1 year of treatment.

Methods. This study is part of the Care in Early RA trial, a prospective, 2-year, investigator-initiated, randomized controlled trial rooted in daily practice and implementing the treat-to-target principle. Short Form 36 (SF-36) health survey and Revised Illness Perception Questionnaire (IPQ-R) data were collected prospectively. We defined 4 clinical response profiles based on speed and consistency of the treatment response within the first year, defined as the Disease Activity Score in 28 joints using the C-reactive protein level <2.6. Linear regression analyses including these response profiles and treatment type were constructed to predict the SF-36 dimensions of vitality, social functioning, role emotional, and mental health, and the IPQ-R illness perception subscales of consequences, treatment control, and illness coherence at year 1.

Results. A total of 333 patients were available for the main analyses, including 140 early persistent responders. Variation in each of the psychosocial outcomes at year 1 was explained mostly by baseline values, followed by the clinical response profiles. Patients with an early persistent response reported significantly higher vitality, more positive beliefs about disease consequences and treatment effect. Treatment type did not matter.

Conclusion. Rapid and persistent disease control and not treatment type were associated with favorable patient-reported health and illness perceptions at year 1, but baseline psychosocial variables mattered most. Our data indicate opportunities to broaden the scope of the treat-to-target principle in early RA.

INTRODUCTION

Rheumatoid arthritis (RA) is a chronic inflammatory disease, traditionally described as progressive and very disabling. This view of an inevitably unfavorable prognosis continues to shape its public perception (1). Today, the future for RA patients is looking much brighter. Aside from the growing availability of effective therapeutic agents, 3 treatment principles have shaped the transformation toward better clinical outcomes: treating early, intensively, and to target (2–4). In view of the window-of-opportunity theory, the goal of early RA treatment is to reach remission as soon as possible, since this initial treatment response is a strong predictor of long-term remission, radiographic damage, and functional capacity (5,6). Pragmatic early RA trials have demonstrated that applying current treatment principles in a daily practice setting is effective in terms of these key outcomes, including the Disease Activity Score in 28 joints using the C-reactive protein level (DAS28-CRP) <2.6 in up to 70% of patients (7).

Despite good disease control at the group level, not all patients with early RA have a sufficient treatment response or enjoy a smooth trajectory toward response, though they prefer to return to normal as quickly as possible (8). A few studies have linked early clinical response to outcomes valued by patients (9,10) and showed that patient-reported outcomes (PROs) of

EudraCT database no. 2008-007225-39.

The CareRA trial was supported by a Flemish governmental grant through the Agency for Innovation by Science and Technology (IWT).

[1]Kristien Van der Elst, PhD, Patrick Verschueren, PhD, René Westhovens, PhD: University Hospitals Leuven, KU Leuven, Catholic University of Leuven, Leuven, Belgium; [2]Veerle Stouten, MSc, Sofia Pazmino, MSc, An De Groef, PhD, Diederik De Cock, PhD: KU Leuven, Catholic University of Leuven, Leuven, Belgium; [3]Johan Joly, MSc: University Hospitals Leuven, Leuven, Belgium; [4]Philip Moons, PhD: KU Leuven, Catholic University of Leuven, Leuven, Belgium, University of Gothenburg, Gothenburg, Sweden, and University of Cape Town, Cape Town, South Africa.

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

Address correspondence to Patrick Verschueren, PhD, Department of Rheumatology, University Hospitals Leuven, Herestraat 49, B-3000 Leuven, Belgium. E-mail: patrick.verschueren@uzleuven.be.

Submitted for publication September 13, 2018; accepted in revised form April 2, 2019.

SIGNIFICANCE & INNOVATIONS

- Time is well established as a valuable factor in rheumatoid arthritis (RA) management, at least in view of the classical outcomes of disease.
- The current study showed that a rapid and persistent initial clinical response is also important from the patient perspective and for patient-valued outcomes, broadening the applicability of the window-of-opportunity theory in early RA.
- Ultimately, psychosocial patient-reported factors were even more related to the overall outcome in an early, intensive, and treat-to-target setting, while treatment regimen was not.
- Our findings could inspire initiatives to broaden the scope of treat-to-target strategies in early RA by integrating the patient's psychosocial profile.

patients in early remission reached values close to normal (11). However, the great heterogeneity in response to treatment, both in terms of clinical outcomes and in how patients perceive their disease and health, cannot be captured at the group level. This diversity among patients entails the need for an alternative way of using PROs, if we want to study questions that are meaningful to patients with early RA (12).

Newly diagnosed patients have preconceived illness beliefs and are not yet familiar with their disease and treatment. The initial clinical response might to a large extent determine how patients adapt to their new illness. For example, patients with an insufficient or delayed clinical response can make at some stage the switch from believing the disease is reversible and manageable to believing it will be irreversible. Illness perceptions have been shown to be barriers or facilitators of adjustment to illness and therefore can be determinants of heterogeneity in functioning and health (13–15). Previous research showed that patient personalities, different stressors, and social support at the early disease stage contributed to future anxiety and depressed mood in early RA (16).

Making use of the Care in Early RA (CareRA) trial, a pragmatic, randomized study implementing current RA treatment recommendations (7,17), we compared the effect of different treatment regimens on PROs. However, the main aim of this study was to explore the relative contributions of the different treatment combinations studied, the initial clinical response, and specific baseline psychosocial characteristics on patient-reported health and illness perceptions at 1 year after treatment initiation for early RA. We assumed that a rapid and stable clinical response would improve future aspects of psychosocial functioning and explored the contributions of personal psychosocial factors on patients' outcome.

PATIENTS AND METHODS

Setting. The present observational study is an integral part of the CareRA trial, a Flemish, multicenter, prospective, 2-year, investigator-initiated, randomized controlled pragmatic trial that implements an early intensive treatment in a treat-to-target approach in all patients with early RA (7,17). Different treatment regimens based on the original Combination therapy for early RA (COBRA) strategy were evaluated. Using a risk stratification algorithm based on classical prognostics factors (i.e., erosions, rheumatoid factor and/or anticitrullinated protein antibody, and DAS28-CRP calculated at screening), patients were randomized to 1 of the following treatment groups (Figure 1): COBRA classic: 15 mg methotrexate (MTX) weekly, 2 gm sulfasalazine daily and a weekly step-down scheme of oral prednisone (60–40–25–20–15–10–7.5 mg daily); COBRA slim: 15 mg MTX weekly with a weekly step-down scheme of oral prednisone (30–20–12.5–10–7.5–5 mg daily); COBRA avant-garde: 15 mg MTX weekly, 10 mg leflunomide daily, and a weekly step-down scheme of oral prednisone (30–20–12.5–10–7.5–5 mg daily), and from week 6 or 7 onwards, a low maintenance dose of prednisone was continued to week 28 and then tapered until discontinuation at week 34; and Tight Step-up: 15 mg MTX weekly, no oral glucocorticoids allowed. Treatment was adjusted to a target of low disease activity (DAS28-CRP $\leq$3.2). The COBRA slim treatment, including monotherapy with MTX and temporary glucocorticoids, provided the best balance between efficacy, safety, and feasibility compared to the other treatment arms in this study setting. This strategy was meanwhile explicitly adopted in the European League Against Rheumatism 2016 recommendations for treating RA (2).

In addition to the traditional disease evaluation measures, implementation aspects were studied (18–20), and patients completed a set of self-reported questionnaires (Figure 1). We focused on data from 2 questionnaires: the Short Form 36 (SF-36) health survey and the Revised Illness Perception Questionnaire (IPQ-R). The CareRA trial was approved by the leading ethics committee (University Hospitals Leuven) after consulting the ethics committees of participating centers. All participants gave written informed consent. Our interdisciplinary research team was strengthened by a patient research partner (ADG), involved in study design, data interpretation, and write-up of the results.

The initial clinical response profiles. The disease activity trajectory in the first year of treatment was determined based on achieving DAS28-CRP <2.6 at week 16 and persistence of this early response until week 52. Four profiles of initial clinical response were defined: 1) the persistent responders (n = 140) had an early clinical response at week 16, which was sustained at every clinical visit between week 16 and week 52; 2) the secondary failures (n = 92) had an early clinical response but lost this response between week 16 and week 52; 3) the delayed responders (n = 71) had a late clinical response, which was obtained between week 16 and week 52; and 4) the nonresponders (n = 30) had no clinical response

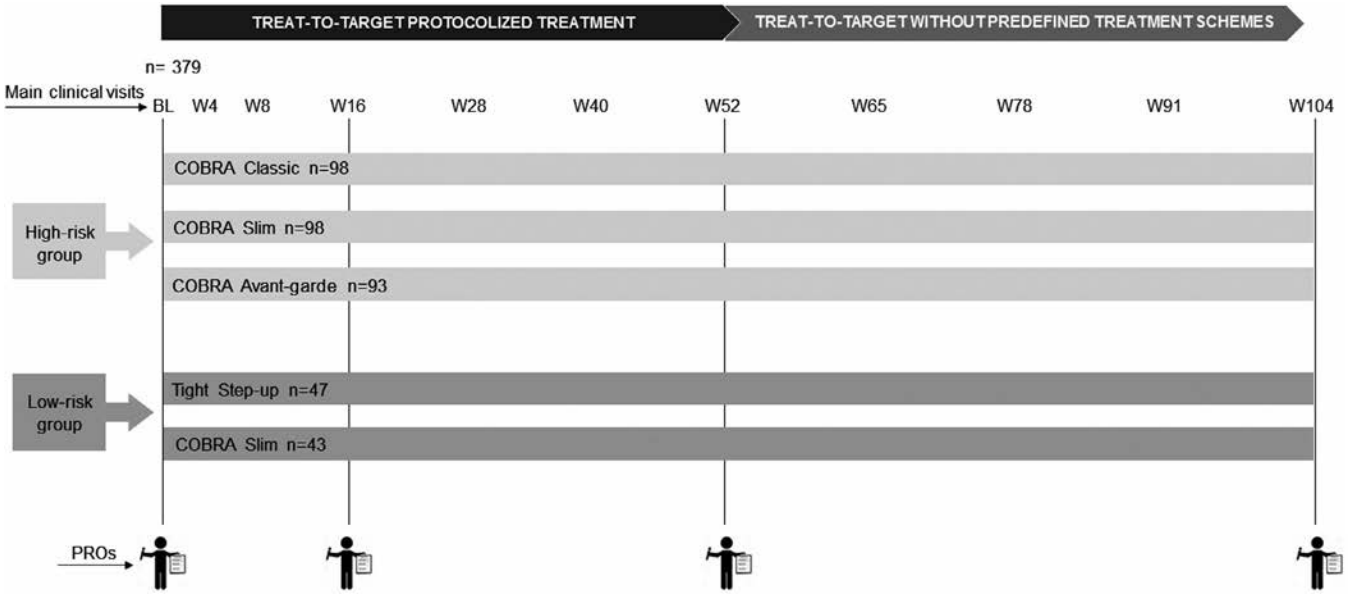

Figure 1. Overview of the Care in Early RA (CareRA) study design. High-risk group was patients with a poor prognostic profile based on the presence of classical prognostic factors (i.e., erosions, rheumatoid factor and/or anticitrullinated protein antibody, and Disease Activity Score in 28 joints using the C-reactive protein level calculated at screening). Low-risk group was patients without a poor prognostic profile. Patients (n = 379) were randomized in the CareRA trial: COBRA classic: 15 mg methotrexate (MTX) weekly, 2 gm sulfasalazine daily, and a daily oral prednisone bridging scheme (weekly step-down from baseline [BL; i.e., randomization to treatment group] until week 7 [60–40–25–20–15–10–7.5 mg], continuing 7.5 mg from week 7 to week 28, and then tapering down weekly until discontinuation at week 34); COBRA slim: 15 mg MTX weekly with a daily oral prednisone bridging scheme (weekly step-down from BL until week 6 [30–20–12.5–10–7.5–5 mg], continuing 5 mg from week 6 to week 28, and then tapering down weekly until discontinuation at week 34); COBRA avant-garde: 15 mg MTX weekly, 10 mg leflunomide daily, and a daily oral prednisone bridging scheme (weekly step-down from BL until week 6 [30–20–12.5–10–7.5–5 mg], continuing 5 mg from week 6 to week 28, and then tapering down weekly until discontinuation at week 34); Tight Step-up: 15 mg MTX weekly, no oral glucocorticoids allowed. W = week; PROs = patient-reported outcomes.

throughout the first treatment year. We selected week 16 to evaluate the speed of response, so that disease activity would not be blurred by the effect of the initial higher dosages of glucocorticoids and because the level of disease activity at this time point is known to be predictive for response to treatment after 1 year (9,21).

Measures/instruments. Descriptive variables included disease activity (DAS28-CRP), allocated treatment (Figure 1), the presence of depression (i.e., having a diagnosis of depression and/or taking antidepressants at baseline), age, sex, body mass index (BMI) at baseline, alcohol intake, smoking, employment status at screening, and perceived social support at baseline using the subscales of supportive interactions and negative interactions of the Social Support List, in which higher scores indicate more experience of supportive or negative interactions (22). The PROs of interest (i.e., perceived health and illness perceptions) were collected through valid and reliable Dutch translations of the SF-36 (version 1) and IPQ-R (23,24). Questionnaires were paper-based and completed by patients at home. For the main research question that focused on psychosocial aspects, we selected the corresponding SF-36 subscales of vitality, social functioning, role emotional, and mental health (25). Higher SF-36 scores indicate better perceived health (26). Furthermore, 3 IPQ-R illness perception subscales were selected based on their relevance for our research question: consequences, treatment control, and illness coherence. Higher scores indicate greater perceived consequences, and a stronger sense of treatment control and illness coherence (27).

Data analysis. *The role of treatment*. This analysis included SF-36 scores from the total CareRA sample according to randomized treatment groups. Mean scores for all 8 dimensions of the SF-36 at baseline, week 16, week 52, and week 104 were plotted on spidergrams for each treatment group. Differences in improvement from baseline at week 16, week 52, and week 104 were tested separately for the 3 high-risk treatment groups using the Kruskal-Wallis test, and for the 2 low-risk treatment groups using the Mann-Whitney U test.

The role of initial clinical response profiles: sample characteristics and baseline differences between initial clinical response profiles. Patients who were allocated to a profile completed the first treatment year and had a DAS28-CRP available at week 16 and at least once between week 16 and week 52 (week 52 included). For the purpose of this analysis, SF-36 and IPQ-R had to be completed minimally on one-fourth of the measurement time points, allowing eventual missing data im-

putation. Descriptive data were reported as absolute numbers and proportions, or as mean ± SDs and medians with interquartile ranges. Patients were not randomized but assigned to 1 of 4 profiles based on their initial clinical response, and therefore we tested baseline differences between profiles on all study variables using 1-way analysis of variance or Kruskal-Wallis tests according to data distribution for the continuous variables and chi-square tests for categorical variables.

The role of initial clinical response profiles: the association between the initial clinical response and psychosocial outcomes after 1 year of treatment. We performed linear regression analyses to investigate the association between the initial clinical response (predictor variable) and the predefined set of SF-36 and IPQ-R subscales (7 outcome variables). Because the clinical response profile is a categorical variable, dummy coding was applied where the persistent responders were chosen as the reference category, against which the secondary failures, delayed responders, and nonresponders were compared (according to our hypothesis). In the multiple regression analyses, the effect of initial clinical response on week 52 psychosocial outcomes was adjusted for age, sex, presence of depression, allocated treatment, baseline DAS28-CRP, and baseline value of the predicted variable. To deal with non-normally distributed data, especially for SF-36 social functioning and SF-36 role emotional, robust regression based on bootstrapping with 1,000 bias-corrected and accelerated replications was applied, providing bootstrap confidence intervals and significance values that give an accurate estimate of the true population beta values (28,29). Evaluation of PROs after 1 year of treatment was chosen as the primary outcome cut point.

Missing data. To deal with missing data, we first applied the method recommended in the SF-36 users' manual (26) and in the Dutch IPQ-R scoring syntax before calculating the subscale scores. If SF-36 and IPQ-R subscale scores could not be calculated because of too many missing items, we used the expectation-maximization algorithm (EM) (30). In our study sample, the proportion of missingness for the 7 selected SF-36 and IPQ-R subscale scores ranged between 9.3% and 10.2% at week 16, 22.2% and 22.8% at week 52, and 30.3% and 31.2% at week 104. These missing scores were imputed using 1 EM imputation model, including available scores of all SF-36 and IPQ-R subscales at baseline, week 16, week 52, and week 104 for the total CareRA sample. No imputations were made for missing scores at baseline (1.8–2.4%). To examine baseline differences between participants with and without the studied PRO data available at week 52, Mann-Whitney U tests were computed. Sensitivity analysis using SF-36 and IPQ-R scores without EM imputation was performed for the regression analyses.

A *P* value of 0.05 was used as the cutoff for statistical significance, with *P* values being corrected per regression model using the Holm-Bonferroni method to reduce the risk for inflating Type I error (31). Data were analyzed using SPSS Statistics for Windows

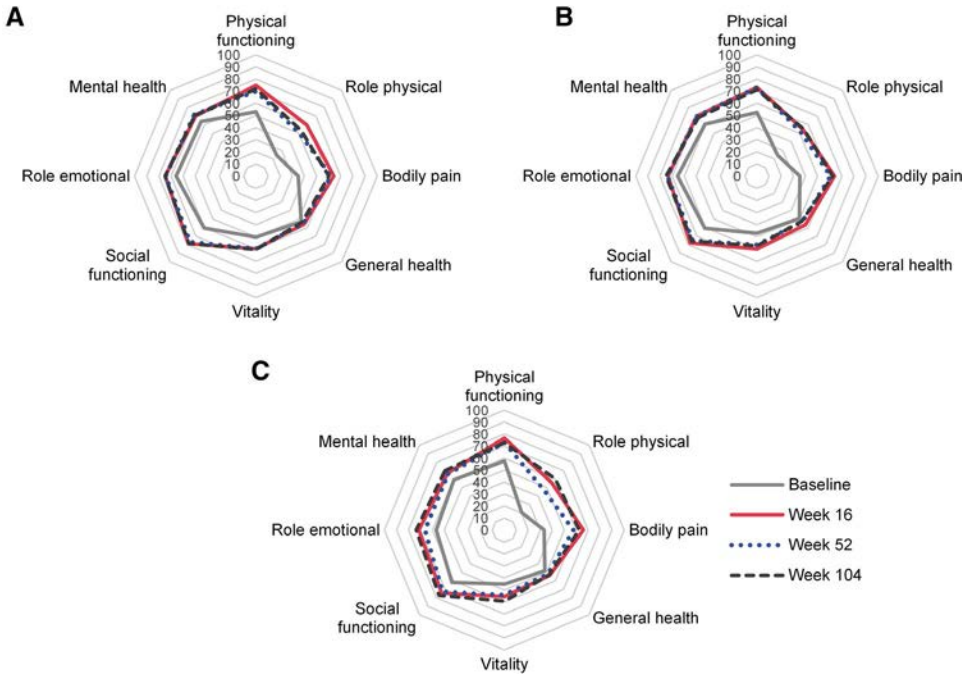

Figure 2. Mean Short Form 36 health survey domain scores at baseline, week 16, week 52, and week 104, plotted on spydergrams for each high-risk treatment group within the Care in Early RA (CareRA) trial. Patients with early rheumatoid arthritis were stratified to a high-risk profile and then randomized to 1 of 3 treatment groups: **A**, COBRA classic; **B**, COBRA slim; **C**, COBRA avant-garde. High-risk group was patients with a poor prognostic profile based on the presence of classical prognostic factors. For weeks 16, 52, and 104, missing data were imputed using the expectation-maximization method.

software, version 25. There was no specific power calculation for this explorative study alongside the CareRA trial.

RESULTS

The role of treatment. Patient characteristics of all CareRA patients have been previously reported (17). All treatments resulted in improved SF-36 domain scores in the high-risk treatment groups (Figure 2) and in the low-risk treatment groups (Figure 3). Change from baseline to week 16, week 52, and week 104 did not significantly differ between the 3 high-risk treatment groups, nor between the 2 low-risk treatment groups for any of the 8 SF-36 domain scores (data on file).

The role of initial clinical response profiles. *Sample characteristics and baseline differences between initial clinical response profiles*. Of the 379 randomized patients in the CareRA trial, 333 (87.9%) had the required data set for this particular analysis available (see Supplementary Figure 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23900/abstract). The selected patients were not significantly different from those excluded (data on file). Of those 333 patients, 259 (77.8%) had SF-36 and IPQ-R data available at week 52. Patients for which these PRO data were missing at week 52 only scored on average 5.5 units less on mental health at baseline than those patients with week 52 PRO data available ($P = 0.042$).

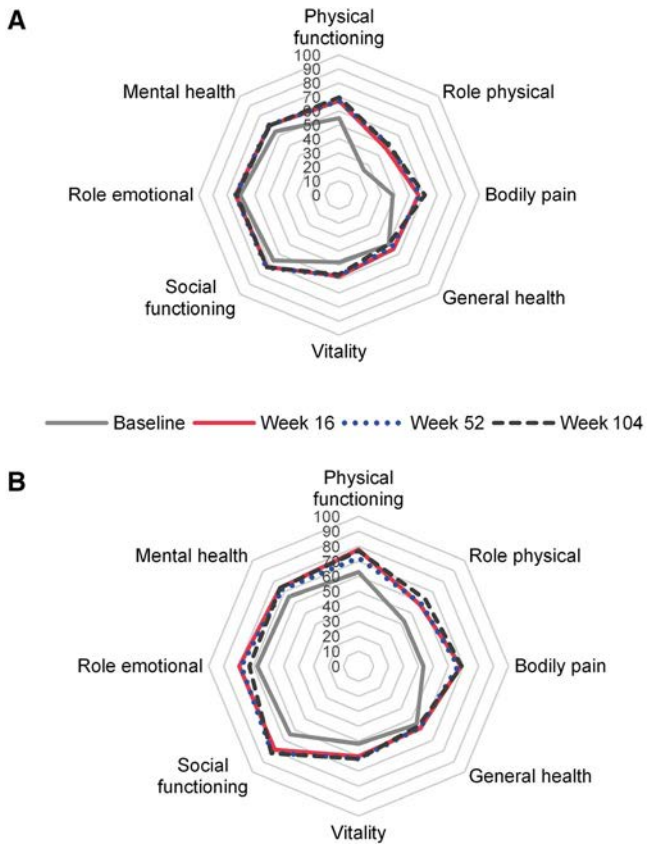

Figure 3. Mean 36-item Short Form health survey domain scores at baseline, week 16, week 52, and week 104 plotted on spydergrams for each low-risk treatment group within the Care in Early RA (CareRA) trial. Patients with early rheumatoid arthritis were stratified to a low-risk profile and then randomized to 1 of 2 treatment groups: **A**, Tight Step-up; **B**, COBRA slim. Low-risk group was patients without a poor prognostic profile based on the presence of classical prognostic factors. For weeks 16, 52, and 104, missing data were imputed using the expectation-maximization method.

Table 1 shows the baseline characteristics and Table 2 the PRO data organized by initial clinical response profile. Because patients were not randomized but subdivided post hoc according to their initial clinical response, baseline differences were observed for BMI, patient-reported vitality, social functioning, mental health, perceived consequences, and negative social interactions, as well as for DAS28-CRP. Separate analysis of the 4 DAS28-CRP components revealed that items based on patient reporting and patient-physician interaction were different between profiles, while the laboratory measure CRP level was not.

The association between the initial clinical response and psychosocial outcomes after 1 year of treatment. Table 2 shows the average week-52 scores for perceived health and illness perceptions by initial clinical response profile. First, the univariate regression results showed that the proportion of the variability explained by the response profiles ranged between 6.3% and 12.6% for perceived health outcomes and between 3.1% and 12.7% for illness perception outcomes (see Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23900/abstract). Second, in almost all multiple regression models the response profiles were independently associated with the psychosocial outcomes at year 1 of treatment (Figure 4) (for details see Supplementary Table 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23900/abstract). Only illness coherence remained unexplained by initial clinical response. Depending on the model, contributing confounders were the baseline value of the outcome variable, baseline disease activity, and the presence of depression. The baseline values had the highest contribution to the overall variability of the week-52 outcome variable, which was in general followed by the response profiles (standardized beta values in Figure 4). The allocated treatment did not contribute at all, nor did sex and age. Alcohol intake, smoking, employment status, and BMI were evaluated as potential confounders, but either were not different at baseline between response groups or did not significantly contribute to the multivariate models we tested.

Figure 4 shows that compared to early and persistent responders, patients without a response scored mostly worse, while the difference in group mean often was smallest between the persistent responders and the secondary failures. Patients

Table 1. Descriptive variables at trial baseline for the sample of patients with early rheumatoid arthritis (n = 333) subdivided according to their initial clinical response profile*

Variables	Persistent responders (n = 140)	Secondary failures (n = 92)	Delayed responders (n = 71)	Nonresponders (n = 30)	*P*
Age, years	52.7 ± 13.4	52.3 ± 13.1	52.0 ± 11.3	51.7 ± 14.8	0.974†
DAS28-CRP (range 0–9.4)	4.5 ± 1.1	4.6 ± 1.4	5.3 ± 1.3	5.2 ± 1.1	<0.001†
28SJC (range 0–28)	6.7 ± 4.9	6.7 ± 5.7	8.7 ± 5.6	8.1 ± 5.4	0.029‡
28TJC (range 0–28)	7.5 ± 5.2	8.4 ± 6.4	11.1 ± 6.8	10.2 ± 6.3	0.001‡
VAS PGA (range 0–100)	50.9 ± 23.8	51.9 ± 21.9	64.6 ± 21.0	61.0 ± 21.6	<0.001‡
CRP, mg/liter	7.7 ± 15.7	11.6 ± 26.0	13.4 ± 28.1	13.5 ± 35.9	0.712‡
Symptom duration, weeks	31.2 ± 30.5	42.2 ± 68.3	33.6 ± 34.5	43.6 ± 76.9	0.773‡
Allocated treatment, no. (%)					0.061§
COBRA classic	42 (30.0)	20 (21.7)	13 (18.3)	9 (30.0)	–
COBRA slim (HR)	33 (23.6)	35 (38.0)	15 (21.1)	5 (16.7)	–
COBRA avant-garde	36 (25.7)	22 (23.9)	19 (26.8)	7 (23.3)	–
Tight step-up	12 (8.6)	8 (8.7)	15 (21.1)	6 (20.0)	–
COBRA slim (LR)	17 (12.1)	7 (7.6)	9 (12.7)	3 (10.0)	–
Women, no. (%)	95 (67.9)	65 (70.7)	45 (63.4)	23 (76.7)	0.570§
Depression, no. (%)	6 (4.3)	6 (6.5)	5 (7.0)	1 (3.3)	0.755§
Alcohol intake, no. (%)¶	88 (62.9)	50 (54.3)	41 (57.7)	12 (40.0)	0.123§
Smoking status (ever), no. (%)	70 (50.0)	58 (63.0)	45 (63.4)	15 (50.0)	0.117§
Employed at screening, no. (%)	68 (48.6)	49 (53.3)	36 (50.7)	16 (53.3)	0.901§
Body mass index, kg/m^2	25.6 ± 3.8	27.0 ± 4.6	27.1 ± 4.2	27.3 ± 4.9	0.037‡
SSL supportive interactions (range 34–136) (n = 324)	79.1 ± 15.7	76.4 ± 15.4	79.0 ± 17.3	80.1 ± 18.6	0.858‡
SSL negative interactions (range 7–28) (n = 324)	9.1 ± 2.9	9.4 ± 3.2	10.9 ± 4.3	9.7 ± 3.5	0.047‡

* Values are the mean ± SD unless indicated otherwise. DAS28-CRP = Disease Activity Score in 28 joints using the C-reactive protein level; 28SJC = 28 swollen joint count; 28TJC = 28 tender joint count; VAS = visual analog scale; PGA = patient global assessment; HR = high-risk treatment group (patients with a poor prognostic profile based on the presence of classical prognostic factors); LR = low-risk treatment group (patients without a poor prognostic profile based on the presence of classical prognostic factors); SSL = Social Support List.
† *P* from 1-way analysis of variance test.
‡ *P* from Kruskal-Wallis test.
§ *P* from chi-square statistic.
¶ Consumption of any form of alcohol.

who enjoyed an early and persistent clinical response reported higher vitality, more positive beliefs about disease consequences, and a higher belief in the effect of treatment at week 52, as compared to patients who had a relapse after the initial response or who had a delayed response or no response at all. Moreover, compared to persistent responders, delayed responders and nonresponders reported more interference with social activities and more problems with work or other daily activities because of emotional problems. Patients with a relapse or no response scored less well on mental health at week 52 compared to persistent responders. Sensitivity analysis using the week-52 PRO scores without EM imputation showed a similar trend in effect sizes and similar study conclusions (see Supplementary Table 3, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23900/abstract).

DISCUSSION

As hypothesized, speed and stability of the initial clinical response were independently associated with aspects of psychosocial functioning at 1 year after RA treatment initiation. In addition, baseline psychosocial characteristics predicted future psychosocial outcomes of patients with early RA, while allocated treatment did not contribute to the variability in patient-reported health and illness perceptions. Evolution of SF-36 scores in the total CareRA patient population did not differ between treatment groups as one might expect based on the similar effect on disease activity in this treat-to-target trial.

It is well-established that there is no time to waste in early RA management. Remarkably, however, limited attention has been given to what the early clinical response profile could mean for long-term outcomes other than disease activity, joint damage, and functionality. Preliminary evidence in RA has shown that time could also be of the essence from the patient's point of view (9–11,32). Our study adds to this evidence by subdefining patients with early RA according to speed and persistence of their clinical response. This methodology allowed us to conclude that patients with a rapid and persistent initial clinical response scored better on aspects of psychosocial functioning after 1 year of treatment. Thus, from the patient's perspective, reaching good disease control is important as well as the clinical trajectory leading to this target. In a previous longitudinal qualitative study, we already demonstrated that a rapid and sustained response was highly valued by patients (8). Similarly,

Table 2. Patient-reported outcomes at trial baseline and at year 1 of treatment for the sample of patients with early rheumatoid arthritis (n = 333), subdivided according to their initial clinical response profile*

Variables	Persistent responders (n = 140)	Secondary failures (n = 92)	Delayed responders (n = 71)	Nonresponders (n = 30)	*P*
SF-36 vitality (range 0–100)					
Baseline (n = 326)	52.9 ± 19.0	46.4 ± 20.2	40.6 ± 19.2	47.0 ± 22.2	<0.001†
Week 52 (n = 333)	64.4 ± 17.0	54.4 ± 17.7	53.8 ± 18.1	46.3 ± 18.0	<0.001†
SF-36 social functioning (range 0–100)					
Baseline (n = 327)	69.1 ± 25.6	60.9 ± 26.3	51.6 ± 27.9	57.6 ± 24.4	<0.001‡
Week 52 (n = 333)	83.0 ± 19.6	74.0 ± 21.2	70.8 ± 20.2	56.4 ± 18.1	<0.001‡
SF-36 role emotional (range 0–100)					
Baseline (n = 324)	69.9 ± 42.9	59.3 ± 42.9	58.5 ± 45.2	61.7 ± 45.0	0.097‡
Week 52 (n = 333)	84.9 ± 29.6	68.7 ± 39.3	61.3 ± 40.2	55.4 ± 37.9	<0.001‡
SF-36 mental health (range 0–100)					
Baseline (n = 327)	65.7 ± 17.3	61.5 ± 18.5	55.8 ± 22.2	63.6 ± 17.9	0.005†
Week 52 (n = 333)	75.2 ± 15.5	67.4 ± 17.3	66.4 ± 16.5	63.4 ± 15.1	<0.001†
IPQ-R consequences (range 6–30)					
Baseline (n = 325)	19.2 ± 4.1	18.9 ± 3.9	21.0 ± 4.4	19.2 ± 4.6	0.010†
Week 52 (n = 333)	15.9 ± 4.2	17.7 ± 4.1	19.2 ± 4.7	20.8 ± 3.7	<0.001†
IPQ-R treatment control (range 5–25)					
Baseline (n = 326)	18.9 ± 2.2	18.7 ± 2.3	18.0 ± 2.7	19.2 ± 2.3	0.055‡
Week 52 (n = 333)	18.7 ± 2.2	17.5 ± 2.4	17.2 ± 2.3	17.3 ± 2.1	<0.001‡
IPQ-R illness coherence (range 5–25)					
Baseline (n = 325)	17.6 ± 3.8	17.1 ± 3.6	16.1 ± 4.9	17.2 ± 3.4	0.175§
Week 52 (n = 333)	18.5 ± 3.4	17.2 ± 3.7	16.9 ± 3.9	17.2 ± 2.7	0.004†

* Values are the mean ± SD unless indicated otherwise. SF-36 = 36-item Short Form health survey; IPQ-R = Revised Illness Perception Questionnaire.
† *P* from 1-way analysis of variance test.
‡ *P* from Kruskal-Wallis test.
§ *P* from Welch's analysis of variance test.

"feeling good as soon as possible for as long as possible" was highlighted in the report of the 10th Outcome Measures in Rheumatology regarding response criteria (33). Although it might seem self-evident that patients with a favorable clinical response profile would also report a beneficial effect for outcomes that matter most to them, unmet needs may also exist in this group requiring further study (34).

Already at baseline, patients differed for DAS28 components other than CRP, depending on their later clinical response profiles, suggesting that disease activity scoring is influenced by a patient-specific reporting style. Recently, Michelsen et al (35) concluded that psychosocial factors (i.e., depression and anxiety) were associated with more subjectively weighted measures of disease activity and not acute-phase reactants and swollen joint count, which was confirmed in an early RA cohort by Boer et al (36). Both studies identified baseline psychosocial distress as a strong negative predictor of remission in early RA. This finding fits with our results, confirming baseline levels as the most powerful predictors of the explored psychosocial outcomes, as previously shown (37,38). Nonetheless, in contrast to these studies, in daily practice, psychosocial variables almost never get the necessary attention in management of disease. Unfortunately, even adherence to guidelines on pharmacologic treatment in RA and the treat-to-target principles is challenging in clinical practice (39).

Awareness is growing that RA pharmacotherapy alone may not be the solution for relevant health outcomes like psychosocial wellbeing (40). Following the definition of precision medicine, treatment should be targeted to a broad range of individual needs, based on genetic, biomarker, phenotypic, and psychosocial characteristics (41). Nevertheless, actual investments are mainly aiming for the development of new diagnostics, prognostics, and therapeutics. Our findings suggest that to benefit from the window of opportunity in all its aspects, timely patient education and counseling might be important as well as personalized support for patients at risk of a more difficult RA trajectory due to personality-related and social aspects. Translating this finding into the therapeutic setting, however, will be a challenge. The response rate in this pragmatic trial illustrates the difficulty of evaluating psychosocial status by means of questionnaires, and a more multidimensional evaluation of newly diagnosed patients, including the chosen measures reported here, can probably only be realized by interdisciplinary teams working toward timely implementation of a wider scope of interventions tailored to patients' needs.

Our results were based on a DAS28-CRP <2.6 and could have been different when using other definitions of disease control. However, applying DAS28-CRP <2.6 resulted in a distribution with good face validity and sufficient patient numbers per response profile. Moreover, in the CareRA trial, the cutoff of DAS28-CRP ≤3.2 was used for therapeutic steering, which results in better effectiveness of the treat-to-target strategy (42),

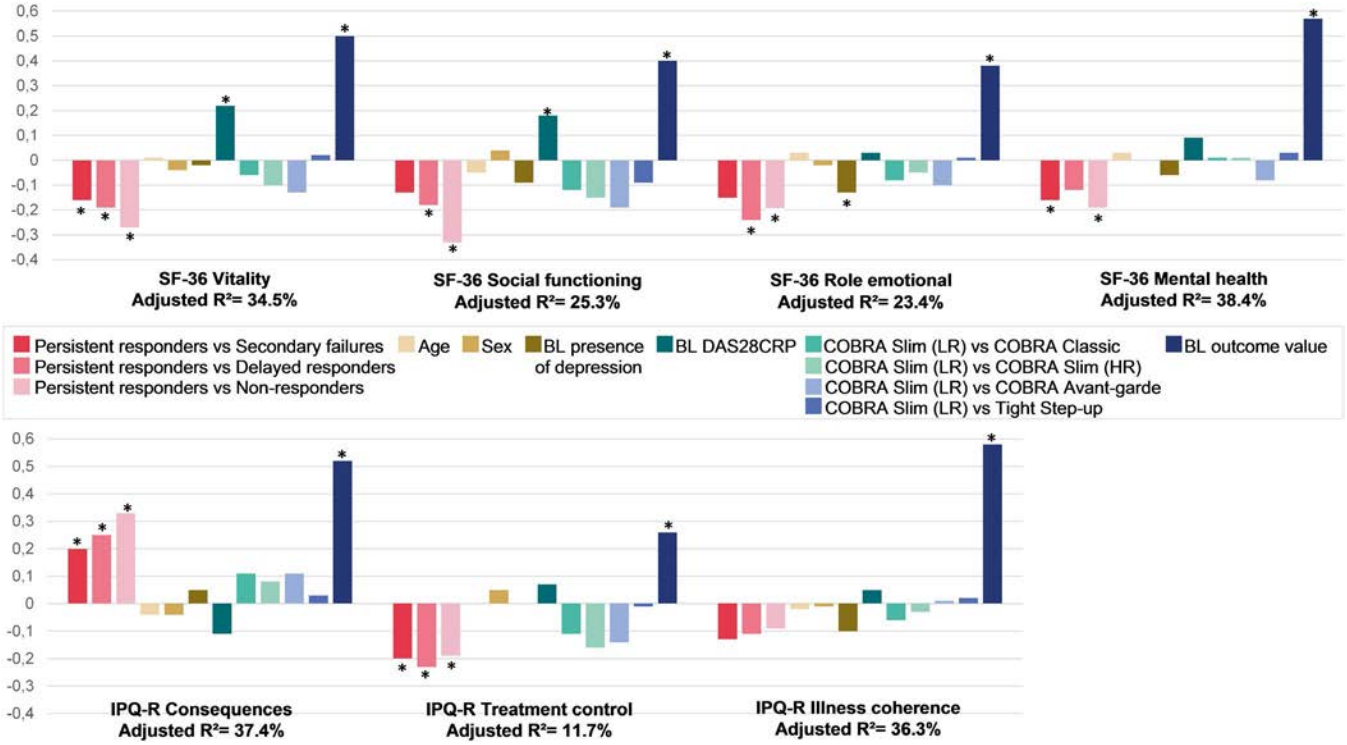

Figure 4. Multivariate linear regression models of the contribution of the initial clinical response and other predictor variables (age, sex, baseline depression, baseline Disease Activity Score in 28 joints using the C-reactive protein level [DAS28-CRP], treatment regimen and baseline outcome value) to psychosocial outcomes after 1 year of early rheumatoid arthritis treatment. Graphs show adjusted standardized beta values and significant *P* values (* = $P < 0.05$ corrected with Holm-Bonferroni method) based on 1,000 bootstrap samples. For details about 95% confidence intervals, see Supplementary Table 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23900/abstract, and for details about the univariate analyses see Supplementary Table 1, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.23900/abstract. SF-36 = Short Form 36 health survey; IPQ-R = Revised Illness Perception Questionnaire; BL = baseline; LR = low-risk; HR = high-risk.

and might correspond more closely to what patients consider to be acceptable. Our study included only particular aspects of the broad concept of psychosocial functioning. For example, work participation, which was added to the most recent treat-to-target recommendations (43), could not reliably be extracted from our data. Because data were collected in a setting mimicking clinical practice, no reminders were sent to encourage patients to return their questionnaires. As in similar studies, we encountered missing PRO data, which might hamper the generalizability of our results. On the other hand, PROs were studied within a pragmatic trial setting close to clinical practice, making study conclusions generally applicable for current disease management. Future studies on the effect of early and sustained disease control on patients' future psychosocial functioning could include even more long-term time points while considering integrating planned missingness in the study design (44), or perhaps computer adaptive testing to improve data completion (45).

The prevalence of depression was low in our early RA sample. There is a risk that we underestimated depression as a potential confounder since we did not formally evaluate patients but used classification criteria with a high specificity but low sensitivity for the diagnosis of depression. We also recognize the potential relevance to our psychosocial outcomes of measures we did not include as confounders, such as socioeconomic status and social support (46,47). However, our study was conducted in Flanders, a region where standards of life are generally good, with a low unemployment rate and only minor socioeconomic inequality. Moreover, the Belgian health care system is easily accessible and includes a comprehensive social security. Eventually, perceived social support was not added as a confounder to the regression models because of the small numerical difference between patients with different clinical response profiles and the lack of evidence about its clinical meaningfulness. A methodologic strength of this study is that we analyzed PRO data by defining clinically relevant response subgroups. For example, our data revealed that experiencing a relapse after a favorable initial response can have implications for future psychosocial functioning and is important to consider in early disease management. Moreover, the patient was at the heart of this study, together with a patient researcher, a condition of clinically useful research (12,48).

In conclusion, this study provides initial evidence in favor of a psychosocial window of opportunity for early RA management. Our data showed that a rapid and sustained

response was independently associated with positive aspects of psychosocial wellbeing at year 1 of treatment. Baseline psychosocial status, however, contributed the most, while treatment type did not. These findings could be a source of inspiration for initiatives to broaden the scope of treat-to-target strategies in early RA.

ACKNOWLEDGMENTS

The authors thank all the patients who were willing to complete the questionnaires during a crucial period of their disease. Furthermore, we thank the local rheumatology staff for data collection, the CareRA team for data monitoring, and Anikó Lovik for providing statistical advice.

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. P. Verschueren 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. Van der Elst, Verschueren, Stouten, Pazmino, De Groef, Moons, Westhovens.

Acquisition of data. Van der Elst, Verschueren, Stouten, Pazmino, De Groef, De Cock, Joly, Moons, Westhovens.

Analysis and interpretation of data. Van der Elst.

REFERENCES

1. Simons G, Mason A, Falahee M, Kumar K, Mallen CD, Raza K, et al. Qualitative exploration of illness perceptions of rheumatoid arthritis in the general public. Musculoskeletal Care 2017;15:13–22.
2. Smolen JS, Landewe R, Bijlsma J, Burmester G, Chatzidionysiou K, Dougados M, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. Ann Rheum Dis 2017;76:960–77.
3. Combe B, Landewe R, Daien CI, Hua C, Aletaha D, Alvaro-Gracia JM, et al. 2016 update of the EULAR recommendations for the management of early arthritis. Ann Rheum Dis 2017;76:948–59.
4. Singh JA, Saag KG, Bridges SL Jr, Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. Arthritis Care Res (Hoboken) 2016;68:1–25.
5. Landewé RB, Boers M, Verhoeven AC, Westhovens R, van de Laar MA, Markusse HM, et al. COBRA combination therapy in patients with early rheumatoid arthritis: long-term structural benefits of a brief intervention. Arthritis Rheum 2002;46:347–56.
6. Markusse IM, Akdemir G, Dirven L, Goekoop-Ruiterman YP, van Groenendael JH, Han KH, et al. Long-term outcomes of patients with recent-onset rheumatoid arthritis after 10 years of tight controlled treatment: a randomized trial. Ann Intern Med 2016;164:523–31.
7. Verschueren P, De Cock D, Corluy L, Joos R, Langenaken C, Taelman V, et al. Effectiveness of methotrexate with step-down glucocorticoid remission induction (COBRA Slim) versus other intensive treatment strategies for early rheumatoid arthritis in a treat-to-target approach: 1-year results of CareRA, a randomised pragmatic open-label superiority trial. Ann Rheum Dis 2017;76:511–20.
8. Van der Elst K, Meyfroidt S, De Cock D, De Groef A, Binnard E, Moons P, et al. Unraveling patient-preferred health and treatment outcomes in early rheumatoid arthritis: a longitudinal qualitative study. Arthritis Care Res (Hoboken) 2016;68:1278–87.
9. Verschueren P, Esselens G, Westhovens R. Predictors of remission, normalized physical function, and changes in the working situation during follow-up of patients with early rheumatoid arthritis: an observational study. Scand J Rheumatol 2009;38:166–72.
10. Kristiansen TM, Primdahl J, Antoft R, Horslev-Petersen K. Everyday life with rheumatoid arthritis and implications for patient education and clinical practice: a focus group study. Musculoskeletal Care 2012;10:29–38.
11. Heimans L, Wevers-de Boer KV, Koudijs KK, Visser K, Goekoop-Ruiterman YP, Harbers JB, et al. Health-related quality of life and functional ability in patients with early arthritis during remission steered treatment: results of the IMPROVED study. Arthritis Res Ther 2013;15:R173.
12. Ioannidis JP. Why most clinical research is not useful. PLoS Med 2016;13:e1002049.
13. Petrie KJ, Jago LA, Devcich DA. The role of illness perceptions in patients with medical conditions. Curr Opin Psychiatry 2007;20:163–7.
14. Groarke A, Curtis R, Coughlan R, Gsel A. The role of perceived and actual disease status in adjustment to rheumatoid arthritis. Rheumatology (Oxford) 2004;43:1142–9.
15. Karademas EC, Dimitraki G, Papastefanakis E, Ktistaki G, Repa A, Gergianaki I, et al. Adaptation to inflammatory rheumatic disease: do illness representations predict patients' physical functioning over time? A complex relationship. J Behav Med 2018;41:232–42.
16. Evers AW, Kraaimaat FW, Geenen R, Jacobs JW, Bijlsma JW. Longterm predictors of anxiety and depressed mood in early rheumatoid arthritis: a 3 and 5 year followup. J Rheumatol 2002;29:2327–36.
17. Verschueren P, De Cock D, Corluy L, Joos R, Langenaken C, Taelman V, et al. Methotrexate in combination with other DMARDs is not superior to methotrexate alone for remission induction with moderate-to-high-dose glucocorticoid bridging in early rheumatoid arthritis after 16 weeks of treatment: the CareRA trial. Ann Rheum Dis 2015;74:27–34.
18. Meyfroidt S, van Hulst L, De Cock D, Van der Elst K, Joly J, Westhovens R, et al. Factors influencing the prescription of intensive combination treatment strategies for early rheumatoid arthritis. Scand J Rheumatol 2014;43:265–72.
19. Meyfroidt S, Van der Elst K, De Cock D, Joly J, Westhovens R, Hulscher M, et al. Patient experiences with intensive combination-treatment strategies with glucocorticoids for early rheumatoid arthritis. Patient Educ Couns 2015;98:384–90.
20. De Cock D, Meyfroidt S, Joly J, Van der Elst K, Westhovens R, Verschueren P, et al. A detailed analysis of treatment delay from the onset of symptoms in early rheumatoid arthritis patients. Scand J Rheumatol 2014;43:1–8.
21. Aletaha D, Funovits J, Keystone EC, Smolen JS. Disease activity early in the course of treatment predicts response to therapy after one year in rheumatoid arthritis patients. Arthritis Rheum 2007;56:3226–35.
22. Van Sonderen E. Het meten van sociale steun met de Sociale Steun Lijst-Interacties (SLL-I) en Sociale Steun Lijst-Discrepanties (SSL-D): een handleiding. Groningen: Noordelijk Centrum voor Gezondheidsvraagstukken, Rijksuniversiteit Groningen; 1993.
23. Heijmans M, Rijken M. Validation of the IPQ-R in a large, representative Dutch sample [abstract]. European Health Psychology Society: European health psychology. Gender, culture and health, Kos, Greece, September 24–27, 2003. p.14.
24. Aaronson NK, Muller M, Cohen PD, Essink-Bot ML, Fekkes M, Sanderman R, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community

and chronic disease populations. J Clin Epidemiol 1998;51:1055–68.

25. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I Conceptual framework and item selection. Med Care 1992;30:473–83.
26. Ware JE Jr, Snow KK, Kosinski M, Gandek B. SF-36 health survey: manual & interpretation guide. Lincoln (RI): QualityMetric; 2005.
27. Moss-Morris R, Weinman J, Petrie KJ, Horne R, Cameron LD, Buick D. The revised illness perception questionnaire (IPQ-R). Psychol Health 2002;17:1–16.
28. Field A. Discovering statistics using IBM SPSS statistics. 4th ed. Los Angeles: Sage Publications; 2013. p. 419–26.
29. Walters SJ, Campbell MJ. The use of bootstrap methods for analysing health-related quality of life outcomes (particularly the SF-36). Health Qual Life Outcomes 2004;2:70.
30. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc 1977;39:1–38.
31. Aickin M, Gensler H. Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm methods. Am J Public Health 1996;86:726–8.
32. Kristiansen TM, Primdahl J, Antoft R, Horslev-Petersen K. It means everything: continuing normality of everyday life for people with rheumatoid arthritis in early remission. Musculoskeletal Care 2012;10:162–70.
33. Strand V, Boers M, Idzerda L, Kirwan JR, Kvien TK, Tugwell PS, et al. It's good to feel better but it's better to feel good and even better to feel good as soon as possible for as long as possible: response criteria and the importance of change at OMERACT 10. J Rheumatol 2011;38:1720–7.
34. Steunebrink LM, Oude Voshaar MA, Taal E, Vonkeman HE, Zijlstra TR, van de Laar M. Determinants of perceived health nonimprovement in early rheumatoid arthritis patients with favorable treatment outcomes. Arthritis Care Res (Hoboken) 2018;70:510–5.
35. Michelsen B, Kristianslund EK, Sexton J, Hammer HB, Fagerli KM, Lie E, et al. Do depression and anxiety reduce the likelihood of remission in rheumatoid arthritis and psoriatic arthritis? Data from the prospective multicentre NOR-DMARD study. Ann Rheum Dis 2017;76:1906–10.
36. Boer AC, Huizinga TW, van der Helm-van Mil AH. Depression and anxiety associate with less remission after 1 year in rheumatoid arthritis. Ann Rheum Dis 2019;78:e1.
37. Benka J, Nagyova I, Rosenberger J, Calfova A, Macejova Z, Middel B, et al. Social support and psychological distress in rheumatoid arthritis: a 4-year prospective study. Disabil Rehabil 2012;34:754–61.
38. Van Lankveld W, Naring G, van 't Pad Bosch P, van de Putte L. The negative effect of decreasing the level of activity in coping with pain in rheumatoid arthritis: an increase in psychological distress and disease impact. J Behav Med 2000;23:377–91.
39. Gvozdenović E, Allaart CF, van der Heijde D, Ferraccioli G, Smolen JS, Huizinga TW, et al. When rheumatologists report that they agree with a guideline, does this mean that they practise the guideline in clinical practice? Results of the International Recommendation Implementation Study (IRIS). RMD Open 2016;2:e000221.
40. Matcham F, Galloway J, Hotopf M, Roberts E, Scott IC, Steer S, et al. The impact of targeted rheumatoid arthritis pharmacological treatment on mental health: a systematic review and network meta-analysis. Arthritis Rheumatol 2018;70:1377–91.
41. Jameson JL, Longo DL. Precision medicine: personalized, problematic, and promising. N Engl J Med 2015;372:2229–34.
42. Verschueren P, Westhovens R. The use of glucocorticoids in early rheumatoid arthritis. Rheumatology (Oxford) 2018;57:1316–7.
43. Smolen JS, Breedveld FC, Burmester GR, Bykerk V, Dougados M, Emery P, et al. Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. Ann Rheum Dis 2016;75:3–15.
44. Little TD, Jorgensen TD, Lang KM, Moore EW. On the joys of missing data. J Pediatr Psychol 2014;39:151–62.
45. Witter JP. The promise of patient-reported outcomes measurement information system. Turning theory into reality: a uniform approach to patient-reported outcomes across rheumatic diseases. Rheum Dis Clin North Am 2016;42:377–94.
46. Puolakka K, Kautiainen H, Mottonen T, Hannonen P, Hakala M, Korpela M, et al. Predictors of productivity loss in early rheumatoid arthritis: a 5 year follow up study. Ann Rheum Dis 2005;64:130–3.
47. Evers AW, Kraaimaat FW, Geenen R, Jacobs JW, Bijlsma JW. Pain coping and social support as predictors of long-term functional disability and pain in early rheumatoid arthritis. Behav Res Ther 2003;41:1295–310.
48. De Wit MP, Berlo SE, Aanerud GJ, Aletaha D, Bijlsma JW, Croucher L, et al. European League Against Rheumatism recommendations for the inclusion of patient representatives in scientific projects. Ann Rheum Dis 2011;70:722–6.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 71, No. 12, December 2019, pp 1576–1582
DOI 10.1002/acr.23801

Qualitative Exploration of Triangulated, Shared Decision-Making in Rheumatoid Arthritis

Pauline Binder-Finnema, Kathryn Dzurilla, Betty Hsiao, and Liana Fraenkel

Objective. Treat-to-target implementation in rheumatoid arthritis (RA) requires a shared decision-making (SDM) process. However, ability to pay is a major determinant of patient choice, but how this factor affects SDM is underexplored.

Methods. Visits at 4 RA clinics during which patients faced a decision to change their treatment were audiotaped between May 2016 and June 2017. Audiotapes were transcribed verbatim and analyzed using qualitative framework analysis.

Results. A total of 156 visits were analyzed. Most patients with RA, except those with effective insurance coverage, had deliberations disrupted or sidelined by third-party insurance providers having power to authorize the preferred disease-modifying antirheumatic drug choice. This triangulated SDM complicated efficiency in deliberations and timely treatment and was a barrier to shared engagement about health risks and symptom improvement typically found in patient-provider dyads.

Conclusion. Rheumatology care providers should aim to incorporate treatment costs and ability to pay into their deliberations so that individualized out-of-pocket estimates can be considered during triangulated SDM at the point-of-care.

INTRODUCTION

Early detection of rheumatoid arthritis (RA) and early intervention with disease-modifying antirheumatic drugs (DMARDs) significantly improve short- and long-term therapeutic outcomes (1–3). Moreover, cardiovascular morbidity and the need for total joint replacement have significantly decreased since aggressive care for RA has become the norm (4). Maintaining tight control of disease activity for all RA patients requires frequent monitoring and adjusting medications to minimize inflammation (5,6). However, numerous contemporary studies on treatment adherence highlight patient-related barriers to DMARD escalation. Patient reluctance to add or change medications is influenced by low motivation or personal beliefs, advanced age, poverty, limited health education, and the inability to cope with high drug costs (2,7–10). Some studies show that patients' reluctance to accept the risks of side effects associated with newer treatments impacts providers' ability to adhere to treat-to-target strategies (11,12). Yet systematic reviews identify cost barriers as the principal factor underlying the reluctance of patients with RA to change treatment, a problem that has worsened as uncertainties related to insurance coverage increase, including deductibles and copays (8,13), as illustrated by the statement of one of the participants in our study.

> Patient: See, I don't dread taking my medicine. My main [concern] is…the insurance company, because you always have to fight with them to get them to give me my etanercept. (patient ID 145; provider ID 5)

Numerous treatment options are now available for patients with RA, and shared decision-making (SDM) is considered best clinical practice. SDM is a process during which patients are informed of all available options and subsequently deliberate with physicians to arrive at treatment choices that are most consistent with their values and priorities (14,15). Despite the apparent influence of cost on patients' adherence to treatment recommendations, few studies have examined this factor at the point-of-care or whether ability to pay influences SDM processes and outcomes.

The drug approval process has been shown in other chronic conditions to be influenced by tiered insurance regulations that may negatively impact preferred treatment choice, SDM, and

Supported by a Rheumatology Research Foundation Innovative Research grant. Dr. Fraenkel's work was supported by the NIH/National Institute of Arthritis and Musculoskeletal and Skin Diseases (grant AR-060231-06).

Pauline Binder-Finnema, PhD, MPP, Kathryn Dzurilla, MSN, AGACNP-BC, APRN, Betty Hsiao, MD, Liana Fraenkel, MD, MPH: Yale University School of Medicine, New Haven, Connecticut.

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

Address correspondence to Pauline Binder-Finnema, PhD, MPP, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 60510. E-mail: Pauline.binder.finnema@gmail.com.

Submitted for publication May 5, 2018; accepted in revised form October 23, 2018.

SIGNIFICANCE & INNOVATIONS

- For patients with limited to no insurance coverage, the focus of deliberation during shared decision-making prioritizes insurance factors over discussions of risk and treatment options and triangulates the decision-making process.
- Triangulated shared decision-making limits the opportunity to embrace assured self-determination and relational autonomy during deliberations, and imposes uncertainty about who takes final responsibility for choice, despite best efforts to involve patients in the treat-to-target process.

timely decision-making (16). O'Connor et al (17) suggest that physicians must confront difficulties in creating a comprehensive treatment plan, despite using SDM, when patients' insurance plans limit options and/or offer no coverage for patient's drug of choice. Although third-party insurance payers and Medicaid reimbursement schemes help cover the high costs of medications, their restrictive drug formularies can decrease rather than expand choice options and have an adverse influence on medication management (18).

The literature on third-party influences and medication communication has advanced in pharmacy research, exploring shared patient/pharmacist decision-making (19). However, the third-party concept in SDM has yet to give widespread attention to the interaction between patients and their physicians.

The objective of this study was to gain deeper understanding of insurance coverage as it relates to treatment escalation in RA, by identifying the perspectives of patients with RA and their rheumatologists at the point-of-contact and while engaged in SDM. The study design supports an assumption that effective deliberation between patients and physicians positively influences decisional determination, i.e., the arrival at a final decision (20). Elwyn and Miron-Shatz (20) explain SDM as gaining sufficient information, appraising new knowledge as sufficient, imagining counterfactual information, and affective (emotional) forecasting, all of which need to be integrated into choosing a treatment option.

PATIENTS AND METHODS

The study was approved by our institution's Human Research Protection Program. Data were collected between May 2016 and June 2017 by an author (PB-F), a medical anthropologist, at 4 academically affiliated rheumatology clinics as part of a broader study on SDM. The parent study involved a 6-month intervention phase in which all patients with RA and their clinicians used an empirically based decision tool meant to encourage open deliberation about medication choices (21). This post-test tool is a paper brochure meant to elicit patient perspectives on such characteristics as onset of action, mode of administration, common and very rare side-effects and quality of life, infection risk, time on market, and treatment affordability. English-speaking patients with RA were approached for consent to be audio-recorded after having been placed in a room by nursing staff for their follow-up rheumatology appointment. The data collection researcher (PB-F) was not in the room at the time of the consultation. The study included 10 attending physcians, 1 advanced practice registered nurse, and 7 rheumatology fellows. All patients initially seen by fellows were also seen by an attending physician. The care providers were recruited together as a group by a different author (LF) after being given information about the study. Participating physicians also gave consent to have their consultations audio-recorded with patients.

The mean ± SD length of time of recorded consultations was 29.9 ± 11.6 minutes in the pretest phase and 25.1 ± 10.7 minutes in the post-test brochure phase. Recordings were transcribed verbatim by an author (KD), a master's degree–level nurse practitioner student with several years of experience assisting in clinical research. Transcripts were initially screened by an author (PB-F) for the presence of DMARD names used during the visits and for patient biometrics and treatment appraisal by an author (LF), a rheumatologist with substantial research experience in the field. Biometrics were validated from the medical record by an author (BH), a rheumatology fellow. A number of transcripts were excluded due to the presence of a coexisting condition with systemic lupus erythematosus or dementia. The mean ± SD disease duration was 12.5 ± 10.2 years, and the majority of patients had moderate disease activity, with a mean ± SD Clinical Disease Activity Index score of 18 ± 9.6 and a mean ± SD Routine Assessment of Patient Index Data score of 12.5 ± 7.6, as described by Hsiao et al (21). Among the participants considered for this study, 153 refused participation and did not give their consent nor were they recorded. The reasons given for withholding consent were wanting to maintain privacy or because they were not feeling well.

In-depth analysis of the transcripts began with reading and rereading by all members of the research team. Data-driven phrase codes for DMARD type, escalation, and cost factors were identified by interrater agreement and placed onto a spreadsheet matrix using Microsoft Excel. Examples of treatment escalation phrase codes included "discussion of DMARD pros and cons," "change of DMARD type was deliberated," "DMARD type," and "patient expressed disfavor of DMARD based on cost." Examples of cost factors included "ability to pay," "insurance coverage," "Medicaid," "copay," and "afford." The matrix was then analyzed using framework analysis (22), which is a method that maintains the hermeneutic perspectives of study participants and interprets those perspectives within the context of a guideline or policy, in this case, the professional ethos in rheumatology to adjust treatment medications

Table 1. Demographic information on patients with rheumatoid arthritis (n = 156) and rheumatology care providers (n = 18)*

	Value
Patients	
Sex	
Female	133 (85.3)
Male	23 (14.7)
Age, years	
26–44	18 (11.5)
45–64	86 (55.1)
65–85	52 (33.3)
Race	
African American/black	26 (16.7)
Asian	1 (0.60)
Latino	16 (10.3)
American Indian/First Nation	3 (1.90)
White	102 (65.4)
Unknown	8 (5.10)
Ethnicity	
Hispanic	29 (18.6)
Non-Hispanic	119 (76.3)
Unknown	8 (5.10)
Employment	
Full time	30 (19.2)
Part time	11 (7.10)
Student	3 (1.90)
Retired	36 (23.1)
Unemployed	41 (26.3)
On disability	30 (19.2)
Unknown	5 (3.20)
Insurance type	
Private	47 (30.1)
Medicare	59 (37.8)
Medicaid	32 (20.5)
Military (Tricare/VA)	6 (3.80)
Indian Health Service	3 (1.90)
Inactive coverage	4 (2.60)
Unknown	5 (3.20)
Rheumatology care providers, no.	
Senior attending physicians	10
Fellows	7
Advanced practice registered nurse	1

* Values are the number (%), unless indicated otherwise. VA = Veterans Administration.

to minimize inflammation and erosive activity using best-practice SDM (14). Constant comparison descriptively identified emerging patterns (23) that were interpreted as support of a preferred DMARD option according to key principles of SDM, i.e., the presence of supported autonomy, self-determination, and option deliberation, during discussions (24).

RESULTS

A total of 156 patient recordings advanced to analysis because the consultation involved statements from the rheumatologist about worsening symptoms and a need to escalate treatment. Participation of rheumatology providers was arbitrary according to their schedule of patients with RA, influencing the relatively low number of patients seen by fellows with a senior attending physician versus the number of patients seen by the senior physicians and advanced practice registered nurse alone. Table 1 shows the overview of participants. The number of female to male participants was greater, at nearly 6-fold, and the total age range was 26–85 years, with 55% of patients ages 45–64 years. The majority of participants were white (65.4%) followed by African American/black (16.7%) and Latino (10.3%). Nearly 19% of total participants identified themselves as Hispanic. An equal number of participants were employed full time as were on disability (19.2%), and 23% were retired. Approximately 30% had private insurance. Those participants with Medicare and/or Medicaid alone comprised the majority of participants, almost 38% and 21%, respectively. Nearly 8% of participants were underinsured or had no known source of insurance.

Insurance coverage and SDM. Certainty about insurance coverage for potential new DMARDs significantly influenced the SDM process (Figure 1). Patient-physician dyads expressing confidence about the patient's insurance coverage experienced advantaged SDM, and deliberations did not appear impacted by issues of affordability. Most patients with drug coverage plans, however, were somewhat less certain about future third-party approval and engaged in contingent SDM with their rheumatologists. Although deliberations did involve the topic of cost, they remained constructive as both patients and physicians expressed confidence their preferred choice would be approved. The patients having authorization restrictions from third-party payers expressed worry

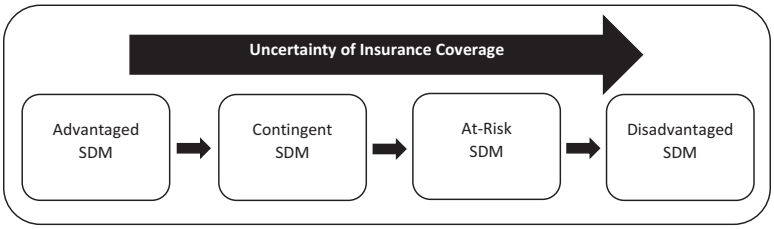

Figure 1. Insurance coverage negatively influences the effectiveness of autonomous shared decision-making (SDM) (**small black arrows**). As cost coverage becomes less certain, the patient-physician dyad becomes less autonomous and increasingly dependent on third-party decision-making for a preferred disease-modifying antirheumatic drug.

that their insurance coverage would limit the treatment options available to them, experiencing at-risk SDM, where the quality of SDM was at risk because choices were restricted. Disrupted SDM occurred for patients having inadequate or no insurance coverage. Physicians were unable to offer options based on medical necessity and deferred discussing appropriate treatment until a payment option was found. There was little deliberation during these visits, despite the need to escalate care.

Advantaged SDM. Approximately one-fourth of patients were actively engaged in employment or education despite their worsening RA symptoms. Most of these patients engaged in in-depth conversations with their physicians about DMARDs, which included detailed information about risks and side-effects and the patients' preferences. Because the majority of patients had secure private insurance coverage, the deliberations involved little to no discussion about DMARD affordability. Also missing was uncertainty about whether a prescription might be denied in favor of the pros and cons of available treatment options. One dyad engaged a 12-minute conversation about possible options that allowed the patient to autonomously express curiosity and ask meaningful questions, with no expressed concerns about costs. The positive deliberation appeared strengthened by the manageability of the DMARD cost.

Doctor: Okay, we will stop the certolizumab. I will give you some information on tofacitinib… Tofacitinib is the name of the [stat-free] drug. Your [employer] health plan is generally pretty good about this stuff. So it is one 5-milligram tablet twice a day.
Patient: Those are actually expensive?
Doctor: Oh, yeah.
Patient: Is it more expensive than injections? I do not know if it is, since it is a pill, I guess I would expect it is not more expensive.
Doctor: They are all expensive.
Patient: Mm.
Doctor: Tofacitinib. Let's see…60, for 1 month: $2,713.
Patient: Oh my God.
Doctor: And 32 cents. That is why insurance exists. (patient ID 67; provider ID 10)

Patients articulated insights into symptoms and treatments, with physicians responding in-kind with open, matter-of-fact information. Importantly, discussions could include options recommended in practice guidelines to each patient. In a few instances, mandated drug lists defined by the patients' insurance carrier restricted choice. Nevertheless, there appeared to be no negative influence of cost on SDM, because the insurance coverage included both physicians' and patients' preferred choices.

Contingent SDM. Other patients with good but less certain insurance coverage had to devote variable amounts of time to discussing the issue of cost.

Patient: And then, I guess, after [we learn it is approved] that it would still be the cost because if I cannot afford it, I cannot take it. So, that would be a factor.
Doctor: Right.
Patient: And then?
Doctor: Right, so… Right. We have to make sure whatever we end up going with, the first thing will be get an authorization. (patient ID 125; provider ID 10)

Yet attitudes remained positive.

Doctor: So we are going to start you on the etanercept and the pharmacy is going to help us with this. They will do our authorization stuff for us. And so, that is that. Etanercept and click [the button], you are good. With those 2 insurance cards, you should be okay.
Patient: Okay. (patient ID 150; provider ID 16 and 8)

Deliberations were supported by presumed adequate insurance coverage. This confidence also afforded time to engage the patient in deeper discussions about options.

Doctor: So infliximab and abatacept. They were before we switched to rituxan, right?
Patient: Yes.
Doctor: So was that an insurance issue that you got the infusion rather than the injectable? It was better covered?
Patient: I guess so. I mean, I did not know I had a choice with the infliximab.
Doctor: Yeah. Well, there are now 5 different medications that work like infliximab.
Patient: Mm.
Doctor: Adalimumab, there is also certolizumab and golimumab, in addition to the infliximab. So they work like infliximab, but each is a little different… But they all do basically the same thing. So you're on Medicare now?
Patient: Yes.
Doctor: Okay. We have to deal with the injectables, because that's the insurance wrinkle here. The injectables go on your Part D plan.
Patient: Okay.
Doctor: You have the Medicare drug plan?
Patient: I have [private insurance] with the drug plan, yeah.
Doctor: Okay. So we will have to find out through the prior authorization process whether it is covered

or whether there are alternatives. Everything else looks pretty good. (patient ID 100; provider ID 10)

Knowing that pre-authorization was required was viewed as less straightforward, and took more time, compared to patients in the advantaged group. Nevertheless, the deliberations still addressed deeper question/answer sessions about DMARD options and about risks and side-effects.

Doctor: And then as far as adding another medicine to the methotrexate and sulfasalazine… Usually the first one we go to would be something like an etanercept.
Patient: Yeah.
Doctor: It is a type of medicine that suppresses part of your immune system that causes inflammation.
Patient: Mm…but is it dangerous with all those [side-effects]?
Doctor: The biggest risk is infection.
Patient: Right. All that stuff makes me nervous, but I just don't know why I'm not getting better.
Doctor: Yeah, the disease is just very active right now. It is really common. A lot of people need this kind of combination, need something like etanercept.
Patient: So you recommend it.
Doctor: I would give it a try.
Patient: I just want something safe. I'm afraid of all those side effects.
Doctor: I would probably recommend, you know, trying the etanercept first… Tofacitinib [costs] a little bit more. What we usually do is we send it to the outpatient pharmacy, and then they get the authorization for you. We could try the tofacitinib first and see if your insurance will cover that. (patient ID 121; provider ID 15)

At-risk SDM. Uncertain insurance coverage threatened high-quality SDM. However, some patients were not aware of gaps in coverage or high deductibles, or how these factors could limit their choices, which contributed to uncertainty about whether a preferred DMARD would be approved.

Doctor: Tofacitinib is very expensive.
Patient: Oh, it is? Will my insurance cover it?
Doctor: It should. But, you know, it will throw you right into a donut hole.
Patient: What do you mean? (patient ID 114; provider ID 12)

Because cost overshadowed these deliberations, physicians reduced time from presenting appropriate DMARDs or discussing benefits and risks to explain the Medicaid coverage gap. The following dialog is from a 6-minute conversation in which the physician ended up providing literature instead of engaging in SDM.

Doctor: I will give you some literature, maybe on that hydroxychloroquine, and let's see what the apothecary says about the etanercept copay. I will try to research to see if there is anything else, any other assistance programs to see if there is anything else… Let's continue etanercept for now.
Patient: No, I am just, I am a little nervous about what that is going to cost me, if I am in the donut hole now.
Doctor: Yeah. So, when you are usually in the donut hole, do you usually come out of it eventually after a few months? Like once you paid for your medications?
Patient: Once I get up to my limit.
Doctor: $6,800 you said it was?
Patient: Yeah, up to $6,800.
Doctor: Yeah, yeah. That just stinks. I am sorry you have to go through that. I wish things were different. (patient ID 99; provider ID 10)

Uncertain insurance coverage restricted autonomy and choice, as illustrated in the passage below.

Doctor: Right. And then, did you also lose insurance coverage for the etanercept?
Patient: No, but I think it is after 65 or 66, it is not covered anymore. There is no [insurance].
Doctor: Oh, because of the…
Patient: Medicaid.
Doctor: Okay.
Patient: You know, so I would have to pay full price for that, and I cannot afford it.
Doctor: Right, right.
Patient: $1,000 a month, if not more.
Doctor: If we could get you the… so you did well on the etanercept.
Patient: Yup.
Doctor: Okay, would you be, if insurance issue was not an issue, would you be willing to go back on that?
Patient: Oh yeah. I did not want to go off of it. (patient ID 10; provider ID 15)

For participants with uncertain, limited insurance coverage, the problem of cost dominated available consultation time, deprioritizing effective deliberation about risks, drug side-effects, or the patient's concerns about treatment escalation. The need to defer to third-party authorization initiated the consultations rather than being a tentative end-product of deliberation.

Disrupted SDM. SDM and deliberative choice-making were essentially absent for patients with little or no drug coverage. This disruption influenced delivery of care by eliminating choice and

deliberation about competing harms and benefits. In these cases, the ultimate choice was fully dependent upon the third-party payer.

Doctor: So, I want you to please, please hear me, and think about going to a TNF inhibitor.
Patient: I'm really trying.
Doctor: Etanercept or adalimumab.
Patient: Listen, when my pharmacist called me I was considering that I might take it, but he said it was gonna cost me $1,700 a month. He asked, "Do you really want me to fill this prescription?"
Patient companion: Doesn't insurance cover any of that?
Doctor: Of course, it does.
Patient: My pharmacist actually called me. I was shocked. I thought, I don't think so.
Doctor: [We will] do an etanercept pre-authorization. We specifically look to see if those side effects happen, and then we decrease the dose of methotrexate or stop it. We've had to do that once.
Patient: No, I decreased it myself.
Doctor: You decreased it yourself…
Patient: (laughs)
Doctor: Actually, if you think about it, [etanercept] is a really much better option for you. I can't push your methotrexate [because of the liver problem].
Patient: Okay, I'm interested, in what that insurance coverage is. Cause that was a lot of money. (patient ID 94; provider ID 1 and 8)

Whereas physicians typically entered the consultation room asking about how a patient was doing and their levels of pain and potential treatment options, the conversation was soon hijacked by restrictions imposed by cost. Some physicians took a relatively large portion of consultation time to encourage and educate patients on how to sort out their financial issues so they could access treatment.

Doctor: And they would all be available to you based on your insurance. We are going to check the insurance thing first.
Patient: What insurance? I had to drop my [insurance]. But I think I am getting a prescription plan from Medicaid. But I had to drop the secondary insurer because I need to see this [other doctor regularly for another condition].
Doctor: Oh okay. Well then, maybe we should talk about that. Do you not have a drug plan?
Patient: I don't.
Doctor: If you don't have a Part D plan that covers tofacitinib, then you may not be able to get it because obviously you won't be able to afford it out of pocket.
Patient: Oh, yeah? How much is it out of pocket? Too much?
Doctor: I would say something on the order of $3,000 to $4,000 a month. I said that with a straight face, but it's not funny.
Patient: No, it isn't. (patient ID 96; provider ID 10)

Emphasis on affordability shortened visits and appeared to negatively influence deliberations. Those participants experiencing disrupted SDM also appeared to have lower expectations for being able to identify a suitable means of escalating care.

DISCUSSION

SDM for treatment escalation in patients with RA can be negatively influenced by uncertain, or inadequate, insurance coverage. As the need to deliberate about insurance factors increases, the more intrusive and triangulated the decision-making becomes. In a triangulated scenario, a third party, e.g., a health insurance payer, significantly contributes to the final treatment decision and has authority to prevent access to the preferred choice. Triangulated SDM created a barrier that limited the ability of the rheumatologists to fully implement SDM.

Triangulated SDM identifies new areas for policy debate about the influence of third-party decision-making on patient adherence and might also influence the patients' perspectives about who takes final responsibility for the treatment choice. The latter is shown by Nota et al (25) to be the physician. Our findings suggest the third-party decision-maker is conceptually situated above both patient and rheumatologist during the clinical encounter, acting as a top-down decisive voice. This influence can significantly and adversely impose on the physicians' best efforts to involve patients in the treat-to-target process.

There is also an apparent danger in the possibility that both patients and clinicians will feel ignored, which is antithetical to patient-centered care and can erode trust (24,26). In addition, triangulated decision-making is routinely required for some biologic DMARDs but may erode the foundations of health equity for patients with uncertain or limited insurance coverage, because decision interference occurs before real deliberation about treatment choices takes place.

The necessity of adequate, stable, and predictable insurance coverage defines effective decisional deliberation and, by proxy, best clinical practice. Appeals for greater cost transparency as it supports effective SDM at the point-of-care should be made. One-size-fits-all managed-care drug formularies may lack necessary sensitivity to capture individual RA patient needs, which these findings suggest can seriously dismantle SDM. Our findings support patient uncertainty about affordability as a problem that can substantially work against effective SDM during treatment deliberations and can upend the overall quality of the clinical encounter in RA.

One limitation of this study is that the post-test analysis included only data from patients with RA confronting escalation of treatment, and the opportunity to evaluate the brochure for nonbiologics was missed. Nevertheless, the randomization procedure of asking all patients to participate may still be considered a strength, mainly because patients were not preselected, with unknown insight into their current disease status. In addition, the final treatment choice was not assessed, which is a potential limitation, since better follow-up as to whether the medication was approved by the insurance payor could have provided an accurate estimate of out-of-pocket costs during the SDM process (27). This qualitative exploration nevertheless offers a baseline for future studies on triangulated SDM in other settings to demonstrate coherence with and validate its findings (23).

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. Binder-Finnema 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. Binder-Finnema, Hsiao, Fraenkel.
Acquisition of data. Binder-Finnema, Dzurilla, Fraenkel.
Analysis and interpretation of data. Binder-Finnema, Dzurilla, Hsiao, Fraenkel.

REFERENCES

1. Wong PK. Medication adherence in patients with rheumatoid arthritis: why do patients not take what we prescribe? Rheumatol Int 2016;36:1535–42.
2. Goh H, Kwan YH, Seah Y, Low LL, Fong W, Thumboo J. A systematic review of the barriers affecting medication adherence in patients with rheumatic diseases. Rheumatol Int 2017;37:1619–28.
3. Mena-Vazquez N, Manrique-Arija S, Yunquera-Romero L, Ureña-Garnica I, Rojas-Gimenez M, Domic C, et al. Adherence of rheumatoid arthritis patients to biologic disease-modifying antirheumatic drugs: a cross-sectional study. Rheumatol Int 2017;37:1709–18.
4. Cordtz RL, Hawley S, Prieto-Alhambra D, Højgaard P, Zobbe K, Overgaard S, et al. Incidence of hip and knee replacement in patients with rheumatoid arthritis following the introduction of biological DMARDs: an interrupted time-series analysis using nationwide Danish healthcare registers. Ann Rheum Dis 2018;77:684–9.
5. Grigor C, Capell H, Stirling A, McMahon AD, Lock P, Vallance R, et al. Effect of a treatment strategy of tight control for rheumatoid arthritis (the TICORA study): a single-blind randomised controlled trial. Lancet 2004;364:263–9.
6. Hetland ML, Hørslev-Petersen K. The CIMESTRA study: intra-articular glucocorticosteroids and synthetic DMARDs in a treat-to-target strategy in early rheumatoid arthritis. Clin Exp Rheumatol 2012;30 Suppl:S44–9.
7. Adams RJ. Improving health outcomes with better patient understanding and education. Risk Manag Healthc Policy 2010;3:61–72.
8. Van Den Bemt BJ, Zwikker HE, van den Ende CH. Medication adherence in patients with rheumatoid arthritis: a critical appraisal of the existing literature. Expert Rev Clin Immunol 2012;8:337–51.
9. Soini EJ, Hallinen TA, Puolakka K, Vihervaara V, Kauppi MJ. Cost-effectiveness of adalimumab, etanercept, and tocilizumab as first-line treatments for moderate-to-severe rheumatoid arthritis. J Med Econ 2012;15:340–51.
10. Pascual-Ramos V, Contreras-Yáñez I. Motivations for inadequate persistence with disease modifying anti-rheumatic drugs in early rheumatoid arthritis: the patient's perspective. BMC Musculoskelet Disord 2013;14:336.
11. Harrold LR, Reed GW, John A, Barr CJ, Soe K, Magner R, et al. Cluster-randomized trial of a behavioral intervention to incorporate a treat-to-target approach to care of US patients with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2018;70:379–87.
12. Constantinescu F, Goucher S, Weinstein A, Smith W, Fraenkel L. Understanding why rheumatoid arthritis patient treatment preferences differ by race. Arthritis Rheum 2009;61:413–8.
13. Vangeli E, Bakhshi S, Baker A, Fisher A, Bucknor D, Mrowietz U, et al. A systematic review of factors associated with non-adherence to treatment for immune-mediated inflammatory diseases. Adv Ther 2015;32:983–1028.
14. Singh JA, Saag KG, Bridges SL Jr, Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. Arthritis Care Res (Hoboken) 2015;67:1–25.
15. Smolen JS, Landewé R, Breedveld FC, Dougados M, Emery P, Gaujoux-Viala C, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs. Ann Rheum Dis 2010;69:964–75.
16. Bourdette DN, Hartung DM, Whitham RH. Practices of US health insurance companies concerning MS therapies interfere with shared decision-making and harm patients. Neurol Clin Pract 2016;6:177–82.
17. O'Connor AM, Llewellyn-Thomas HA, Flood AB. Modifying unwarranted variations in health care: shared decision making using patient decision aids. Health Aff (Millwood) 2004;23 Suppl:VAR63.
18. Chewning B, Sleath B. Medication decision-making and management: a client-centered model. Soc Sci Med 1996;42:389–98.
19. Murad MS, Chatterley T, Guirguis LM. A meta-narrative review of recorded patient–pharmacist interactions: exploring biomedical or patient-centered communication? Res Social Adm Pharm 2014;10:1–20.
20. Elwyn G, Miron-Shatz T. Deliberation before determination: the definition and evaluation of good decision making. Health Expect 2010;13:139–47.
21. Hsiao B, Binder-Finnema P, Nowell WB, Michel G, Wiedmeyer C, Fraenkel L. Preference phenotypes in support of shared decision-making at point-of-care for patients with rheumatoid arthritis: a proof-of-concept study. Arthritis Care Res (Hoboken) 2019;71:629–37.
22. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Huberman AM, Miles MB, editors. The qualitative researcher's companion. Thousand Oaks (CA): SAGE Publications; 2002. p. 305–29.
23. Lincoln YS, Guba EG. Naturalistic inquiry. Newbury Park (CA): SAGE Publications; 1985.
24. Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: a model for clinical practice. J Gen Intern Med 2012;27:1361–7.
25. Nota I, Drossaert CH, Taal E, van de Laar MA. Arthritis patients' motives for (not) wanting to be involved in medical decision-making and the factors that hinder or promote patient involvement. Clin Rheumatol 2016;35:1225–35.
26. Berrios-Rivera JP, Street RL Jr, Garcia Popa-Lisseanu MG, Kallen MA, Richardson MN, Janssen NM, et al. Trust in physicians and elements of the medical interaction in patients with rheumatoid arthritis and systemic lupus erythematosus. Arthritis Rheum 2006;55:385–93.
27. Meluch AL, Oglesby WH. Physician–patient communication regarding patients' healthcare costs in the US: a systematic review of the literature. J Commun in Healthc 2015;8:151–160.

Arthritis Care & Research
Vol. 71, No. 12, December 2019, pp 1583–1592
DOI 10.1002/acr.23820

American College of Rheumatology
Empowering Rheumatology Professionals

Impact of Cyclic Citrullinated Peptide Antibody Level on Progression to Rheumatoid Arthritis in Clinically Tested Cyclic Citrullinated Peptide Antibody–Positive Patients Without Rheumatoid Arthritis

Julia A. Ford,[1] Xinyi Liu,[1] Allison A. Marshall,[2] Alessandra Zaccardelli,[1] Maria G. Prado,[1] Charlene Wiyarand,[1] Bing Lu,[1] Elizabeth W. Karlson,[1] Peter H. Schur,[1] Kevin D. Deane,[3] and Jeffrey A. Sparks[1]

Objective. To investigate the risk of progression to rheumatoid arthritis (RA) in patients who were cyclic citrullinated peptide (CCP) antibody positive without RA at initial presentation.

Methods. We performed a retrospective cohort study of CCP+ individuals seen at a US tertiary care system between 2009 and 2018 who were without RA or other systemic rheumatic disease by medical record review at the time of CCP antibody positivity. Progression to classifiable RA was determined through medical record review. We investigated the risk of progression to RA overall and stratified by CCP antibody level (low: >1 to 2× the upper limit of normal [ULN]; medium: >2 to 3× ULN; high: >3× ULN). Multivariable Cox regression estimated the hazard ratio (HR) and 95% confidence interval (95% CI) for RA by CCP antibody level.

Results. We identified 340 CCP+ patients who were without RA or other rheumatic disease at baseline. During 1,047 person-years of follow-up, 73 patients (21.5%) developed RA. The risk of progression to RA increased with CCP antibody level, with 46.0% (95% CI 34.7–55.3) of patients with high-level CCP antibodies progressing to RA by 5 years. Compared to low CCP antibody level, medium (HR 3.00 [95% CI 1.32–6.81]) and high (HR 4.83 [95% CI 2.51–9.31]) CCP antibody levels were strongly associated with progression to RA, adjusting for age, sex, body mass index, smoking, family history of RA, and rheumatoid factor level.

Conclusion. Among CCP+ patients without RA, the risk for progression to RA increased substantially with increasing CCP antibody level. This study provides further support for close monitoring for development of RA among CCP+ patients and identifying strategies to mitigate this risk.

INTRODUCTION

Rheumatoid arthritis (RA) develops through preclinical phases prior to onset of classifiable RA (1). Previous studies have demonstrated the presence of RA-specific antibodies like cyclic citrullinated peptide (CCP) antibody in the serum several years prior to RA onset (2–4). To date, much of what is known about CCP+ individuals without classifiable RA comes from blood bank studies (5–7), studies of unaffected family members of patients with RA (8–15), and cohort studies of patients recruited from European arthralgia clinics (16–19).

A prospective cohort study (16) of undifferentiated arthritis patients showed that CCP+ status was a significant risk factor for RA compared to CCP antibody negativity. Similarly, another prospective cohort study of seropositive arthralgia patients (17) showed that CCP+ status predicted arthritis development compared to being CCP– and that arthritis risk increased with a high level of CCP antibodies. While these European cohort studies have been instrumental in enhancing knowledge of how RA develops, the findings may not be generalizable to the US (where early arthritis or arthralgia clinics are uncommon), and typically the studies were performed only among patients with undifferentiated arthritis

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or Harvard University or its affiliated academic health care centers.

Dr. Sparks' work was supported by the National Institute of Arthritis and Skin and Musculoskeletal Diseases (grants K23-AR-069688, L30-AR-066953, P30-AR-070253, and P30-AR-072577).

[1]Julia A. Ford, MD, Xinyi Liu, MS, Alessandra Zaccardelli, BS, Maria G. Prado, MPH, Charlene Wiyarand, BS, Bing Lu, MD, DrPH, Elizabeth W. Karlson, MD, MS, Peter H. Schur, MD, Jeffrey A. Sparks, MD, MMSc: Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; [2]Allison A. Marshall, MPH: Brigham and Women's Hospital, Harvard Medical School, and Tufts University School of Medicine, Boston, Massachusetts; [3]Kevin D. Deane, MD, PhD: University of Colorado, Denver, Aurora.

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

Address correspondence to Julia A. Ford, MD, Department of Medicine, Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, 60 Fenwood Road, Boston, MA 02115. E-mail: jrford@partners.org.

Submitted for publication September 19, 2018; accepted in revised form December 11, 2018.

SIGNIFICANCE & INNOVATIONS

- We performed a retrospective cohort study of cyclic citrullinated peptide (CCP) antibody–positive patients without systemic rheumatic disease, including rheumatoid arthritis (RA), to investigate progression to RA in this population.
- Overall risk of progression to RA in the study sample was 21.5% during a median of 2.7 years of follow-up.
- Patients with CCP antibody levels 3-fold or higher than normal had a 5-fold increased RA risk compared to patients with low-level CCP antibody positivity (between 1- and 2-fold higher than normal). Approximately 46% of patients with high CCP antibody levels progressed to RA within 5 years.
- These results quantify the risk of RA associated with an elevated CCP antibody level and other clinical characteristics and provide a rationale for close monitoring of CCP+ patients for progression to RA.

or arthralgias who agreed to participate in research. Furthermore, prior studies compared the presence of CCP antibodies to the absence of CCP antibodies, so less is known about the effect of CCP antibody level among a population who are all CCP+.

Therefore, we aimed to investigate the risk for progression to RA in a clinical population in the US of CCP+ individuals without classifiable RA at the time of initial CCP antibody positivity. We first aimed to quantify the absolute risk of progression to RA among these patients. We then aimed to identify predictors at the time of initial CCP+ status for subsequent progression to RA. We hypothesized that increasing CCP antibody levels and the presence of other arthritis-related traits would increase the risk for progression to RA.

SUBJECTS AND METHODS

Study design and population. We performed a retrospective cohort study among outpatients or inpatients seen at Partners HealthCare, a tertiary health care system in Boston, Massachusetts. In August 2016, we queried the Partners Research Patient Data Registry, a research repository of all patients seen at Partners hospitals since 1990, to identify all individuals who tested positive for CCP antibody (greater than the upper limit of normal [ULN] of the laboratory assay) between 2009 and 2016. All aspects of this study were approved by the Partners HealthCare Institutional Review Board.

To be included in the study, CCP+ individuals age ≥18 years had to be free of RA or other systemic rheumatic diseases at the index date, defined as the date of the first positive CCP antibody result in the medical record. RA status at the index date according to 2010 American College of Rheumatology (ACR)/European League Against Rheumatism (EULAR) criteria was determined by medical record review (20). We excluded patients with other systemic rheumatic diseases at the index date: systemic lupus erythematosus, scleroderma, spondyloarthritis (including ankylosing spondylitis, reactive arthritis, psoriatic arthritis), antiphospholipid syndrome, mixed connective tissue disease, Sjögren's syndrome, systemic vasculitis, polymyalgia rheumatica, dermatomyositis, polymyositis, and juvenile idiopathic arthritis. We performed an initial brief medical record screen to filter out these conditions and then performed a more detailed review to confirm eligibility. Conditions permitted in the study at the index date were gout, pseudogout, osteoarthritis, inflammatory bowel disease, psoriasis, fibromyalgia, and palindromic rheumatism. We required sufficient detail in the medical record at the index date related to possible RA and ≥1 follow-up visit for individuals to be included in the study. If a diagnosis of RA was made within 28 days of the index date, the individual was considered as having prevalent RA at the index date and was excluded. All individuals included in the study were reviewed independently by 2 rheumatologists (JAF, JAS), who agreed that all analyzed patients were without RA or systemic rheumatic disease at the index date. Selection of the final analyzed study sample is shown in Figure 1.

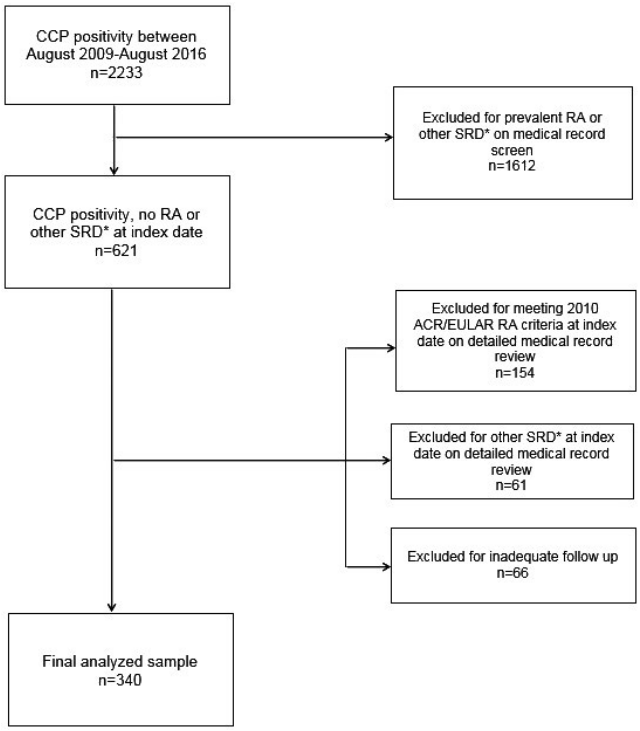

Figure 1. Flow diagram of the study sample. CCP = cyclic citrullinated peptide; RA = rheumatoid arthritis; SRD = systemic rheumatic disease; ACR/EULAR = American College of Rheumatology/European League Against Rheumatism. * = other systemic rheumatic disease: systemic lupus erythematosus, scleroderma, dermatomyositis, polymyositis, seronegative spondyloarthropathies (including ankylosing spondylitis, reactive arthritis, psoriatic arthritis), antiphospholipid syndrome, mixed connective tissue disease, Sjögren's syndrome, systemic vasculitis (including Takayasu arteritis, giant cell arteritis, polyarteritis nodosa, Behçet's disease, IgA vasculitis purpura, antineutrophil cytoplasmic antibody–associated vasculitis), polymyalgia rheumatica, dermatomyositis, and juvenile idiopathic arthritis.

Primary exposure: CCP antibody level. The primary exposure was CCP antibody level at the index date, measured as the fold increase above the ULN of the assay used. Because the study population drew from different hospitals whose laboratories used different CCP assays over time, the ULN of assays varied by site and year. Therefore, we standardized all CCP antibody results by dividing by the ULN of the assay used to obtain the fold above the ULN. We stratified the CCP antibody level as low (>1 to 2× ULN), medium (>2 to 3× ULN), and high (>3× ULN) as clinically relevant cut points based on prior literature (21).

Outcome: RA diagnosis. The primary outcome was incident diagnosis of RA meeting 2010 ACR/EULAR criteria and occurring after the index date, as determined by medical record review. All cases of incident RA and date of RA diagnosis were adjudicated by 2 rheumatologists (JAF, JAS).

Presence and date of diagnosis of other systemic rheumatic diseases were identified and reviewed independently by both rheumatologists, and the date of death was recorded. We also recorded the date of the last clinical follow-up, defined as the last clinical note from any physician (regardless of specialty) with a problem and medication list. Medical records were reviewed for all clinical follow-up as of February 2018.

Covariates. We identified possible confounders based on their associations with CCP and RA in prior literature (9,22–26). Data were collected as of the index date by medical record review. Body mass index (BMI as kg/m^2) was calculated based on measured height and weight through clinical care within 6 months of the index date and further categorized as <25, 25–29.9, or ≥30. Race was dichotomized as white or nonwhite and education level as high school or less versus some college or higher. Smoking status by medical record review was categorized as current smoker within 1 year of the index date versus not current smoker (including never smoker, former smoker, or unknown smoking status). Family history (first-and/or second-degree relatives) was categorized as present versus absent family history of RA (the latter included unknown family history). We recorded rheumatoid factor (RF) and antinuclear antibody levels that were measured clinically within 1 year of the index date. The erythrocyte sedimentation rate and C-reactive protein level were recorded if they were measured clinically within 3 months of the index date. Comorbidities (hypertension, thyroid disease, interstitial lung disease, and osteoarthritis) were ascertained from the medical record.

Secondary exposures. We collected data on factors important in care delivery in these patients or early symptoms or signs of RA. We collected the reason for ordering the initial CCP test as described by the ordering clinician (categorized as arthralgia, lung disease, abnormal laboratory result, axial pain, fatigue, fever, or other clinical conditions). We noted the specialty of the ordering physician (rheumatologist versus nonrheumatologist), duration of joint symptoms (when available), and diagnosis of palindromic rheumatism (defined as diagnosis per treating physician of any specialty) prior to the index date. Symptoms (pain, stiffness, or both reported by the patient) and signs (swelling, tenderness, or both noted by the treating physician) in RA-specific and non-RA–specific joints at the index date were also collected through medical record review. RA-specific joints were defined as metacarpophalangeal joints, proximal interphalangeal joints, metatarsophalangeal joints, thumb interphalangeal joints, wrists, and elbows, as previously defined (27). Non-RA–specific peripheral joints were defined as shoulders, knees, and ankles.

We performed additional medical record review in the subset of patients whose reason for testing CCP was for lung disease. The nature of lung disease at the index date (interstitial lung abnormalities, obstructive disease, nodules/lesions, or other) was determined by medical record review and agreed on by 2 rheumatologists (JAF, JAS).

Statistical analysis. We reported baseline characteristics among the entire study sample and stratified by subsequent progression or nonprogression to RA using descriptive statistics: mean ± SD for normal continuous variables, median and interquartile range (IQR) for non-normal continuous variables, and frequency and proportion for categorical variables. We tested for statistical differences between RA progressors and nonprogressors using univariate tests (*t*-test for normally distributed continuous variables, Wilcoxon's rank sum test for non-normally distributed continuous variables, the chi-square test for categorical variables, and Fisher's exact test for categorical variables with low cell size). We also reported these baseline characteristics in the subset of patients whose reason for ordering CCP testing was for lung disease. We reported the reason for ordering the initial CCP test in all patients and stratified by subsequent progression or nonprogression to RA. We tested for statistical differences between RA progressors and nonprogressors in this subset using Fisher's exact test.

We created Kaplan-Meier curves to visualize RA-free survival after the index date according to CCP antibody level, CCP antibody level and RF status, CCP antibody level and family history of RA, and CCP antibody level and the presence of symptoms in RA-specific joints. We used log-rank tests to test for statistical differences between Kaplan-Meier curves.

We calculated the absolute risk of progressing to RA at fixed time intervals of 1 year, 3 years, and 5 years, due to their use in prior literature (28). The absolute risk of progression to RA was calculated in all patients and stratified by CCP antibody level (low, medium, high), RF status (negative or not sent, low/medium positive, high positive), family history of RA, and the presence of symptoms in RA-specific joints. We also calculated the absolute risk of RA by combinations of dichotomized CCP antibody level (low/medium versus high) and RF status, family history, or symptom status. We obtained estimates for RA risk and 95% confidence interval (95% CI) bounds at each time point using the Kaplan-Meier curves.

We used Cox proportional hazards models to investigate the risk for progression to RA by CCP antibody level. Person-time accrued from the start of the index date. Censoring variables were according to the following dates, whichever came first: incident RA diagnosis (the primary outcome), diagnosis of other systemic rheumatic disease, death, or last documented follow-up note in the electronic medical record (end of follow-up). Therefore, we determined whether or not a patient had RA at all follow-up time analyzed. Initial models for CCP antibody levels, potential confounders, and the secondary exposures were unadjusted to obtain hazard ratios (HRs) and 95% CIs. We used Cox regression to estimate the effect of CCP antibody level on RA risk independent of potential confounders of age, sex, BMI, smoking, family history, and RF level, chosen based on prior literature (9,22–26). We did not include the following variables for the multivariable model, because we considered them to be related to early

Table 1. Characteristics at index date (initial positive CCP result) in the entire study sample and stratified by those who later progressed or did not progress to RA (n = 340)*

Characteristic	All patients (n = 340)	Progressed to RA (n = 73)	Nonprogressors (n = 267)	*P*
Demographics				
Age, mean ± SD years	55.0 ± 15.3	53.0 ± 14.2	55.6 ± 15.6	0.19
Female	223 (65.6)	56 (76.7)	167 (62.5)	0.02
White	254 (74.7)	49 (67.1)	205 (76.8)	0.09
Some college education or higher	148 (43.5)	25 (34.2)	123 (46.1)	0.07
Laboratory data				
RF level by category				
Not sent	14 (4.1)	0 (0.0)	14 (5.2)	<0.0001
Negative	242 (71.2)	39 (53.4)	203 (76.0)	<0.0001
>1 to 3× ULN (low/medium positive)	38 (11.2)	13 (17.8)	25 (9.4)	<0.0001
>3× ULN (high positive)	46 (13.5)	21 (28.8)	25 (9.4)	<0.0001
CCP antibody level × ULN, mean ± SD	4.2 ± 4.7	7.5 ± 6.9	3.3 ± 3.5	<0.0001
CCP antibody level by category				
>1 to 2× ULN (low positive)	167 (49.1)	14 (19.2)	153 (57.3)	<0.0001
>2 to 3× ULN (medium positive)	52 (15.3)	10 (13.7)	42 (15.7)	<0.0001
>3× ULN (high positive)	121 (35.6)	49 (67.1)	72 (27.0)	<0.0001
ANA titer				
Not sent	39 (11.5)	6 (8.2)	33 (12.4)	0.31
Negative	04 (30.6)	18 (24.7)	86 (32.2)	0.31
1:40–1:160	147 (43.2)	35 (47.9)	112 (41.9)	0.31
>1:160	50 (14.7)	14 (19.2)	36 (13.5)	0.31
ESR, mean ± SD mm/hour (n = 269)†	22.5 ± 22.7	23.4 ± 20.6	22.2 ± 23.3	0.36
CRP, mean ± SD mg/liter (n = 231)†	14.2 ± 38.3	16.5 ± 33.4	13.6 ± 39.6	0.03
Lifestyle and family history				
Current smoker	40 (11.8)	14 (19.2)	26 (9.7)	0.03
Body mass index, kg/m^2				
<25	136 (40.0)	30 (41.1)	106 (39.7)	0.97
25 to <30	101 (29.7)	21 (28.8)	80 (30.0)	0.97
≥30	103 (30.3)	22 (30.1)	81 (30.3)	0.97
Family history of RA	50 (14.7)	18 (24.7)	32 (12.0)	0.01
Comorbidities				
Hypertension	129 (37.9)	21 (28.8)	108 (40.4)	0.07
Osteoarthritis, any joint	91 (26.8)	16 (21.9)	75 (28.1)	0.29
Hypothyroidism or hyperthyroidism	40 (11.8)	10 (13.7)	30 (11.2)	0.56
Interstitial lung disease	28 (8.3)	0 (0.0)	28 (10.5)	0.004
Clinical presentation				
CCP ordered by rheumatologist	159 (46.8)	44 (60.3)	115 (43.1)	0.01
Palindromic rheumatism	32 (9.4)	18 (24.7)	14 (5.2)	<0.0001
Symptom duration, mean ± SD weeks (n = 254)†	110.7 ± 216.6	115.9 ± 277.7	108.8 ± 191.4	0.69
Symptoms in RA-specific joint‡	200 (58.8)	54 (74.0)	146 (54.7)	0.003
Signs in RA-specific joint§	81 (23.8)	26 (35.6)	55 (20.6)	0.008
Swelling in RA-specific joint¶	46 (13.5)	13 (17.8)	33 (12.4)	0.23

* Values are the number (%) unless indicated otherwise. CCP = cyclic citrullinated peptide; RA = rheumatoid arthritis; RF = rheumatoid factor; ULN = upper limit of normal; ANA = antinuclear antibody; ESR = erythrocyte sedimentation rate; CRP = C-reactive protein.
† No. of patients with available data.
‡ Symptoms include pain, stiffness, or both. RA-specific joint was defined as metacarpophalangeal joints, proximal interphalangeal joints, metatarsophalangeal joints, thumb interphalangeal joints, wrists, and elbows.
§ Symptoms include pain, stiffness, or both. Signs are tenderness, swelling, or both.
¶ Symptoms include pain, stiffness, or both.

symptoms and signs of RA, and not true confounders: specialty of the ordering physician, testing for indication of arthralgia, palindromic rheumatism, and symptoms/signs/swelling involving RA-specific joints.

We tested for the proportional hazards assumption by including an interaction term between time after the index date and CCP antibody level for RA risk and verifying that there was no statistically significant interaction. The proportional hazards assumption was met in all analyses. Analyses were performed using SAS software, version 9.4. We set the threshold for statistical significance as a 2-sided *P* value of less than 0.05.

RESULTS

Baseline characteristics. Among a total of 340 CCP+ patients who were without prevalent RA or systemic rheumatic disease at baseline, we identified 73 incident RA cases (21.5%) during 1,047 person-years of follow-up (incidence rate 69.7 per 1,000 person-years), with a median follow-up of 2.7 years/patient (IQR 1.1–4.6). Eleven patients (2.9%) developed systemic rheumatic diseases other than RA during follow-up (consisting of systemic lupus erythematosus, antineutrophil cytoplasmic antibody–associated vasculitis, polymyalgia rheumatica, and spondyloarthritis). Twenty patients (5.9%) died during follow-up.

Baseline characteristics of the study sample, overall and stratified by progressors to RA versus nonprogressors, are shown in Table 1. The study sample overall was 65.6% female and 74.7% white with a mean ± SD age of 55.0 ± 15.3 years.

Reason for ordering initial CCP test. The reasons for ordering the initial CCP testing are shown in Table 2. Arthralgia (75.9%) was the most common reason, followed by lung disease (10.0%). Of the patients for whom the reason for checking CCP antibody status was an abnormal laboratory value, axial joint pain, fatigue, or fever, none progressed to RA.

Subgroup of CCP+ ordered for lung disease. Characteristics of the subgroup of patients who had CCP antibodies checked for lung disease evaluation (n = 34) are shown in Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23820/abstract. The median duration of follow-up in this subgroup was 25.0 months (IQR 10.8–48.0 months), and 7 patients (20.6%) died during follow-up. Only 38.2% were female, mean ± SD age was 66.9 ± 12.8 years, and most were white (76.5%). Current smokers comprised 8.8% of this subgroup, and 58.8% were former smokers. The proportion of patients with low-level (47.1%), medium-level (20.6%), or high-level (32.4%) CCP antibodies was similar to the overall study sample. The underlying lung diseases were interstitial lung abnormalities (70.6%), pleural effusion (8.8%), nodules/masses (5.9%), pulmonary hypertension (5.9%), chronic obstructive pulmonary disease (2.9%), bronchiectasis (2.9%), and hemoptysis (2.9%). Only 1 of these 34 patients (2.9%) progressed to RA during follow-up, and 1 patient developed another systemic rheumatic disease (spondyloarthritis).

Absolute risk of progression to RA. The absolute risk of progression to RA at 1, 3, and 5 years is shown in Table 3, and corresponding Kaplan-Meier curves showing RA-free survival are in Figure 2. For all patients, the 5-year risk was 25.3% (95% CI 19.6–30.6). Absolute risk of progression to RA increased with increasing CCP antibody level; for example, among patients with high-level CCP antibodies, 1-year risk was 24.9% (95% CI 16.4–32.6), 3-year risk was 41.5% (95% CI 30.9–50.4), and 5-year risk was 46.0% (95% CI 34.7–55.3). The absolute risk of progression to RA also increased with increased RF level, having a family history of RA, and the presence of symptoms in RA-specific joints.

The risk of progression to RA according to the CCP antibody level in combination with RF status (positive or negative), family history of RA (positive or negative), or symptoms in RA-specific joints (present or absent) is shown in Table 3 and Figure 2. For example, with respect to CCP antibody level and a family history of RA, patients with high-level CCP antibodies and a positive family history were at the highest risk (5-year risk 61.8% [95% CI 32.5–78.4]).

Cox regression models. Unadjusted and multivariable HRs for RA are shown in Table 4. Compared to low-

Table 2. Clinical reason for ordering initial CCP test (n = 340)*

Reason	All patients	Progressed to RA	Nonprogressors	*P*
Arthralgia	258 (75.9)	70 (95.9)	188 (70.4)	<0.0001
Lung disease	34 (10.0)	1 (1.4)	33 (12.4)	<0.0001
Abnormal laboratory results†	16 (4.7)	0 (0.0)	16 (6.0)	<0.0001
Axial pain (neck, back, or hip)	4 (1.2)	0 (0.0)	4 (1.5)	0.06
Fatigue	3 (0.9)	0 (0.0)	3 (1.1)	0.13
Fever	2 (0.6)	0 (0.0)	2 (0.8)	0.25
Miscellaneous symptoms or other clinical conditions‡	20 (5.9)	2 (2.7)	18 (6.7)	0.0002

* Values are the number (%) unless indicated otherwise. CCP = cyclic citrullinated peptide; RA = rheumatoid arthritis.
† Includes one or more of following: antinuclear antibody, erythrocyte sedimentation rate, C-reactive protein level, rheumatoid factor, creatine phosphokinase, platelets, hepatitis C positivity, cryoglobulinemia.
‡ Includes rash, sicca symptoms, myalgia, paresthesia/neuropathy, scleritis, Raynaud's phenomenon, cirrhosis, pericarditis.

Table 3. Absolute risk of progression to RA among patients with CCP antibody positivity but no RA by specified follow-up*

	1-year risk	3-year risk	5-year risk
All patients	12.8 (9.1–16.4)	21.0 (16.0–25.6)	25.3 (19.6–30.6)
CCP >1 to 2× ULN (low)	5.6 (2.0–9.2)	7.3 (3.0–11.5)	10.2 (4.3–15.7)
CCP >2 to 3× ULN (medium)	8.2 (0.1–15.6)	18.0 (4.2–29.8)	27.4 (8.6–42.3)
CCP >3× ULN (high)	24.9 (16.4–32.6)	41.5 (30.9–50.4)	46.0 (34.7–55.3)
RF negative or not sent	9.1 (5.4–12.6)	14.9 (10.0–19.6)	18.3 (12.4–23.7)
RF >1 to 3× ULN (low/medium)	16.2 (3.4–27.3)	32.3 (14.2–46.6)	37.9 (16.9–53.7)
RF >3× ULN (high)	32.1 (15.8–45.2)	44.1 (25.5–58.1)	51.8 (31.2–66.2)
No family history of RA	11.0 (7.2–14.6)	17.9 (12.9–22.6)	21.3 (15.6–26.6)
Family history of RA	24.2 (10.5–35.9)	40.1 (21.6–54.2)	53.6 (25.8–71.0)
No symptoms in RA-specific joint	8.0 (3.1–12.6)	13.5 (6.7–19.8)	16.8 (8.7–24.1)
Symptoms in RA-specific joint†	16.1 (10.8–21.2)	25.9 (19.0–32.1)	30.6 (22.9–37.6)
CCP low or medium and RF–	6.5 (2.8–9.9)	9.4 (4.8–13.8)	12.8 (6.9–18.3)
CCP low or medium and RF+	5.0 (0.0–14.1)	12.3 (0.0–27.2)	22.1 (0.0–42.1)
CCP high and RF–	17.5 (7.0–26.9)	33.0 (18.4–45.0)	36.1 (20.6–48.6)
CCP high and RF+	31.4 (18.2–42.5)	47.7 (32.3–59.7)	53.3 (36.7–65.6)
CCP low or medium, negative family history of RA	5.9 (2.5–9.2)	8.8 (4.4–13.1)	12.0 (6.3–17.4)
CCP low or medium, positive family history of RA	10.5 (0.0–23.4)	18.0 (0.0–34.9)	38.5 (0.0–66.6)
CCP high, negative family history of RA	21.6 (12.5–29.8)	36.6 (25.0–46.5)	40.3 (28.0–50.5)
CCP high, positive family history of RA	34.4 (13.1–50.4)	55.5 (27.9–72.5)	61.8 (32.5–78.4)
CCP low or medium, no symptoms in RA-specific joint	2.3 (0.0–5.3)	7.2 (0.7–13.2)	11.8 (2.8–20.0)
CCP low or medium, symptoms in RA-specific joint	9.6 (4.0–14.8)	11.9 (5.5–17.8)	15.5 (7.5–22.8)
CCP high, no symptoms in RA-specific joint	22.3 (7.3–34.8)	29.1 (11.9–43.0)	29.1 (11.9–43.0)
CCP high, symptoms in RA-specific joint	25.6 (15.2–34.7)	45.9 (32.7–56.6)	52.0 (37.9–62.9)

* Values are the percentage (95% confidence interval). RA = rheumatoid arthritis; CCP = cyclic citrullinated peptide; ULN = upper limit of normal; RF = rheumatoid factor.
† RA-specific joints were defined as metacarpophalangeal joints, proximal interphalangeal joints, metatarsophalangeal joints, thumb interphalangeal joints, wrists, or elbows.

level CCP antibodies, medium-level (HR 2.75 [95% CI 1.22–6.19]) and high-level (HR 6.18 [95% CI 3.40–11.2]) antibodies were associated with increased RA risk in the unadjusted analysis. An increased CCP antibody level remained associated with progression to RA when adjusted for age, sex, BMI, smoking, and family history (medium CCP: HR 2.94 [95% CI 1.30–6.69]; high CCP: HR 5.86 [95% CI 3.20–10.7]; reference: low CCP). When also adjusted for RF level, CCP antibody level remained predictive of RA (medium CCP: HR 3.00 [95% CI 1.32–6.81]; high CCP: HR 4.83 [95% CI 2.51–9.31]) compared to low CCP.

Unadjusted HRs for RA with respect to the specialty of the ordering physician, whether CCP testing was ordered for arthralgias, the presence of symptoms in RA-specific joints, signs in RA-specific joints, swelling in RA-specific joints, and palindromic rheumatism are shown in Supplementary Table 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23820/abstract. Testing of CCP for arthralgias was associated with increased risk for RA (HR 7.84 [95% CI 2.47–24.9]), as was the presence of signs (swelling or tenderness) in RA-specific joints (HR 1.91 [95% CI 1.18–3.09]).

DISCUSSION

In a US hospital-based cohort of 340 CCP+ patients who were without RA or other systemic rheumatic disease at the time of initial CCP antibody positivity, 21.5% went on to develop RA according to 2010 ACR/EULAR criteria. The strongest predictor of RA risk was CCP antibody level, with high-level CCP antibodies conferring a 5-fold increased hazard for RA independent of age, sex, BMI, smoking, family history, and RF level. Our results suggest that approximately 46% of patients with high-level CCP antibodies develop RA within 5 years.

Studies involving blood banks or asymptomatic first-degree relatives of patients with RA have demonstrated that CCP antibody positivity is significantly associated with increased risk of progression to RA. In a Swedish blood bank case–control study comparing 83 RA cases to age- and sex-matched controls, Rantapää-Dahlqvist et al (2) found that at any time prior to RA diagnosis, CCP antibodies were positive prior to diagnosis in 34% of RA patients and were predictive of RA development (HR 16.1 [95% CI 3.3–76.7]) compared to CCP antibody negativity. Ramos-Remus et al (8) tested 819 healthy relatives of RA patients for CCP antibodies and RF and followed them longitudinally for 5 years, with 2% developing RA during follow-up; CCP+ relatives had an HR of 223.1 (95% CI 63.8–779.9) for RA compared to CCP– relatives. Because these studies investigated RA risk for seropositivity compared to seronegativity in a largely asymptomatic population, these findings may not be applicable to a clinical population who all have seropositivity and a clinical indication for CCP antibody testing.

Cohort studies of patients recruited from arthralgia referral clinics in Europe have also advanced the understanding of the

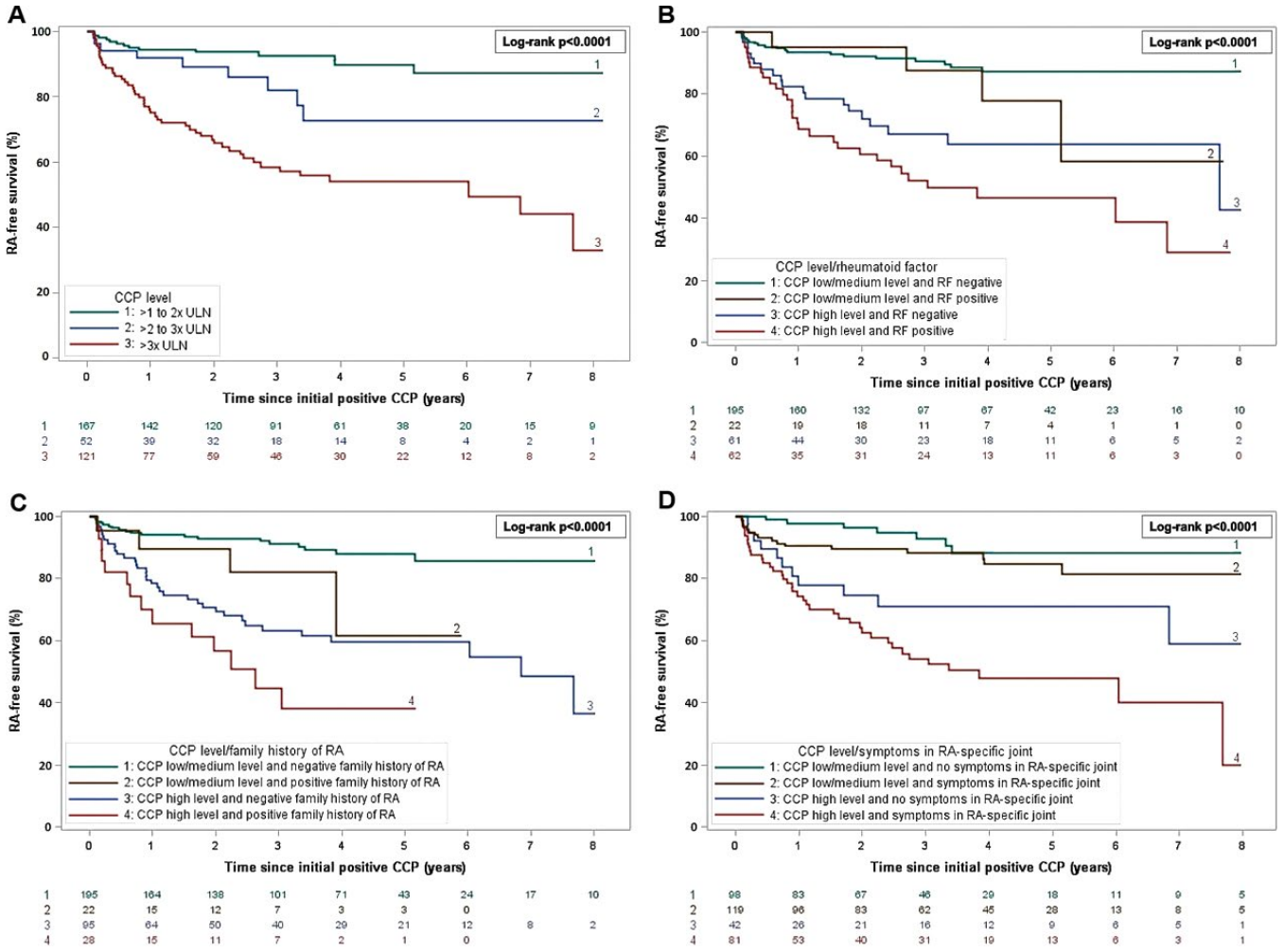

Figure 2. Kaplan-Meier curves for rheumatoid arthritis–free survival after the index date of initial cyclic citrullinated peptide (CCP) antibody positivity according to: **A**, CCP antibody in 3 levels. **B**, CCP antibody binary level and rheumatoid factor (RF) status. **C**, CCP antibody binary level and presence/absence of family history of rheumatoid arthritis (RA). **D**, CCP antibody binary level and presence/absence of symptoms (pain, stiffness, or both) in RA-specific joints. RA-specific joints were defined as metacarpophalangeal joints, proximal interphalangeal joints, metatarsophalangeal joints, thumb interphalangeal joints, wrists, and elbows. ULN = upper limit of normal.

impact of CCP antibodies on RA risk. A prospective cohort study by van Gaalen et al (16) of patients with undifferentiated arthritis recruited to an early arthritis clinic in The Netherlands showed that 93% of CCP+ patients with undifferentiated arthritis developed classifiable RA according to 1987 ACR criteria within 3 years of follow-up (29). Being CCP+ was a significant risk factor for RA, with an odds ratio of 37.8 (95% CI 13.8–111.9) compared to being CCP–. However, this study was performed prior to development of the 2010 ACR/EULAR classification criteria (20), which allowed for earlier RA detection, so some of those with undifferentiated arthritis who were CCP+ likely had RA at baseline under the new criteria. We did not limit our study to patients with articular signs or symptoms, which may explain why the absolute RA risk was higher in that cohort compared to our study. In a prospective cohort study of 147 seropositive patients with arthralgias recruited in The Netherlands (17), CCP antibody positivity was associated with arthritis development (HR 6.0 [95% CI 1.8–19.8]) compared to CCP antibody negativity. Among CCP+ patients, RA risk increased with higher CCP antibody levels. Rakieh et al (19) prospectively followed 100 CCP+ patients with arthralgia in the UK for a median of 20 months, and 43% developed RA according to 2010 ACR/EULAR criteria. The majority (>80%) of that cohort had high-level CCP antibodies, and all had articular symptoms that may explain this higher RA incidence compared to our study. Rakieh et al analyzed a combined variable of high level RF and/or CCP (defined as >3× ULN for either assay) in multivariable analyses, and there was an HR of 4.52 (95% CI 1.07–19.15) for inflammatory arthritis compared to lower levels of these autoantibodies, similar to our findings. Overall, our study adds to the literature by investigating a CCP+ population tested through routine clinical care in the US in the current era of RA diagnosis. We were able to directly compare high- to low-level CCP antibody positivity and also had detailed granular data available related to the reason for ordering the test and clinical

Table 4. Unadjusted and multivariable hazard ratios for progression to rheumatoid arthritis among patients with CCP antibody positivity (n = 340)*

	Cases/ person-years	IR per 1,000 person-years	Unadjusted	Model 1†	Model 2‡
CCP antibody level					
>1 to 2× ULN (low)	14/589	23.8	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)
>2 to 3× ULN (medium)	10/143	69.9	2.75 (1.22–6.19)	2.94 (1.30–6.69)	3.00 (1.32–6.81)
>3× ULN (high)	49/315	155.6	6.18 (3.40–11.2)	5.86 (3.20–10.7)	4.83 (2.51–9.31)
Age, continuous per year	NA	–	0.99 (0.98–1.01)	1.00 (0.98–1.01)	1.00 (0.98–1.01)
Male	17/359	47.4	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)
Female	56/687	81.5	1.75 (1.01–3.01)	1.79 (1.01–3.16)	1.84 (1.04–3.25)
Body mass index, kg/m^2					
<25	30/401	74.8	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)
25 to <30	21/333	63.1	0.89 (0.51–1.55)	1.15 (0.65–2.06)	1.00 (0.55–1.83)
≥30	22/312	70.5	0.96 (0.55–1.66)	0.91 (0.51–1.62)	0.86 (0.48–1.53)
Not current smoker	59/918	64.3	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)
Current smoker	14/129	108.5	1.78 (0.99–3.19)	1.68 (0.92–3.07)	1.43 (0.76–2.69)
No family history of RA	55/946	58.1	1.00 (Ref.)	1.00 (Ref.)	1.00 (Ref.)
Family history of RA	18/100	180.0	2.62 (1.53–4.51)	1.76 (1.00–3.10)	1.63 (0.92–2.91)
RF level					
Negative or not sent	39/819	47.6	1.00 (Ref.)	–	1.00 (Ref.)
>1 to 3× ULN (low/medium)	13/12	107.4	2.26 (1.20–4.23)	–	1.20 (0.60–2.39)
>3× ULN (high)	21/106	198.1	3.80 (2.22–6.49)	–	1.84 (0.95–3.56)

* Values are the hazard ratio (95% confidence interval), unless indicated otherwise. CCP = cyclic citrullinated peptide; IR = incidence rate; ULN = upper limit of normal; Ref. = reference; NA = not applicable; RA = rheumatoid arthritis; RF = rheumatoid factor.
† Multivariable, adjusted for age, sex, body mass index, smoking, and family history of RA.
‡ Multivariable, adjusted for variables in model 1 and rheumatoid factor level.

characteristics at the time of CCP antibody positivity for predicting subsequent progression to RA.

Our study highlights the importance of CCP antibody level on the risk for progression to RA, independent of other known risk factors. Because patients in our study had CCP testing ordered as part of routine clinical care, our findings may be helpful to clinicians confronted with interpretation of a positive CCP test. Knowledge of the absolute risk of progressing to RA among CCP+ patients without RA based on their CCP antibody level and other clinical factors could encourage lifestyle changes and affect subsequent screening. For example, at-risk patients could be encouraged (30,31) to quit smoking (24,25,27), increase fish intake (32–34), lose weight (24,35,36), and improve dental hygiene (37). Furthermore, there is increasing interest in primary prevention of RA in at-risk patient populations (38,39).

We found that only 1 of the 34 patients who had CCP tested for lung disease went on to develop RA. This group had high mortality (20.6% versus 5.9% in the overall study sample), so they had a shorter duration of follow-up (median 25.0 months versus 32.4 months). Citrullination of proteins can occur in the lung and is influenced by smoking, which suggests that autoimmunity may begin in respiratory mucosa prior to articular involvement (40,41). Whether CCP antibody positivity in patients with interstitial lung disease represents a pre-RA state, or CCP antibody positivity is unrelated to future articular manifestations requires further study. Few patients in our study who had CCP tested for a clinical indication other than arthralgia went on to develop RA during follow-up. While CCP antibody has high specificity for RA (42), we demonstrated that the clinical context in which CCP is tested clearly affects the risk for progression to RA.

Our study has several limitations inherent to a retrospective cohort design. While we collected detailed data on clinical and demographic characteristics, there remains the possibility of unmeasured confounding. We had to rely on clinical notes that may not have included completely accurate details, particularly for articular symptoms and signs. We were unable to perform analyses based on hand or foot radiographs since these were only performed in a minority of patients. Further, we had to rely on the treating clinician's impression and choice not to diagnose and treat for RA at the index date. We mitigated the possibility that patients may have had untreated RA at the index date by excluding patients who quickly progressed to RA, because they were likely to have had prevalent RA at baseline. Diagnostic uncertainty is inherent in a complex disease such as RA, so our study may aid clinicians faced with patients who have symptoms and are CCP+ but without clear RA. Because our medical record is exclusive to our health care system, we may not have captured whether RA was diagnosed outside of our system. However, we assessed all notes and problem lists for RA or RA-related medications and censored at the last RA-free note from a clinician regardless of specialty to maximize the likelihood that patients were free of RA or other systemic rheumatic diseases during all analyzed follow-up. While all patients in our study were CCP+, not all patients were tested using the same CCP assay. Therefore, we standardized CCP antibody level based on the fold above ULN. However, the type of CCP assay used may affect diagnostic performance and titer (43). Therefore, possibly some patients with low CCP antibody positivity were false posi-

tives or may have been negative on other CCP assays. In addition, our small sample size had limited power to detect associations of elevated BMI and smoking for progression to RA. Finally, our study sample was drawn from a single tertiary care system in Boston, so it may not be generalizable to community care settings or other geographic regions. We did not collect detailed geographic information on included subjects, but most likely resided in the greater Boston area. Further research is needed to extend these observations to other settings. Furthermore, while we feel that studying patients when CCP was tested for clinical reasons is a strength of the study, we would emphasize that the predictive value of CCP antibody level demonstrated in our study may not apply to other CCP+ populations outside of a clinical setting.

A major strength of our study is the clinical setting in the US with detailed data on the indication for testing CCP as well as demographics, lifestyle, symptoms, signs, comorbidities, laboratory results, and family history. Thus, we were able to quantify the risk for progression to RA for a variety of clinical scenarios pertinent for treating clinicians. We were able to estimate the absolute risk based on CCP antibody level and combinations of RA risk factors in this real-world population that may be applied to other patients who present clinically in the US. Unlike other studies, we did not restrict our study to patients referred to a specialty clinic with articular signs or symptoms. Furthermore, for all outcomes in our study, 2 rheumatologists agreed on RA outcome, date of diagnosis, and the absence of prevalent RA at the index date, to ensure that we truly captured incident RA in our cohort. Finally, we applied the 2010 ACR/EULAR criteria to determine prevalent and incident RA, whereas some prior studies used 1987 ACR criteria and most of the follow-up in our study occurred after 2010. Thus, our study quantifies RA risk for a clinical US CCP+ population in the current era of RA research.

In conclusion, we found that CCP antibody level was predictive of progression to RA in a clinical US cohort of CCP+ patients without classifiable RA. These findings provide evidence for close monitoring for development of RA in this population. Further research is needed to identify pharmacologic and nonpharmacologic strategies to prevent progression to RA among these patients at very elevated risk for RA.

ACKNOWLEDGMENT

The authors thank the staff of the Clinical Immunology Laboratory at the Brigham and Women's Hospital for their helpful assistance.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. Ford 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. Ford, Sparks.
Acquisition of data. Ford, Marshall, Zaccardelli, Prado, Wiyarand, Sparks.
Analysis and interpretation of data. Ford, Liu, Marshall, Zaccardelli, Prado, Wiyarand, Lu, Karlson, Schur, Deane, Sparks.

REFERENCES

1. Deane KD. Can rheumatoid arthritis be prevented? Best Pract Res Clin Rheumatol 2013;27:467–85.
2. Rantapää-Dahlqvist S, de Jong BA, Berglin E, Hallmans G, Wadell G, Stenlund H, et al. Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. Arthritis Rheum 2003;48:2741–9.
3. Nielen MM, van Schaardenburg D, Reesink HW, van de Stadt RJ, van der Horst-Bruinsma IE, de Koning MH, et al. Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors. Arthritis Rheum 2004;50:380–6.
4. Kokkonen H, Mullazehi M, Berglin E, Hallmans G, Wadell G, Ronnelid J, et al. Antibodies of IgG, IgA and IgM isotypes against cyclic citrullinated peptide precede the development of rheumatoid arthritis. Arthritis Res Ther 2011;13:R13.
5. Chibnik LB, Mandl LA, Costenbader KH, Schur PH, Karlson EW. Comparison of threshold cutpoints and continuous measures of anti-cyclic citrullinated peptide antibodies in predicting future rheumatoid arthritis. J Rheumatol 2009;36:706–11.
6. Arkema EV, Goldstein BL, Robinson W, Sokolove J, Wagner CA, Malspeis S, et al. Anti-citrullinated peptide autoantibodies, human leukocyte antigen shared epitope and risk of future rheumatoid arthritis: a nested case-control study. Arthritis Res Ther 2013;15:R159.
7. Sokolove J, Bromberg R, Deane KD, Lahey LJ, Derber LA, Chandra PE, et al. Autoantibody epitope spreading in the pre-clinical phase predicts progression to rheumatoid arthritis. Matloubian M, ed. PLoS One 2012;7:e35296.
8. Ramos-Remus C, Castillo-Ortiz JD, Aguilar-Lozano L, Padilla-Ibarra J, Sandoval-Castro C, Vargas-Serafin CO, et al. Autoantibodies in prediction of the development of rheumatoid arthritis among healthy relatives of patients with the disease. Arthritis Rheumatol 2015;67:2837–44.
9. Sparks JA, Chen CY, Hiraki LT, Malspeis S, Costenbader KH, Karlson EW. Contributions of familial rheumatoid arthritis or lupus and environmental factors to risk of rheumatoid arthritis in women: a prospective cohort study. Arthritis Care Res (Hoboken) 2014;66:1438–46.
10. Young KA, Deane KD, Derber LA, Hughes-Austin JM, Wagner CA, Sokolove J, et al. Relatives without rheumatoid arthritis show reactivity to anti–citrullinated protein/peptide antibodies that are associated with arthritis-related traits: studies of the etiology of rheumatoid arthritis. Arthritis Rheum 2013;65:1995–2004.
11. Demoruelle MK, Parish MC, Derber LA, Kolfenbach JR, Hughes-Austin JM, Weisman MH, et al. Performance of anti–cyclic citrullinated peptide assays differs in subjects at increased risk of rheumatoid arthritis and subjects with established disease. Arthritis Rheum 2013;65:2243–52.
12. Hughes-Austin JM, Deane KD, Derber LA, Kolfenbach JR, Zerbe GO, Sokolove J, et al. Multiple cytokines and chemokines are associated with rheumatoid arthritis-related autoimmunity in first-degree relatives without rheumatoid arthritis: Studies of the Aetiology of Rheumatoid Arthritis (SERA). Ann Rheum Dis 2013;72:901–7.
13. Koppejan H, Trouw LA, Sokolove J, Lahey LJ, Huizinga TJ, Smolik IA, et al. Role of anti–carbamylated protein antibodies compared to anti–citrullinated protein antibodies in indigenous North Americans with rheumatoid arthritis, their first-degree relatives, and healthy controls. Arthritis Rheumatol 2016;68:2090–8.
14. Frisell T, Saevarsdottir S, Askling J. Family history of rheumatoid arthritis: an old concept with new developments. Nat Rev Rheumatol 2016;12:335–43.

15. Jiang X, Frisell T, Askling J, Karlson EW, Klareskog L, Alfredsson L, et al. To what extent is the familial risk of rheumatoid arthritis explained by established rheumatoid arthritis risk factors? Arthritis Rheumatol 2015;67:352–62.
16. Van Gaalen FA, Linn-Rasker SP, van Venrooij WJ, de Jong BA, Breedveld FC, Verweij CL, et al. Autoantibodies to cyclic citrullinated peptides predict progression to rheumatoid arthritis in patients with undifferentiated arthritis: a prospective cohort study. Arthritis Rheum 2004;50:709–15.
17. Bos WH, Wolbink GJ, Boers M, Tijhuis GJ, de Vries N, van der Horst-Bruinsma IE, et al. Arthritis development in patients with arthralgia is strongly associated with anti-citrullinated protein antibody status: a prospective cohort study. Ann Rheum Dis 2010;69:490–4.
18. Bizzaro N, Bartoloni E, Morozzi G, Manganelli S, Riccieri V, Sabatini P, et al. Anti-cyclic citrullinated peptide antibody titer predicts time to rheumatoid arthritis onset in patients with undifferentiated arthritis: results from a 2-year prospective study. Arthritis Res Ther 2013;15:R16.
19. Rakieh C, Nam JL, Hunt L, Hensor EM, Das S, Bissell LA, et al. Predicting the development of clinical arthritis in anti-CCP positive individuals with non-specific musculoskeletal symptoms: a prospective observational cohort study. Ann Rheum Dis 2015;74:1659–66.
20. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis 2010;69:1580–8.
21. Peoples C, Valiyil R, Davis RB, Shmerling RH. Clinical use of anti-cyclic citrullinated peptide antibody testing [abstract]. J Clin Rheumatol 2013;19:351–2.
22. Gabriel SE, Crowson CS, O'Fallon WM. The epidemiology of rheumatoid arthritis in Rochester, Minnesota, 1955–1985. Arthritis Rheum 1999;42:415–20.
23. Crowson CS, Matteson EL, Myasoedova E, Michet CJ, Ernste FC, Warrington KJ, et al. The lifetime risk of adult-onset rheumatoid arthritis and other inflammatory autoimmune rheumatic diseases. Arthritis Rheum 2011;63:633–9.
24. De Hair MJ, Landewé RB, van de Sande MG, van Schaardenburg D, van Baarsen LG, Gerlag DM, et al. Smoking and overweight determine the likelihood of developing rheumatoid arthritis. Ann Rheum Dis 2013;72:1654–8.
25. Karlson EW, Chang SC, Cui J, Chibnik LB, Fraser PA, De Vivo I, et al. Gene-environment interaction between HLA-DRB1 shared epitope and heavy cigarette smoking in predicting incident rheumatoid arthritis. Ann Rheum Dis 2010;69:54–60.
26. Hemminki K, Li X, Sundquist J, Sundquist K. Familial associations of rheumatoid arthritis with autoimmune diseases and related conditions. Arthritis Rheum 2009;60:661–8.
27. Sparks JA, Chang SC, Deane KD, Gan RW, Kristen Demoruelle M, Feser ML, et al. Associations of smoking and age with inflammatory joint signs among unaffected first-degree relatives of rheumatoid arthritis patients: results from Studies of the Etiology of Rheumatoid Arthritis. Arthritis Rheumatol 2016;68:1828–38.
28. Van de Stadt LA, Witte BI, Bos WH, van Schaardenburg D. A prediction rule for the development of arthritis in seropositive arthralgia patients. Ann Rheum Dis 2013;72:1920–6.
29. 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.
30. Sparks JA, Iversen MD, Yu Z, Triedman NA, Prado MG, Miller Kroouze RM, et al. Disclosure of personalized rheumatoid arthritis risk using genetics, biomarkers, and lifestyle factors to motivate health behavior improvements: a randomized controlled trial. Arthritis Care Res (Hoboken) 2018;70:823–33.
31. Prado MG, Iversen MD, Yu Z, Miller Kroouze R, Triedman NA, Kalia SS, et al. Effectiveness of a web-based personalized rheumatoid arthritis risk tool with or without a health educator for knowledge of rheumatoid arthritis risk factors. Arthritis Care Res (Hoboken) 2018;70:1421–30.
32. Shapiro JA, Koepsell TD, Voigt LF, Dugowson CE, Kestin M, Nelson JL. Diet and rheumatoid arthritis in women: a possible protective effect of fish consumption. Epidemiology 1996;7:256–63.
33. Rosell M, Wesley AM, Rydin K, Klareskog L, Alfredsson L, and the EIRA study group. Dietary fish and fish oil and the risk of rheumatoid arthritis. Epidemiology 2009;20:896–901.
34. Gan RW, Demoruelle MK, Deane KD, Weisman MH, Buckner JH, Gregersen PK, et al. Omega-3 fatty acids are associated with a lower prevalence of autoantibodies in shared epitope-positive subjects at risk for rheumatoid arthritis. Ann Rheum Dis 2017;76:147–52.
35. Lu B, Hiraki LT, Sparks JA, Malspeis S, Chen CY, Awosogba JA, et al. Being overweight or obese and risk of developing rheumatoid arthritis among women: a prospective cohort study. Ann Rheum Dis 2014;73:1914–22.
36. Qin B, Yang M, Fu H, Ma N, Wei T, Tang Q, et al. Body mass index and the risk of rheumatoid arthritis: a systematic review and dose-response meta-analysis. Arthritis Res Ther 2015;17:86.
37. De Pablo P, Dietrich T, McAlindon TE. Association of periodontal disease and tooth loss with rheumatoid arthritis in the US population. J Rheumatol 2008;35:70–6.
38. Gonzalez-Lopez L, Gamez-Nava JI, Jhangri G, Russell AS, Suarez-Almazor ME. Decreased progression to rheumatoid arthritis or other connective tissue diseases in patients with palindromic rheumatism treated with antimalarials. J Rheumatol 2000;27:41–6.
39. Burgers LE, Allaart CF, Huizinga TW, van der Helm-van Mil AH. Clinical trials aiming to prevent rheumatoid arthritis cannot detect prevention without adequate risk stratification: a trial of methotrexate versus placebo in undifferentiated arthritis as an example. Arthritis Rheumatol 2017;69:926–31.
40. Aubart F, Crestani B, Nicaise-Roland P, Tubach F, Bollet C, Dawidowicz K, et al. High levels of anti-cyclic citrullinated peptide autoantibodies are associated with co-occurrence of pulmonary diseases with rheumatoid arthritis. J Rheumatol 2011;38:979–82.
41. Makrygiannakis D, Hermansson M, Ulfgren AK, Nicholas AP, Zendman AJ, Eklund A, et al. Smoking increases peptidylarginine deiminase 2 enzyme expression in human lungs and increases citrullination in BAL cells. Ann Rheum Dis 2008;67:1488–92.
42. Avouac J, Gossec L, Dougados M. Diagnostic and predictive value of anti-cyclic citrullinated protein antibodies in rheumatoid arthritis: a systematic literature review. Ann Rheum Dis 2005;65:845–51.
43. Van Hoovels L, Jacobs J, Vander Cruyssen B, Van den Bremt S, Verschueren P, Bossuyt X. Performance characteristics of rheumatoid factor and anti-cyclic citrullinated peptide antibody assays may impact ACR/EULAR classification of rheumatoid arthritis. Ann Rheum Dis 2018;77:667–77.

Arthritis Care & Research
Vol. 71, No. 12, December 2019, pp 1593–1599
DOI 10.1002/acr.23810

American College of Rheumatology
Empowering Rheumatology Professionals

Screening of Hyperlipidemia Among Patients With Rheumatoid Arthritis in the United States

Iris Navarro-Millán,[1] Shuo Yang,[2] Lang Chen,[2] Huifeng Yun,[2] Aprajita Jagpal,[2] Christie M. Bartels,[3] Liana Fraenkel,[4] Monika M. Safford,[5] and Jeffrey R. Curtis[2]

Objective. To determine the proportion of primary lipid screening among patients with rheumatoid arthritis (RA) and compare it with those among patients with diabetes mellitus (DM) and patients with neither RA nor DM, and to assess whether primary lipid screening varied according to the health care provider (rheumatologist versus non-rheumatologist).

Methods. We analyzed claims data from US private and public health plans from 2006–2010. Eligibility requirements included continuous medical and pharmacy coverage for ≥12 months (baseline period) and ≥2 physician diagnoses and relevant medications to define RA, DM, RA and DM, or neither condition. Among the 330,695 eligible participants, we calculated the proportion with a lipid profile ordered during the 2 years following baseline. Time-varying Cox proportional hazard models were used to determine the probability of hyperlipidemia screening in participants with RA according to provider specialty.

Results. More than half of the patients were ages 41–71 years. Among patients with RA (n = 12,182), DM (n = 62,834), RA and DM (n = 1,082), and those who did not have either condition (n = 167,811), the proportion screened for hyperlipidemia was 37%, 60%, 55%, and 41%, respectively. Patients with RA who visited a rheumatologist and a non-rheumatology clinician during follow-up had a 55% (95% confidence interval 1.36–1.78) higher screening probability than those who only visited a rheumatologist.

Conclusion. Primary lipid screening was suboptimal among patients with RA. It was also lower for patients with DM and minimally different from the general population. Screening was higher for RA patients who received care from both a rheumatologist and a non-rheumatologist (e.g., primary care physician).

INTRODUCTION

Rheumatoid arthritis (RA) is associated with an increased risk for cardiovascular disease (CVD), which involves disease-specific mechanisms that are both similar and different from CVD in the general population (1,2). Compared with age- and sex-adjusted individuals without RA, patients with RA have more CVD events and higher mortality (3). Indeed, a large meta-analysis of 24 cohorts showed a 50% higher risk for CVD mortality in patients with RA compared to those without RA (4). Evidence suggests that both traditional and nontraditional risk factors play a role in the development of CVD in patients with RA (5,6). Traditional risk factors such as hyperlipidemia, hypertension, diabetes mellitus (DM), cigarette smoking, and obesity are also associated with an increased risk for CVD in the RA population (7). Systemic inflammation was also associated with an increased risk for CVD events and mortality among patients with RA. High erythrocyte sedimentation rate, high C-reactive protein levels, and other RA-related biomarkers have been found to be associated with an increased risk of myocardial infarction (MI) in patients with RA (1,8).

Statins are a major focus of CVD risk reduction in the general population, but to determine if these are indicated, a lipid profile is required to estimate 10-year CVD risk. In patients with RA, statins have been associated with a reduction in cardiovascular risk (9–11).

Dr. Navarro-Millán's work was supported by the NIH (grant K23-AR-068449 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases). Dr. Fraenkel's work was supported by the NIH (award AR-060231 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases). Dr. Safford's work was supported by the NIH (grant K24-HL-111154 from the National Heart, Lung, and Blood Institute). Dr. Curtis' work was supported in part by a contract from the Actuarial Research Corporation on behalf of the Department of Health and Human Services.

[1]Iris Navarro-Millán, MD, MSPH: Weill Cornell Medicine and the Hospital for Special Surgery, New York, New York; [2]Shuo Yang, MS, Lang Chen, PhD, Huifeng Yun, PhD, Aprajita Jagpal, MD, Jeffrey R. Curtis, MD, MS, MPH: University of Alabama at Birmingham; [3]Christie M. Bartels, MD, MS: University of Wisconsin, Madison; [4]Liana Fraenkel, MD, MPH: Yale University, New Haven, Connecticut; [5]Monika M. Safford, MD: Weill Cornell Medicine, New York, New York.

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

Address correspondence to Jeffrey R. Curtis, MD, MS, MPH, 510 20th Street South, Faculty Office Tower 802D, Birmingham, AL 35294. E-mail: jrcurtis@uabmc.edu.

Submitted for publication May 11, 2018; accepted in revised form November 6, 2018.

SIGNIFICANCE & INNOVATIONS

- Despite the generally recognized increased risk for rheumatoid arthritis (RA) that is associated with cardiovascular disease, in this population-based study, patients with RA were underscreened for primary hyperlipidemia.
- The proportion of patients with RA screened for primary hyperlipidemia was comparable to the general population and lower than that of patients with diabetes mellitus.
- Co-management, defined as patients managed both by rheumatologists and by specialists who predominantly provide primary care services, increased the likelihood of primary hyperlipidemia screening in patients with RA by 55%.

Additionally, it has been reported that, compared with patients with RA who continued statins, those who discontinued this medication had a 67% increased risk for MI (12,13). Despite the known higher risk for CVD events and mortality, screening for hyperlipidemia among patients with RA has been observed to be low (14–16).

Previous studies using Medicare data among patients older than 65 years of age showed that primary lipid screening occurred in only 45% of patients with RA over a 3-year period from 2004–2006 (15). The pattern of primary lipid screening among individuals with RA who are younger than 65 years of age is still unclear. Hence, the first goal of our study was to determine the proportion of primary lipid screening among patients with RA ages >40 years and compare these results to the primary lipid screening for patients with neither RA nor DM and patients with DM only. The second goal was to determine whether the proportion of patients with RA who received primary lipid screening varied based on the specialty of the health care provider encountered (rheumatologist versus non-rheumatologist).

PATIENTS AND METHODS

Participants. All patients were 41–85 years of age. This retrospective study used claims data from a US commercial health plan and public health plans (Medicare and Medicaid) from 2006 to 2010 as a single data set, a multi-payer claims database (MPCD) (17,22). This data set included beneficiaries from 4 main regions in the US, including the Northeast, South, West, and Midwest. Thirty-two percent of beneficiaries were commercially insured, 64% were enrolled in Medicare, and 4% in Medicaid. Eligible participants were required to have 12 months of continuous medical and pharmacy coverage within the data set (baseline period) to ensure a complete claims history and characterize participants' characteristics. They were also required to satisfy additional criteria that enabled patients to be placed into 1 of 4 mutually exclusive cohorts, including patients with RA only, patients with DM, patients with both RA and DM, or patients with neither condition. RA was defined based on the presence of ≥2 International Classification of Diseases, Ninth Revision (ICD-9) codes for RA (714.xx) associated with an encounter with a physician and ≥1 filled prescription for a disease-modifying antirheumatic drug (DMARD) (e.g., methotrexate, sulfasalazine, hydroxychloroquine, biologic therapies, and leflunomide) during baseline (18,19). The positive predictive value of this definition is >85% to identify RA compared to medical record review (20). We excluded patients with inflammatory arthritis other than RA (e.g., psoriatic arthritis and ankylosing spondylitis), lupus, Sjögren's syndrome, malignancy, or HIV infection by ICD-9 diagnoses from all 4 cohorts.

Patients with DM were identified by ≥2 physician diagnosis codes for DM (ICD-9 250.xx) or a filled prescription for DM-specific medication during baseline (21). The cohort of patients with neither RA nor DM with commercial insurance, Medicare, or Medicaid is described hereafter as the general population. The general population comparator reflected individuals without physician diagnosis codes or medications for either RA or DM and was a random sample of insured individuals in the US from the MPCD (17,22). The cohort of patients with RA and DM were required to meet criteria for both RA and DM separately. Patients who had partial evidence for either RA and/or DM (e.g., only 1 physician diagnosis code) did not contribute person-time to the analysis until they met the criteria above. Thus, membership in the 4 disease cohorts was classified in a time-varying fashion.

In all 4 cohorts, we excluded patients with prevalent inpatient or outpatient MI, stroke, or coronary heart disease, and those with a lipid profile tested and/or use of statins during the 12-month baseline. Follow-up began after the baseline year. We only included participants with 2-year follow-up data consistently available after the baseline year. We conducted a subgroup analysis for the second goal of this study that included only patients with RA regardless of having comorbid DM. The inclusion criteria for this subgroup of RA patients were those with an encounter with either a rheumatologist and/or non-rheumatology practitioner, which was defined as internal medicine physicians, family medicine physicians, nurse practitioners (NP), or physician assistants (PA). All of these patients also had 2 ICD-9 diagnosis codes for RA and DMARD medications. The Institutional Review Board of the University of Alabama at Birmingham approved this project.

Variables. The variables for this analysis included age, sex, and hypertension (determined by ICD-9 codes). We had pharmacy data on RA medications, statins, and other non-statin lipid-lowering therapy. We identified physician specialties through evaluation and management codes associated with claims for services within the database, and we classified the types of outpatient physician evaluation and management encounters into evaluation and management with a rheumatologist, or evaluation and management with another physician or provider. Non-rheumatology practitioners of interest focused on specialties commonly providing primary care, including internal medicine physicians, family medicine physicians, NPs, and

PAs. Although NPs and PAs may have provided primary care services, they may have actually been associated with a rheumatology clinic. This data source could not differentiate their practice setting. Evaluation and management by other physician specialties (e.g., cardiologist) was not examined in this analysis.

Outcome. The outcome for the primary objective of this study was to determine the proportion of participants who were screened for hyperlipidemia in the cohorts of RA, DM, both RA and DM, and the general population. The outcome for the second objective was the likelihood of a patient with RA being screened for lipids, based on co-management (defined as patients who had been evaluated or followed by a rheumatologist and a non-rheumatology practitioner, e.g., primary care) versus management by only a rheumatologist. This study did not determine whether there was communication or coordinated care between these practitioners, but only that patients were evaluated in an ambulatory setting (i.e., had outpatient visits).

Statistical analysis. We used descriptive statistics to examine the baseline characteristics of the patients in this study. We determined the proportion of primary lipid screening during the 2-year follow-up period in each condition. We used chi-square tests to determine differences in the proportion of patients with a lipid profile between the different cohorts. We used Cox proportional hazard ratios to determine the likelihood of lipid screening only among patients with RA regardless of having comorbid DM based on visiting a non-rheumatology practitioner, a rheumatologist, or both (subgroup analysis for patients with RA). The type of provider visit (rheumatologist or non-rheumatologist practitioner) was time variant, while we controlled for other demographic and RA-related variables at baseline (nonvariant). The Cox model allowed for time-varying evaluation of co-management between rheumatologist and non-rheumatologist providers such that patients could be referred to a primary care physician after the start of follow-up and be correctly classified over time. This type of analysis allowed for a more accurate categorization of the main exposure (co-management between rheumatologist and non-rheumatology providers) for the longitudinal analysis.

RESULTS

Overall, 243,909 participants met the eligibility criteria for the 4 disease-specific groups: 12,182 patients had RA only, 62,834 had DM only, 1,082 had RA and DM (reflecting 8.9% of all RA patients), and 167,811 had neither condition (Figure 1). As part of cohort selection, 27% of patients with RA only, 25% of patients with DM only, and 22% of patients from the general population were excluded due to baseline use of statin therapy

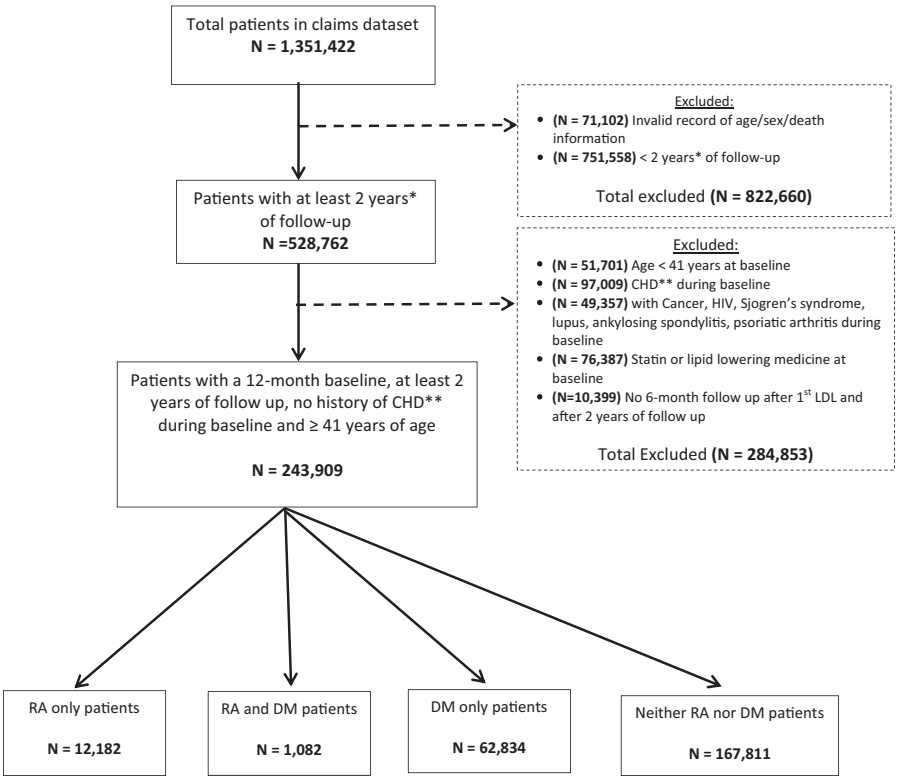

Figure 1. Construction of patient cohorts. LDL = low-density lipoprotein; RA = rheumatoid arthritis; DM = diabetes mellitus.
*Patients with exactly 2 years of follow-up. Those with >2 years of follow-up were included, but only 2 years of follow-up were used.
**Patients with myocardial infarction, stroke, or coronary heart disease (CHD) during the 12-month baseline period.

or lipid screening. More than half of the patients were 41–70 years of age. The age distribution by disease is presented in Table 1, along with other demographic and clinical patient characteristics. The prevalence of hypertension was similar between patients with RA and patients with neither RA nor DM (41% and 39%, respectively) and similar between those who had both RA and DM and only DM (70% and 79%, respectively). Table 1 also describes the pattern of provider visits (rheumatologist or non-rheumatology practitioner) during the 12-month baseline period.

Lipid screening. Among the patients with RA, DM, RA and DM, and neither condition, 37%, 60%, 55%, and 41%, respectively, were screened over the 2-year follow-up period (RA versus neither, $P < 0.0001$; RA versus DM only, $P < 0.0001$) (Figure 2).

Screening for lipids based on patterns of care. Table 2 describes the rheumatologist or non-rheumatology practitioner encounters for patients with only RA. Twenty-two percent of the patients with RA saw only a rheumatologist and 56% visited both a non-rheumatology practitioner and a rheumatologist during the 12-month baseline.

In a multivariable-adjusted model including age, sex, race, comorbidities, and RA medications, the likelihood of hyperlipidemia screening was 55% higher for patients who visited both a rheumatologist and a non-rheumatology practitioner during the 2-year follow-up than for those who only visited a rheumatologist (Table 3). Hyperlipidemia screening was 21% higher for patients who only visited a non-rheumatology practitioner than for those who only visited a rheumatologist.

DISCUSSION

Our study identified a low frequency of primary lipid screening over 2 years among patients with RA. If we consider

Table 1. Baseline characteristics of study participants*

	Neither (n = 167,811)	DM only (n = 62,834)	RA only (n = 12,182)	RA and DM (n = 1,082)
Demographics				
Age range, years				
41–50	14	8	11	8
51–60	15	15	18	18
61–70	23	30	30	33
71–85	48	47	42	42
Women	60	59	82	80
Race/ethnicity				
Caucasian	65	66	78	66
African American	12	16	9	16
Hispanic	4	5	4	6
Asian	3	3	1	2
Other†	16	11	9	10
Clinical				
Hypertension	39	79	41	70
Charlson comorbidities index				
0	76	0.8	0.3	0.1
1–2	21	80	90	52
≥3	3	19	9	48
Inpatient stay (any)	8	15	14	23
Physician visits				
Rheumatologist visits				
≤1	98.3	99.1	44.8	51.5
2–4	1.5	0.7	43.4	38.4
>5	0.2	0.1	11.8	10.2
Other physician visits				
≤1	58.3	28.0	42.6	29.6
2–4	26.9	34.7	36.9	34.2
>5	14.8	37.3	20.5	36.2
RA medications				
MTX monotherapy	0	0	27	30
TNFi	0	0	26	18
Non-TNFi biologic	0	0	5	5
MTX combination‡	0	0	13	14

* Values are the percent. DM = diabetes mellitus; RA = rheumatoid arthritis; MTX = methotrexate; TNFi = tumor necrosis factor inhibitor.
† Unknown and other race combined.
‡ Methotrexate combined with either nonbiologic disease-modifying antirheumatic drug or with TNF biologic or non-TNF biologic.

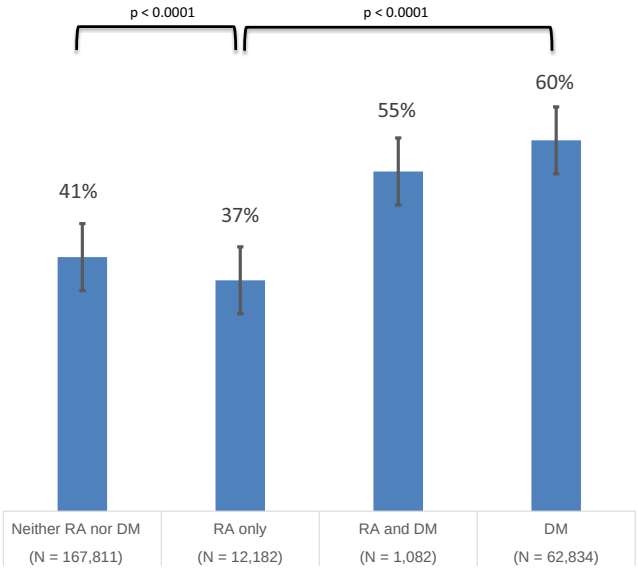

Figure 2. Proportion of patients with primary screening for low-density lipoprotein by different diseases during the 2 years of follow-up.

the 27% of patients with RA that were screened or treated for hyperlipidemia at baseline (and thus excluded from the cohort sample), slightly less than two-thirds of all patients with only RA (64%) identified in this data set were screened. If we consider the 25% of patients with DM only that were excluded at baseline because they were already screened or on treatment for hyperlipidemia, most patients with DM only were screened for hyperlipidemia in this data set. Patients with RA visited rheumatologists more frequently than they visited a non-rheumatology practitioner. Co-management with 1 of these non-rheumatology practitioners increased the likelihood of a patient being screened for hyperlipidemia. Compared with other studies that have also studied hyperlipidemia screening in patients with RA in the US, our study includes a population-based sample of patients between the ages of 40–85 years, whereas previous studies were limited to patients ages 65 years and older, only insured by Medicare, or were even focused on secondary CVD prevention (14,15).

Some health care systems that exist in Europe may provide advantages regarding CVD risk factor assessment in patients with RA. Recently, several published studies have found that European patients with RA are equally or more likely to be

Table 3. Probability of screening for hyperlipidemia among patients with RA*

Variable	Multivariable
Physician visit	
Rheumatologist	Referent
Non-rheumatology practitioner†	1.21 (1.03–1.41)‡
Non-rheumatology practitioner and rheumatologist	1.55 (1.36–1.78)‡
Age range, years	
41–50	0.86 (0.74–1.00)
51–60	Referent
61–70	0.96 (0.86–1.08)
71–85	0.74 (0.66–0.83)‡
Men vs. women	0.94 (0.85–1.05)
Race	
White	Referent
African American	1.14 (0.99–1.31)
Other	1.12 (0.99–1.26)
Charlson comorbidity index§	
0	Referent
1–2	0.74 (0.31–1.77)
≥3	0.68 (0.28–1.65)
Diabetes mellitus	1.48 (1.26–1.74)‡
Hypertension	1.02 (0.94–1.11)
TNFi biologic	1.09 (0.98–1.22)
Non-TNFi biologic	1.03 (0.86–1.24)
MTX combination¶	1.07 (0.93–1.22)
MTX monotherapy	1.06 (0.95–1.19)

* Values are the hazard ratio (HR) (95% confidence interval). RA = rheumatoid arthritis; TNFi = tumor necrosis factor inhibitor; MTX = methotrexate.
† Includes internal medicine, family medicine, nurse practitioner, and physician assistant.
‡ Statistically significant.
§ Charlson scoring in this model did not include RA.
¶ Methotrexate combined with either nonbiologic disease-modifying antirheumatic drug or with TNF biologic or non-TNF biologic.

Table 2. Baseline pattern of physician visits among patients with RA*

	All RA (n = 8,606)	Non-rheumatology practitioner only (n = 1,934)†	Rheumatologist only (n = 1,872)	Non-rheumatology practitioner and rheumatologist (n = 4,800)
Average no. of outpatient visits during 12-month baseline period, mean ± SD	8.4 ± 5.6	8.0 ± 5.4	5.6 ± 4.0	9.8 ± 5.8
≤1 visit	187 (2.2)	55 (2.8)	131 (7.0)	1 (0.0)
2–4 visits	1,927 (22.4)	483 (25.0)	799 (42.7)	645 (13.4)
>5 visits	6,492 (75.4)	1,396 (72.2)	942 (50.3)	4,154 (86.5)

* Values are the number (%) unless indicated otherwise. Baseline was defined as the documentation of non-rheumatology practitioner/rheumatologist visits and disease-modifying antirheumatic drug prescription, no low-density lipoprotein (LDL) tests during the 12-month baseline, and no statin or lipid-lowering medication use before the first LDL test during follow-up. The rheumatoid arthritis (RA) population differs from that in Table 1 in that patients must have received RA diagnoses from a rheumatologist or specialist who typically provides primary care (e.g., family practice, internal medicine, nurse practitioner, or physician assistant).
† Non-rheumatology practitioner consisted of a visit with either an internal medicine doctor, family medicine doctor, nurse practitioner, or physician assistant in the outpatient setting. Note that this data set did not distinguish if the nurse practitioner or the physician assistant was with a rheumatology clinic.

treated for hyperlipidemia and other modifiable risk factors, such as hypertension and smoking cessation, when compared with American patients with RA (23,24). Many US rheumatologists are still reluctant to take responsibility to assess and mitigate (if needed) CVD risk for patients with RA (25). In contrast, the European League Against Rheumatism (EULAR) recommendations for CVD risk management in inflammatory arthritis, and in particular in RA, have contributed to better management of CVD risk in patients with RA in some European countries (26–28). Whereas EULAR emphasizes that CVD risk is the rheumatologist's responsibility, our study, as well as others conducted in the US, showed that coordinated care between rheumatologists and non-rheumatology practitioners increased the likelihood of primary lipid screening (15,27).

Our data suggest that rheumatologists are less likely than other non-rheumatology practitioners to initiate primary lipid screening in patients with RA. Two qualitative studies that evaluated possible reasons for such hesitation among physicians found that this inaction resulted primarily from the perceived role boundaries between specialty doctors, including rheumatologists, and primary care providers, and, secondarily, from the lack of familiarity with CVD treatment guidelines. A third reason for the decreased likelihood of rheumatologists to initiate the screening comes from challenges in communication between physicians, and the fourth reason is because of the misalignment in the perceived responsibility of who should be in charge of screening and management of hyperlipidemia in patients with RA (25,29).

Our study has several strengths, including a population-based sample that included not only patients with RA but also patients with DM, a disease with a very high risk for CVD, and a random sample of the patients enrolled in similar health insurance programs. However, our general population cohort consisted of Medicare, Medicaid, and commercially insured individuals and did not include individuals that were uninsured. The benefit of the MPCD for population-based research has been previously described (17,22). Regarding the limitations of this study, our results may not affect delivery of care in light of current CVD screening guidelines since these recommend screening for hyperlipidemia every 5 years and our ascertainment period spanned only 3 years (1 year baseline, 2 year follow-up). However, EULAR CVD management guidelines for inflammatory arthritis from 2009, relevant at the time of the study, recommended annual CVD risk assessment (28). Still, our results do serve as a useful starting point to assess the US experience as a springboard to inform future CVD management practices and interventions to mitigate CVD risk among patients with RA. The data source available did not have information needed in order to determine smoking status, body mass index, or familial CVD history. The specialty of the clinicians caring for these individuals may have been misclassified, as suggested by the observation that in these data, 22% of patients with RA did not visit a rheumatologist. Indeed, it is likely that some of the NPs and PAs were providing rheumatology-specific care (likely in collaboration with a rheumatologist), but NPs and PAs attached to a rheumatology clinic could not be distinguished from those attached to a primary care clinic. We also recognize that internists recently completing fellowship and transitioning to become rheumatologists may still be classified as internists in the health insurance claims data that we used. Finally, while we recognize that the RA and DM cohort assignments were derived from administrative data, we note that our definitions included a combination of 2 ICD-9 codes for 714.xx plus condition-specific medications, which makes misclassification less likely. However, it still may have missed some cases, particularly for underrecognized diseases or DM managed only with lifestyle modification (20).

In conclusion, reducing modifiable CVD risk factors should be a priority in patients with RA. Measures to achieve this goal must be implemented and may include defining specific roles for rheumatologists, non-rheumatology practitioners, and patients to determine who should be responsible for hyperlipidemia screening and treatment for patients with RA.

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. Navarro-Millán 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. Navarro-Millán, Yang, Chen, Bartels, Fraenkel, Safford, Curtis.

Acquisition of data. Curtis.

Analysis and interpretation of data. Navarro-Millán, Yang, Chen, Yun, Jagpal, Bartels, Fraenkel, Safford, Curtis.

REFERENCES

1. Navarro-Millán I, Yang S, DuVall SL, Chen L, Baddley J, Cannon GW, et al. Association of hyperlipidaemia, inflammation and serological status and coronary heart disease among patients with rheumatoid arthritis: data from the National Veterans Health Administration. Ann Rheum Dis 2016;75:341–7.
2. Innala L, Möller B, Ljung L, Magnusson S, Smedby T, Sodergren A, et al. Cardiovascular events in early RA are a result of inflammatory burden and traditional risk factors: a five year prospective study. Arthritis Res Ther 2011;13:R131.
3. Watson DJ, Rhodes T, Guess HA. All-cause mortality and vascular events among patients with rheumatoid arthritis, osteoarthritis, or no arthritis in the UK General Practice Research Database. J Rheumatol 2003;30:1196–1202.
4. Aviña-Zubieta JA, Choi HK, Sadatsafavi M, Etminan M, Esdaile JM, Lacaille D. Risk of cardiovascular mortality in patients with rheumatoid arthritis: a meta-analysis of observational studies. Arthritis Rheum 2008;59:1690–7.
5. Solomon DH, Kremer J, Curtis JR, Hochberg MC, Reed G, Tsao P, et al. Explaining the cardiovascular risk associated with rheumatoid arthritis: traditional risk factors versus markers of rheumatoid arthritis severity. Ann Rheum Dis 2010;69:1920–5.
6. Del Rincón I, Freeman GL, Haas RW, O'Leary DH, Escalante A. Relative contribution of cardiovascular risk factors and rheumatoid arthritis clinical manifestations to atherosclerosis. Arthritis Rheum 2005;52:3413–23.

7. Baghdadi LR, Woodman RJ, Shanahan EM, Mangoni AA. The impact of traditional cardiovascular risk factors on cardiovascular outcomes in patients with rheumatoid arthritis: a systematic review and meta-analysis. PLoS One 2015;10:e0117952.
8. Gonzalez-Gay MA, Gonzalez-Juanatey C, Lopez-Diaz MJ, Piñeiro A, Garcia-Porrua C, Miranda-Filloy JA, et al. HLA-DRB1 and persistent chronic inflammation contribute to cardiovascular events and cardiovascular mortality in patients with rheumatoid arthritis. Arthritis Rheum 2007;57:125–32.
9. Danninger K, Hoppe UC, Pieringer H. Do statins reduce the cardiovascular risk in patients with rheumatoid arthritis? Int J Rheum Dis 2014;17:606–11.
10. An J, Alemao E, Reynolds K, Kawabata H, Solomon DH, Liao KP, et al. Cardiovascular outcomes associated with lowering low-density lipoprotein cholesterol in rheumatoid arthritis and matched nonrheumatoid arthritis. J Rheumatol 2016;43:1989–96.
11. Schoenfeld SR, Lu L, Rai SK, Seeger JD, Zhang Y, Choi HK. Statin use and mortality in rheumatoid arthritis: a general population-based cohort study. Ann Rheum Dis 2016;75:1315–20.
12. De Vera MA, Choi H, Abrahamowicz M, Kopec J, Lacaille D. Impact of statin discontinuation on mortality in patients with rheumatoid arthritis: a population-based study. Arthritis Care Res (Hoboken) 2012;64:809–16.
13. De Vera MA, Bhole V, Burns LC, Lacaille D. Impact of statin adherence on cardiovascular disease and mortality outcomes: a systematic review. Br J Clin Pharmacol 2014;78:684–98.
14. Bartels CM, Kind AJ, Thorpe CT, Everett CM, Cook RJ, McBride PE, et al. Lipid testing in patients with rheumatoid arthritis and key cardiovascular-related comorbidities: a medicare analysis. Semin Arthritis Rheum 2012;42:9–16.
15. Bartels CM, Kind AJ, Everett C, Mell M, McBride P, Smith M. Low frequency of primary lipid screening among medicare patients with rheumatoid arthritis. Arthritis Rheum 2011;63:1221–30.
16. Desai SS, Myles JD, Kaplan MJ. Suboptimal cardiovascular risk factor identification and management in patients with rheumatoid arthritis: a cohort analysis. Arthritis Res Ther 2012;14:R270.
17. Department of Health and Human Services. Multi-Payer Claims Database /Task 12: Summary Report and Recommended Design Option (Award No. HHSP233201000162G); 2010. URL: https://www.apcdcouncil.org/publication/multi-payer-claims-database-task-12-summary-report-and-recommended-design-option.
18. Gabriel SE. The sensitivity and specificity of computerized databases for the diagnosis of rheumatoid arthritis. Arthritis Rheum 1994;37:821–3.
19. Singh JA, Holmgren AR, Noorbaloochi S. Accuracy of Veterans Administration databases for a diagnosis of rheumatoid arthritis. Arthritis Rheum 2004;51:952–7.
20. Kim SY, Servi A, Polinski JM, Mogun H, Weinblatt ME, Katz JN, et al. Validation of rheumatoid arthritis diagnoses in health care utilization data. Arthritis Res Ther 2011;13:R32.
21. Miller DR, Safford MM, Pogach LM. Who has diabetes? Best estimates of diabetes prevalence in the Department of Veterans Affairs based on computerized patient data. Diabetes Care 2004;27 Suppl 2:B10–21.
22. Porter J, Love D, Costello A, Peters A, Rudolph B. All-payer claims database development manual: establishing a foundation for health care transparency and informed decision making. APCD Council and West Health Policy Center; 2015 Feb.:1–98.
23. Jafri K, Bartels CM, Shin D, Gelfand JM, Ogdie A. Incidence and management of cardiovascular risk factors in psoriatic arthritis and rheumatoid arthritis: a population-based study. Arthritis Care Res (Hoboken) 2017;69:51–7.
24. Alemao E, Cawston H, Bourhis F, Al M, Rutten-van Molken MP, Liao KP, et al. Cardiovascular risk factor management in patients with RA compared to matched non-RA patients. Rheumatology (Oxford) 2016;55:809–16.
25. Bartels CM, Roberts TJ, Hansen KE, et al. Rheumatologist and primary care management of cardiovascular disease risk in rheumatoid arthritis: patient and provider perspectives. Arthritis Care Res (Hoboken) 2016;68:415–23.
26. Agca R, Heslinga SC, Rollefstad S, Heslinga M, McInnes IB, Peters MJ, et al. EULAR recommendations for cardiovascular disease risk management in patients with rheumatoid arthritis and other forms of inflammatory joint disorders: 2015/2016 update. Ann Rheum Dis 2017;76:17–28.
27. Weijers JM, Rongen-van Dartel SAA, Hoevenaars D, Rubens M, Hulscher M, van Riel P. Implementation of the EULAR cardiovascular risk management guideline in patients with rheumatoid arthritis: results of a successful collaboration between primary and secondary care. Ann Rheum Dis 2018;77:480–3.
28. Peters MJ, Symmons DP, McCarey D, Dijkmans BA, Nicola P, Kvien TK, et al. EULAR evidence-based recommendations for cardiovascular risk management in patients with rheumatoid arthritis and other forms of inflammatory arthritis. Ann Rheum Dis 2010;69:325–31.
29. Ab E, Denig P, van Vliet T, Dekker JH. Reasons of general practitioners for not prescribing lipid-lowering medication to patients with diabetes: a qualitative study. BMC Fam Pract 2009;10:24.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 71, No. 12, December 2019, pp 1600–1605
DOI 10.1002/acr.23840

BRIEF REPORT

Association of Short-Term Ultraviolet Radiation Exposure and Disease Severity in Juvenile Dermatomyositis: Results From the Childhood Arthritis and Rheumatology Research Alliance Legacy Registry

Jessica Neely,[1] Craig S. Long,[2] Hugh Sturrock,[1] and Susan Kim,[1] for the Childhood Arthritis and Rheumatology Research Alliance Registry Investigators

Objective. Ultraviolet (UV) radiation is considered to be an important environmental factor in the clinical course of children with juvenile dermatomyositis (DM). We aimed to evaluate the association between UV radiation and severe disease outcomes in juvenile DM.

Methods. This is a cross-sectional study of patients with juvenile DM enrolled in the US multicenter Childhood Arthritis and Rheumatology Research Alliance (CARRA) Legacy Registry from 2010 to 2015. The mean UV index (UVI) in the calendar month prior to symptom onset in each subject's zip code was calculated from daily satellite solar noon measurements. Multivariable logistic regression was used to model the relationship between the mean UVI and calcinosis as well as other outcomes of severe disease. Covariates included sex, race, age, time to diagnosis, disease duration, and latitude.

Results. In a multivariable model, there was no association between the mean UVI and calcinosis. African American race was associated with a 3-fold greater odds of calcinosis. However, there was a significant statistical interaction between race and mean UVI. Accounting for this interaction, the odds of calcinosis markedly decreased in African American subjects and steadily increased in non–African American subjects over a range of increasing the mean UVI. Higher mean UVI was associated with decreased odds of using biologics or nonmethotrexate disease-modifying antirheumatic drugs and skin ulceration.

Conclusion. We described a novel association between UV radiation, calcinosis, and race in a large cohort of patients with juvenile DM. This study furthers our knowledge of the role of UV radiation in the clinical course of juvenile DM and highlights the complex interplay between genes and environment in the clinical phenotypes and development of calcinosis in children with juvenile DM.

INTRODUCTION

Juvenile dermatomyositis (DM) is the most common subtype of the juvenile idiopathic inflammatory myopathies (IIMs), a group of heterogeneous autoimmune disorders characterized by muscle inflammation. Juvenile DM is distinguished from other juvenile IIMs by distinctive, photo-distributed skin rashes, and ultraviolet (UV) radiation has been postulated to play a role in disease pathogenesis (1–4). Photosensitivity is reported in nearly half of patients with juvenile myositis (JM) (5), and exacerbations of skin disease following sun exposure have been described. Once DM is established, UV radiation appears to be a strong trigger: laboratory testing of non-irradiated skin of adults with DM determined increased sensitivity to ultraviolet B (UVB) radiation compared to

The Childhood Arthritis and Rheumatology Research Alliance (CARRA) Legacy Registry was supported by a grant from the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the NIH (award RC2-AR-058934). The content herein is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The CARRA Legacy Registry was also supported by CARRA, Friends of CARRA, the Arthritis Foundation, and the Duke Clinical Research Institute. Dr. Neely's work was supported by the Cure JM Foundation (Fellowship award).

[1]Jessica Neely, MD, Hugh Sturrock, PhD, Susan Kim, MD, MSSc: University of California, San Francisco; [2]Craig S. Long, MS: National Oceanic and Atmospheric Administration, College Park, Maryland.

Drs. Sturrock and Kim contributed equally to this work.

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

Address correspondence to Jessica Neely, MD, University of California, San Francisco, 550 16th Street, 5th Floor, San Francisco, CA 94158. E-mail: Jessica.Neely@ucsf.edu.

Submitted for publication October 25, 2018; accepted in revised form January 29, 2019.

SIGNIFICANCE & INNOVATIONS

- This study explores the association between the environmental factor ultraviolet (UV) radiation and disease severity outcomes in a large registry of patients with juvenile dermatomyositis (DM) by integrating clinical and demographic data from the Childhood Arthritis and Rheumatology Research Alliance (CARRA) Legacy Registry with historical National Aeronautics and Space Administration satellite measurements.
- Mean UV index (UVI) exposure in the month prior to disease onset was associated with development of calcinosis; however, the directionality of this relationship was dependent on race.
- Mean UVI exposure was not associated with other features representative of severe disease, including skin ulceration, a Childhood Health Assessment Questionnaire score >1, use of biologics or nonmethotrexate disease-modifying antirheumatic drugs, use of intravenous immunoglobulin, or persistent skin disease, muscle weakness, or steroid use beyond 2 years of disease duration.
- These results further our knowledge of the role of UV radiation in the clinical course of juvenile DM and highlight the need for clinicians and researchers to be aware of the complex interplay of genes and environment in the clinical phenotypes of children with juvenile DM.

healthy controls, many of whom also reported photosensitivity and disease exacerbation following sun exposure (6), and questionnaire data from patients with DM and JM suggested that UV exposure is an important environmental exposure correlated with disease flares (7).

Prior research investigating the link between UV radiation and IIMs suggests that UV radiation may modulate myositis phenotypes and autoantibody profiles. A global study showed that the proportion of individuals with DM relative to polymyositis rose incrementally with increasing UV radiation across diverse geographic regions worldwide, and these differences could not be explained by variation in population-specific genetic structure (1). In a study of 298 patients with JM, those with juvenile DM who had higher UV radiation exposure in the month prior to symptom onset were more likely to have anti-p155/140 antibodies (4), which have been associated with a chronic disease course (2,5). In addition, they were less likely to have anti-MJ antibodies, which have been associated with a monocyclic disease course (5). Among this cohort, the strongest association between UV radiation and anti-p155/140 antibodies was observed in white males, indicating that there may be differential effects of UV radiation based on sex and race.

Collectively, these findings suggest that UV radiation may modulate clinical phenotypes. However, it is unclear if initial UV radiation exposure has systemic effects on the immune system that result in a more severe disease course. The development of severe disease features, such as calcinosis or skin ulcerations, and need for stronger immune suppressive agents can cause significant morbidity in juvenile DM. Calcinosis, in particular, may cause infection, pain, limited joint mobility, and physical disfigurement. Risk factors for severe disease and calcinosis are not well understood but have been associated with a delay in diagnosis, longer disease duration, race, and myositis-specific autoantibodies (5,8).

In this study, we investigated the association between the mean UV index (UVI) in the month prior to symptom onset and disease severity in a large cohort of patients with juvenile DM from the Childhood Arthritis and Rheumatology Research Alliance (CARRA) Legacy Registry. We used calcinosis and other measures collected in the registry, including skin ulceration, a Childhood Health Assessment Questionnaire (C-HAQ) score >1, and second-line medication use as proxies for severe disease. We hypothesized that higher UV radiation exposure would be associated with more severe disease outcomes. An enhanced understanding of the role of UV radiation in juvenile DM can help clinicians to develop interventions to attenuate this exposure and improve long-term outcomes in juvenile DM.

PATIENTS AND METHODS

Patients. This is a cross-sectional study of patients enrolled in the CARRA Legacy Registry who met definite or probable diagnostic criteria for juvenile DM according to the modified criteria proposed by Bohan and Peter (9). This is a US multicenter registry, which enrolled subjects with a variety of childhood rheumatic diseases between 2010 and 2015. A subset of patients with juvenile DM enrolled in this registry has been described previously (10). Subjects were in various stages of disease at the time of enrollment. Those with incomplete data for the variables date of symptom onset and zip code were excluded. Data were abstracted from the enrollment visit or at the subsequent visit when enrollment visit data were missing.

Methods. We determined the mean UVI based on the US zip code of subjects in the calendar month prior to symptom onset. This time frame was selected in order to evaluate the effect of short-term UV radiation in accordance with methods used by Shah et al (4) and the seasonal variation of UV radiation. UVI is an internationally standardized unit on a linear scale, ranging from 0 to the mid-teens, which quantifies the amount of skin-damaging erythema when the sun is highest in the sky (i.e., solar noon). A higher number indicates a shorter amount of time to skin erythema, which is also influenced by skin color and tendency to burn. The amount of UV radiation reaching the earth's surface is affected by total column ozone, elevation, surface reflectivity, cloud transmissivity, and tropospheric aerosol loading (pollutants or dust), all of which are accounted for in the UVI calculation.

The mean UVI in the calendar month prior to symptom onset was obtained from the National Aeronautics and Space Administration Total Ozone Mapping Spectrometer and Ozone Monitoring Instrument, satellite instruments that record daily solar noon estimates of erythema dose rate across the US. Calcinosis, a hallmark morbidity in juvenile DM that reflects severe disease and damage, was used as the primary outcome. Secondary outcomes included history of skin ulceration, a C-HAQ disability index score >1, treatment with biologics or nonmethotrexate disease-modifying antirheumatic drugs (DMARDs), and treatment with intravenous immune globulin (IVIG). In the subset of patients with disease duration of >2 years, we assessed additional outcomes, including persistent weakness, persistent skin rash (malar, Gottron's papules, heliotrope, or V-/Shawl sign), and persistent steroid use.

Statistical analysis. Subjects were stratified by mean UVI quartiles and comparisons were made using chi-square tests for categorical variables and Kruskal-Wallis tests for continuous variables, which were non-normal. Characteristics were also evaluated and stratified by race. Based on our understanding of the relationship between UV radiation and disease severity outcomes, we included the following covariates in our model: sex, race, age at disease onset, time to diagnosis, and disease duration. We evaluated for interactions between sex and mean UVI and between race and mean UVI based on results of prior literature (3,4). We dichotomized race as African American and non–African American based on prior work related to race and calcinosis as well as differential risk for damage from UV radiation due to skin pigmentation.

Multivariable logistic regression was used to evaluate the association between mean UVI and each disease outcome. We evaluated for linearity in the predictors mean UVI and latitude by including these terms as polynomial terms and evaluating if the polynomial terms enhanced fit. We evaluated interactions between mean UVI and sex and mean UVI and race for each model. We included the interaction terms in the model if they were statistically significant. To test for spatial autocorrelation, we ran Moran's I test on the residuals from the model.

RESULTS

Patient characteristics. A total of 522 subjects were included. The median age at disease onset was 5 years (interquartile range [IQR] 3–9 years), 71.8% of subjects were female, and 11% were African American. Among non–African American subjects, 89% identified as white, and the remaining 11% identified as either Asian, American Indian or Alaskan Native, Native Hawaiian or Pacific Islander, or multiracial/other. Median time to diagnosis was 3.1 months (IQR 1.6–7.2 months), and median disease duration was 1.9 years (IQR 0.5–4.4 years). Eleven percent developed calcinosis, 5.6% had skin ulcerations, 15.5% had a C-HAQ score >1, 24.9% received biologics or nonmethotrexate DMARDs, and 34.1% received IVIG. There were 247 patients with disease duration >2 years, of whom 26% had persistent rash, 14% persistent weakness, and 25% persistent steroid use. The mean UVI was mean ± SD 4.9 ± 2.6. Stratified by mean UVI quartiles, clinical and demographic characteristics were similar except for a higher proportion of individuals with skin ulceration (P = 0.03) and history of treatment with biologics or DMARDs (P = 0.02) in lower quartiles of mean UVI. Patient characteristics stratified by race were similar except for a greater prevalence of calcinosis among African American subjects (24.5%) compared to 9.2% in non–African American subjects (P = 0.002).

Mean UVI as a predictor of calcinosis. In a multivariable logistic regression model, there was no significant association between mean UVI and calcinosis (adjusted P = 0.64) (Table 1). African American race was associated with a 3-fold greater odds of calcinosis. However, there was significant statistical interaction between race and mean UVI. Accounting for this interaction, the odds of calcinosis markedly decreased in African American subjects and steadily increased in non–African American subjects over the range of increasing the mean UVI. This interaction is shown in Figure 1 for a representative female at mean values of all other covariates in the model. There was no interaction between mean UVI and sex. Therefore, this term was not included in the model. Moran's I test revealed no significant residual spatial autocorrelation (P = 0.36). Additional risk

Table 1. Multivariable logistic regression model of the mean UVI as a predictor of calcinosis*

Predictor	Unadjusted			Adjusted		
	OR	95% CI	*P*	OR	95% CI	*P*
Mean UVI	0.94	0.84–1.05	0.25	1.03	0.90–1.18	0.64
African American race	3.22	1.54–6.41	0.001†	3.36	1.22–8.10	0.01†
Mean UVI × African American race interaction	0.69	0.48–0.95	0.03†	0.67	0.45–0.94	0.03†
Age at disease onset, year	1.01	0.94–1.09	0.71	1.10	1.01–1.20	0.03†
Female sex	0.62	0.34–1.14	0.11	0.48	0.25–0.95	0.03†
Diagnosis interval, month	1.03	1.01–1.05	0.005†	1.04	1.02–1.06	< 0.001†
Disease duration, year	1.23	1.13–1.33	< 0.001†	1.30	1.18–1.44	< 0.001†

* Model adjusted for sex, age at disease onset, time to diagnosis, disease duration, and race. Mean UVI = mean ultraviolet index; OR = odds ratio; 95% CI = 95% confidence interval.
† P < 0.05.

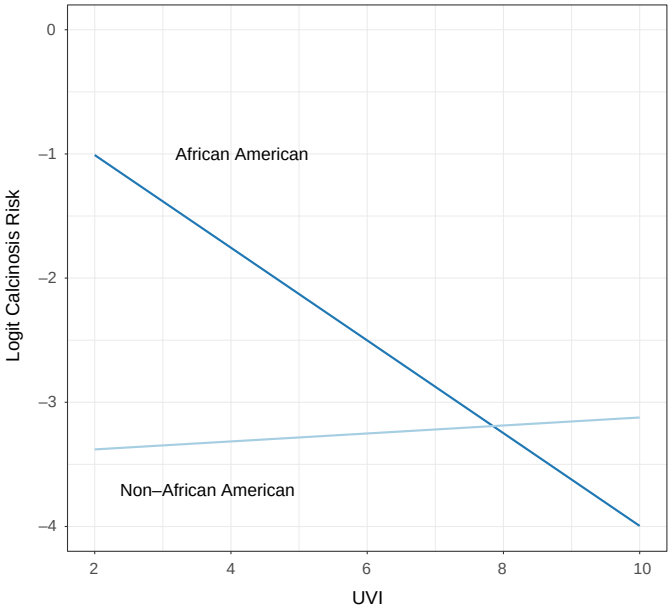

Figure 1. Log odds of calcinosis by race. UVI = ultraviolet index.

factors for calcinosis included male sex, older age at disease onset, longer disease duration, and delay in diagnosis (Table 1).

Secondary outcomes. In a univariable model, the mean UVI was associated with decreased odds of developing skin ulceration and treatment with biologics or nonmethotrexate DMARDs (Table 2). After adjustment for covariates, this relationship remained significant for the outcome of treatment with biologics or nonmethotrexate DMARDs and trended toward significance for the outcome of skin ulceration. In the multivariable models, there was also a trend toward increased odds of having a C-HAQ disability index score >1, but this did not meet statistical significance.

DISCUSSION

In this study of 522 subjects with juvenile DM enrolled in the CARRA Legacy Registry, we describe the association between short-term UV radiation exposure in the month prior to disease onset and several outcomes representative of severe disease. We found a novel association between mean UVI and calcinosis that was dependent on race. Consistent with prior studies, which identified African American race as a risk factor for calcinosis in juvenile DM (8), we found that African American subjects living in areas with a lower mean UVI had a 3-fold greater odds of calcinosis compared to non–African American subjects. However, when accounting for interaction between race and mean UVI, we were surprised to find a striking negative correlation between calcinosis and the mean UVI in African American subjects, suggesting a protective effect of higher UV radiation exposure on calcinosis risk in this subgroup. Non–African American subjects had increased risk of calcinosis beyond those of African American subjects at higher levels of mean UVI, suggesting a correlation between higher UV radiation exposure and development of calcinosis in subjects with lighter skin. These findings help to confirm the need for a personalized, differential approach to treatment and monitoring recommendations in patients with juvenile DM.

We would like to emphasize that these findings are correlative, and future research is needed to better understand the causative influence of UV radiation in myositis. However, our results are in accordance with prior studies of juvenile DM and DM in the US that have shown associations between UV radiation and clinical phenotypes to be significant only in white individuals (3,4). Skin color is a key factor in determining time to burn, and thus may also determine the susceptibility of an individual to the effects of UV radiation on immune responses. The field of photoimmunology has shown that in addition to local immune responses, UV radiation causes systemic immunomodulatory effects, which have been theorized to play a role in human autoimmune diseases (11). In addition, there is growing evidence that vitamin D levels are associated with disease activity in autoimmune diseases, including juvenile DM (12).

Genetic factors may also play a role in predisposing individuals with juvenile DM to UV radiation sensitivity as well as calcinosis. The tumor necrosis factor (TNF)-308A polymorphism is a risk

Table 2. Mean UVI as a predictor of secondary outcomes of disease severity*

Outcome	Unadjusted			Adjusted		
	OR	95% CI	*P*	OR	95% CI	*P*
All patients (n = 522)						
Skin ulceration	0.85	0.72–0.99	0.04†	0.86	0.72–1.00	0.05
C-HAQ score >1	1.07	0.98–1.18	0.14	1.08	0.98–1.19	0.08
Biologics or non-MTX DMARDs	0.88	0.82–0.96	0.003†	0.87	0.80–0.95	0.003†
IVIG	1.02	0.95–1.09	0.64	1.00	0.93–1.08	0.91
Disease duration >2 years (n = 247)						
Persistent weakness	0.96	0.83–1.12	0.61	0.97	0.83–1.13	0.65
Persistent rash	0.95	0.86–1.06	0.39	0.97	0.87–1.08	0.61
Persistent steroid use	0.91	0.80–1.02	0.11	0.92	0.81–1.04	0.20

* Model adjusted for sex, age at disease onset, disease duration, and race. UVI = ultraviolet index; OR = odds ratio; 95% CI = 95% confidence interval; C-HAQ = Childhood Health Assessment Questionnaire; MTX = methotrexate; DMARDs = disease-modifying antirheumatic drugs; IVIG = intravenous immunoglobulin.
† $P < 0.05$. Significant.

factor for juvenile DM (13) and DM (14) in the white population. Stimulation of skin keratinocytes and fibroblasts with UVB causes increased transcription of the cytokine TNF. This cytokine triggers cell apoptosis and release of intracellular autoantigens, which may contribute to disease pathogenesis. Children with juvenile DM who have the TNF-308A polymorphism have increased risk of developing calcinosis (14,15). Consequently, it is conceivable that non–African American subjects in our study, the majority of which were white, are enriched for this genetic risk factor, which may make them more susceptible to both the effects of UV radiation on immune responses, possibly via increased TNF production, and the development of calcinosis.

We did not find additional evidence to support our hypothesis when evaluating additional outcomes representative of severe disease. In fact, we found decreased odds of treatment with biologics and nonmethotrexate DMARDs and skin ulceration with increasing mean UVI. There are several possible explanations for these unexpected results. One possibility is that children who are photosensitive and living in regions with higher UV radiation are more conscious of the need for sun protection and modify this risk factor. Another possibility is that genetic factors and/or myositis autoantibodies moderate the effect of UV radiation on disease phenotypes. Mamyrova et al showed that while TNF-308A is a genetic risk factor for calcinosis it is not a risk factor for skin or gastrointestinal ulcerations (15). Likewise, prior research shows that anti-MJ antibodies are associated with calcinosis (5) and anti-p155/140 antibodies are associated with skin ulcerations in the juvenile DM population (16). Although UV radiation is an important environmental factor in the course of juvenile DM, it is possible that genetic and serologic profiles are more influential in determining disease severity.

Prior studies investigating the role of UV radiation in IIMs include a more heterogeneous group of patients with various types of IIMs. Our study is strengthened by focusing on a large cohort of patients with juvenile DM from the CARRA Registry who reside in diverse geographic regions across the US. The CARRA Registry highlights the versatility of large patient registries in research and demonstrates how patient registries can be integrated with other publicly available datasets to answer important questions. Furthermore, we utilized individual zip codes for calculation of the mean UVI, which we believe provides a more relevant and resolute estimate of UV radiation exposure.

There are limitations to this study to consider. Individual exposure data, such as the duration and time of day spent outdoors and use of sun protective measures (e.g., sunscreen, hats) represent unmeasured confounders. Furthermore, behaviors modify the risk associated with UV radiation exposure, and there may be regional differences in awareness and use of these interventions. Historically, individual behavior and exposure data have not been practical to collect, and questionnaires are subject to recall bias. Emerging technologies may allow for in-depth studies regarding UV radiation and disease on the individual level in the future. In addition, we used the mean UVI to estimate UV radiation exposure, which is heavily weighted in UVB. Unfortunately, it is not possible to separate UVA and UVB exposures in historical UV radiation recordings. Future studies that are able to distinguish these exposures are needed.

We were also limited in the outcomes that we could assess and realize that the outcomes in our study may not definitively identify subjects with severe disease. Overall, patients in the CARRA Registry trended toward milder disease with a low prevalence of severe features (10), and since screening methods are not standardized, we may not have captured the true incidence of calcinosis. In addition, the median disease duration in this cohort was 1.9 years, which may not have been long enough to identify all patients at risk for developing calcinosis.

In summary, we describe a novel association between UV radiation and the development of calcinosis dependent upon race in a large cohort of patients with juvenile DM. This study furthers our knowledge of the role of UV radiation in the clinical course of juvenile DM and highlights the need for clinicians and researchers to be aware of the complex interplay of genes and environment in the clinical phenotypes of children with juvenile DM. As we continue to study this complex autoimmune disease, it is imperative that we consider the combination of multiple data types, including the exposome, demographics, genomics, serologic patterns, and clinical phenotypes, in order to personalize our treatment approach and improve outcomes for each child affected by juvenile DM.

ACKNOWLEDGMENTS

The authors thank all participants and hospital sites that recruited patients for the CARRA Registry, as well as the Cure JM Foundation. The authors also thank the following CARRA Registry site principal investigators and research coordinators: L. Abramson, E. Anderson, M. Andrew, N. Battle, M. Becker, H. Benham, T. Beukelman, J. Birmingham, P. Blier, A. Brown, H. Brunner, A. Cabrera, D. Canter, D. Carlton, B. Caruso, L. Ceracchio, E. Chalom, J. Chang, P. Charpentier, K. Clark, J. Dean, F. Dedeoglu, B. Feldman, P. Ferguson, M. Fox, K. Francis, M. Gervasini, D. Goldsmith, G. Gorton, B. Gottlieb, T. Graham, T. Griffin, H. Grosbein, S. Guppy, H. Haftel, D. Helfrich, G. Higgins, A. Hillard, J. R. Hollister, J. Hsu, A. Hudgins, C. Hung, A. Huttenlocher, N. Ilowite, A. Imlay, L. Imundo, C. J. Inman, J. Jaqith, R. Jerath, L. Jung, P. Kahn, A. Kapedani, D. Kingsbury, K. Klein, M. Klein-Gitelman, A. Kunkel, S. Lapidus, S. Layburn, T. Lehman, C. Lindsley, M. Macgregor-Hannah, M. Malloy, C. Mawhorter, D. McCurdy, K. Mims, N. Moorthy, D. Morus, E. Muscal, M. Natter, J. Olson, K. O'Neil, K. Onel, M. Orlando, J. Palmquist, M. Phillips, L. Ponder, S. Prahalad, M. Punaro, D. Puplava, S. Quinn, A. Quintero, C. Rabinovich, A. Reed, C. Reed, S. Ringold, M. Riordan, S. Roberson, A. Robinson, J. Rossette, D. Rothman, D. Russo, N. Ruth, K. Schikler, A. Sestak, B. Shaham, Y.

Sherman, M. Simmons, N. Singer, S. Spalding, H. Stapp, R. Syed, E. Thomas, K. Torok, D. Trejo, J. Tress, W. Upton, R. Vehe, E. von Scheven, L. Walters, J. Weiss, P. Weiss, N. Welnick, A. White, J. Woo, J. Wootton, A. Yalcindag, C. Zapp, L. Zemel, and A. Zhu.

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. Neely 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. Neely, Sturrock, Kim.
Acquisition of data. Long, Kim.
Analysis and interpretation of data. Neely, Sturrock, Kim.

REFERENCES

1. Okada S, Weatherhead E, Targoff IN, Wesley R, Miller FW, for the International Myositis Collaborative Study Group. Global surface ultraviolet radiation intensity may modulate the clinical and immunologic expression of autoimmune muscle disease. Arthritis Rheum 2003;48:2285–93.
2. Habers GE, Huber AM, Mamyrova G, Targoff IN, O'Hanlon TP, Adams S, et al. Association of myositis autoantibodies, clinical features, and environmental exposures at illness onset with disease course in juvenile myositis. Arthritis Rheumatol 2016;68:761–8.
3. Love LA, Weinberg CR, McConnaughey DR, Oddis CV, Medsger TA Jr, Reveille JD, et al. Ultraviolet radiation intensity predicts the relative distribution of dermatomyositis and anti–Mi-2 autoantibodies in women. Arthritis Rheum 2009;60:2499–504.
4. Shah M, Targoff IN, Rice MM, Miller FW, Rider LG, with the Childhood Myositis Heterogeneity Collaborative Study Group. Ultraviolet radiation exposure is associated with clinical and autoantibody phenotypes in juvenile myositis. Arthritis Rheum 2013;65:1934–41.
5. Rider LG, Shah M, Mamyrova G, Huber AM, Rice MM, Targoff IN, et al. The myositis autoantibody phenotypes of the juvenile idiopathic inflammatory myopathies. Medicine (Baltimore) 2013;92:223–43.
6. Dourmishev L, Meffert H, Piazena H. Dermatomyositis: comparative studies of cutaneous photosensitivity in lupus erythematosus and normal subjects. Photodermatol Photoimmunol Photomed 2004;20:230–4.
7. Mamyrova G, Rider LG, Ehrlich A, Jones O, Pachman LM, Nickeson R, et al. Environmental factors associated with disease flare in juvenile and adult dermatomyositis. Rheumatology (Oxford) 2017;56:1342–7.
8. Hoeltzel MF, Oberle EJ, Robinson AB, Agarwal A, Rider LG. The presentation, assessment, pathogenesis, and treatment of calcinosis in juvenile dermatomyositis. Curr Rheumatol Rep 2014;16:467.
9. Bohan A, Peter JB. Polymyositis and dermatomyositis (first of two parts). N Engl J Med 1975;292:344–7.
10. Robinson AB, Hoeltzel MF, Wahezi DM, Becker ML, Kessler EA, Schmeling H, et al. Clinical characteristics of children with juvenile dermatomyositis: the Childhood Arthritis and Rheumatology Research Alliance Registry. Arthritis Care Res (Hoboken) 2014;66:404–10.
11. Norval M. The mechanisms and consequences of ultraviolet-induced immunosuppression. Prog Biophys Mol Biol 2006;92:108–18.
12. Robinson AB, Thierry-Palmer M, Gibson KL, Rabinovich CE. Disease activity, proteinuria, and vitamin D status in children with systemic lupus erythematosus and juvenile dermatomyositis. J Pediatr 2012;160:297–302.
13. Pachman LM, Liotta-Davis MR, Hong DK, Kinsella TR, Mendez EP, Kinder JM, et al. TNFα-308A allele in juvenile dermatomyositis: association with increased production of tumor necrosis factor α, disease duration, and pathologic calcifications. Arthritis Rheum 2000;43:2368–77.
14. Werth VP, Callen JP, Ang G, Sullivan KE. Associations of tumor necrosis factor α and HLA polymorphisms with adult dermatomyositis: implications for a unique pathogenesis. J Invest Dermatol 2002;119:617–20.
15. Mamyrova G, O'Hanlon TP, Sillers L, Malley K, James-Newton L, Parks CG, et al. Cytokine gene polymorphisms as risk and severity factors for juvenile dermatomyositis. Arthritis Rheum 2008;58:3941–50.
16. Gunawardena H, Wedderburn LR, North J, Betteridge Z, Dunphy J, Chinoy H, et al. Clinical associations of autoantibodies to a p155/140 kDa doublet protein in juvenile dermatomyositis. Rheumatology (Oxford) 2007;47:324–8.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 71, No. 12, December 2019, pp 1606–1610
DOI 10.1002/acr.23808

Risk of Ocular Anomalies in Children Exposed In Utero to Antimalarials: A Systematic Literature Review

Rouan Gaffar, Christian A. Pineau, Sasha Bernatsky, Susan Scott, and Évelyne Vinet

Objective. To determine whether offspring from mothers with systemic lupus erythematosus (SLE), exposed in utero to antimalarials, have an increased risk of ocular anomalies during childhood versus unexposed SLE offspring.

Methods. We systematically performed searches of PubMed, Embase, and Web of Science databases for original human data on fetal and/or child ocular outcomes following exposure to antimalarials during pregnancy and/or lactation, from their inception until March 2017.

Results. A total of 10 cohort studies and 2 randomized controlled trials, ranging in size from 6 to 444 exposed infants studied, and 3 case reports met the inclusion criteria for our systematic review. Collectively, 1,477 infants were studied, 789 of which were exposed to hydroxychloroquine or chloroquine. In all, 563 exposed infants had follow-up visits after delivery (ranging from <3 months to 19 years), and 331 of these exposed infants underwent ophthalmologic examinations during the follow-up period. Our review of the literature suggests a low-to-nonexistent risk of visual abnormalities in offspring exposed to antimalarials.

Conclusion. In children exposed to appropriate doses of antimalarials antenatally, the risk of ocular toxicity appears low to nonexistent. The potential benefits and risks of antimalarials should be discussed in all SLE pregnancies, and high dosages should continue to be avoided.

INTRODUCTION

Hydroxychloroquine (HCQ) and chloroquine (CQ) are antimalarial agents used to treat systemic lupus erythematosus (SLE). HCQ is effective in reducing lupus flares and promotes remission (1), but it can cause ocular toxicity when taken over long periods of time and/or at high doses, leading in some cases to irreversible retinopathy (2). Moreover, HCQ concentrations in the cord blood have been demonstrated to be nearly as high as in maternal blood (3). Animal studies have shown an accumulation of the drug in fetal mice eyes (4).

On the other hand, withdrawing from the drug could lead to a 2.5-fold increase in the risk of a clinical flare-up (5), which during pregnancy can result in maternal and fetal complications, including fetal loss, preterm birth, low birth weight, and preeclampsia (6). Antimalarials have been proven to decrease the risk of flare in both pregnant (7) and nonpregnant patients (1). For these reasons, many researchers recommend that the patient continue antimalarials throughout gestation. Prior literature reviews suggest no increased risk of congenital birth defects, spontaneous abortions, premature births, or visual or hearing impairments in women taking antimalarials during pregnancy (8,9). However, the risk of toxicity remains uncertain in the context of individual case reports showing retinal damage in infants prenatally exposed to antimalarials (10,11). Therefore, we performed a systematic literature review to determine whether children exposed in utero to antimalarials have an increased risk of developing ocular anomalies during childhood compared to children who were not exposed to this medication.

MATERIALS AND METHODS

Search strategy. Our systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement (12) and followed a prespecified protocol. To identify all studies with original human data on fetal and/or child ocular outcomes following exposure to antimalarials during pregnancy and/or lactation, searches of PubMed, Embase, and Web of Science were conducted in March 2017. Our search strategy was restricted to primary studies published in English or French and included

Supported by a McGill University Department of Medicine Research Award. Dr. Vinet's work was supported by a Fonds de Recherche du Québec Santé Clinical Research Scholar Junior 1 Award.

Rouan Gaffar, MD, Christian A. Pineau, MD, Sasha Bernatsky, MD, PhD, Susan Scott, MSc, Évelyne Vinet, MD, PhD: McGill University Health Centre, Montreal, Quebec, Canada.

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

Address correspondence to Évelyne Vinet, MD, PhD, Research Institute of the McGill University Health Centre, 5252 Boulevard de Maisonneuve Ouest, Office 3D.57, Montréal H4A 3S5, Quebec, Canada. E-mail: evelyne.vinet@mcgill.ca.

Submitted for publication April 29, 2018; accepted in revised form November 6, 2018.

SIGNIFICANCE & INNOVATIONS

- We conducted a systematic review to quantify the potential risks of antimalarial exposures on ocular complications in offspring of women with systemic lupus erythematosus (SLE). Our results suggest a low-to-nonexistent risk of visual abnormalities in offspring exposed to antimalarials.
- The potential benefits for pregnant women with SLE to continue antimalarials should be weighed against the possible risk, and high dosages should be avoided.

the following key words combined with relevant Boolean operators: antimalarials, hydroxychloroquine, chloroquine, infant, pregnancy, lactation, breastfeeding, retinal diseases, eye abnormalities, retinopathy, and ocular anomalies (for a full list of MeSH words see Supplementary Appendix 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23808/abstract). Reference lists of the primary studies and previous review articles were also manually searched for relevant articles not already captured by the electronic searches mentioned above. Any pertinent secondary references were also reviewed.

Study selection. We included randomized controlled trials, observational studies, and case reports on women taking antimalarials in pregnancy. We excluded all abstracts and conference proceedings, editorials, reviews, and commentaries. We also excluded studies that did not show ocular abnormalities as an outcome, as well as basic science and animal models. Finally, inclusion was limited to publications in English and French.

One reviewer (RG) screened the citations (titles and abstracts) identified from all of the sources. Subsequently, full-text articles of the studies selected in the initial screen were reviewed to identify the final set of relevant studies (Figure 1). Data were then extracted from included articles using a data collection spreadsheet. Collected data included type of study, year of publication, maternal disease type, pharmacologic therapy (quinine, HCQ, and/or CQ), sample size (mothers, live exposed infants, and live unexposed infants), reported ocular anomalies, and the effect estimate for risk of ocular anomalies.

RESULTS

Our initial search yielded 345 articles (Figure 1). Five additional articles were identified via a hand search of reference lists. After duplicates were removed, 266 titles and abstracts were screened, of which 27 were retrieved for full-text review. A total of 10 cohort studies (7,13–21) and 2 randomized-controlled trials (22,23), ranging in size from 6 to 444 exposed infants studied, and 3 case reports (10,11,24) met the inclusion criteria for our systematic review (Table 1).

Nine studies included infants born to mothers with rheumatic diseases (7,15–21,23), and 5 studies involved mothers taking antimalarials for malaria prophylaxis (18,20,22) or treatment (13,14). The 3 case reports featured offspring born to mothers receiving antimalarial therapy for discoid lupus erythematous (24), malaria prophylaxis (10), and rheumatoid arthritis (11). Collectively, 1,477 infants were studied, 789 of which were exposed to HCQ or CQ. In total, 2 infants exposed to HCQ were born with retinal hemorrhages that had healed by the first month (18). No other clinically evident ocular anomalies were noted at birth. Only 2 studies (15,18) conducted systematic ophthalmologic evaluations (e.g., retinal exam, electroretinogram [ERG]) during the neonatal period on a total of 30 infants exposed to HCQ.

Of the 563 exposed infants who had follow-up visits after delivery (ranging from <3 months to 19 years) (7,13,14,16–20,22,23), 331 of these exposed infants underwent ophthalmologic examinations during the follow-up period (17–19,22,23,25). In 1 study (25), ERGs and retinal examinations were conducted on 4 infants exposed to HCQ, all of which were normal. The randomized controlled trial by Villegas et al (22) only followed up a subset of exposed infants (251 of 444), all of which received visual acuity examinations at age 1 year and were found to be normal.

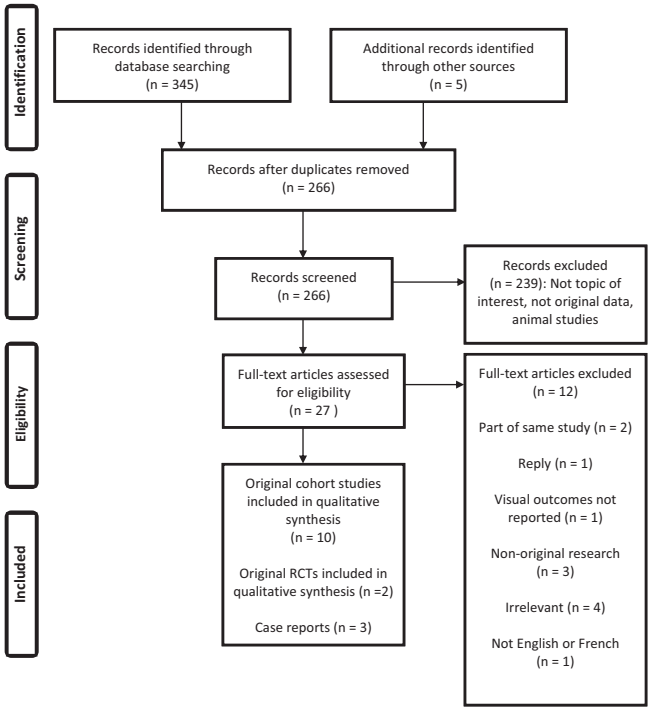

Figure 1. Flow diagram of the systematic literature review based on PRISMA guidelines (26). (For the PRISMA checklist, see Supplementary Appendix 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23808/abstract). RCTs = randomized controlled trials.

Table 1. Characteristics of studies included in the systematic literature review (n = 15)*

Author (ref.)	Type of study	Maternal disease type	Type of antimalarial	Live infants exposed, no.	Live infants unexposed, no.	Live infants exposed with ocular anomalies, no.	Risk of ocular anomalies†
Adam et al (13)	Cohort, prospective	Malaria treatment	Quinine	25	NA	None	NA
Adam et al (14)	Cohort, prospective	Malaria treatment	Quinine	33	NA	None	NA
Cimaz et al (15)	Cohort	SLE, DLE, undifferentiated CTD	HCQ	6	NA	None	NA
Costedoat et al (16)	Cohort, prospective	SLE, SS, APS, miscellaneous or unclassified CTD	HCQ	117	59	None	NA
Hart et al (24)	Case report	DLE	CQ	3	3	3 exposed: retinal pigmentary changes with no visual symptoms, observed with congenital defects; 3 unexposed: no retinal examination done, clinically healthy	NA
Paufique et al (10)	Case report	Malaria prophylaxis	Quinine	2	NA	2: retinal lesions originating from birth and not having evolved	NA
Renault et al (17)	Cohort, prospective	SLE or other CTD	HCQ	21	NA	No clinical visual abnormalities; 6 infants had abnormal ERGs, but normal fundus at ophthalmologic examination before age 4 years	Duration of exposure to HCQ longer (41.57 weeks) for infants with normal ERG than infants with abnormal ERG (36.37 weeks); $P = 0.007$‡
Clowse et al (6)	Cohort, prospective	SLE	HCQ	47	145	None	NA
Motta et al (18)	Cohort, prospective	SLE, MCTD, SCLE, undifferentiated CTD, APS, RA, SS	HCQ	40	NA	2 had retinal hemorrhages that resolved by 1 month; no evidence of retinal disease at follow-up	NA
Villegas et al (22)	RCT, double blind	Malaria prophylaxis	CQ	444	431	None	NA
Klinger et al (19)	Cohort, prospective	SLE, RA	HCQ and CQ	21	NA	None	NA
Levy et al (20)	Cohort, prospective	SLE (n = 11), RA (n = 3), malaria prophylaxis (n = 4)	HCQ and CQ	14	NA	None	NA
Levy et al (23)	RCT, double blind	SLE, DLE	HCQ	10	9	None	NA
Buchanan et al (21)	Cohort, retrospective	SLE	HCQ	32	44	None	NA
Mulholland et al (11)	Case report	RA	HCQ (MTX, folic acid)	1	NA	Congenital anterior chamber dysgenesis (Peters anomaly)	NA

* NA = not applicable; SLE = systemic lupus erythematosus; DLE = discoid lupus erythematosus; HCQ = hydroxychloroquine; CTD = connective tissue diseases; SS = Sjögren's syndrome; APS = antiphospholipid syndrome; CQ = chloroquine; ERG = electroretinogram; MCTD = mixed connective tissue disease; SCLE = subacute cutaneous lupus erythematosus; RA = rheumatoid arthritis; RCT = randomized controlled trial; MTX = methotrexate.
† Effect estimate.
‡ Statistically significant.

In the remaining studies (17–19,22,23), ophthalmologic examinations were conducted systematically for all exposed patients at varying time points during the follow-up period. In 1 study, 6 infants were found to have had abnormal ERGs but had a normal fundus on ophthalmologic examination before age 4 years (17). The ophthalmologic examinations in the remaining studies were otherwise normal.

Three case reports showing ocular anomalies following in utero exposure to antimalarials were also found in our literature search. In 1964, a woman with discoid lupus erythematosus was reported to have taken CQ phosphate during 4 of her 7 pregnancies, 1 of which ended in miscarriage. The remaining 3 exposed offspring were born with various congenital defects, including retinal pigmentary changes of unknown significance, while the 3 unexposed infants were born healthy (24). In another report published in 1969, 2 sisters exposed in utero to an excess daily dose of quinine (a drug from which CQ was derived) for malaria prophylaxis were documented with retinal lesions appearing to have originated from birth and not having evolved (10). Notably, in all 5 cases of retinal changes, the mothers were consistently taking excess doses of the drug. Finally, in 2011 a neonate exposed in utero to HCQ, methotrexate (a proven teratogen), and folic acid supplementation presented with the Peters anomaly, a congenital ocular defect consisting of anterior chamber dysgenesis (11). The mother had rheumatoid arthritis and discontinued methotrexate 8 weeks into her pregnancy after learning that she was pregnant. She continued to take HCQ and folic acid.

DISCUSSION

We aimed to determine whether offspring from mothers with SLE exposed in utero to antimalarials have an increased risk of ocular anomalies during childhood compared to unexposed offspring from mothers with SLE. To our knowledge, this is the most current and comprehensive systematic review assessing potential fetal ocular toxicity in infants exposed in utero to antimalarials.

Of the 1,477 infants studied, 789 were exposed to HCQ or CQ, of which only 2 had clinically evident ocular anomalies noted at birth, namely retinal hemorrhages that healed by the first month (18). Among the 6 studies (17–19,22,23,25) in which ophthalmologic examinations were done during follow-up, only 1 study showed potential abnormalities. The study by Renault et al (17) raised concern after 6 of 21 infants exposed to HCQ antenatally were found to have abnormal ERGs. On repeat assessment at ages between 6.5 and 9.5 months, the ERG abnormalities persisted in 5 of the infants, after normalizing in 1. All 6 infants underwent fundoscopy examinations before the age of 4, all of which were normal, and received no further ERGs. The significance of these ERG abnormalities and the correlation to HCQ exposure is uncertain. Mean duration of exposure to HCQ was close to 5 weeks longer in the 15 infants with normal ERG findings than in the 6 with abnormal ERGs. Furthermore, there were no unexposed infants in the study to which to compare the incidence of abnormal ERG results. Finally, 2 studies previously published had shown normal ERG results in 10 exposed infants (15,25). Overall, no other studies showed ERG abnormalities, and no studies showed clinical evidence of visual defects after antenatal HCQ exposure, which is reassuring.

The remaining reports of ocular problems after antenatal antimalarial exposure in the literature were from case reports. In the first, 3 exposed infants were found to have retinal changes of uncertain significance in the context of various congenital defects (24), and in the second, 2 exposed infants had retinal lesions since birth (10). Notably, in all 5 cases of retinal changes, the mothers were consistently taking excess doses of the drug. Finally, in the case report by Mulholland et al (11), an infant was born with anterior chamber dysgenesis in the context of in utero exposure to HCQ and methotrexate, a known teratogen, in the first 8 weeks of life. The significance of this last report remains unclear, because this is the first documented case of Peters anomaly following intrauterine exposure to either of those medications.

For the most part, reports of ocular abnormalities in offspring with antenatal exposure to antimalarials in the literature are few. Among the cohort and randomized controlled studies, which included a total number of 789 exposed infants, there was a low incidence of abnormal ocular findings and no documented defects in visual function. Isolated reports of ocular abnormalities in children exposed in utero to antimalarials are documented in the literature, but in many cases the exposures were not typical for what is used in SLE. Therefore, high dosages of these medications should be avoided, particularly in pregnancy. Staying within the current maximum daily dosage of 5.0 mg/kg of HCQ (2) may reduce the risk of ocular abnormalities in antenatally exposed infants.

In conclusion, our results suggest a low-to-nonexistent risk of visual abnormalities in offspring exposed to antimalarials. The potential benefits for pregnant women with SLE to continue antimalarials should be weighed against the possible risk. Long-term population-based studies of ocular anomalies at birth and throughout childhood in children exposed in utero to antimalarials are required to confirm the safety of these drugs with respect to retinal complications.

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. Vinet 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. Gaffar, Pineau, Bernatsky, Scott, Vinet.
Acquisition of data. Gaffar, Vinet.
Analysis and interpretation of data. Gaffar, Vinet.

REFERENCES

1. Kuhn A, Aberer E, Bata-Csörgő Z, Caproni M, Dreher A, Frances C, et al. The diagnosis and treatment of systemic lupus erythematosus. Deutsches Arzteblatt Int 2015;112:423–32.
2. Marmor MF, Kellner U, Lai TY, Melles RB, Mieler WF. Recommendations on screening for chloroquine and hydroxychloroquine retinopathy (2016 revision). Ophthalmol 2016;123:1386–94.
3. Costedoat-Chalumeau N, Amoura Z, Aymard G, Hong DL, Wechsler B, Vauthier D, et al. Evidence of transplacental passage of hydroxychloroquine in humans. Arthritis Rheum 2002;46:1123–4.
4. Ullberg S, Lindquist NG, Sjostrand SE. Accumulation of chorioretinotoxic drugs in the foetal eye [letter]. Nature 1970;227:1257–8.
5. Canadian Hydroxychloroquine Study Group. A randomized study of the effect of withdrawing hydroxychloroquine sulfate in systemic lupus erythematosus. N Engl J Med 1991;324:150–4.
6. Clowse ME. Lupus activity in pregnancy. Rheum Dis Clin North Am 2007;33:237–52.
7. Clowse ME, Magder L, Witter F, Petri M. Hydroxychloroquine in lupus pregnancy. Arthritis Rheum 2006;54:3640–7.
8. Costedoat-Chalumeau N, Amoura Z, Huong DL, Lechat P, Piette JC. Safety of hydroxychloroquine in pregnant patients with connective tissue diseases: review of the literature. Autoimmun Rev 2005;4:111–5.
9. Osadchy A, Ratnapalan T, Koren G. Ocular toxicity in children exposed in utero to antimalarial drugs: review of the literature. J Rheumatol 2011;38:2504–8.
10. Paufique L, Magnard P. Retinal degeneration in 2 children following preventive antimalarial treatment of the mother during pregnancy. Bull Soc Ophtalmol Fr 1969;69:466–7. In French.
11. Mulholland CP, Pollock TJ. The Peters anomaly following antenatal exposure to methotrexate and hydroxychloroquine [letter]. Can J Ophthalmol 2011;46:289–90.
12. Moher D, Liberati A, Tetzlaff J, Altmanl DC, and the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009;62:1006–12.
13. Adam I, Idris HM, Elbashir MI. Quinine for chloroquine-resistant falciparum malaria in pregnant Sudanese women in the first trimester. East Mediterr Health J 2004;10:560–5.
14. Adam I, Mirghani OA, Saed OK, Ahmed SM, Mohamadani AA, Ahmed HM, et al. Quinine therapy in severe Plasmodium falciparum malaria during pregnancy in Sudan. East Mediterr Health J 2004;10:159–66.
15. Cimaz R, Brucato A, Meregalli E, Muscara M, Sergi P. Electroretinograms of children born to mothers treated with hydroxychloroquine during pregnancy and breast-feeding: comment on the article by Costedoat-Chalumeau et al [letter]. Arthritis Rheum 2004;50:3056–7.
16. Costedoat-Chalumeau N, Amoura Z, Duhaut P, Huong DL, Sebbough D, Wechsler B, et al. Safety of hydroxychloroquine in pregnant patients with connective tissue diseases: a study of one hundred thirty-three cases compared with a control group. Arthritis Rheum 2003;48:3207–11.
17. Renault F, Flores-Guevara R, Renaud C, Renault C, Richard P, Vermersh AI, et al. Visual neurophysiological dysfunction in infants exposed to hydroxychloroquine in utero. Acta Paediatr 2009;98:1500–3.
18. Motta M, Tincani A, Faden D, Zinzini E, Lojacono A, Marxhesi A, et al. Follow-up of infants exposed to hydroxychloroquine given to mothers during pregnancy and lactation. J Perinatol 2005;25:86–9.
19. Klinger G, Morad Y, Westall CA, Laskin C, Spitzer KA, Koren G, et al. Ocular toxicity and antenatal exposure to chloroquine or hydroxychloroquine for rheumatic diseases [letter]. Lancet 2001;358:813–4.
20. Levy M, Buskila D, Gladman DD, Urowitz MB, Koren G. Pregnancy outcome following first trimester exposure to chloroquine. Am J Perinatol 1991;8:174–8.
21. Buchanan NM, Toubi E, Khamashta MA, Lima F, Kerslake S, Hughes GR. Hydroxychloroquine and lupus pregnancy: review of a series of 36 cases. Ann Rheum Dis 1996;55:486–8.
22. Villegas L, McGready R, Htway M, Paw MK, Pimanpanarak M, Arunjerdja R, V et al. Chloroquine prophylaxis against vivax malaria in pregnancy: a randomized, double-blind, placebo-controlled trial. Trop Med Int Health 2007;12:209–18.
23. Levy RA, Vilela VS, Cataldo MJ, Ramos RC, Duarte JL, Tura BR, et al. Hydroxychloroquine (HCQ) in lupus pregnancy: double-blind and placebo-controlled study. Lupus 2001;10:401–4.
24. Hart CW, Naunton RF. The ototoxicity of chloroquine phosphate. Arch Otolaryngol 1964;80:407–12.
25. Costedoat-Chalumeau N, Amoura Z, Sebbough D, Piette JC. Electroretinograms of children born to mothers treated with hydroxychloroquine during pregnancy and breast feeding [letter]. Arthritis Rheum 2004;50:3056–7.
26. Moher D, Liberati A, Tetzlaff J, Altman DG, and the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6:e1000097.

Arthritis Care & Research
Vol. 71, No. 12, December 2019, pp 1611–1620
DOI 10.1002/acr.23818

American College of Rheumatology
Empowering Rheumatology Professionals

Physical Fitness in Patients With Oligoarticular and Polyarticular Juvenile Idiopathic Arthritis Diagnosed in the Era of Biologics: A Controlled Cross-Sectional Study

Kristine Risum,[1] Elisabeth Edvardsen,[2] Kristin Godang,[3] Anne M. Selvaag,[3] Bjørge H. Hansen,[4] Øyvind Molberg,[1] Jens Bollerslev,[1] Inger Holm,[1] Hanne Dagfinrud,[5] and Helga Sanner[6]

Objective. To perform a comprehensive evaluation of and identify correlates for physical fitness in consecutive patients with juvenile idiopathic arthritis (JIA) who have been diagnosed in the era of biologics and to compare the results with those obtained in healthy controls.

Methods. The study cohort included 60 patients with JIA (50 girls) ages 10–16 years and 60 age- and sex-matched controls. The JIA group included 30 patients with persistent oligoarticular JIA and 30 patients with extended oligoarticular or polyarticular disease. Measures of physical fitness included cardiorespiratory fitness (CRF) by peak oxygen uptake (Vo_{2peak}) during a continuous graded treadmill exercise test, muscle strength by isokinetic and isometric knee and hand grip evaluations, and bone mineral density (BMD) and body composition by dual-energy x-ray absorptiometry. Physical activity was assessed by accelerometry.

Results. Forty-two percent of the patients were being treated with biologic drugs. Patients with JIA demonstrated lower muscle strength and total body BMD compared to controls, but there were no differences in CRF and body composition. Physical fitness was comparable between the persistent oligoarticular and extended oligoarticular/polyarticular-JIA groups. In patients with JIA, we identified associations between higher vigorous physical activity and higher CRF and muscle strength, but did not find any association between physical fitness and disease variables.

Conclusion. In this cohort of patients with JIA, we found suboptimal muscle strength and BMD compared to controls, but no differences in CRF and body composition. Vigorous physical activities appeared important for optimizing muscle strength and CRF in patients with JIA; the importance of such activities should be highlighted in patient education.

INTRODUCTION

Early aggressive medical treatment, including biologic therapy, has dramatically improved the overall outcome in patients with juvenile idiopathic arthritis (JIA). This has allowed for a shift in focus regarding patient outcome measures, with less focus on disability and more on physical function (1,2). One of the most important and complex measures of physical function in health and disease is physical fitness. Physical fitness is defined as a set of attributes, including cardiorespiratory fitness (CRF), muscle strength and endurance, bone density, and body composition, that people have or achieve, which relates to the ability to perform physical activity (PA) (3). The individual components of physical fitness have been assessed previously in small cohorts of patients with JIA, but there is limited knowledge about this outcome measure in patient cohorts subjected to early and aggressive medical treatment.

Previous cohort studies have found that patients with JIA have impaired CRF, mostly measured as peak oxygen uptake (Vo_{2peak}), compared to healthy peers (4–10); with the lowest CRF values reported in patients having polyarticular disease (6). Previous studies have also reported lower muscle strength in patients with JIA compared to controls (11–14). However, knowledge about isokinetic knee muscle strength, resembling muscle activity important for activities of daily living, is sparse in patients with JIA.

Supported by The Sophies Minde Foundation.

[1]Kristine Risum, PT, MSc, Øyvind Molberg, MD, PhD, Jens Bollerslev, MD, PhD, Inger Holm, PT, PhD: Oslo University Hospital, Rikshospitalet and University of Oslo, Oslo, Norway; [2]Elisabeth Edvardsen, PhD: Oslo University Hospital, Ullevål and Norwegian School of Sports Sciences, Oslo, Norway; [3]Kristin Godang, BSc, Anne M. Selvaag, MD, PhD: Oslo University Hospital, Rikshospitalet, Oslo, Norway; [4]Bjørge H. Hansen, PhD: Norwegian School of Sports Sciences, Oslo, Norway; [5]Hanne Dagfinrud, PT, PhD: Diakonhjemmet Hospital and University of Oslo, Oslo, Norway; [6]Helga Sanner, MD, PhD: Oslo University Hospital, Rikshospitalet and Bjørknes University College, Oslo, Norway.

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

Address correspondence to Kristine Risum, PT, MSc, Section for Orthopaedic Rehabilitation, Oslo University Hospital, Rikshospitalet, Postboks 4950 Nydalen, 0424 Oslo, Norway. E-mail: krisum@ous-hf.no.

Submitted for publication July 20, 2018; accepted in revised form November 20, 2018.

SIGNIFICANCE & INNOVATIONS

- Cardiorespiratory fitness and body composition in patients with oligoarticular and polyarticular juvenile idiopathic arthritis (JIA) that have been diagnosed in the era of biologics are comparable with controls.
- Muscle strength and bone health are both suboptimal in patients with oligoarticular and polyarticular JIA compared to controls.
- In this well-treated JIA cohort, there were no associations between disease variables and the different components of physical fitness.
- Higher vigorous physical activity is associated with better cardiorespiratory fitness and muscle strength in patients.

In healthy children, low CRF is associated with subclinical cardiovascular disease (CVD) (15,16). Adequate CRF is important in children with JIA as the chronic inflammation is associated with early subclinical signs of atherosclerosis (17) and, moreover, many have active disease into adulthood (18). Also, higher muscle strength is associated with lower CVD risk and better bone health in healthy children and adolescents (16,19).

A recent review concluded that most studies show lower bone mineral density (BMD) in JIA patients when compared to reference values (20), whereas another study demonstrated normal BMD in patients with JIA (21). Reduced lean mass has been reported in JIA patients when compared to controls (22,23), while both higher (24) and comparable (25) percentage total body fat have been reported.

Notably, the various measurements of physical fitness have not been systematically evaluated using optimized methods in the same JIA cohort, as well as in patients diagnosed in the biologic era. The latter is potentially important as advances in medical therapies could have an impact on physical fitness. Contrary to previous studies, our research group recently reported that objectively measured PA levels were comparable between patients with JIA and controls from the general population (26); indicating that it would be informative to evaluate physical fitness in the same cohort. The aims of the present study were 3-fold and included comparing physical fitness in patients with JIA who were diagnosed in the era of biologics with age- and sex-matched controls from the general population, comparing physical fitness in patients with persistent oligoarticular JIA and those who had polyarticular disease to examine if disease severity is still of importance, and exploring associations between physical fitness and different disease variables.

PATIENTS AND METHODS

Study participants. In this controlled cross-sectional study, we included patients with JIA, ages 10–16 years, with a planned routine visit at Oslo University Hospital (OUS) and with a home address in the geographic area served by the South-Eastern Norway Regional Health Authority. This area has a denominator population of 2.8 million (57% of the Norwegian population). The patients were classified with persistent oligoarthritis or polyarticular disease (extended oligoarthritis and polyarticular rheumatoid factor positive/negative) according to the International League of Associations for Rheumatology criteria (27) with a disease duration >6 months. Patients were excluded if they had comorbidities associated with, or potentially associated with, impaired cardiopulmonary fitness (e.g., heart or lung disease, severe orthopedic conditions or recent surgery). In addition, individually age- and sex-matched controls from the general population (living in or nearby Oslo) were randomly selected from the National Registry. Exclusion criteria for the controls were inflammatory rheumatic or autoimmune disease, severe heart or lung disease, or other diseases involving mobility problems.

The study complied with the Declaration of Helsinki. All participants provided written informed consent/assent. The study was approved by the Norwegian South East Regional Ethics Committee for Medical Research (2014/188).

Data collection including clinical and laboratory measures. All study participants were examined at OUS and included patients in conjunction with their routine visit between January and August 2015 and controls during a 1-day program between November 2015 and March 2016. Height, body weight, and waist circumference were measured. Body mass index (BMI) was calculated, and age- and sex-specific BMI cutoff values were used to categorize the children as normal weight, overweight, or obese (28). Pubertal status was self-reported using Tanner stages 1–5 (29,30). The puberty stages were then categorized into pre-puberty (Tanner stage 1), mid-puberty (Tanner stages 2–4), and post-puberty (Tanner stage 5). Girls were asked about age at menarche. Smoking and other tobacco-related habits were registered during an interview without the parents present. Current pain, and pain and fatigue during the previous week were assessed by numerical rating scale 0–10 (31). The erythrocyte sedimentation rate (ESR), C-reactive protein (CRP) levels and hemoglobin were analyzed. In patients, we performed a total joint count (71 joints) and reported separately the count for the lower extremities (including hip, knee, ankle, tarsus, metatarsophalangeal, and interphalangeal joints). Disease activity was assessed by the Juvenile Arthritis Disease Activity Score in 71 joints (JADAS-71) (32). The Wallace criteria were used to determine if patients had clinical inactive disease (referred to as inactive disease) or active disease (33). The Childhood Health Assessment Questionnaire (C-HAQ) was used to measure functional disability (34,35).

CRF. CRF was directly measured as peak oxygen uptake ($V_{O_{2peak}}$) during a continuous graded exercise test on a treadmill

Table 1. Treadmill protocol used for all study participants

Time (mins.)	Speed (km/hour)	Inclination (%)
1	4	4
2	5	4
3	5	6
4	6	6
5	6	8
6	7	8
7	7	10
8	8	10
9	8	12
10	9	12
11	9	14
12	10	14

(Woodway) until exhaustion. The test protocol was specifically developed for this study (Table 1). The test was terminated when the participant was unable to continue, after strong encouragement from the test leader.

Gas exchange and ventilator variables were measured continuously, breath-by-breath, as the participants breathed into a 2-way breathing mask (2700 series; Hans Rudolph, Inc.). The gas analyzer (Vmax SensorMedics; Yorba Linda) was calibrated for volume daily, and calibrated for gas before each test. The gas-exchange variables were reported as 30-second averages. The highest achieved oxygen uptake averaged over a 30-second period was defined as $V_{O_{2peak}}$. The highest respiratory exchange ratio (RER) measured before or corresponding to the highest minute ventilation was reported. The heart rate (HR) was recorded every minute using Polar Sports Watch (Polar Electro) and the peak HR was also reported.

The perceived exertion (RPE) was rated by Borg scale 6–20 (36), and the participants also gave a reason for terminating the test. As no consensus of criteria for maximal test exists in children, the exercise test was considered maximal by evaluation from the test leader, including signs of rapid breathing, facial flushing, unsteady running and high RPE, HR, and RER in the participants. $V_{O_{2peak}}$ expressed as ml × kg^{-1} × $minutes^{-1}$ is our main CRF outcome. $V_{O_{2peak}}$ refers to $V_{O_{2peak}}$ (ml × kg^{-1} × $minutes^{-1}$), unless stated otherwise. A poor $V_{O_{2peak}}$ was defined as values below 85% of the mean in controls, for girls and boys separately. CRF variables also included absolute $V_{O_{2peak}}$ (liters × $minutes^{-1}$), O_2 pulse (ml × $minutes^{-1}$/peak HR) and ventilatory efficiency slope (V_E/V_{CO_2} slope).

Muscle strength. Knee extension and flexion were tested isokinetically using a Cybex 6000 (Cybex-Lumex, Inc.). Both legs were tested in a sitting position, with the dominant leg first.

The children performed 5 repetitions at angular velocities of 60°/second and 30 repetitions at 240°/second, respectively. Results are given as peak torque (Nm) and work (Joule). The values were normalized for body weight (absolute values divided by body weight × 100) for comparison between groups. In linear multiple regression analysis, the absolute value for maximal knee extension in the dominant leg was used. The protocol included 4 trial repetitions at both velocities bilaterally.

Grip strength was measured bilaterally using the Baseline dynamometer (Fabrication Enterprises). The participants were seated with their arm alongside the trunk and elbow in 90° of flexion. The first and second positions out of the 5 positions of the dynamometer were used. The dominant hand was tested first. The mean value (in kilograms) of 2 trials for each hand was used in the analyses.

Bone mineral density and body composition. The pediatric BMD and total body composition were determined by dual-energy x-ray absorptiometry, anterior–posterior projection, at the lumbar spine (L2–L4) and total body. A GE Healthcare Lunar Prodigy narrow fan beam from (Lunar Corp.) densitometer was used and all the scans were analyzed using enCORE software version 14.10 (Lunar Corp.). Standard imaging and positioning protocols were used to scan the subjects. Absolute BMD values (g/cm^2) and BMD Z scores were estimated by comparison to the Lunar-reference database incorporated in the software suitable for clinical use in the Norwegian population (37). Z score values ≤2 were defined as below the expected range for age and sex.

Physical activity. Objective registration of PA was measured with an accelerometer (ActiGraph GT3X) for 7 consecutive days during waking hours. We applied recommended thresholds for PA intensities in children (38,39), including sedentary time (<101 cpm), light PA (LPA) (≥101 to ≤2,295 cpm), moderate PA (MPA) (≥2,296 to ≤4,011 cpm), and vigorous PA (VPA) (>4,011 cpm).

Statistical analysis. A power analysis was conducted based on previous research on differences in $V_{O_{2peak}}$ between patients with JIA and controls (6,7), and with an alpha of 5% and power of 80% we would need at least 13 participants in each group. To be able to compare patient subgroups and to study secondary outcomes, we aimed at recruiting 60 patients (30 patients with oligoarticular JIA and 30 patients with polyarticular disease) and 60 controls.

Continuous data were expressed as mean ± SD or median (interquartile range), as appropriate. Independent sample *t*-tests, Mann-Whitney U tests, or chi-square tests were used to assess differences between patients and controls and between patient subgroups (oligoarticular versus polyarticular and active versus inactive disease), as appropriate. Linear regression analyses were conducted to identify correlates of physical fitness in patients. Variables that were associated ($P < 0.15$) with the outcome variables ($V_{O_{2peak}}$, maximal knee extension, and total body BMD [TBBMD]) in univariate linear analyses were evaluated in the multiple linear analyses. Highly correlated independent variables ($r > 0.7$) were avoided. Potential correlates of physical fitness

included the use of any medication, use of biologic medication, CRP, ESR, hemoglobin, active joints, active joints in the lower extremities, JADAS-71, C-HAQ, disease duration, disease state, current pain, pain and fatigue during previous week, age, sex, BMI, puberty status, sedentary time, LPA, MPA, and VPA. Because the sample size did not allow us to simultaneously include all potential variables in the multiple linear regression model, we preselected a maximum of 7 variables. Statistical tests were conducted using SPSS version 23.0. *P* values < 0.05 were considered statistically significant.

RESULTS

Characteristics of patients and controls. Some of the demographic characteristics of the study population and disease characteristics of the cohort of patients with JIA have previously been published (26), but are shown in Table 2 for the sake of clarity and completeness. No significant difference was found in puberty stages (Table 2) or in age at menarche among the girls. All participants had normal hemoglobin values. None of the participants were regular smokers or had other tobacco-related habits or used systemic corticosteroids. Forty patients (67%) had active disease and 20 (33%) had inactive disease.

Cardiorespiratory fitness. The exercise test was considered maximal in all participants, and all participants reported general tiredness as the reason for terminating the exercise test. We found no significant differences in $V_{O_{2peak}}$ and in minute ventilation (liters × minutes^{-1}) between patients and controls (Table 3). Also, the peak HR, RER, and RPE were comparable between these groups. We did, however, find that patients had lower absolute $V_{O_{2peak}}$ (liters × minutes^{-1}) and O_2 pulse than controls. There was no significant difference between the proportion of patients and controls defined as having poor $V_{O_{2peak}}$ (stratified by sex) (Figure 1).

Ventilatory efficiency (VE/VCO_2 slope), was higher in patients with polyarticular compared to oligoarticular disease, but all values were within normal cutoff values (Table 3). The other CRF variables did not differ significantly between oligoarticular and polyarticular JIA. No CRF variables were significantly different between patients with active and inactive disease (Figure 2A; data not shown).

Muscle strength. Patients showed lower grip strength and maximal quadriceps and hamstring strength bilaterally compared to controls (Table 3). Patients also had lower knee flexion endurance bilaterally. No significant muscle strength differences were found between patient subgroups; oligoarticular versus polyarticular (Table 3) and active versus inactive disease (Figure 2B; data not shown).

BMD and total body composition. Patients had lower TBBMD for both absolute values and Z scores compared to

Table 2. Characteristics of patients with JIA and controls*

	JIA total (n = 60)	Controls (n = 60)	Oligoarticular JIA (n = 30)	Polyarticular JIA (n = 30)
Age, mean ± SD years	13.6 ± 2.2	13.5 ± 2.6	13.5 ± 2.2	13.7 ± 2.2
Female sex	50 (83)	50 (83)	27 (90)	23 (77)
Height, mean ± SD cm	157.9 ± 12.6	161.2 ± 12.6	157.1 ± 11.8	158.7 ± 13.6
Weight, mean ± SD kg	49.3 ± 13.8	53.5 ± 15.4	47.0 ± 10.1	51.5 ± 16.2
BMI, mean ± SD kg/m^2	19.4 ± 3.5	20.2 ± 3.5	18.8 ± 2.1	20.1 ± 4.4
Pubertal status (pre-/mid-/postpubertal), %	23/62/17	17/68/15	20/63/17	27/60/13
NRS current pain (0–10), with score >0	23 (38)	18 (30)	12 (40)	11 (37)
NRS pain previous week (0–10), median (IQR)	1.0 (0.0–3.0)	1.0 (0.0–3.0)	0.0 (2.0–3.3)	1.0 (0.0–3.5)
NRS fatigue previous week (0–10), median (IQR)	3.0 (2.0–6.0)	3.0 (1.0–3.5)	3.5 (2.0–6.3)	3.0 (2.0–5.3)
CRP level >4 mg/liter	3 (5)	0 (0)	1 (3)	2 (7)
Physiotherapy regularly	13 (22)	4 (7)†	3 (10)	10 (33)†
Disease duration, mean ± SD years	7.5 ± 3.8	NA	7.6 ± 3.9	7.3 ± 4.0
JADAS-71 (0–101), median (IQR)	3.3 (1.1–4.8)	NA	3.3 (0.8–4.8)	3.2 (1.4–4.6)
C-HAQ score (0–3), median (IQR)	0.0 (0.0–0.4)	NA	0.1 (0.0–0.3)	0.0 (0.0–0.4)
Off medication	12 (20)	NA	10 (33)	2 (7)†
Synthetic DMARDs	40 (67)	NA	18 (60)	22 (73)
Biologic DMARDs	25 (42)	NA	5 (17)	20 (67)‡
Synthetic + biologic DMARDs	19 (32)	NA	5 (17)	14 (47)†
Active disease	40 (67)	NA	18 (60)	22 (73)
Clinical inactive disease	20 (33)	NA	12 (40)	8 (27)

* Values are the number (%) unless indicated otherwise. JIA = juvenile idiopathic arthritis; BMI = body mass index; NRS = numerical rating scale; CRP = C-reactive protein; JADAS-71 = Juvenile Arthritis Disease Activity Score in 71 joints; IQR = interquartile range; C-HAQ = Childhood Health Assessment Questionnaire; DMARDs = disease-modifying antirheumatic drugs; NA = not applicable.
† = $P < 0.05$.
‡ = $P < 0.001$ when comparing JIA total versus controls or oligoarticular JIA versus polyarticular JIA.

controls (Table 3). Two patients had lumbar spine BMD Z score ≤2. No significant differences were found regarding body composition, but there was a tendency for patients with JIA to have lower total body lean mass and total body bone mineral content than controls. There were no significant differences in any of the BMD and body composition measurements between JIA subgroups; oligoarticular versus polyarticular (Table 3) and active versus inactive disease (Figures 2C and D; data not shown).

Correlates of physical fitness in patients. Since we found no differences in Vo_{2peak}, maximal knee extension dominant leg and TBBMD between any patient subgroups (oligoarticular versus polyarticular and active versus inactive disease), we included all patients in the linear regression analyses. Correlates for physical fitness in patients are shown in Table 4. Male sex, lower BMI, and higher VPA were identified as correlates for higher Vo_{2peak}. Male sex, higher BMI, age, and VPA were identified as correlates for higher muscle strength. Muscle strength and total lean mass were highly correlated (r = 0.79). For higher TBBMD, we identified higher lean mass, age, and female sex as correlates. Age and puberty status were highly correlated (r = 0.74), and age was chosen in the best fit models for muscle strength and TBBMD, as this variable provided more to the explanatory power of the models.

Table 3. Physical fitness in patients with JIA and controls*

	JIA total (n = 59–60)†	Controls (n = 59–60)	Oligoarticular JIA (n = 30)‡	Polyarticular JIA (n = 29–30)§
Cardiorespiratory fitness				
Vo_{2peak} (liters × minutes^{-1})	2.2 ± 0.6	2.4 ± 0.7¶	2.1 ± 0.5	2.2 ± 0.6
Vo_{2peak} (ml × kg^{-1} × minutes^{-1})	45.1 ± 8.5	46.5 ± 8.5	45.0 ± 7.6	45.3 ± 9.4
Minute ventilation (liters × minutes^{-1})	80.3 ± 21.6	87.2 ± 22.2	77.8 ± 21.1	82.8 ± 22.2
VE/VCO_2 slope	27.0 ± 2.5	27.4 ± 2.6	26.2 ± 2.6	27.9 ± 2.1†
O_2 pulse (ml/beat)	10.9 ± 2.8	12.3 ± 3.6¶	10.7 ± 2.7	11.1 ± 2.9
Respiratory exchange ratio	1.27 ± 0.12	1.23 ± 0.10	1.27 ± 0.12	1.26 ± 0.13
Peak heart rate (beats/minute)	196 ± 9	197 ± 7	195 ± 10	197 ± 7
Borg scale (6–20)	18.9 ± 1.9	18.5 ± 1.0	19.0 ± 1.4	18.9 ± 2.4
Running distance (m)	909 ± 236	968 ± 190	905 ± 228	913 ± 247
Test time (secs)	527 ± 99	554 ± 76	526 ± 97	529 ± 103
Grip strength (kg)				
Dominant hand strength	23.2 ± 7.4	27.9 ± 8.4¶	22.9 ± 7.0	23.5 ± 7.8
Nondominant hand strength	23.2 ± 6.6	25.9 ± 7.7¶	23.3 ± 6.1	23.0 ± 7.2
Maximal muscle strength#				
Dominant knee extension	124 ± 38	146 ± 34¶	131 ± 33	118 ± 42
Nondominant knee extension	131 ± 38	154 ± 32**	138 ± 37	125 ± 38
Dominant knee flexion	65 ± 26	83 ± 22**	64 ± 25	66 ± 28
Nondominant knee flexion	72 ± 28	86 ± 23¶	74 ± 27	68 ± 29
Muscle endurance††				
Dominant knee extension	2,470 ± 633	2,660 ± 466	2,516 ± 650	2,424 ± 624
Nondominant knee extension	2,528 ± 609	2,565 ± 477	2,596 ± 583	2,461 ± 636
Dominant knee flexion	1,065 ± 461	1,376 ± 409**	1,109 ± 470	1,018 ± 455
Nondominant knee flexion	1,023 ± 489	1,249 ± 404¶	1,106 ± 511	940 ± 459
Bone mineral density (gm/cm^2)				
TBBMD	0.971 ± 1.127	1.033 ± 0.128¶	0.970 ± 0.128	0.972 ± 0.129
TBBMD Z score	0.152 ± 0.810	0.712 ± 0.910¶	0.200 ± 0.909	0.103 ± 0.709
Lumbar spine BMD	0.998 ± 0.189	1.0366 ± 0.174	0.998 ± 0.174	0.979 ± 0.206
Lumbar spine BMD Z score	−0.362 ± 1.021	−0.008 ± 1.024	−0.243 ± 0.931	−0.480 ± 1.107
Body composition				
Total body mass (kg)	49.5 ± 13.6	53.9 ± 15.3	47.3 ± 10.2	51.7 ± 16.2
Total body lean mass (kg)	33.4 ± 8.5	36.7 ± 10.0	32.5 ± 7.9	34.4 ± 9.1
Total body BMC (kg)	1.8 ± 0.5	2.2 ± 1.5	1.8 ± 0.4	1.9 ± 0.5
Total body fat mass (kg)	14.2 ± 7.1	15.2 ± 6.7	13.0 ± 3.7	15.4 ± 9.3
Total body fat (%)	29.3 ± 7.6	28.7 ± 6.4	28.8 ± 6.4	29.8 ± 8.6

* Values are the mean ± SD. JIA = juvenile idiopathic arthritis; Vo_{2peak} = peak oxygen uptake; VE/VCO_2 = ventilation/carbon dioxide production; TBBMD = total body bone mineral density; BMD = bone mineral density; BMC = bone mineral content.
† Consisted of patients with persistent oligoarticular JIA, extended oligoarticular JIA, polyarticular JIA rheumatoid factor positive/negative.
‡ Consisted of patients with persistent oligoarticular JIA.
§ Consisted of patients with extended oligoarticular JIA, polyarticular JIA rheumatoid factor positive/negative.
¶ = $P < 0.05$.
# Newton meter divided by body weight × 100.
** $P < 0.001$ when comparing JIA total versus controls or oligoarticular JIA versus polyarticular JIA.
†† Joule divided by body weight × 100.

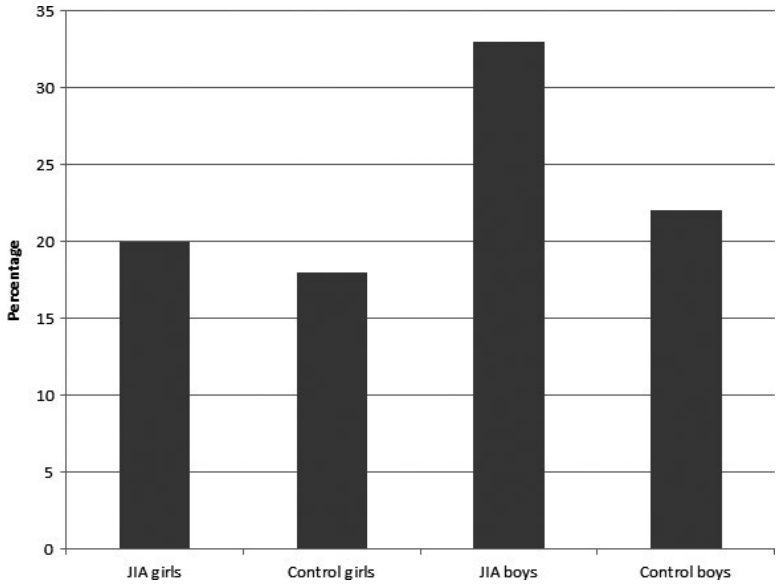

Figure 1. There was no significant difference between the proportion of patients and controls defined as having poor VO_{2peak} (stratified by sex). JIA = juvenile idiopathic arthritis.

DISCUSSION

Studies that concurrently evaluate all aspects of physical fitness are lacking in JIA, and there are sparse physical fitness data from contemporary cohorts. Here, we showed that patients with JIA who were diagnosed in the era of biologics had CRF and body composition comparable to age- and sex-matched controls from the general population, but lower muscle strength and TBBMD. Furthermore, higher CRF and muscle strength were associated with higher VPA levels, which reinforces the view that high intensity activities are desirable in JIA. No disease-related variables were identified as correlates for physical fitness in patients. To our knowledge, this is the first study to assess physical fitness comprehensively with state-of-the-art measurements in a contemporary JIA patient cohort with good access to biologic therapies.

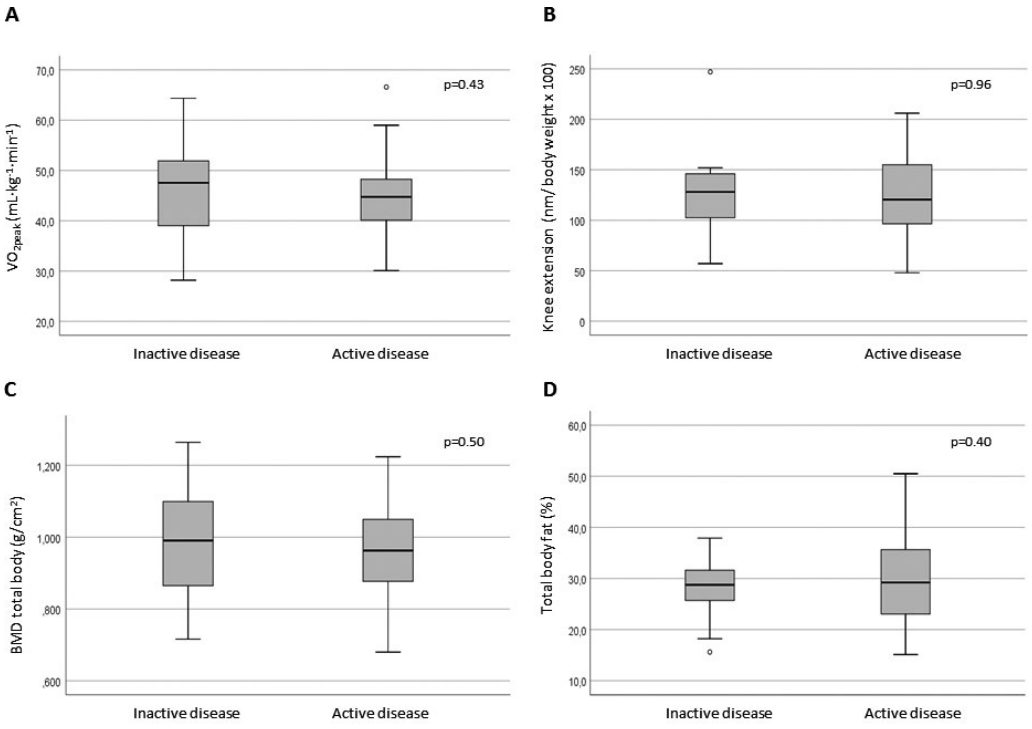

Figure 2. Differences in measures of cardiorespiratory fitness (**A**), muscle strength (**B**), bone mineral density (**C**), and body composition (**D**) between patients with juvenile idiopathic arthritis with active disease versus those with inactive disease. VO_{2peak} = peak oxygen uptake; BMD = bone mineral density.

Table 4. Correlates for physical fitness in patients with JIA*

	Univariate regression analyses		Multiple regression analyses	
	Unstandardized B (95% CI)	*P*	Unstandardized B (95% CI)	*P*
Cardiorespiratory fitness (Vo_{2peak} ml × kg^{-1} × $minutes^{-1}$)				
Female sex	−8.12 (−13.93, −2.30)	0.007	−9.0 (−13.4, −4.6)	<0.001
BMI	−1.49 (−2.14, −0.85)	<0.001	−1.4 (−1.9, −1.0)	<0.001
MPA	0.23 (0.04, 0.42)	0.02		
VPA	0.29 (0.11, 0.47)	0.002	0.3 (0.2, 0.5)	<0.001
Pain previous week (NRS 0–10)	−1.15 (−2.19, −0.11)	0.03		
Fatigue previous week (NRS 0–10)	−0.88 (−1.75, −0.02)	0.05		
R^2 adjusted			0.55	
Maximal knee extension, dominant leg (Nm)				
Age	6.8 (4.3, 9.4)	<0.001	6.0 (3.6, 8.3)	<0.001
Women	−20.3 (−37.7, −2.9)	0.02	−21.5 (−34.9, −8.2)	0.002
BMI	2.9 (1.1, 4.7)	0.002	1.8 (0.3, 3.2)	0.02
VPA	0.7 (0.05, 1.3)	0.04	0.8 (0.3, 1.2)	0.001
Sedentary time	0.07 (−0.2, 0.17)	0.12		
Fatigue previous week (NRS 0–10)	40.2 (−5.8, 86.2)	0.09		
C-HAQ	−23.6 (−43, −4.2)	0.02		
R^2 adjusted			0.54	
TBBMD (gm/cm^2)				
Age	0.04 (0.03, 0.05)	<0.001	0.04 (0.02, 0.05)	<0.001
Women	0.02 (−0.07, 0.11)	0.59	0.01 (0.05, 0.15)	<0.001
BMI	0.02 (0.01, 0.03)	<0.001		
Total lean mass	0.012 (0.009, 0.014)	<0.001	0.01 (0.007, 0.013)	<0.001
Total fat mass	0.008 (0.004, 0.012)	0.001		
C-HAQ	−0.13 (−0.22, −0.04)	0.007		
R^2 adjusted			0.74	

* JIA = juvenile idiopathic arthritis; 95% CI = 95% confidence interval; Vo_{2peak} = peak oxygen uptake; BMI = body mass index; MPA = moderate physical activity; VPA = vigorous physical activity; NRS = numeric rating scale; C-HAQ = Childhood Health Assessment Questionnaire; TBBMD = total body bone mineral density.

The current JIA cohort included patients with oligoarticular and polyarticular disease. These 2 JIA subgroups constitute ~75% of the JIA population included in the JIA patient registry run at our hospital since 1999; thus, our results cannot be generalized to JIA categories that were not included. Since we have no data available on the 37% of patients who declined to participate in the current study, we cannot rule out that the patients enrolled might be biased towards more physically fit patients. However, 42% of the enrolled patients were treated with biologic medication, indicating that we were not biased towards patients with a milder disease. The patients were matched with controls randomly drawn from the National Registry and the same equipment was used for all measurements, avoiding measurement errors. Additionally, the controls were tested within a year after the patients, which avoided changes in PA patterns that might influence the physical fitness levels. Importantly, the results on CRF and muscle strength from the controls are in accordance with data from Norwegian population studies (40,41).

Our findings of comparable CRF between patients and controls are in contrast to previous studies, all showing poorer CRF in patients with JIA than in controls (4–10). Only 1 of the previous studies assessed CRF with direct measurement of oxygen uptake by treadmill testing (8). Comparison should be made with caution, because our patients are older. Nevertheless, our patients had higher $Vo_{2peak,}$ while the Vo_{2peak} in the 2 control groups was comparable (8). The mean values for HR, RER, and RPE in our study indicate that the study participants exercised at their maximum levels, underlined also by the fact that all participants reported exhaustion as the reason for terminating the test. Maggio et al (8) suggested that the low RER in their patients possibly could be explained by physical limitations making it difficult to complete maximal exercise testing. Interestingly, compared to our patient cohort, they found higher mean active joint count in their patients, possibly explaining the differences in achieving maximal exercise tests. The previous studies measuring Vo_{2peak} on a cycle ergometer are not directly comparable to treadmill testing, since cycling gives a lower cardiopulmonary stress and is thought to underestimate Vo_{2peak} by more than 8–10% (42).

Our encouraging results of comparable CRF between patients and controls could possibly be explained by advances in the multidisciplinary management of JIA, including biologic therapy and individualized tailored patient education on PA emphasizing participation in physical activities like healthy peers that patients find enjoyable, without any general restrictions.

There was a significant difference in absolute Vo_{2peak} and O_2 pulse between patients and controls. However, these CRF var-

iables are highly influenced by body weight, thus, the differences are most likely explained by the tendency for patients to have lower body weight than controls.

We found that male sex, lower BMI, and higher VPA were correlates for higher CRF in patients. No disease variables were associated with CRF, which is in contrast to previous studies reporting associations with disease activity (6,8,9), articular limitation (8), swollen, limited, and active joints, CRP, ESR, and hemoglobin (9). In univariate analyses, we found associations only between CRF and pain and fatigue during the previous week. However, these associations did not remain significant in the multiple linear regression analyses, indicating that they were weak or possibly mediated by other variables. The lack of association with disease variables in our study might be explained by low disease activity and functional disability, indicating that our patients were well treated.

We did observe a large variation in CRF among both patients and controls. In the clinical setting, the 20–30% of patients with poor CRF should be identified, and subsequently encouraged to improve their CRF by participating in enjoyable physical activities, preferably also of vigorous intensity.

Patients had lower muscle strength in most of the variables tested, particularly for maximal muscle strength. This is in line with previous findings (11–13), although direct comparison is difficult due to differences regarding the applied methods and the muscle groups that were tested. We observed a trend for patients with JIA to have lower body weight and total lean mass than controls. Total fat mass and percentage body fat were, however, comparable. The total body composition in our cohort of patients with JIA is more favorable than that which has been described in previous studies (24). Advances in the multidisciplinary management of JIA might explain this result, subsequently leading to PA levels being relatively comparable between patients and controls (26), which is important for healthy body composition.

Our patients had decreased TBBMD (absolute values and Z scores) compared to controls, which is in line with previous findings (20). Contrary to our study, most of these studies were conducted before the introduction of biologics, in a time when systemic corticosteroids were more commonly used. We have no data on the cumulative corticosteroid use in our patients, but none of our patients currently used systemic corticosteroids. In contrast, a recent study found that patients with JIA had TBBMD comparable to a reference group (21).

In patients, we found that older age, male sex, higher BMI, and higher VPA were associated with higher muscle strength, similar to previous findings in healthy children (43). Lean mass was not included as an explanatory variable in the model for muscle strength because it was highly associated with muscle strength; this association probably reflects that lean mass can be considered as an indirect measure of muscle strength. To our knowledge, correlates for muscle strength in patients with JIA have not previously been reported. We identified higher lean mass, higher age, and female sex as correlates for higher TBBMD. Higher lean mass is important for higher TBBMD in healthy children as well (19). Disease-related variables were not identified as correlates for TBBMD in our study, and similar findings were demonstrated in a recent study on bone health in patients with JIA (21). However, older studies have shown the opposite results (20,44), indicating that disease-related variables might be less important with advanced treatment resulting in less active disease.

Exercise may improve muscle strength and bone health in patients with JIA (14,21). In healthy children, maximal muscle strength training is suggested as a safe and efficient method to increase muscle strength. Furthermore, maximal muscle strength training also increases muscle endurance, improves bone health, and is an independent predictor of CVD (16,43,45). The current PA recommendations for patients with JIA, with 60 minutes of daily MVPA are in accordance with the World Health Organization's (WHO) general recommendations for children (46). Additionally, the WHO recommends VPA at least 3 times a week, including activities that strengthen muscle and bone. Our results suggest that these WHO recommendations should be applied for patients with JIA as well, preferably through activities that patients find enjoyable.

We found comparable physical fitness between JIA subgroups (oligoarticular versus polyarticular and active versus inactive disease). As mentioned above, we found no associations between disease variables and any components of physical fitness. We believe our findings suggest that JIA-related disease factors no longer have an impact on physical fitness when multidisciplinary patient care is given in the era of biologics, but further studies are needed. Studying the same cohort, we have previously found comparable overall physical activity levels between patients and controls (26), which we believe contribute to our findings of no associations between disease variables and physical fitness.

Study limitations include the cross-sectional design not allowing for causal conclusions. The proportion of boys was low, and sex differences should therefore be interpreted with caution.

In conclusion, patients with oligoarticular and polyarticular JIA who have been diagnosed in the era of biologics have similar CRF and body composition, but lower muscle strength and bone health than matched healthy controls. All components of physical fitness are comparable between patients with persistent oligoarthritis and polyarticular disease. Higher levels of VPA are associated with higher CRF and muscle strength in patients with JIA. Patient education should include specific advice on increasing physical activities of vigorous intensity, including PA that strengthens muscle and bone, preferably through activities that patients find enjoyable.

ACKNOWLEDGMENTS

The authors thank the children, adolescents, and parents participating in this study, and the staff at the Unit of Paediatric

Rheumatology and at the Children's Outpatient Clinic at Rikshospitalet, Oslo University Hospital, and to the Clinical Research Unit, Department of Pharmacology, Rikshospitalet, Oslo University Hospital, for help with drawing and processing the blood samples from the controls. The authors also thank Iren Lindbak Matthews, PhD, MD, Department of Paediatric Allergy and Pulmonology, Oslo University Hospital, for invaluable access to the exercise physiology laboratory.

AUTHOR CONTRIBUTIONS

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

Acquisition of data. Risum, Godang.

Analysis and interpretation of the data. Risum, Edvardsen, Godang, Selvaag, Hansen, Molberg, Bollerslev, Holm, Dagfinrud, Sanner.

REFERENCES

1. Prakken B, Albani S, Martini A. Juvenile idiopathic arthritis. Lancet 2011;377:2138–49.
2. Beukelman T, Patkar NM, Saag KG, Tolleson-Rinehart S, Cron RQ, DeWitt EM, et al. 2011 American College of Rheumatology recommendations for the treatment of juvenile idiopathic arthritis: initiation and safety monitoring of therapeutic agents for the treatment of arthritis and systemic features. Arthritis Care Res (Hoboken) 2011;63:465–82.
3. Bar-Or O, Rowland T. Pediatric Exercise Medicine: From Physiologic Principles to Health Care Application. Champaign, Illinois: Human Kinetics; 2004.
4. Takken T, Hemel A, van der Net J, Helders PJ. Aerobic fitness in children with juvenile idiopathic arthritis: a systematic review. J Rheumatol 2002;29:2643–7.
5. Metin G, Ozturk L, Kasapcopur O, Apelyan M, Arisoy N. Cardiopulmonary exercise testing in juvenile idiopathic arthritis. J Rheumatol 2004;31:1834–9.
6. Lelieveld OT, van Brussel M, Takken T, van Weert E, van Leeuwen MA, Armbrust W. Aerobic and anaerobic exercise capacity in adolescents with juvenile idiopathic arthritis. Arthritis Rheum 2007;57:898–904.
7. Van Brussel M, Lelieveld OT, van der Net J, Engelbert RH, Helders PJ, Takken T. Aerobic and anaerobic exercise capacity in children with juvenile idiopathic arthritis. Arthritis Rheum 2007;57:891–7.
8. Maggio AB, Hofer MF, Martin XE, Marchand LM, Beghetti M, Farpour-Lambert NJ. Reduced physical activity level and cardiorespiratory fitness in children with chronic diseases. Eur J Pediatr 2010;169:1187–93.
9. Van Pelt PA, Takken T, van Brussel M, de Witte I, Kruize AA, Wulffraat NM. Aerobic capacity and disease activity in children, adolescents and young adults with juvenile idiopathic arthritis (JIA). Pediatr Rheumatol Online J 2012;10:27.
10. Norgaard M, Twilt M, Andersen LB, Herlin T. Accelerometry-based monitoring of daily physical activity in children with juvenile idiopathic arthritis. Scand J Rheumatol 2016;45:179–87.
11. Saarinen J, Lehtonen K, Malkia E, Lahdenne P. Lower extremity isometric strength in children with juvenile idiopathic arthritis. Clin Exp Rheumatol 2008;26:947–53.
12. Brostrom E, Nordlund MM, Cresswell AG. Plantar- and dorsiflexor strength in prepubertal girls with juvenile idiopathic arthritis. Arch Phys Med Rehabil 2004;85:1224–30.
13. Lindehammar H, Lindvall B. Muscle involvement in juvenile idiopathic arthritis. Rheumatology (Oxford) 2004;43:1546–54.
14. Sandstedt E, Fasth A, Eek MN, Beckung E. Muscle strength, physical fitness and well-being in children and adolescents with juvenile idiopathic arthritis and the effect of an exercise programme: a randomized controlled trial. Pediatr Rheumatol Online J 2013;11:7.
15. Melo X, Santa-Clara H, Santos DA, Pimenta NM, Minderico CS, Fernhall B, et al. Linking cardiorespiratory fitness classification criteria to early subclinical atherosclerosis in children. Appl Physiol Nutr Metab 2015;40:386–92.
16. Ortega FB, Ruiz JR, Castillo MJ, Sjostrom M. Physical fitness in childhood and adolescence: a powerful marker of health. Int J Obes (Lond) 2008;32:1–11.
17. Bohr AH, Fuhlbrigge RC, Pedersen FK, de Ferranti SD, Muller K. Premature subclinical atherosclerosis in children and young adults with juvenile idiopathic arthritis. A review considering preventive measures. Pediatr Rheumatol Online J 2016;14:3.
18. Selvaag AM, Aulie HA, Lilleby V, Flato B. Disease progression into adulthood and predictors of long-term active disease in juvenile idiopathic arthritis. Ann Rheum Dis 2016;75:190–5.
19. Smith JJ, Eather N, Morgan PJ, Plotnikoff RC, Faigenbaum AD, Lubans DR. The health benefits of muscular fitness for children and adolescents: a systematic review and meta-analysis. Sports Med 2014;44:1209–23.
20. Huber AM, Ward LM. The impact of underlying disease on fracture risk and bone mineral density in children with rheumatic disorders: A review of current literature. Semin Arthritis Rheum 2016;46:49–63.
21. Sandstedt E, Fasth A, Fors H, Beckung E. Bone health in children and adolescents with juvenile idiopathic arthritis and the influence of short-term physical exercise. Pediatr Phys Ther 2012;24:155–61.
22. Lofthouse CM, Azad F, Baildam EM, Akobeng AK. Measuring the nutritional status of children with juvenile idiopathic arthritis using the bioelectrical impedance method. Rheumatology (Oxford) 2002;41:1172–7.
23. Lien G, Selvaag AM, Flato B, Haugen M, Vinje O, Sorskaar D, et al. A two-year prospective controlled study of bone mass and bone turnover in children with early juvenile idiopathic arthritis. Arthritis Rheum 2005;52:833–40.
24. Caetano MC, Sarni RO, Terreri MT, Ortiz TT, Pinheiro M, de Souza FI, et al. Excess of adiposity in female children and adolescents with juvenile idiopathic arthritis. Clin Rheumatol 2012;31:967–71.
25. Jednacz E, Rutkowska-Sak L. Assessment of the body composition and parameters of the cardiovascular risk in juvenile idiopathic arthritis. BioMed Res Int i 2015;2015:619023.
26. Risum K, Hansen BH, Selvaag AM, Molberg O, Dagfinrud H, Sanner H. Physical activity in patients with oligo- and polyarticular juvenile idiopathic arthritis diagnosed in the era of biologics: a controlled cross-sectional study. Pediatr Rheumatol Online J 2018;16:64.
27. 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.
28. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. BMJ 2000;320:1240–3.
29. Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. Arch Dis Child 1970;45:13–23.
30. Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. Arch Dis Child 1969;44:291–303.

31. Weiss JE, Luca NJ, Boneparth A, Stinson J. Assessment and management of pain in juvenile idiopathic arthritis. Paediatr Drugs 2014;16:473–81.
32. Consolaro A, Ruperto N, Bazso A, Pistorio A, Magni-Manzoni S, Filocamo G, et al. Development and validation of a composite disease activity score for juvenile idiopathic arthritis. Arthritis Rheum 2009;61:658–66.
33. Wallace CA, Giannini EH, Huang B, Itert L, Ruperto N. American College of Rheumatology provisional criteria for defining clinical inactive disease in select categories of juvenile idiopathic arthritis. Arthritis Care Res (Hoboken) 2011;63:929–36.
34. Singh G, Athreya BH, Fries JF, Goldsmith DP. Measurement of health status in children with juvenile rheumatoid arthritis. Arthritis Rheum 1994;37:1761–9.
35. Flato B, Sorskaar D, Vinje O, Lien G, Aasland A, Moum T, et al. Measuring disability in early juvenile rheumatoid arthritis: evaluation of a Norwegian version of the childhood Health Assessment Questionnaire. J Rheumatol 1998;25:1851–8.
36. Borg GA. Perceived exertion. Exerc Sport Sci Rev 1974;2:131–53.
37. Gjesdal CG, Aanderud SJ, Haga HJ, Brun JG, Tell GS. Femoral and whole-body bone mineral density in middle-aged and older Norwegian men and women: suitability of the reference values. Osteoporos Int 2004;15:525–34.
38. Evenson KR, Catellier DJ, Gill K, Ondrak KS, McMurray RG. Calibration of two objective measures of physical activity for children. J Sports Sci 2008;26:1557–65.
39. Trost SG, Loprinzi PD, Moore R, Pfeiffer KA. Comparison of accelerometer cut points for predicting activity intensity in youth. Med Sci Sports Exerc 2011;43:1360–8.
40. Kolle E, Stokke JS, Hansen BH, Anderssen S. Physical activity in 6-, 9- and 15-year-olds in Norway. Results from a survey in 2011; 2012.
41. Holm I, Fredriksen P, Fosdahl M, Vollestad N. A normative sample of isotonic and isokinetic muscle strength measurements in children 7 to 12 years of age. Acta Paediatr 2008;97:602–7.
42. Armstrong N, Welsman JR. Assessment and interpretation of aerobic fitness in children and adolescents. Exerc Sport Sci Rev 1994;22:435–76.
43. Myers AM, Beam NW, Fakhoury JD. Resistance training for children and adolescents. Transl Pediatr 2017;6:137–43.
44. Burnham JM, Shults J, Dubner SE, Sembhi H, Zemel BS, Leonard MB. Bone density, structure, and strength in juvenile idiopathic arthritis: importance of disease severity and muscle deficits. Arthritis Rheum 2008;58:2518–27.
45. Faigenbaum AD, Kraemer WJ, Blimkie CJ, Jeffreys I, Micheli LJ, Nitka M, et al. Youth resistance training: updated position statement paper from the national strength and conditioning association. J Strength Cond Res 2009;23:S60–79.
46. World Health Organization. Global Recommendations on Physical Activity for Health 5-17 years old; 2011. URL: https://www.who.int/dietphysicalactivity/factsheet_young_people/en/.

Arthritis Care & Research
Vol. 71, No. 12, December 2019, pp 1621–1629
DOI 10.1002/acr.23803

American College of Rheumatology
Empowering Rheumatology Professionals

Cost-Effectiveness of Alternative Anticoagulation Strategies for Postoperative Management of Total Knee Arthroplasty Patients

Savannah R. Smith,[1] Jeffrey N. Katz,[2] and Elena Losina[3]

Objective. To evaluate the cost-effectiveness of prolonged (35-day) and standard-duration (14-day) anticoagulation therapy following total knee arthroplasty (TKA).

Methods. Using Markov modeling, we assessed clinical and economic outcomes of 14-day and 35-day anticoagulation therapy following TKA with rivaroxaban, low molecular weight heparin (LMWH), fondaparinux, warfarin, and aspirin. Incidence of complications of TKA and anticoagulation therapy (deep vein thrombosis [DVT], pulmonary embolism [PE], prosthetic joint infection [PJI], and bleeding) were derived from published literature. Daily costs ranged from $1 (aspirin) to $43 (fondaparinux). Primary outcomes included quality-adjusted life years (QALYs), direct medical costs, and incremental cost-effectiveness ratios (ICERs) at 1 year post-TKA. The preferred regimen was the regimen with highest QALYs maintaining an ICER below the willingness-to-pay threshold ($100,000/QALY). We conducted probabilistic sensitivity analyses, varying complication incidence and anticoagulation efficacy, to evaluate the impact of parameter uncertainty on model results.

Results. Aspirin resulted in the highest cumulative incidence of DVT and PE, while prolonged fondaparinux led to the largest reduction in DVT incidence (15% reduction compared to no prophylaxis). Despite differential bleeding rates (ranging from 3% to 6%), all strategies had similar incidence of PJI (1% to 2%). Prolonged rivaroxaban was the least costly strategy ($3,300 at 1 year post-TKA) and the preferred regimen in the base case. In sensitivity analyses, prolonged rivaroxaban and warfarin had similar likelihoods of being cost-effective.

Conclusion. Extending postoperative anticoagulation therapy to 35 days increases QALYs compared to standard 14-day prophylaxis. Prolonged rivaroxaban and prolonged warfarin are most likely to be cost-effective post-TKA; the costs of fondaparinux and LMWH precluded their being preferred strategies.

INTRODUCTION

Patients with total knee arthroplasty (TKA) are at risk of deep vein thrombosis (DVT) and pulmonary embolism (PE). Approximately 60% of patients without prophylaxis develop DVT (1,2). While many of these thromboses are isolated to calf veins, resolving without complications, nearly 12% of patients after TKA without preventative treatment present with evidence of proximal thromboses early in the postoperative course (2,3). Proximal DVTs are more likely to be clinically significant and can spontaneously break free, resulting in PE, which contributes to 100,000 deaths annually and increases the risk of recurrent DVT (4–6).

To reduce the risk of DVT and PE, TKA patients are prescribed anticoagulants, as recommended by American College of Chest Physicians (ACCP) and the American Academy of Orthopedic Surgeons (AAOS) (7,8). While both professional societies suggest chemoprophylaxis, the guidelines are unclear regarding the specific agent and appropriate duration. ACCP recommends therapy for a minimum of 10–14 days and up to 35 days, whereas AAOS leaves the duration of postoperative anticoagulants to physician discretion (9,10). The absence of guidance on duration and regimen selection has resulted in high variability in the postoperative care of TKA patients, with various anticoagulants employed from 7 days to 6 weeks (7). Further,

Supported by the Rheumatology Research Foundation Medical Student Preceptorship and the NIH (grant K24-AR-057827 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases).

[1]Savannah R. Smith, BA: George Washington University School of Medicine and Health Sciences, Washington, DC, and Brigham and Women's Hospital, Boston, Massachusetts; [2]Jeffrey N. Katz, MD, MSc: Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; [3]Elena Losina, PhD: Brigham and Women's Hospital, Harvard Medical School, and Boston University School of Public Health, Boston, Massachusetts.

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

Address correspondence to Elena Losina, PhD, Brigham and Women's Hospital, Orthopedic and Arthritis Center for Outcomes Research and Policy and Innovation Evaluation in Orthopedic Treatments Center, 75 Francis Street, Building for Transformative Medicine, Room 5-016, Boston, MA 02115. E-mail: elosina@bwh.harvard.edu.

Submitted for publication April 5, 2018; accepted in revised form October 23, 2018.

SIGNIFICANCE & INNOVATIONS

- For all anticoagulants, prolonging the duration of postoperative prophylaxis to 35 days increases quality-adjusted life years compared to standard 14-day prophylaxis.
- Prolonged therapy with warfarin and rivaroxaban are cost-effective strategies following total knee arthroplasty. The cost of fondaparinux and low molecular weight heparin precluded their cost-effectiveness.
- As prolonged prophylaxis with rivaroxaban and warfarin are comparable from a cost-effectiveness standpoint, patient preferences regarding the higher incidence of bleeding with rivaroxaban versus the increased incidence of deep vein thrombosis with warfarin can help inform the choice of postoperative anticoagulation therapy.

the daily cost of anticoagulation regimens varies from $1 to $43 (9–12).

Some research has suggested that longer anticoagulation regimens can substantially reduce the risk of DVT and PE (10,13); however, in the absence of definitive recommendations, physicians are left weighing the risks of DVT and PE against those of anticoagulation therapy, including hemorrhage of gastrointestinal (GI) and central nervous system (CNS) sites, as well as a higher likelihood of prosthetic joint infection (PJI) (14). Prior cost-effectiveness analyses evaluating anticoagulation therapy after joint arthroplasty have compared only 2 agents and few studies have considered the duration of therapy (15,16). One analysis evaluating prolonged (42-day) versus standard (12-day) treatment with enoxaparin in total hip arthroplasty (THA) or TKA patients suggested that prolonged therapy was cost-effective in THA patients (17). However, this study did not account for increased cumulative bleeding risk for prolonged therapy, thereby minimizing the potential adverse effects of extending the duration of anticoagulation therapy.

Given the lack of guidance regarding the specific agent and duration of prophylaxis and the wide range in the cost of anticoagulants, we sought to weigh the clinical benefits of prolonged (35-day) and standard-duration (14-day) anticoagulation therapy, including reduced likelihood of DVT and PE, against the increased risks, including bleeding and PJI, taking into consideration the regimen costs, of commonly used post-TKA anticoagulants.

MATERIALS AND METHODS

Analytic overview. We conducted a formal cost-effectiveness analysis to evaluate 10 plausible clinical strategies: 14-day (standard) and 35-day (prolonged) postoperative anticoagulation therapy with rivaroxaban, low molecular weight heparin (LMWH), warfarin, fondaparinux, and aspirin in patients undergoing TKA. The primary outcomes were quality-adjusted life years (QALYs) and costs at 1-year post-TKA. While TKA and anticoagulation complications most commonly present within the first few postoperative months, the associated decrease in quality of life (QoL) can persist for the first postoperative year; we therefore chose a 1-year time horizon to capture "important differences in consequences, both intended and unintended," as recommended by the Panel of Cost-Effectiveness in Health and Medicine (18). We report quality-adjusted life days (QALDs) in each health state, which represent the number of days and proportion of the first postoperative year spent without complications and in each complication state. The cost-effectiveness of each strategy was expressed in terms of incremental cost-effectiveness ratios (ICERs), defined as the increase in cost per increase in QALY. We adopted a societal perspective and assumed a willingness-to-pay threshold of $100,000/QALY (19,20); strategies with ICERs below willingness-to-pay were considered cost-effective. Strategies that increased cost while reducing QALYs were deemed dominated. Because there is no standard of care for post-TKA anticoagulation therapy, we identified comparator strategies by considering the most commonly used clinically relevant DVT chemoprophylactic regimens for TKA patients; we therefore conducted cost-effectiveness analyses with 2 strategies as the reference case: standard-duration warfarin and LMWH.

Model structure. We constructed a probabilistic, Markov, state-transition computer simulation model (TreeAge Pro 2017). Following TKA, subjects transitioned between the following major health states (Figure 1): proximal DVT, PE, bleeding, PJI, or no postoperative complications. We focused on proximal (versus distal) DVT, because this type of DVT is more likely to manifest clinically (21–24). Each health state is associated with a unique health-related QoL and cost. Death can occur at any state. The model follows each simulated subject from the time of TKA (model entry) to death or the end of the first postoperative year, whichever occurs first.

Subjects developing proximal DVT were stratified by symptomatic status (symptomatic or asymptomatic), each carrying a risk of propagation to PE. Those who did not progress to PE within 30 days were considered resolved and at risk of recurrent DVT. Those who developed PE experienced an increased risk of death for 30 days following PE diagnosis; subjects who did not die from PE entered a post-PE state where they were at risk of recurrent DVT.

Major bleeding was categorized as operative- or nonoperative-site bleeding. Operative-site bleeding increased the likelihood of PJI, while nonoperative-site bleeding (GI or CNS) increased mortality. We assumed that the increased risks of PJI and death following operative- and nonoperative-site bleeding persisted for 30 days.

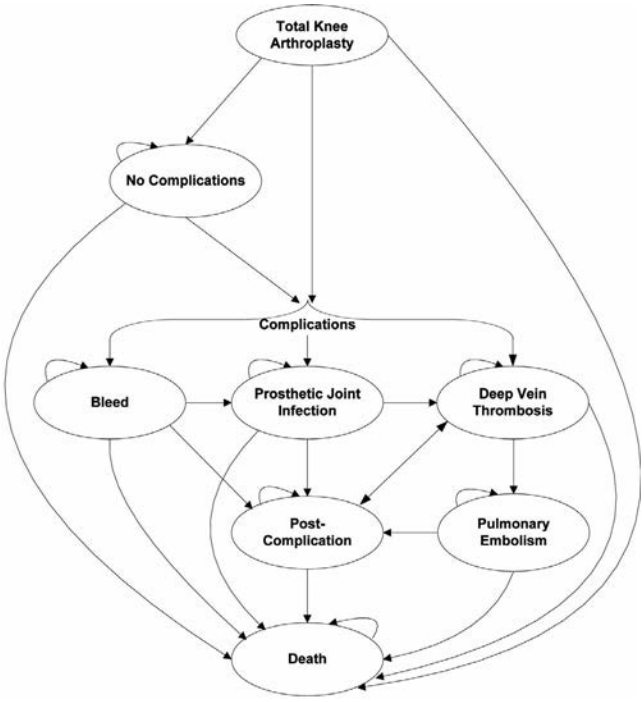

Figure 1. Model structure used to assess anticoagulation strategies after total knee arthroplasty (TKA). Following TKA, subjects can experience deep vein thrombosis, prosthetic joint infection, or bleeding or can proceed with no complications. Those who experience a complication and do not die enter a post-complication state, which is specific to the complication experienced (post–deep vein thrombosis, post–pulmonary embolism, post-bleeding, and post–prosthetic joint infection). Subjects are at continued risk of deep vein thrombosis in the post-complication state. Death can occur at any point.

Input parameters. *Transition probabilities*. Background mortality in each state represented the likelihood of death unrelated to TKA or anticoagulation complications. We derived age-stratified background mortality from the Centers for Disease Control 2013 Life Tables (25). Complications of TKA and anticoagulation therapy included proximal DVT, PE, major bleeding, and PJI. The likelihood of each complication was dependent on days post-TKA.

The underlying incidence of proximal DVT in the immediate postoperative period (days 0–14 post-TKA) was estimated from a published medical record review of 517 TKA patients (638 total TKAs) with average age 66 years; 78% were female, and the majority (85%) were white. Within this study population, indications for TKA were primarily degenerative arthritis (73%) and rheumatoid arthritis (15%). DVT was established via venography in the first postoperative week and was defined as an intraluminal filling defect within the calf, popliteal, or thigh veins. Of the 49 patients without anticoagulation therapy (62 TKAs), 36 had evidence of thrombosis in the calf, and 7 had evidence of thrombosis in the popliteal or thigh veins. We defined proximal DVTs as thromboses within the popliteal or thigh veins, resulting in 11.29% incidence of proximal DVT following TKA without preoperative anticoagulants (3). Evidence suggests that the increased risk of DVT following orthopedic surgery is highest in the immediate postoperative period and may persist for up to 3 months (13,26). Based on published literature, we derived a likelihood of DVT in the extended post-TKA period (days 15–90) of 6.41% (3,7). At 3 months, we assumed no increased risk of DVT as a direct result of TKA and employed age-based DVT risks for the remainder of the year (0.25% annually) (27).

Based on published evidence, we estimated 68.75% of proximal DVTs were symptomatic (28). Patients with DVTs were at risk of developing PE. Because symptomatic patients are more likely to present to care and receive treatment, the likelihood of progression to PE from symptomatic DVT is lower than asymptomatic DVT. We derived the likelihood of progression to PE from asymptomatic and symptomatic DVT as 50% and 6.25%, respectively (21,28).

We incorporated the risk of recurrent DVT for those who experienced resolution of the primary DVT or PE. Several patient factors are associated with recurrence of DVT (6,24,29–31). Postsurgical patients have an increased risk of primary DVT but a lower risk of DVT recurrence (hazard ratio 0.36) (30), whereas

Table 1. Model input data complication characteristics*

	Proximal DVT	PE	Bleeding	PJI
Probability (no prophylaxis), %				
Days 1–14	11.29 ± 5.49 (ref. 3)	–	1.50 ± 0.75 (ref. 7)	0.43 ± 0.22 (refs. 35,36)
Days 15+	6.41 ± 3.26/0.25 ± 0.12 (refs. 3,7,27)†	–	0.50 ± 0.25 (ref. 7)	1.06 ± 0.53 (refs. 35,36)
Cost, $ (refs. 9, 11, 12, 48–50)	9,600	11,000	13,000‡	48,300
Utility	0.661 (ref. 41)§	0.499 (ref. 41)	0.573 (refs. 44,48)‡	0.44 (ref. 43)¶
Mortality, %	–	3.00/7.70 (ref. 32)#	9.26 (ref. 34)‡	RR = 7.20 (ref. 38)

* All costs are in 2016 US$. DVT = deep vein thrombosis; PE = pulmonary embolism; PJI = prosthetic joint infection; ref. = reference; RR = risk ratio.
† Values for days 15–90/days 91+.
‡ Applied to nonoperative-site bleeding only.
§ Applied to symptomatic proximal DVT only.
¶ Applied to acute PJI state; utility of post-PJI = 0.72.
# Days post-PE diagnosis 1–5/6–30.

Table 2. Model input data treatment characteristics*

	Fondaparinux	Rivaroxaban	LMWH	Warfarin	Aspirin
Daily cost, $ (refs. 9–12)	43†	8	37†	6/3‡	1
RR DVT (ref. 39)	0.08 (0.03)	0.12 ± 0.04	0.20 ± 0.07	0.36 ± 0.12	0.62 ± 0.21
RR bleeding (ref. 39)	2.21 (0.74)	2.12 ± 0.71	1.23 ± 0.41	1.21 ± 0.40	1.0 ± 0.15

* All costs are in 2016 US$. LMWH = low molecular weight heparin; refs. = references; RR = risk ratio; DVT = deep vein thrombosis.
† Includes cost of injection administration ($20).
‡ Week 1/weeks 2+, includes cost of monitoring.

those patients with primary DVT who progress to PE have a nearly 3-fold (6) increased risk of recurrent DVT. Using data from Prandoni et al (30), we derived a 1-year cumulative incidence of recurrent DVT among those with resolved primary DVT of 17.68%. The likelihood of recurrent DVT in those with resolved PE was increased by a relative risk (RR) of 2.77 (6).

The likelihood of death from PE was time-dependent, with greater risk immediately following diagnosis. Data from Smith et al (32) suggest 3.00% in-hospital and 7.70% 30-day mortality risk for PE patients receiving early medical intervention. The median hospital stay for PE patients was 4.6 days; we therefore applied 3.00% mortality risk for the first 5 days following PE diagnosis.

The risk of post-TKA bleeding without anticoagulation therapy was derived from ACCP data: 1.50% for early bleeding (0–14 days) and 0.50% for late bleeding (15+ days) (7). Bleeding was differentiated by location (operative or nonoperative site). Based on published data, we assumed 68.7% of bleeding would occur at the operative site, and the remaining 31.3% at nonoperative sites, including GI and CNS (33). Operative-site bleeding increased the risk of PJI development (RR 9.8) (14), while nonoperative-site bleeding carried a case fatality of 9.26% (34), which was applied for 90 days following the hemorrhagic event.

We estimated the annual incidence of PJI from a published medical record review of 1,214 primary TKA recipients with median age of 72 years. The majority of the patients were female (63%), and 59% were considered obese or morbidly obese. A total of 92% of patients underwent TKA due to osteoarthritis, 8% due to rheumatoid arthritis, and 1% due to osteonecrosis or trauma. There were 18 prosthetic infections identified within the first postoperative year, resulting in 1.48% annual incidence of PJI (35). Based on a retrospective review of PJI in post-TKA patients, we assumed 28.87% of PJIs would occur in the first 3 months

Table 3. Base case cost-effectiveness analysis*

Strategy	Cost, $	QALY	ICER	DVT, %	Bleeding, %
Warfarin†					
Prolonged rivaroxaban	3,279	0.733	Cost saving	18.0	6.0
Prolonged warfarin	3,291	0.732	Cost saving	21.9	4.0
Standard rivaroxaban	3,416	0.732	Cost saving	22.8	5.4
Standard warfarin	3,551	0.732	Reference	25.6	3.9
Prolonged aspirin	3,689	0.731	Dominated	25.7	3.5
Standard aspirin	3,777	0.731	Dominated	28.4	3.4
No prophylaxis	3,869	0.726	Dominated	32.1	3.3
Standard LMWH	3,898	0.732	Dominated	23.9	3.9
Standard fondaparinux	3,932	0.732	$977,100	22.3	5.6
Prolonged LMWH	4,375	0.733	Dominated	19.5	4.1
Prolonged fondaparinux	4,529	0.733	$1,085,600	17.3	6.2
LMWH†					
Prolonged rivaroxaban	3,279	0.733	Cost saving	18.0	6.0
Prolonged warfarin	3,291	0.732	Cost saving	21.9	4.0
Standard rivaroxaban	3,416	0.732	Cost saving	22.8	5.4
Standard warfarin	3,551	0.732	Dominated	25.6	3.9
Prolonged aspirin	3,689	0.731	Dominated	25.7	3.5
Standard aspirin	3,777	0.731	Dominated	28.4	3.4
No prophylaxis	3,869	0.726	Dominated	32.1	3.3
Standard LMWH	3,898	0.732	Reference	23.9	3.9
Standard fondaparinux	3,932	0.732	$243,500	22.3	5.6
Prolonged LMWH	4,375	0.733	Dominated	19.5	4.1
Prolonged fondaparinux	4,529	0.733	$1,085,600	17.3	6.2

* A strategy that leads to greater cost without clinical benefit is deemed dominated. Deep vein thrombosis (DVT) and bleeding are shown as cumulative incidence. Standard strategies are 14 days, and prolonged strategies are 35 days. QALYs = quality-adjusted life years; ICER = incremental cost-effectiveness ratio; LMWH = low molecular weight heparin.
† Standard-duration standard of care.

(36), which led to a cumulative incidence of PJI of 0.43% in the first 3 months and 1.06% in the subsequent 9 months post-TKA. Mortality was derived using Nosocomial Infection National Surveillance Service, resulting in increased odds of death for patients with deep infection compared to no infection post-TKA of 7.2 (37,38). The increased mortality from PJI was applied for 120 days following diagnosis.

Regimen-specific anticoagulation efficacy and toxicity. We estimated the QoL from complications using data from the National Institute for Health and Clinical Excellence (NICE) guidelines on reducing venous thromboembolic events in hospitalized patients (39). Using network meta-analysis of anticoagulation studies for TKA, NICE derived RRs of DVT and major bleeding events (Tables 1 and 2), which were applied for 35 days for prolonged and 14 days for standard strategies.

QoL utilities. We assumed a QoL utility of 0.69 in the first 3 months following TKA for those patients without complications, reflecting the pain and functional limitation in the immediate recovery period. The utility of the post-TKA state with no complications was increased to 0.76 after 3 months (40). Symptomatic proximal DVT, PE, nonoperative-site bleeding, and PJI carried reduced QoL utilities of 0.66, 0.49, 0.57, and 0.44, respectively, which were applied for the duration of the complication state (described above) (41–44). Those patients with asymptomatic proximal DVT or nonoperative-site bleeding, as well as those with resolved complications, acquired the same utility as patients with no complications. Patients with resolved PJI were assigned a lower post-PJI utility (0.72), due to the necessity of revision and associated recovery (43).

Costs. Daily costs of anticoagulants were derived by converting average wholesale prices from Red Book Online to average sales prices by discounting brand name drugs by 26% and generic drugs by 68% (45,46). The resulting average sales prices were weighted as 93% generic and 7% brand name, representing the proportion of generic versus brand name drugs prescribed in the US (47). We included the cost of injection administration ($25), estimated from Medicare reimbursement schedules, for LMWH and fondaparinux (10). Additionally, the cost of monitoring ($20/week), consisting of prothrombin time (international normalized ratio), and an established patient visit were incorporated into the cost of warfarin treatment (9–12). We assumed 2 monitoring sessions within the first week of therapy and weekly for all subsequent weeks.

Complication costs were derived from the Healthcare Cost and Utilization Project 2014 and inflated to 2016 US$ (9,12,48). Because PJI frequently requires revision surgery, the acute cost of PJI included the cost of revision TKA, estimated at $25,000 (9,10,12,48–50). Final costs (2016 US$) were $9,600 for DVT, $11,000 for PE, $48,400 for PJI, and $13,000 for nonoperative-site bleeding.

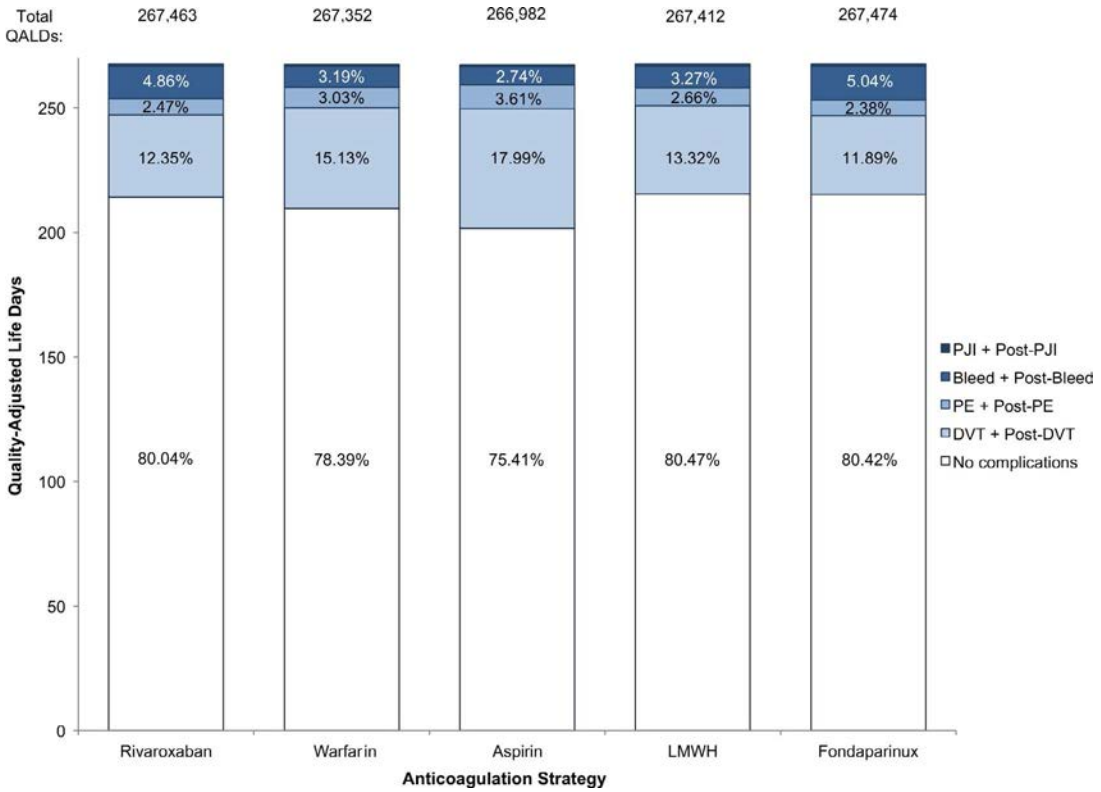

Figure 2. The total quality-adjusted life days and proportion of the first postoperative year spent in each health state is depicted for each prolonged strategy. QALDs = quality-adjusted life days; PJI = prosthetic joint infection; PE = pulmonary embolism; DVT = deep vein thrombosis, LMWH = low molecular weight heparin.

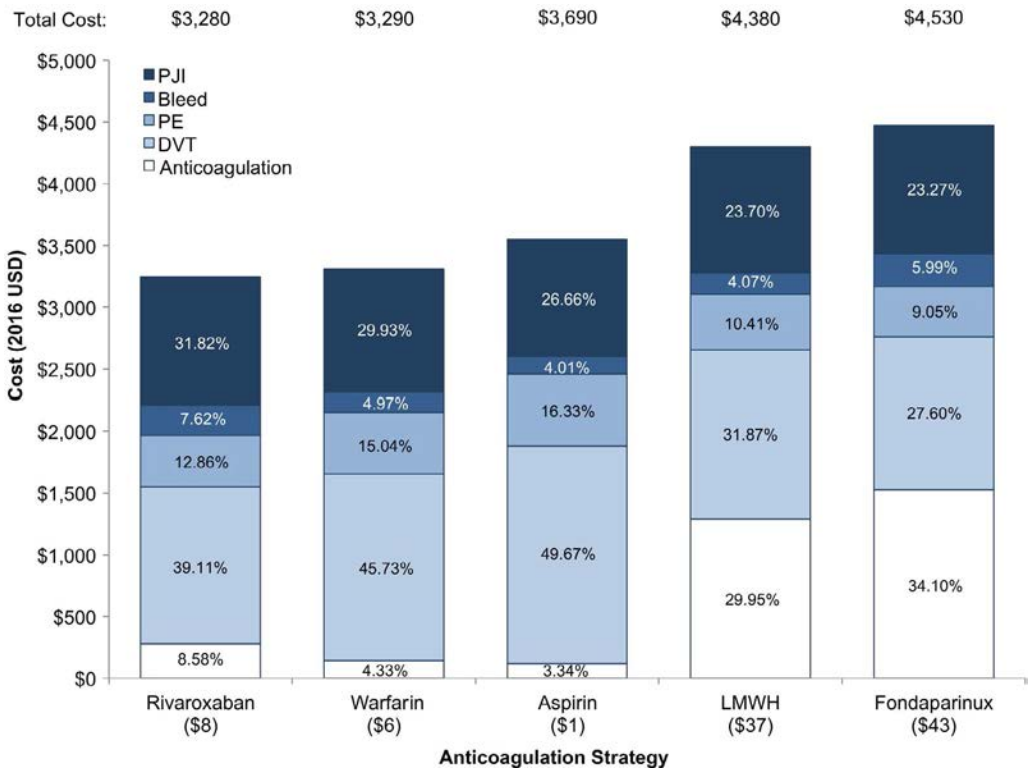

Figure 3. The individual components of the annual cost for each prolonged strategy are shown. Anticoagulation includes the cost of the anticoagulant and any monitoring (warfarin) or administration (low molecular weight heparin [LMWH] and fondaparinux), as appropriate. The cost of anticoagulation in the prolonged aspirin strategy is minimal and not indicated here. PJI = prosthetic joint infection; PE = pulmonary embolism; DVT = deep vein thrombosis. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.23803/abstract.

Sensitivity analyses. We conducted a sensitivity analysis removing the cost of administration, to represent self-injection, of fondaparinux and LMWH. We performed probabilistic sensitivity analyses incorporating the variability in efficacy of each anticoagulation agent (QoL with DVT and bleeding) and baseline estimates of DVT, bleeding, and PJI incidences. Distributions used in probabilistic sensitivity analyses are shown in Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23803/abstract. Results of probabilistic sensitivity analyses are represented by a cost-effectiveness acceptability curve, which gives the proportion of 10,000 iterations for which each strategy was the preferred regimen over a range of willingness-to-pay thresholds. We repeated all analyses modeling a cohort of obese patients with increased baseline risk of DVT; the results of these analyses are shown in Supplementary Tables 2a and 2b and Supplementary Figure 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23803/abstract.

RESULTS

Aspirin (taken for either 14 or 35 days) was associated with the highest projected 1-year cumulative incidence of DVT and PE and the fewest bleeding events (Table 3). Prolonged rivaroxaban resulted in a 14% absolute reduction in proximal DVTs compared to no prophylaxis, and prolonged fondaparinux reduced proximal DVT incidence by an additional 1%. Prolonged rivaroxaban and fondaparinux led to bleeding rates of 6%. All strategies resulted in similar rates of PJI in the year following TKA.

Figure 2 shows the number of QALDs and the proportion of the year spent in each health state. All prolonged strategies were associated with approximately 267 QALDs at 1 year post-TKA; however, the distribution of the number QALDs varied between strategies. Prolonged rivaroxaban, fondaparinux, and LMWH resulted in an average 215 QALDs (approximately 80%) without complications. Prolonged aspirin had the fewest QALDs without complications (201) and the largest proportion spent in the DVT state (18%). Strategies resulting in the fewest QALDs with DVT (prolonged fondaparinux and rivaroxaban) also resulted in the highest average QALDs with bleeding.

Prolonged rivaroxaban was the least costly strategy, accumulating $3,280 in 1 year post-TKA (Figure 3 and Table 3). Anticoagulation cost was the largest contributor to overall cost for prolonged LMWH and fondaparinux, the most costly strategies (30% for LMWH, 34% for fondaparinux). In both standard and prolonged aspirin therapy, the cost of DVT treatment was the largest contributor to overall cost (50% for prolonged, 53% for

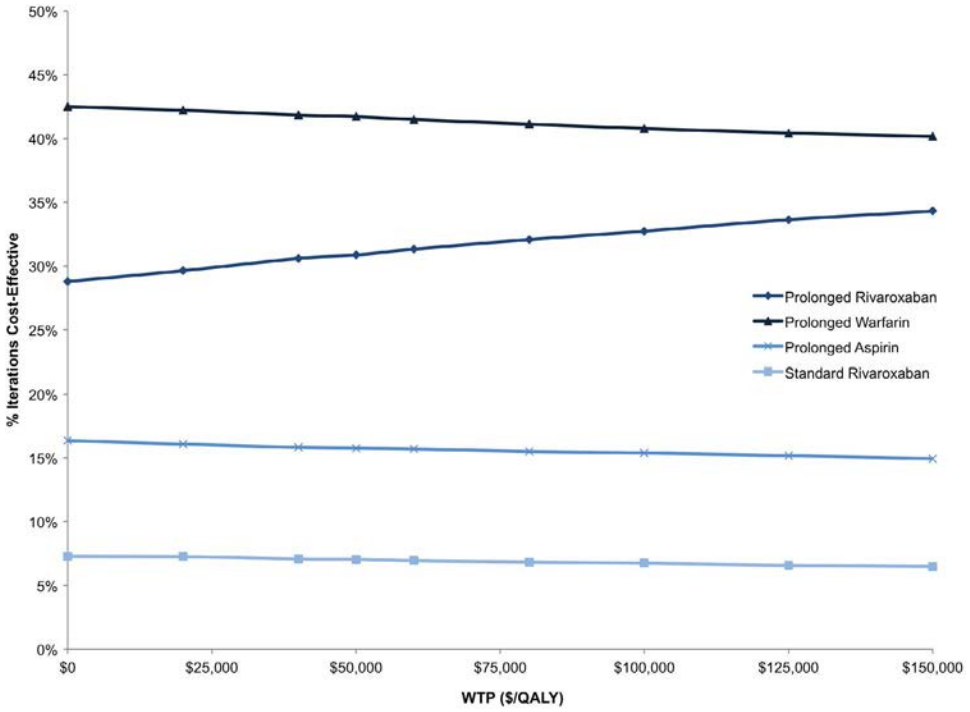

Figure 4. Proportion of iterations where a given strategy was the cost-effective option at various willingness-to-pay (WTP) thresholds. Strategies with probabilities of cost-effectiveness <5% are not shown. QALY = quality-adjusted life year. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.23803/abstract.

standard). For all prolonged strategies, 1–2% of the cohort experienced a PJI and 4–6% a bleeding event, leading to an average per-person annual cost of $1,000 and $200 on PJI and bleeding, respectively. Extending the duration of therapy saved an average of $300, $240, $170, $280, and $320 per person per year on DVT treatment for rivaroxaban, warfarin, aspirin, LMWH, and fondaparinux strategies, respectively. The cost saved from preventing DVT by prolonging therapy exceeded the additional cost of anticoagulation therapy for rivaroxaban, warfarin, and aspirin, but not for LMWH or fondaparinux.

Table 3 presents the base case cost-effectiveness analysis, comparing standard-duration warfarin and standard-duration LMWH as standard of care. All strategies increased QALYs compared to no prophylaxis. Assuming either standard-duration warfarin or LMWH as the standard of care, both durations of fondaparinux increased the cost with minimal increases in QALY, leading to ICERs more than $200,000/QALY. Prolonged warfarin and both durations of rivaroxaban were cost saving in both reference cases. Prolonged rivaroxaban was the preferred strategy using either standard-duration warfarin or LMWH as the standard of care.

The results of probabilistic sensitivity analyses are shown in Figure 4. Across all willingness-to-pay thresholds, prolonged warfarin was the cost-effective strategy in approximately 40% of iterations. Prolonged rivaroxaban reached maximum likelihood of cost-effectiveness (34%) at willingness-to-pay of $150,000/QALY. Prolonged aspirin was cost-effective in 15% of iterations across all willingness-to-pay thresholds.

DISCUSSION

We built a probabilistic state-transition computer simulation model to weigh the benefits, reduction in DVT and PE, against the harms, including increased bleeding and PJI, of standard or prolonged-duration therapy of 5 commonly prescribed anticoagulants following TKA. Our results show that extending the duration of therapy increases QALYs. In the base case, prolonged rivaroxaban was the preferred strategy; however, in sensitivity analyses, incorporating the uncertainty surrounding the efficacy of each anticoagulant, we found that prolonged rivaroxaban and warfarin were likely to be cost-effective in nearly equal proportions. The cost of LMWH and fondaparinux prohibited them from being cost-effective.

While minor differences may be seen between overall QALYs in the first postoperative year, all prolonged strategies increased the number no-complication days in the first postoperative year compared to standard-duration therapies, ranging from 10 QALDs for rivaroxaban, LMWH, and fondaparinux to 6 QALDs for aspirin. Additionally, the increase in postoperative bleeding events and PJIs was outweighed by the decrease in QALDs associated with DVT and PE for all prolonged strategies compared to their standard-duration counterparts.

To our knowledge, this is the first cost-effectiveness analysis to compare multiple anticoagulants at several durations post-TKA. Previous studies of anticoagulation therapy in TKA or THA recipients have focused on comparative analyses of 2 strategies or a single agent prescribed at different doses or durations (15,16). A

cost-effectiveness analysis by Schousboe and Brown (33) showed that, when compared to LMWH, aspirin was the preferred strategy in THA patients but the cost-effectiveness in TKA patients was dependent on age and DVT risk, with aspirin assuming a higher likelihood of cost-effectiveness in older patients without high DVT risk. Another analysis evaluating aspirin and warfarin showed that aspirin was the preferred strategy in THA patients of all ages, regardless of the rate of venous thromboembolism. In TKA patients, warfarin was the preferred strategy in patients with low risks of bleeding and a high risk of venous thromboembolism (51). The results of the base case analysis presented here contrast with those of Tabatabaee et al and Schousboe and Brown, with warfarin and LMWH leading to higher QALYs than aspirin. This difference could be explained by our use of a shorter time horizon, emphasizing the immediate QoL decrements of complications, notably DVT.

Few studies have formally assessed the clinical and economic outcomes associated with extending thromboprophylaxis in a TKA population. Haentjens et al (17) evaluated 12-day versus 42-day enoxaparin therapy following total joint arthroplasty and reported ICERs (2016 US$) of $8,900/QALY for THA and $83,200/QALY for TKA for 42-day compared to 12-day therapy. This analysis, however, assessed only enoxaparin and did not increase the risk of bleeding for extended therapy, potentially limiting the utility of the results.

There are important limitations to our analyses. We used the best currently available data to inform model inputs, but data on the continued efficacy of anticoagulants in the extended prophylactic period were limited. We based our efficacy estimates on network meta-analyses combining trials of both THA and TKA patients, which may overestimate the efficacy. Based on network meta-analytic results, we assumed that aspirin did not increase the risk of bleeding; however, some studies have observed similar incidence of bleeding between aspirin and LMWH or warfarin. We addressed the uncertainty in this parameter in probabilistic sensitivity analyses and found the results robust to plausible ranges in bleeding risk while taking aspirin.

In conclusion, we found prolonged therapies to increase QALYs compared to standard-duration therapies, supporting the extension of anticoagulation therapy post-TKA. Prolonged prophylaxis with warfarin and rivaroxaban emerged as cost-effective strategies in this setting. Because these 2 agents are comparable from a cost-effectiveness standpoint, patient preferences can help inform the choice of the appropriate postoperative anticoagulation strategy.

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. Losina 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, Katz, Losina.
Acquisition of data. Smith.
Analysis and interpretation of data. Smith, Katz, Losina.

REFERENCES

1. Anderson FA Jr, Spencer FA. Risk factors for venous thromboembolism. Circulation 2003;107 Suppl:I9–16.
2. Brookenthal KR, Freedman KB, Lotke PA, Fitzgerald RH, Lonner JH. A meta-analysis of thromboembolic prophylaxis in total knee arthroplasty. J Arthroplasty 2001;16:293–300.
3. Stulberg BN, Insall JN, Williams GW, Ghelman B. Deep-vein thrombosis following total knee replacement: an analysis of six hundred and thirty-eight arthroplasties. J Bone Joint Surg Am 1984;66: 194–201.
4. Goldhaber SZ, Visani L, De Rosa M. Acute pulmonary embolism: clinical outcomes in the International Cooperative Pulmonary Embolism Registry (ICOPER). Lancet 1999;353:1386–9.
5. Beckman MG, Hooper WC, Critchley SE, Ortel TL. Venous thromboembolism: a public health concern. Am J Prev Med 2010;38 Suppl:S495–501.
6. Eichinger S, Heinze G, Jandeck LM, Kyrle PA. Risk assessment of recurrence in patients with unprovoked deep vein thrombosis or pulmonary embolism: the Vienna prediction model. Circulation 2010;121:1630–6.
7. Falck-Ytter Y, Francis CW, Johanson NA, Curley C, Dahl OE, Schulman S, et al. Prevention of VTE in orthopedic surgery patients: antithrombotic therapy and prevention of thrombosis, 9th ed. American College of Chest Physicians evidence-based clinical practice guidelines. Chest 2012;141 Suppl:e278–325.
8. Mont MA, Jacobs JJ, Boggio LN, Bozic KJ, Della Valle CJ, Goodman SB, et al. Preventing venous thromboembolic disease in patients undergoing elective hip and knee arthroplasty. J Am Acad Orthop Surg 2011;19:768–76.
9. Centers for Medicare and Medicaid Services. Physician fee schedule. URL: https://www.cms.gov/medicare/medicare-fee-for-service-payment/physicianfeesched/.
10. Forster R, Stewart M. Anticoagulants (extended duration) for prevention of venous thromboembolism following total hip or knee replacement or hip fracture repair. Cochrane Database Syst Rev 2016;3:CD004179.
11. Centers for Medicare and Medicaid Services. National health expenditure data. URL: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/index.html.
12. Federal Reserve Bank of St. Louis. Personal consumption expenditures. URL: https://fred.stlouisfed.org/series/PCE.
13. Kearon C. Duration of venous thromboembolism prophylaxis after surgery. Chest 2003;124 Suppl:386–92.
14. Galat DD, McGovern SC, Hanssen AD, Larson DR, Harrington JR, Clarke HD. Early return to surgery for evacuation of a postoperative hematoma after primary total knee arthroplasty. J Bone Joint Surg Am 2008;90:2331–6.
15. Kapoor A, Chuang W, Radhakrishnan N, Smith KJ, Berlowitz D, Segal JB, et al. Cost effectiveness of venous thromboembolism pharmacological prophylaxis in total hip and knee replacement: a systematic review. Pharmacoeconomics 2010;28:521–38.
16. Brockbank J, Wolowacz S. Economic evaluations of new oral anticoagulants for the prevention of venous thromboembolism after total hip or knee replacement: a systematic review. Pharmacoeconomics 2017;35:517–35.
17. Haentjens P, De Groote K, Annemans L. Prolonged enoxaparin therapy to prevent venous thromboembolism after primary hip or knee replacement: a cost-utility analysis. Arch Orthop Trauma Surg 2004;124:507–17.
18. Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG. Cost-effectiveness in health and medicine. 2nd ed. Oxford: Oxford University Press; 2016.

19. Ubel PA, Hirth RA, Chernew ME, Fendrick AM. What is the price of life and why doesn't it increase at the rate of inflation? Arch Intern Med 2003;163:1637–41.
20. Ryen L, Svensson M. The willingness to pay for a quality adjusted life year: a review of the empirical literature. Health Econ 2015;24: 1289–301.
21. Kearon C. Natural history of venous thromboembolism. Circulation 2003;107 Suppl:I22–30.
22. Sule AA, Chin TJ, Handa P, Earnest A. Should symptomatic, isolated distal deep vein thrombosis be treated with anticoagulation? Int J Angiol 2009;18:83–7.
23. Galanaud JP, Sevestre-Pietri MA, Bosson JL, Laroche JP, Righini M, Brisot D, et al. Comparative study on risk factors and early outcome of symptomatic distal versus proximal deep vein thrombosis: results from the OPTIMEV study. Thromb Haemost 2009;102: 493–500.
24. Hansson PO, Sörbo J, Eriksson H. Recurrent venous thromboembolism after deep vein thrombosis: incidence and risk factors. Arch Intern Med 2000;160:769–74.
25. Arias E, Heron M, Xu J. National vital statistics reports; United States life tables, 2013. 2017. URL: https://www.cdc.gov/nchs/data/nvsr/nvsr66/nvsr66_03.pdf.
26. Heit JA, O'Fallon WM, Petterson TM, Lohse CM, Silverstein MD, Mohr DN, et al. Relative impact of risk factors for deep vein thrombosis and pulmonary embolism: a population-based study. Arch Intern Med 2002;162:1245–8.
27. Tsai AW, Cushman M, Rosamond WD, Heckbert SR, Polak JF, Folsom AR. Cardiovascular risk factors and venous thromboembolism incidence: the longitudinal investigation of thromboembolism etiology. Arch Intern Med 2002;162:1182–9.
28. Moser KM, LeMoine JR. Is embolic risk conditioned by location of deep venous thrombosis? Ann Intern Med 1981;94:439–44.
29. Kyrle PA, Rosendaal FR, Eichinger S. Risk assessment for recurrent venous thrombosis. Lancet 2010;376:2032–9.
30. Prandoni P, Lensing AW, Cogo A, Cuppini S, Villalta S, Carta M, et al. The long-term clinical course of acute deep venous thrombosis. Ann Intern Med 1996;125:1–7.
31. Wallace G, Judge A, Prieto-Alhambra D, de Vries F, Arden NK, Cooper C. The effect of body mass index on the risk of postoperative complications during the 6 months following total hip replacement or total knee replacement surgery. Osteoarthritis Cartilage 2014;22:918–27.
32. Smith SB, Geske JB, Maguire JM, Zane NA, Carter RE, Morgenthaler TI. Early anticoagulation is associated with reduced mortality for acute pulmonary embolism. Chest 2010;137:1382–90.
33. Schousboe JT, Brown GA. Cost-effectiveness of low-molecular-weight heparin compared with aspirin for prophylaxis against venous thromboembolism after total joint arthroplasty. J Bone Joint Surg Am 2013;95:1256–64.
34. Linkins LA, Choi PT, Douketis JD. Clinical impact of bleeding in patients taking oral anticoagulant therapy for venous thromboembolism: a meta-analysis. Ann Intern Med 2003;139:893–900.
35. Dowsey MM, Choong PF. Obese diabetic patients are at substantial risk for deep infection after primary TKA. Clin Orthop Relat Res 2009;467:1577–81.
36. Peersman G, Laskin R, Davis J, Peterson M. Infection in total knee replacement: a retrospective review of 6489 total knee replacements. Clin Orthop Relat Res 2001;392:15–23.
37. Public Health England. Protocol for surveillance of surgical site infection: surgical site infection surveillance service. Version 6. 2013. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/633775/surgical_site_infections_protocol_version_6.pdf.
38. Coello R, Charlett A, Wilson J, Ward V, Pearson A, Borriello P. Adverse impact of surgical site infections in English hospitals. J Hosp Infect 2005;60:93–103.
39. National Clinical Guideline Centre–Acute and Chronic Conditions (UK). Venous thromboembolism: reducing the risk of venous thromboembolism (deep vein thrombosis and pulmonary embolism) in patients admitted to hospital. London: Royal College of Physicians; 2010.
40. 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.
41. Melnikow J, Birch S, Slee C, McCarthy TJ, Helms LJ, Kuppermann M. Tamoxifen for breast cancer risk reduction: impact of alternative approaches to quality-of-life adjustment on cost-effectiveness analysis. Med Care 2008;46:946–53.
42. Bamber L, Muston D, McLeod E, Guillermin A, Lowin J, Patel R. Cost-effectiveness analysis of treatment of venous thromboembolism with rivaroxaban compared with combined low molecular weight heparin/vitamin K antagonist. Thromb J 2015;13:20.
43. Skaar DD, Park T, Swiontkowski MF, Kuntz KM. Cost-effectiveness of antibiotic prophylaxis for dental patients with prosthetic joints: comparisons of antibiotic regimens for patients with total hip arthroplasty. J Am Dent Assoc 2015;146:830–9.
44. You JH, Tsui KK, Wong RS, Cheng G. Cost-effectiveness of dabigatran versus genotype-guided management of warfarin therapy for stroke prevention in patients with atrial fibrillation. PloS One 2012;7:e39640.
45. Kimberlin DW, Brady MT, Jackson MA, Long SS. Red book: 2018–2021 report on the Committee on Infectious Diseases. 31st ed. Itasca (IL): American Academy of Pediatrics; 2018.
46. Department of Health and Human Services. Medicaid drug price comparison: average sales price to average wholesale price. 2005. URL: https://oig.hhs.gov/oei/reports/oei-03-05-00200.pdf.
47. IMS Institute for Healthcare Informatics. The use of medicines in the United States: review of 2010. Norwalk (CT): IMS Institute for Healthcare Informatics; 2011.
48. Long EF, Swain GW, Mangi AA. Comparative survival and cost-effectiveness of advanced therapies for end-stage heart failure. Circ Heart Fail 2014;7:470–8.
49. Centers for Medicare and Medicaid Services. Acute inpatient PPS. URL: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/index.html.
50. Centers for Medicare and Medicaid Services. Medicare anesthesia fee schedule. URL: https://www.cms.gov/medicare/medicare-fee-for-service-payment/physicianfeesched/.
51. Tabatabaee RM, Rasouli MR, Maltenfort MG, Parvizi J. Cost-effective prophylaxis against venous thromboembolism after total joint arthroplasty: warfarin versus aspirin. J Arthroplasty 2015;30:159–64.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 71, No. 12, December 2019, pp 1630–1639
DOI 10.1002/acr.23797

Psychometric Evaluation of the National Institutes of Health Patient-Reported Outcomes Measurement Information System in a Multiracial, Multiethnic Systemic Lupus Erythematosus Cohort

Patricia Katz,[1] Jinoos Yazdany,[1] Laura Trupin,[1] Stephanie Rush,[1] Charles G. Helmick,[2] Louise B. Murphy,[2] Cristina Lanata,[1] Lindsey A. Criswell,[1] and Maria Dall'Era[1]

Objective. We examined psychometric performance of Patient-Reported Outcomes Measurement Information System (PROMIS) measures in a racially/ethnically and linguistically diverse cohort with systemic lupus erythematosus (SLE).

Methods. Data were from the California Lupus Epidemiology Study, a multiracial/multiethnic cohort of individuals with physician-confirmed SLE. The majority (n = 332) attended in-person research visits that included interviews conducted in English, Spanish, Cantonese, or Mandarin. Up to 12 PROMIS short forms were administered (depending on language availability). An additional 99 individuals completed the interview by phone only. Internal consistency was examined with Cronbach's alpha and item-total correlations. Correlations with the Short Form 36 subscales and both self-reported and physician-assessed disease activity assessed convergent validity. All analyses were repeated within each racial/ethnic group. Differences in scores by race/ethnicity were examined in bivariate analyses and by multiple regression analyses controlling for age, sex, disease duration, and disease damage and activity.

Results. The total sample was 30.0% white, 22.3% Hispanic, 10.9% African American, 33.7% Asian, and 3.0% other race/ethnicity. Seventy-seven percent of interviews were conducted in-person. Non-English interviews were conducted in 26.0% of the Hispanic subjects and 18.6% of the Asian subjects. Each scale demonstrated adequate reliability and validity overall and within racial/ethnic groups. Minimal floor effects were observed, but ceiling effects were noted. Missing item responses were minimal for most scales, except for items related to work. No differences were noted by mode of administration or by language of administration among Hispanics and Asians. After accounting for differences in disease status, age, and sex, few differences in mean scores between whites and other racial/ethnic groups were noted.

Conclusion. PROMIS measures appear reliable and valid in persons with lupus across racial/ethnic groups.

INTRODUCTION

Lupus is a disease with extreme biologic and clinical heterogeneity that makes measurement of outcomes challenging in clinical research. The complexity of lupus is evident in the clinical measures of disease damage and activity, which assess diverse manifestations across multiple organ systems. The range and complexity of patient-reported outcomes (PROs) parallel that of the clinical outcomes (1). Multiple measures of lupus-specific quality of life have been published (1), but none are routinely used in clinical trials, observational studies, or clinical practice. The importance of including PROs as end points in clinical trials of novel therapies is gaining momentum and is recognized by the lupus community and the Food and Drug Administration (2,3). However, there is no consensus on which PROs should be used.

The National Institutes of Health Patient-Reported Outcomes Measurement Information System (PROMIS) initiative was undertaken to improve and standardize measurement of PROs (4). The

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

Supported by the Centers for Disease Control and Prevention (grants A114297 and U01DP005120).

[1]Patricia Katz, PhD, Jinoos Yazdany, MD, MPH, Laura Trupin, MPH, Stephanie Rush, BA, Cristina Lanata, MD, Lindsey A. Criswell, MD, MPH, DSc, Maria Dall'Era, MD: University of California San Francisco; [2]Charles G. Helmick, MD, Louise B. Murphy, PhD: Centers for Disease Control and Prevention, Atlanta, Georgia.

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

Address correspondence to Patricia Katz, PhD, University of California San Francisco, Arthritis Research Group, 3333 California Street, San Francisco, CA 94143. E-mail: Patti.katz@ucsf.edu.

Submitted for publication May 22, 2018; accepted in revised form October 16, 2018.

SIGNIFICANCE & INNOVATIONS

- The current study shows the first examination of Patient-Reported Outcomes Measurement Information System short forms across different racial/ethnic and language groups in a diverse lupus cohort, showing adequate reliability and validity, and minimal floor effects for each of 4 racial/ethnic groups.
- No differences were noted by mode of administration (in-person versus telephone) or by language among Hispanic and Asian participants.
- After controlling for differences in disease status, age, and sex, few differences existed between whites and other racial/ethnic groups, suggesting that differences in scale scores may be primarily attributable to differences in disease and demographics rather than race/ethnicity per se.

PROMIS measures reflect the broad view of health proposed by the World Health Organization (http://www.who.int/about/definition/en/print.html), covering physical, mental, and social health. PROMIS measures were developed using state-of-the-art psychometric techniques and may be administered via computer adaptive testing (CAT) or through static short forms that range from 4 to 20 items. Item banks and short forms exist for 20+ domains that represent a comprehensive model of health and health-related quality of life (HRQoL) that includes physical, social, and mental health. This theoretical framework is in accord with domains that have been reported to be important and meaningful to individuals with lupus (1,5). Notably, many of these content domains are among those of greatest concern to patients with lupus (e.g., cognition, sleep) (1,5), yet are not measured in current generic HRQoL questionnaires.

To date, only a handful of published reports have examined the psychometric characteristics of PROMIS measures in cohorts of individuals with systemic lupus erythematosus (SLE) (6–9). In all 4 of these studies, participants were exclusively English-speaking. In 2 studies, participants were primarily white, and in the remaining 2, PROMIS responses by race/ethnicity were not examined. In this study, we address this gap in the literature by examining the reliability, validity, and usefulness of the PROMIS measures in a racially/ethnically and linguistically diverse cohort of individuals with SLE.

SUBJECTS AND METHODS

Subjects. Subjects were participants in the California Lupus Epidemiology Study (CLUES), a multiracial/multiethnic cohort of individuals with physician-confirmed SLE. Participants were recruited from the California Lupus Surveillance Project, a population-based cohort of individuals with SLE living in San Francisco County from 2007 to 2009 (10). Additional participants who resided in the geographic region were recruited through local academic and community rheumatology clinics and through existing local research cohorts.

Study procedures involved an in-person research clinic visit, which included collection and review of medical records prior to the visit, a history and physical examination conducted by a physician specializing in lupus, collection of biospecimens for clinical and research purposes, and completion of a structured interview administered by an experienced research assistant. All SLE diagnoses were confirmed by study physicians, according to any of the following definitions: 1) meeting ≥4 of the 11 American College of Rheumatology (ACR) revised criteria for the classification of SLE as defined in 1982 and updated in 1997 (11,12),

Table 1. Patient-Reported Outcomes Measurement Information System (PROMIS) measures administered in the California Lupus Epidemiology Study

PROMIS short form	No. items	English	Spanish	Chinese
Physical health				
Physical function	10	✓	✓	✓
Pain interference*	4	✓	✓	✓
Fatigue*	4	✓	✓	✓
Sleep disturbance	4	✓	✓	✓
Sleep impairment	8	✓	✓	
Mental health				
Applied cognition, abilities	4	✓	✓	
Psychosocial illness impact, negative	8†	✓		
Psychosocial illness impact, positive	8†	✓		
Social health				
Ability to participate in social roles and activities	4	✓	✓	
Satisfaction with participation in discretionary social activities	7	✓	✓	
Satisfaction with participation in social roles	7	✓	✓	
Social isolation	4	✓	✓	

* In the Chinese version of the PROMIS short form, the number of items for pain interference = 6 and the number of items for fatigue = 7.
† Only 4 items are scored.

2) meeting 3 of the 11 ACR criteria plus a rheumatologist's documented diagnosis of SLE, or 3) a confirmed diagnosis of lupus nephritis (10). A subgroup of participants was unable to attend the in-person visit. For these individuals, medical records were collected and reviewed, and the same structured interview was administered by telephone. Diagnoses were confirmed through medical record review.

CLUES specifically aimed to include a diverse patient sample, with representation from multiple racial/ethnic groups speaking multiple languages. Study interviews were conducted in English, Spanish, Mandarin, or Cantonese. Data for these analyses included a total of 431 individuals, 332 of whom participated in an in-person visit.

Variables. *PROMIS*. The PROMIS short forms shown in Table 1 were administered as part of the structured interviews. All scales were scored as recommended and converted to T scores, with a mean ± SD population of 50 ± 10, using PROMIS scoring documentation (available at http://assessmentcenter.net). For all PROMIS scales, higher scores reflect "more" of the construct being measured. For example, higher physical function and satisfaction with social roles scores would reflect better functioning and satisfaction, and thus would be considered to be better scores; higher fatigue, pain interference, sleep disturbance, depression, and anxiety scores would be considered to be worse.

As noted above, PROMIS measures can be administered as static short forms or through CAT. CAT is intended to administer items that are targeted to the individual respondent, which may lead to greater measurement precision (4). However, the PROMIS item banks that support CAT are available only in English and Spanish, while the short forms are available in additional languages, including Mandarin and Cantonese. Because we wanted to use the same mode of administration for all CLUES participants, we chose to administer short forms.

Other patient-reported outcomes. Three other instruments were used to measure PROs. The Medical Outcomes Study Short Form 36 (SF-36) is a widely used PRO measure and includes 8 subscales, including physical function, role physical, role emotional, vitality, mental health, social function, and bodily pain (13). Scores for each scale range 0–100, with a mean ± SD population of 50 ± 10. Higher scores for each scale, except bodily pain, reflect better outcomes. SLE disease activity was measured with the Systemic Lupus Activity Questionnaire (SLAQ) (14,15), a validated, self-report measure of SLE disease activity. SLAQ scores range 0–44, with higher scores reflecting more disease activity. The SLAQ also includes an item for respondents to rate their lupus disease activity over the past 3 months (where 0 = no activity and 10 = high activity). The Brief Index of Lupus Damage (BILD) was used to estimate organ damage (16). The BILD is based on the Systemic Lupus International Cooperating Clinics/ACR Damage Index (SDI) (17), and consists of 28 items capturing information on 26 SDI items, including determinations of important comorbid conditions such as cardiovascular disease and events and diabetes mellitus. It has been shown to be predictive of hospitalizations and mortality (18).

Physician-reported measures and covariates. The Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) (19) and SDI (17) were completed by study rheumatologists during the in-person study visit. Race, ethnicity, age, age at lupus onset, household income, and education level were self-reported. Language was categorized by the language in which interviews were conducted (English, Spanish, Mandarin, or Cantonese). Current medications were recorded during interviews.

Statistical analysis. *Descriptive analyses*. Descriptive statistics were calculated for the total sample and for each racial/ethnic group. Differences in characteristics of groups were tested using chi-square analyses and *t*-tests. The percentage of respondents with missing items and scale scores was calculated. Distributions of PROMIS scores were examined. Because of the difference in the direction of scores (i.e., high scores reflected better health states for some scales and worse health states for other scales), we modified the standard terminology with floor referring to the worst score, and ceiling referring to the best. Chi-square analyses and *t*-tests were used to compare characteristics and PROMIS scores of individuals completing in-person versus telephone interviews.

Reliability and validity. Internal consistency was assessed by examining item-total correlations and Cronbach's alpha. Item-total correlations ≥0.4 and α ≥0.80 are considered acceptable (20). For assessment of convergent validity, Pearson's correlation and Spearman's correlation analyses were used to examine associations of PROMIS scale scores with PROs for similar domains and measures of disease activity and damage. Correlations of 0.3–0.5 were considered low, 0.5–0.7 moderate, and ≥0.7 high (21).

Psychometric analyses by racial/ethnic group. All descriptive, reliability, and validity analyses were repeated within each racial/ethnic group. Within the relevant racial/ethnic group, *t*-tests compared PROMIS scores of individuals completing interviews in English or another language.

Differential scores by race/ethnicity. Differences in PROMIS scores by race/ethnicity were examined to determine if there appeared to be systematic differences in scores that were not attributable to differences in lupus severity, health status, or socioeconomic status. Differences in PROMIS scores between whites and other racial/ethnic groups were examined using multiple linear regression analyses, first with no covariates, and then controlling for age, sex, disease duration, SLEDAI, and SDI to determine if systematic differ-

Table 2. Baseline characteristics of the California Lupus Epidemiology Study cohort (n = 431)*

	Total	White	Hispanic	African American	Asian	Other	*P*
Number	431	130 (30.0)	96 (22.3)	47 (10.9)	145 (33.7)	13 (3.0)	
In-person interview	332 (77.0)	96 (73.9)	76 (79.2)	36 (76.6)	118 (81.4)	6 (46.2)	0.05
Female	387 (89.8)	116 (89.2)	83 (86.5)	46 (97.9)	129 (89.0)	13 (100)	0.19
Age, mean ± SD years	46.6 ± 14.3	51.4 ± 12.3	42.7 ± 14.1	52.8 ± 14.8	42.7 ± 14.0	48.7 ± 13.3	<0.0001
Below poverty	75 (19.4)	7 (5.7)	25 (29.1)	15 (37.5)	24 (19.1)	4 (36.4)	<0.0001
Low education	94 (22.1)	11 (8.6)	29 (30.5)	18 (38.3)	34 (23.6)	2 (16.7)	<0.0001
Non-English interview	52 (12.1)	0	25 (26.0)	0	27 (18.6)	0	<0.0001
Disease duration, mean ± SD years	17.7 ± 11.1	22.1 ± 10.7	15.0 ± 9.9	18.9 ± 13.1	14.6 ± 9.8	24.3 ± 11.0	<0.0001
Current GC use	210 (48.7)	50 (38.5)	50 (52.1)	26 (55.3)	77 (53.1)	7 (53.9)	0.09
High dose GC use (≥7.5 mg for ≥ 3 months in past year)	95 (22.4)	27 (20.8)	22 (23.2)	10 (21.7)	32 (22.5	34 (3.3)	0.90
Current non-GC immunosuppressant	200 (46.4)	39 (30.0)	48 (50.0)	25 (53.2)	84 (57.9)	4 (30.8)	<0.0001
SLEDAI (n = 330), mean ± SD	3.0 ± 3.1	2.4 ± 3.0	3.6 ± 3.6	2.3 ± 2.0	3.1 ± 2.9	4.2 ± 4.5	0.07
SDI (n = 331), mean ± SD	1.8 ± 2.0	1.9 ± 2.2	1.9 ± 2.0	2.4 ± 2.2	1.6 ± 1.8	2.8 ± 2.4	0.21
BILD, mean ± SD	2.1 ± 2.3	2.1 ± 2.3	2.3 ± 2.6	2.4 ± 2.3	1.8 ± 2.0	3.1 ± 2.0	0.09
SLAQ, mean ± SD	8.8 ± 7.3	9.4 ± 7.5	9.2 ± 7.2	11.3 ± 7.8	7.0 ± 6.6	12.4 ± 8.6	0.0007
SLE activity (0–10 rating), mean ± SD	3.3 ± 2.7	3.1 ± 2.7	3.7 ± 2.6	4.3 ± 3.0	2.7 ± 2.5	4.5 ± 2.6	0.001

* Values are the number (%) of patients unless indicated otherwise. GC = glucocorticoid; SLEDAI = Systemic Lupus Erythematosus Disease Activity Index; SDI = Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index; BILD = Brief Index of Lupus Damage; SLAQ = Systemic Lupus Activity Questionnaire; SLE = systemic lupus erythematosus.

ences among the groups remained. Multiple regression analyses were repeated using self-reported disease activity (SLAQ) and damage (BILD) so that telephone-only participants could be included in the analysis. Individuals categorized as other race/ethnicity were omitted from the race- or ethnic-stratified analyses because of the small number (n = 13). All analyses were performed with SAS, version 9.4.

RESULTS

Descriptive, total sample. Descriptive characteristics of the CLUES sample (n = 431) are shown in Table 2. Approximately 90% were female, the mean age was 46.6 years, 19.5% had household incomes below the poverty level, and 22.1% had education at the high school level or lower. Thirty percent were white, 22.3% Hispanic, 10.9% African American, 33.7% Asian/Pacific Islander, and 3.0% other race/ethnicity.

Missing item data were greatest in the satisfaction with social roles scale, ranging from 1.1% of items for Hispanic Spanish-speaking respondents to 4.9% for white respondents (Table 3). Specific items with the greatest number of missing values queried were satisfaction with the amount of work one could do (9.0% missing), ability to work (6.6% missing), and ability to meet the needs of those who depend on the respondent (3.8% missing). The item in the ability to participate in social roles scale, which deals with work, also had a relatively large number of missing responses (5.9%). The 2 psychosocial impact of illness scales and the social isolation scales also had a relatively large number of missing item data, ranging from 0.5% to 2.8% of items, although missingness was not concentrated on specific items.

Mean scores of all PROMIS measures were within one-half SD of the population mean (50.0) (Table 4). Floor effects (worst scores) were minimal, with the largest effect seen for fatigue (5.4%). In contrast, more than 20% of the cohort scored at the ceiling (best scores) for physical function, positive psychosocial impact of illness, and all 4 social health scales.

Ninety-nine of the CLUES participants completed the PROMIS measures by telephone. The telephone completion group was older (mean age of participants via phone of 52 years versus the mean age of those in-person of 45 years; $P < 0.0001$), more likely to complete the interview in English (95% versus 86%; $P = 0.02$), and had longer disease duration (23 years versus 16 years; $P < 0.0001$) and higher BILD scores (2.6 versus 1.9; $P = 0.03$). There were no other significant differences between the 2 groups in sex, race/ethnicity, education, income, SLAQ, or PROMIS scale scores (data not shown).

Reliability and validity, total sample. All item-total correlations were >0.50, except for 3 individual items, which were all >0.40 (Table 4). Cronbach's alpha was acceptable across domains: ≥0.80 for all scales except negative psychosocial impact ($\alpha = 0.78$) and positive psychosocial impact ($\alpha = 0.79$).

PROMIS scores demonstrated moderate to high correlations with SF-36 scores that measured similar constructs

Table 3. Missing data in PROMIS items and scale scores in CLUES cohort by race/ethnicity*

	White (n = 130)	Hispanic, Spanish (n = 25)	Hispanic, English (n = 71)	African American (n = 47)	Asian, Chinese (n = 27)	Asian, English (n = 118)
Physical health						
Physical function (10 items)						
Items, no.	1	1	2	0	1	0
Items, %	0.1	0.4	0.3	0	0.4	0
Scale scores, no.	0	0	0	0	0	0
Pain interference (4 items)†						
Items, no.	1	1	0	6	5	0
Items, %	0.2	1.0	0	3.2	3.1	0
Scale scores, no.	1	0	0	4	0	0
Fatigue (4 items)†						
Items, no.	0	0	0	2	5	4
Items, %	0	0	0	1.1	2.7	0.9
Scale scores, no.	0	0	0	1	0	3
Sleep disturbance (4 items)						
Items, no.	1	0	0	0	1	1
Items, %	0.2	0	0	0	0.9	0.2
Scale scores, no.	1	0	0	0	1	1
Sleep impairment (8 items)						
Items, no.	2	0	1	3	–	1
Items, %	0.2	0	0.2	0.8	–	0.1
Scale scores, no.	0	0	0	0	–	0
Mental health						
Cognitive (4 items)						
Items, no.	0	1	0	0		1
Items, %	0	1.0	0	0	–	0.2
Scale scores, no.	0	1	0	0	–	1
Psychosocial impact, neg. (4 items scored)						
Items, no.	12	–	8	1	–	13
Items, %	2.3	–	2.8	0.5	–	2.8
Scale scores, no.	6	–	5	1	–	8
Psychosocial impact, pos. (4 items scored)						
Items, no.	13	–	8	3	–	11
Items, %	2.5	–	2.8	1.6	–	2.3
Scale scores, no.	7	–	5	3	–	7
Social health						
Ability to participate social roles, activities (4 items)						
Items, no.	16	0	8	9	–	6
Items, %	3.1	0	2.8	4.8	–	1.2
Scale scores, no.	10	0	5	6	–	6
Satisfaction, discretionary social activities (7 items)						
Items, no.	16	0	0	9	–	2
Items, %	1.8	0	0	2.7	–	0.2
Scale scores, no.	2	0	0	1	–	0
Satisfaction social roles (7 items)						
Items, no.	44	2	13	14	–	26
Items, %	4.9	1.1	2.6	4.3	–	3.2
Scale scores, no.	3	1	0	1	–	1
Social isolation (4 items)						
Items, no.	9	1	1	0	–	3
Items, %	1.7	1.0	0.4	0	–	0.6
Scale scores, no.	3	1	1	0	–	2

* Other race/ethnicity excluded because of small sample size. The percentage of items missing is calculated as (number of items in scale with missing responses) / (number of items in scale × number of respondents). PROMIS = Patient-Reported Outcomes Measurement Information System; CLUES = California Lupus Epidemiology Study.
† In the Chinese version of the PROMIS short form, the number of items for pain interference = 6 and the number of items for fatigue = 7.

Table 4. PROMIS scale T-score characteristics in CLUES cohort*

	Mean ± SD	At floor (%)†	At ceiling (%)‡	Cronbach's alpha
Physical health				
Physical function	47.4 ± 9.9	0.2	20.2	0.94
Pain interference	52.4 ± 10.0	0.5	3.3	0.95
Fatigue	52.5 ± 11.6	5.4	0.5	0.96
Sleep disturbance	52.7 ± 9.1	2.3	3.7	0.80
Sleep impairment	52.9 ± 10.7	0.3	4.5	0.92
Mental health				
Cognition, ability	48.7 ± 8.5	2.2	14.7	0.90
Psychosocial illness impact, neg.	52.1 ± 8.2	0.3	12.3	0.78
Psychosocial illness impact, pos.	48.2 ± 9.1	0.6	21.7	0.79
Social health				
Ability to participate social roles, activities	50.5 ± 10.0	3.2	24.8	0.96
Satisfaction, discretionary social activities	52.8 ± 10.0	1.8	20.0	0.95
Satisfaction, social roles	51.1 ± 10.7	2.3	24.1	0.96
Social isolation	46.3 ± 9.4	0.5	27.5	0.90

* For Patient-Reported Outcomes Measurement Information System (PROMIS) scales, higher T scores reflect more of the domain being measured (i.e., better physical function, more pain interference). Because the directionality of scores is not consistent in terms of best or worst scores, floor and ceiling were defined to have a consistent meaning. All item-total correlations were r > 0.50 except physical function (limitations in vigorous activity [r = 0.48]), sleep impairment ("I felt alert when I woke up" [r = 0.39]), and PIN ("I worry about the future" [r = 0.44]). neg. = negative; pos. = positive.
† Floor = worst score.
‡ Ceiling = best score.

(Table 5). The highest correlations were noted for the scales with the most similar content (i.e., physical function, pain, and fatigue). PROMIS scores had moderate correlations with patient-reported disease activity (SLAQ), and low correlations with patient-reported disease damage (BILD) (Table 5). However, there were no associations between PROMIS measures and physician-assessed disease activity (SLEDAI), and only minimal associations with physician-assessed disease damage (SDI).

Descriptive, by racial/ethnic group. White participants were significantly older, with longer disease duration, and were significantly less likely to have below-poverty incomes or low education (Table 2). There were no significant differences among groups in the physician-assessed measures of disease activity and damage. In contrast, Asian patients had significantly lower scores on the self-reported disease activity measures. Twenty-six percent of the Hispanic participants and 19% of the Asian participants completed the PROMIS measures in Spanish or Chinese (Cantonese or Mandarin), respectively. There were no significant differences in PROMIS scores by language for these 2 groups (data not shown). As noted in Table 1, only 10 of 12 PROMIS measures were available in Spanish and only 4 in Chinese languages.

Psychometric analysis, by race/ethnicity. There were no appreciable racial/ethnic differences in the percentage of scores at the floor (see Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23797/abstract). However, racial/ethnic differences were apparent in the percentage of ceiling responses, with the most notable pattern being a lower percentage of African American patients at the ceiling for a number of scales. All Cronbach's α were ≥0.80 when examined by racial/ethnic group, except for sleep disturbance (whites, African Americans), psychosocial impact of illness negative (Hispanics, African Americans, Asians), and psychosocial impact of illness positive (African Americans, Asians); each of these α coefficients was ≥0.70. Item-total correlations were ≥0.50 for all groups, with a few exceptions, most notably the physical function item regarding limitations in vigorous activity, the sleep impairment item regarding feeling alert upon awakening, and the psychosocial impact of illness item regarding worry about the future.

There were no substantive differences in correlations with SF-36 scales by race/ethnicity (see Supplementary Table 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23797/abstract). As with the total cohort, correlation coefficients of PROMIS scores with physician-assessed disease activity were minimal for all groups, and correlations with patient-reported disease activity were generally moderate. Among white participants only, correlations with SDI and BILD were significant for almost all PROMIS scales, although most correlations were small. No consistent patterns of differences among racial/ethnic groups were noted (see Supplementary Table 3, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.23797/abstract).

Differences in scores by race/ethnicity. In bivariate analyses, Asians had significantly better scores than whites for

Table 5. PROMIS scale correlations with SF-36 and measures of disease activity and damage*

	Patient-reported measures										Physician-reported measures	
PROMIS scales	SF-36 PF	SF-36 Pain	SF-36 Vit	SF-36 SF	SF-36 RP	SF-36 RE	SF-36 MH	SLAQ†	Lupus activity (0–10)	BILD	SLEDAI	SDI
Physical health												
Physical function	0.94‡							−0.60‡	−0.48‡	−0.38‡	0.09	−0.32‡
Pain interference		−0.79‡						0.65‡	0.56‡	0.28‡	−0.05	0.19§
Fatigue			−0.80‡					0.67‡	0.56‡	0.17§	−0.03	0.13
Sleep disturbance			−0.55‡					0.44	0.37‡	0.08	−0.01	0.01
Sleep impairment			−0.78‡					0.60‡	0.51‡	0.16§	−0.03	0.07
Mental health												
Cognition, ability						0.41‡	0.42‡	−0.44‡	−0.35‡	−0.20‡	−0.01	−0.14§
Psychosocial illness impact, neg.						−0.43‡	−0.60‡	0.41‡	0.34‡	0.12§	−0.03	0.08
Psychosocial illness impact, pos.						0.47‡	0.53‡	−0.38‡	−0.29‡	−0.15§	−0.01	−0.10
Social health												
Ability to participate social roles, activities				0.51‡	0.72‡			−0.59‡	−0.48‡	−0.27‡	0.05	−0.22§
Satisfaction, discretionary social activities				0.67‡	0.68‡			−0.58‡	−0.49‡	−0.22‡	0.01	−0.20§
Satisfaction, social roles				0.66‡	0.74‡			−0.58‡	−0.48‡	−0.23‡	0.01	−0.19§
Social isolation				−0.53‡	−0.51‡			0.45‡	0.34‡	0.18§	−0.04	0.10

* For Patient-Reported Outcomes Measurement Information System (PROMIS) scales, higher T-scores reflect more of the domain being measured (i.e., better physical function, more pain interference). Short Form 36 (SF-36) scales include physical function (PF), pain, vitality (Vit), social functioning (SF), role physical (RP), role emotional (RE), and mental health (MH). SLAQ = Systemic Lupus Activity Questionnaire; BILD = Brief Index of Lupus Damage; SLEDAI = Systemic Lupus Erythematosus Disease Activity Index; SDI = Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index; neg. = negative; pos. = positive.
† From SLAQ, rating of lupus over the past 3 months (range 0–10, where 0 = no activity and 10 = high activity).
‡ $P < 0.0001$.
§ $P < 0.05$.

PROMIS physical function, pain interference, fatigue, ability to participate in social roles and activities, satisfaction with discretionary activities, and satisfaction with social roles (Table 6). African Americans had significantly worse physical function scores. In analyses adjusting for age, sex, disease duration, SLEDAI, and SDI, differences between whites and other race/ethnicity groups were seen only for fatigue and satisfaction with discretionary activities. Again, Asians had significantly better scores than whites for each of these scales (Table 6, Model 1). After further adjustment for obesity and smoking, these differences remained (Table 6, Model 2). Further adjustment for income did not change results substantively (data not shown). In sensitivity analyses using patient-reported disease activity and damage instead of SLEDAI and SDI in order to include subjects who participated by phone only, similar results were noted (Table 6, Model 3).

DISCUSSION

To our knowledge, this study shows the first examination of PROMIS short forms across different racial/ethnic and language groups in a diverse lupus cohort. Overall, each of the scales demonstrated adequate reliability (internal consistency) and validity (correlations with similar measures). Minimal floor effects were observed, but ceiling effects were noted, particularly in social health measures, which could limit responsiveness to change. Missing item responses and resulting missing scale scores were minimal and random for most scales. The notable exceptions were the satisfaction with social roles, participation in social roles and activities, and the 2 psychosocial impact of illness scales; for the first 2 of these, items dealing with work accounted for the majority of missing item responses. It is possible that individuals who were not working felt these items were not applicable to them. For the psychosocial impact of illness scales, there were no clear patterns of missing items. Item-total correlations were lower than optimal for a few items, including ability to engage in vigorous physical activities, feeling alert upon awakening, and worry about the future. The relatively poor performance of these items within their respective scales may reflect lupus-specific biases; i.e., the domains addressed by these items may be affected by lupus in a manner different than that in the general population. Further work is needed to determine the underlying reasons for these anomalies and the usefulness of these items among

Table 6. PROMIS scores in CLUES cohort by race/ethnicity*

	Unadjusted				Model 1†			Model 2‡			Model 3§		
	White (n = 127)	Hispanic (n = 96)	AA (n = 47)	Asian (n = 145)	Hispanic	AA	Asian	Hispanic	AA	Asian	Hispanic	AA	Asian
Physical health													
Physical function	47.2 ± 10.2	47.6 ± 10.0	42.1 ± 9.7¶	49.7 ± 8.9¶	–1.5	–3.3	0.7	–0.8	–1.5	–0.2	–0.5	–3.2¶	–0.5
Pain interference	52.8 ± 9.8	53.1 ± 9.6	55.7 ± 10.3	50.0 ± 9.7¶	0.9	1.3	–2.6	0.3	–0.3	–1.8	–0.0	1.3	–1.0
Fatigue	54.4 ± 11.8	52.1 ± 11.3	55.4 ± 10.4	49.4 ± 11.2¶	–2.5	0.4	–5.0¶	–3.2	–1.7	–4.3¶	–3.7¶	–2.1	–4.0¶
Sleep disturbance	52.2 ± 8.6	54.2 ± 9.3	53.9 ± 8.4	51.6 ± 9.4	1.5	0.9	–0.7	1.0	–0.4	–0.3	1.7	0.1	0.3
Sleep impairment	53.3 ± 11.1	52.3 ± 10.7	54.4 ± 10.1	51.8 ± 10.4	–1.6	–0.3	–2.1	–2.3	–2.4	–1.5	–2.4	–1.5	–0.7
Mental health													
Cognition, ability	48.9 ± 9.3	48.6 ± 8.1	46.8 ± 6.7	49.3 ± 8.6	0.1	1.0	1.1	0.6	2.5	0.6	–0.2	–0.5	–0.6
Psychosocial illness impact, negative	52.7 ± 8.1	51.5 ± 8.0	52.3 ± 9.5	51.6 ± 7.9	–2.2	–0.2	–1.5	–2.2	–1.2	–1.3	–1.9	–1.8	–1.0
Psychosocial illness impact, positive	48.8 ± 8.9	47.2 ± 10.0	46.0 ± 9.0	49.0 ± 9.0	–1.6	–1.3	–0.1	–1.5	–0.2	–0.5	–2.3	–1.8	–1.3
Social health													
Ability to participate in social roles, activities	49.6 ± 10.2	51.2 ± 10.4	47.1 ± 8.6	52.9 ± 9.4¶	2.2	1.1	2.6	2.8	2.4	2.1	2.2	0.6	1.1
Satisfaction, discretionary	51.3 ± 10.6	52.5 ± 10.0	52.2 ± 9.1	55.7 ± 9.3¶	1.4	3.0	4.4¶	2.0	4.7¶	3.7¶	1.7	3.2¶	2.6¶
Satisfaction, social roles	50.3 ± 10.8	50.8 ± 10.5	49.2 ± 11.0	53.8 ± 10.3¶	0.5	0.5	3.0	1.2	2.5	2.2	0.8	1.1	1.4
Social isolation	46.5 ± 8.9	45.4 ± 10.2	48.3 ± 8.9	45.3 ± 9.3	–1.1	1.6	–1.0	–1.6	0.3	–0.6	–1.7	0.4	–0.2

* Values are the mean ± SD unless indicated otherwise. Values for Models 1, 2, and 3 are from multiple linear regression analysis, and indicate difference from score of whites (i.e., reference group = white). "Other" racial/ethnic group omitted because of small sample (n = 13). AA = African American. See Table 5 for other definitions.
† Model 1 adjusted for age, sex, disease duration, and physician-reported disease activity and damage (SLEDAI and SDI).
‡ Model 2 adjusted for age, sex, disease duration, obesity, smoking, and physician-reported disease activity and damage (SLEDAI and SDI). Further adjustment for income did not change results substantively.
§ Model 3 adjusted for age, sex, disease duration, and patient-reported disease activity and damage (systemic lupus erythematosus activity [0 – 10 rating] and BILD).
¶ $P < 0.05$, statistically significant, compared to whites.

individuals with lupus, particularly those who may be work-disabled.

All of the PROMIS short forms demonstrated consistent reliability and validity across racial/ethnic groups. We found no differences by mode of administration (in-person versus telephone) or by language of administration among the Hispanic and Asian participants. Bivariate analyses showed significant differences in mean scores by race/ethnicity for more than half of the scales. However, after accounting for differences in disease status, age, and sex, few differences remained between whites and other racial/ethnic groups, suggesting that differences in scale scores may be attributable to differences in disease and demographics rather than race/ethnicity (per se).

We found that in our population-based cohort of individuals with lupus, scores were generally reflective of better health than have been reported in a sample of lupus patients recruited from a clinical setting (6). That study also reported fewer ceiling effects, possibly due to the shift toward lower mean scores. The difference between that study and the data reported here may be because individuals are less likely to attend a research visit during episodes of poor health or a flare, while clinical visits are more likely to occur during those times. However, the study of clinic patients also used the CAT versions of the PROMIS scales, so it is possible that the CAT version produces greater precision of item selection and yields fewer ceiling effects, as has been suggested in studies of the PROMIS measures of depressive symptoms (22) and physical function (23).

We found no association of PROMIS scores with physician-assessed disease activity and only minimal correlations with physician-assessed disease damage. A similar lack of correlation with the SLEDAI and low correlation with SDI was noted in a previous study by Kasturi et al (6). However, the lack of correspondence between physician-completed and patient-reported measures in lupus has been well-documented (24), so this finding is not surprising.

Strengths of this study include the diverse cohort, with sufficient sample sizes to examine measures by racial/ethnic group. This study included the largest number of PROMIS short forms administered in a lupus cohort. In addition, administration in multiple languages and in both in-person and phone formats permitted comparisons of these subgroups. Only 4 other studies have been published examining PROMIS scales in lupus. Two examined the PROMIS 29-item profile, which includes 4-item short forms for physical function, fatigue, pain interference, sleep disturbance, satisfaction with social role, anxiety, and depression (7,8), in non-clinical study settings among patients who were primarily white and exclusively English-speaking. The remaining 2 publications were based in a clinical setting, with a more diverse, yet exclusively English-speaking, cohort, but did not examine racial/ethnic differences in scale performance (6,9).

Limitations include the underrepresentation of clinically active lupus, as noted above. While the cohort was quite diverse, the number of non-English speakers was relatively small. Yet, this is the first comparison of PROMIS scores of English and non-English speakers. A limited number of legacy measures was available for validity analyses; however the SF-36 is the PRO most commonly used in lupus studies, including clinical trials. A comparison of PROMIS measures with one of the lupus-specific quality of life measures could provide useful information. All questionnaires were interviewer-administered, so results may have been different if self-administered. However, interviewer administration provided consistency in mode of administration between the in-person and phone interviews.

PROMIS measures offer several advantages to existing PRO measures; in particular the SF-36, which is the most commonly used PRO in lupus studies. There is evidence that the SF-36 does not adequately cover the broad range of symptoms and outcomes important in lupus (5,25). With PROMIS, a broader range of domains can be examined, including domains that are relevant to SLE and meaningful to patients, such as sleep quality, cognitive abilities, impact of pain, and satisfaction with social roles, with a relatively small response burden. Like the SF-36, PROMIS short forms have been translated and culturally adapted to multiple languages. PROMIS measures are available in the public domain and are free to use in clinical trials as well as in registries, observational studies, and clinical practice. PROMIS measures have also been adapted to use within many existing electronic health record systems.

It is yet to be seen if the PROMIS measures are responsive to changes in lupus disease activity or severity. The ceiling effects noted in this study may indicate a limitation in responsiveness, but such limitations may not exist in a sample of respondents with more active disease. Additional studies are needed to determine responsiveness to change and minimal clinically important differences in lupus.

In summary, these results add to the growing evidence supporting PROMIS measures as reliable and valid in lupus. This study adds information regarding the performance of PROMIS measures in lupus across racial/ethnic groups, which is particularly important given the high burden of disease among racial/ethnic minorities. Differences that we observed between racial/ethnic groups appeared to be primarily due to differences in the groups' clinical or sociodemographic characteristics rather than differential scale performance. Overall, the PROMIS measures appear to be well-suited to use in lupus and may be particularly useful as outcome measures in clinical trials of targeted therapies and longitudinal studies.

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. Katz 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. Katz, Yazdany, Trupin, Helmick, Murphy, Criswell, Dall'Era.
Acquisition of data. Yazdany, Trupin, Rush, Lanata, Dall'Era.
Analysis and interpretation of data. Katz.

REFERENCES

1. Yazdany J. Health-related quality of life measurement in adult systemic lupus erythematosus: Lupus Quality of Life (LupusQoL), Systemic Lupus Erythematosus-Specific Quality of Life Questionnaire (SLEQOL), and Systemic Lupus Erythematosus Quality of Life Questionnaire (L-QoL). Arthritis Care Res (Hoboken) 2011;63:S413–9.
2. Bertsias G, Gordon C, Boumpas D. Clinical trials in systemic lupus erythematosus (SLE): lessons from the past as we proceed to the future—the EULAR recommendations for the management of SLE and the use of end-points in clinical trials. Lupus 2008;17:437–42.
3. US Department of Health and Human Services. Guidance for industry. Systemic lupus erythematosus: developing medical products for treatment. June 2010. URL: https://www.fda.gov/media/71150/download.
4. Witter J. The promise of Patient-Reported Outcomes Measurement Information System—turning theory into reality: a uniform approach to patient-reported outcomes across rheumatic disease. Rheum Dis Clin North Am 2016;42:377–94.
5. Stamm TA, Bauernfeind B, Coenen M, Feierl E, Mathis M, Stucki G, et al. Concepts important to persons with systemic lupus erythematosus and their coverage by standard measures of disease activity and health status. Arthritis Rheum 2007;57:1287–95.
6. Kasturi S, Szymonifka J, Burket J, Berman JR, Kirou KA, Levine AB, et al. Validity and reliability of patient reported outcomes measurement information system computerized adaptive tests in systemic lupus erythematosus. J Rheumatol 2017;44:1024–31.
7. Lai J, Beaumont J, Jensen S, Kaiser K, Van Brunt DL, Kao AH, et al. An evaluation of health-related quality of life in patients with systemic lupus erythematosus using PROMIS and Neuro-QoL. Clin Rheumatol 2017;36:555–62.
8. Katz P, Andrews J, Yazdany J, Schmajuk G, Trupin L, Yelin E. Is frailty a relevant concept in SLE? Lupus Sci Med 2017;4:3000186.
9. Kasturi S, Szymonifka J, Burket J, Berman JR, Kirou KA, Levine AB, et al. Feasibility, validity, and reliability of the 10-item Patient Reported Outcomes Measurement Information System Global Health short form in outpatients with systemic lupus erythematosus. J Rheumatol 2018;45:397–404.
10. Dall'Era M, Cisternas MG, Snipes K, Herrinton LJ, Gordon C, Helmick CG. The incidence and prevalence of systemic lupus erythematosus in San Francisco County, California: the California Lupus Surveillance Project. Arthritis Rheum 2017;69:1996–2005.
11. Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1982;25:1271–7.
12. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1997;40:1725.
13. Ware JJ, Snow K, Kosinski M, Gandek B. SF-36 health survey: manual and interpretation guide. Boston: The Health Institute, New England Medical Center; 1993.
14. Karlson E, Daltroy L, Rivest C, Ramsey-Goldman R, Wright EA, Partridge AJ, et al. Validation of a systemic lupus activity questionnaire (SLAQ) for population studies. Lupus 2003;12:280–6.
15. Yazdany J, Yelin EH, Panopalis P, Trupin L, Julian L, Katz PP. Validation of the Systemic Lupus Erythematosus Activity questionnaire in a large observational cohort. Arthritis Rheum 2008;59:136–43.
16. Yazdany J, Trupin L, Gansky S, Dall'era M, Yelin EH, Criswell LA, et al. Brief Index of Lupus Damage: a patient-reported measure of damage in systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2011;63:1170–7.
17. Gladman D, Ginzler E, Goldsmith C, Fortin P, Liang M, Urowitz M, et al. The development and initial validation of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index for systemic lupus erythematosus. Arthritis Rheum 1996;39:363–9.
18. 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. Buyon J, Petri M, Kim M, Kalunian KC, Grossman J, Hahn BH, et al. The effect of combined estrogen and progesterone hormone replacement therapy on disease activity in systemic lupus erythematosus: a randomized trial. Ann Intern Med 2005;142:953–62.
20. Nunnally J, Bernstein I. Psychometric theory. 3rd ed. New York: McGraw-Hill; 1994.
21. Hinkle D, Wiersma W, Jurs S. Applied statistics for the behavioral sciences. 5th ed. Boston: Houghton Mifflin; 2003.
22. Choi S, Reise S, Pilkonis P, Hays R, Cella D. Efficiency of static and computer adaptive short form forms compared to full-length measures of depressive symptoms. Qual Life Res 2010;19:125–36.
23. Rose M, Bjorner J, Gandek B, Bruce B, Fries J, Ware J Jr. The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. J Clin Epidemiol 2014;67:516–26.
24. Bertsias G, Ioannidis J, Boletis J, Bombardieri S, Cervera R, Dostal C, et al. EULAR points to consider for conducting clinical trials in systemic lupus erythematosus: literature based evidence for the selection of endpoints. Ann Rheum Dis 2009;68:477–83.
25. Ow YL, Thumboo J, Cella D, Cheung YB, Fong KY, Wee HL. Domains of health-related quality of life important and relevant to multiethnic English-speaking Asian systemic lupus erythematosus patients: a focus group study. Arthritis Care Res (Hoboken) 2011;63:899–908.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 71, No. 12, December 2019, pp 1640–1646
DOI 10.1002/acr.23833

Metabolic and Structural Skeletal Muscle Health in Systemic Lupus Erythematosus–Related Fatigue: A Multimodal Magnetic Resonance Imaging Study

Sai Man Cheung,[1] Karen Keenan,[1] Nicholas Senn,[1] Gayle Hutcheon,[2] Kwokshing Chan,[1] Lars Erwig,[3] Andrew Schrepf,[4] Paula Dospinescu,[2] Stuart Gray,[5] Gordon Waiter,[1] Jiabao He,[1] and Neil Basu[5]

Objective. To investigate the potential structural and metabolic role of skeletal muscle in systemic lupus erythematosus (SLE)–related fatigue.

Methods. A case–control, multimodal magnetic resonance imaging (MRI) study was conducted. Cases were patients with inactive SLE who reported chronic fatigue. Controls were age- and sex-matched healthy members of the general population. Patients were clinically characterized and then underwent a 3T whole-body MRI scan. Resting and dynamic ^{31}P MRI spectroscopy of the calf muscles was applied, from which phosphocreatine (PCr) recovery halftime, a marker of mitochondrial dysfunction, was computed. In addition, microstructural sequences (T1-weighted anatomic images, T2 mapping, and diffusion tensor imaging) were acquired. Descriptive statistics evaluated group differences and within-case physical fatigue correlations were explored.

Results. Of the 37 recruits (mean age 43.8 years, 89.2% female), cases (n = 19) reported higher levels of physical fatigue, pain, depression, and sleep disturbance compared to the control group ($P < 0.0001$). PCr was greater ($P = 0.045$) among cases (mean ± SD 33.0 ± 9.0 seconds) compared to controls (mean ± SD 27.1 ± 6.6 seconds). No microstructural group differences were observed. Within cases, physical fatigue did not correlate with PCr ($r = –0.28$, $P = 0.25$).

Conclusion. We report preliminary data demonstrating greater skeletal muscle mitochondrial dysfunction among fatigued patients with SLE compared to healthy controls.

INTRODUCTION

Patients with systemic lupus erythematosus (SLE) consider fatigue to be one of the most pervasive and disabling aspects of their disease. As many as 85% of patients report significant levels of fatigue (1), a prevalence greater than that observed in the general population and among patients with more common inflammatory rheumatic disorders (2). Moreover, the impact of fatigue permeates all aspects of life, as reflected by its strong associations with impaired quality of life (3) and work disability (4). Despite these significant consequences, little is understood about this symptom. The major challenge in clinical practice is to deliver therapeutic options to those patients whose disease is otherwise in remission and for whom no other reversible causes are apparent (5).

Patients describe multiple dimensions of fatigue, and therefore its etiology is likely to be complex. The predominance of both physical and mental fatigue (6) alludes to a mixture of peripheral and central mechanisms. In terms of investigating the former, skeletal muscle dysfunction has previously been associated with SLE-related fatigue (7), although no studies have investigated whether this observation is underpinned by pathologic abnormalities within the muscles themselves.

Developments in magnetic resonance imaging (MRI) technology offer a noninvasive opportunity to comprehensively quantify skeletal muscle pathology at both metabolic and structural levels. For example, ^{31}P MRI spectroscopy (MRS) allows for the direct measurement of altered metabolic activity, such as levels of phosphocreatine (PCr), in vivo during physical activity, and MRS has previously signaled dysfunction in the muscles of chronic fatigue

Supported by Lupus UK.

[1]Sai Man Cheung, MSc, Karen Keenan, PhD, Nicholas Senn, MSc, Kwokshing Chan, MSc, Gordon Waiter, PhD, Jiabao He, PhD: University of Aberdeen, Aberdeen, UK; [2]Gayle Hutcheon, BSc, Paula Dospinescu, MD: NHS Grampian, Aberdeen, UK; [3]Lars Erwig, MD, PhD: University of Aberdeen, Aberdeen, and GlaxoSmithKline, Stevenage, UK; [4]Andrew Schrepf, PhD: University of Michigan, Ann Arbor; [5]Stuart Gray, PhD, Neil Basu, MD, PhD: University of Glasgow, Glasgow, UK.

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

Address correspondence to Neil Basu, MD, PhD, Institute of Infection, Immunology and Inflammation, University of Glasgow, Glasgow, G12 8TA, UK. E-mail: neil.basu@glasgow.ac.uk.

Submitted for publication May 31, 2018; accepted in revised form January 8, 2019.

SIGNIFICANCE & INNOVATIONS

- Systemic lupus erythematosus (SLE)–related fatigue does not appear to be implicated with abnormal skeletal muscle microstructure.
- Patients with SLE exhibit higher levels of skeletal muscle mitochondrial dysfunction.

syndrome (CFS) populations (8). In contrast to CFS, there is some histologic evidence that at least selected patients with SLE exhibit structural abnormalities in their skeletal muscle (9). Novel methods, such as diffusion tensor imaging (DTI), are sensitive to pathologic abnormalities associated with overall cell geometry and edema (10,11). In addition to high-resolution MRI for the quantification of muscle volume, T2 mapping highlights edema, while Dixon MRI allows quantitative measurement of fat infiltration. To our knowledge, no study has yet to contemporaneously use these methodologic advances to investigate SLE. We aimed to investigate the differences between the metabolic and structural features of skeletal muscle among patients with SLE with idiopathic fatigue and healthy controls using multimodal MRI muscle imaging.

PATIENTS AND METHODS

A case–control study was conducted. Subjects were invited to undertake a multimodal MRI scan of their calf muscles alongside the collection of clinical data. The East Midlands–Leicester Research Ethics Committee (ref: 15/EM/0418) approved the study, and written informed consent was obtained from patients according to the Declaration of Helsinki.

Patients. Cases were patients with SLE, classified according to the 1997 American College of Rheumatology criteria (12), attending secondary care clinics in NHS Grampian. To be eligible, patients were required to report chronic (>3 months), clinically important fatigue (defined as a score of >3 on the Chalder Fatigue Scale [binary scoring]) (13), experience reduced muscle strength (item 6 of the Chalder Fatigue Scale), and have inactive SLE, defined as a British Isles Lupus Assessment Group 2004 score of 0 (excluding the fatigue constitutional domain) (14). In addition, patients were excluded if they had any past history of clinically diagnosed myositis or alternative medical explanations for their fatigue (symptomatic cardiorespiratory disease, a history of cancer in the previous 5 years, unstable thyroid disease, moderate-to-severe chronic kidney disease, moderate-to-severe anemia, a beta-blocker prescription, or fibromyalgia).

Controls, recruited by local advertising, were healthy (no relevant past medical history) subjects who did not report clinically important fatigue (Chalder Fatigue Scale ≤3) or reduced muscle strength (item 6 of the Chalder Fatigue Scale). They were approximately matched to cases by age and sex. In order to offset potential confounding due to deconditioning, controls were additionally required to be sedentary, defined as those having a desk job and undertaking <3 hours of physical activity per week (8). Any potential case or control with a contraindication to MRI (e.g., pacemaker in situ) was excluded.

Clinical characterization. Eligible cases underwent a clinical evaluation that included an assessment of disease damage (according to Systemic Lupus International Collaborating Clinics [SLICC] criteria) (15), previous organ involvement, and disease duration. Erythrocyte sedimentation rate, serum creatinine, and creatine kinase were measured in both cases and controls. All subjects completed a self-reported questionnaire that included the following validated measures and known confounders of fatigue: 1) the Chalder Fatigue Scale is one of the most commonly employed measures of fatigue and has been found to be both feasible and acceptable in SLE (16); of the 11 questions, 7 specifically examine physical fatigue and are scored on a Likert scale (range 0–21), with high scores indicating high levels of physical fatigue; 2) the Hospital Anxiety and Depression Scale is a validated 14-item tool for assessing anxiety and depression in patients with SLE and in the general population (17); this scale also employs a Likert-style scoring system (range 0–21 for each domain); 3) pain severity was measured using a 0–10 numerical rating scale; and 4) Jenkin's Sleep Scale is known to perform well in both nonclinical and clinical populations, succinctly quantifying key sleep dysfunction domains, i.e., difficulties in sleep onset and maintenance, early wakening, and nonrestorative sleep; the domain scores are totaled (range 0–20), with higher scores indicating greater sleep disturbance (18).

Finally, both cases and controls underwent the Siconolfi Step Test. This measure of aerobic fitness (a putative confounder) has been validated in patients with SLE (19). It involved patients stepping up and down from a 10-inch bench for 3 minutes at a rate of 17 steps per minute (guided by a metronome). Heart rate was monitored and the protocol stopped if 65% of the predicted heart rate (220 minus age) was exceeded. If not reached, then a second stage (26 steps per minute) and a third stage (34 steps per minute) were performed, with 1-minute rest between stages. Maximal oxygen uptake was then estimated using the formulas described by Siconolfi et al (20).

MRI acquisition. Images were acquired on a 3T whole-body MRI scanner (Achieva TX, Philips Healthcare) using the body coil for transmission and an 8-channel knee coil as the receiver. In 1 patient with SLE and 2 healthy controls, the diameter of the lower leg was too large for the knee coil, and for these participants a 2-channel flex-M receiver coil was used. The imaging volume was centered at the thickest part of the right calf, with the subject in supine position.

T1-weighted anatomic images were acquired using a standard sequence with repetition time (TR) of 2,700 msec, echo time (TE) of 55 msec, field of view (FOV) of 160 × 160 mm^2, matrix size of 160 × 160, and 48 slices of 1.5-mm thickness. T2 mapping was performed using a gradient and spin-echo sequence with TR of 3,137 msec, and with 12 equally spaced echoes from TE of 10–120 msec (21,22). DTI was acquired using a single-shot pulsed gradient spin-echo sequence with TR/TE of 2,000/53 msec (11,23), 32 diffusion directions, diffusion weighting of 400 seconds/mm^2, and 2 averages (24). Fat mapping was performed using a multislice multi-echo spoiled gradient-echo sequence, with TR of 20 msec, 16 equally spaced echoes from TE of 1.14 to 18.24 msec, and 3° flip angle (25,26). For T2 mapping, DTI, and fat mapping, the imaging volume was set to FOV of 192 × 192 mm^2, and 12 transverse slices of 6-mm thickness. The matrix size was set to 128 × 128 for T2 and fat mapping, and 64 × 64 for DTI, to ensure adequate signal-to-noise ratio (24).

^{31}P-MRS scans were acquired from a 14-cm diameter ^{31}P coil positioned underneath the thickest part of the calf, using a 1-D image-selected in vivo spectroscopy sequence with the detection slab covering the posterior portion of the calf (27). Dynamic spectra were acquired with TR of 5 seconds and 108 dynamics (28), while the subject concurrently performed a plantar flexion exercise protocol at 20% maximal voluntary contraction in synchrony to an audio metronome prompt at 35 beats per minute. Isometric maximal voluntary contraction of the right calf with 90° plantar ankle flexion for each subject was measured (KinCom 500H dynamometer). Measurements at 2-minute intervals were performed until the difference between the last 2 measures was >5% of their average; normally 3–4 repetitions were performed. The highest of the last 2 measures was taken as the maximal voluntary contraction (29). The exercise paradigm contained a 2-minute baseline followed by 2 8-minute cycles, where each cycle was composed of 3 minutes of exercise before a 5-minute recovery period.

Image analysis. The ^{31}P spectra were processed in jMRUI software, version 3.0 (30), and PCr halftime was computed from the PCr time course in the postexercise recovery period as an indicator of muscle energetics (31). DTI analysis was performed in FSL software FMRIB, to derive metrics maps of mean diffusion (MD), radial diffusivity (RD), and fractional anisotropy (FA) (23) as indicators of muscle integrity (24). In 5 subjects, images affected by motion artefact, resulting in failure of the motion correction algorithm, were identified and removed before the calculation of diffusion metrics. T2 maps were computed using in-house software in MATLAB (MathWorks), following standard procedures (21). Fat fraction maps, as the ratio between fat and the sum of fat and water images, were computed using the ISMRM Fat-Water Toolbox in MATLAB (32). Fat and water were separated using a multistep fitting approach (33), incorporating a multifrequency fat-spectrum model (34,35). To avoid confounding factors, patients with nonadherence to the exercise protocol were excluded from the PCr halftime analysis; patients using the flex-M coil or showing severe image artefact were excluded from corresponding image analysis (Table 1).

Regions of interest were manually drawn by a single operator in MRIcron on the central 10 slices of the image acquired at a TE of 10 msec from T2 mapping, to delineate soleus and exclude subcutaneous fat or blood vessels. The binary masks were subsequently applied on maps of T2, MD, FA, RD, and fat

Table 1. Magnetic resonance imaging results*

Result	SLE	Healthy controls	*t*-test score†	*P*
Metabolism‡				
PCr halftime, seconds	33.0 ± 9.0	27.1 ± 6.6	2.087	0.045§
End-exercise pH	7.00 ± 0.01	7.01 ± 0.01	0.704	0.488
Muscle integrity¶				
MD ($\times 10^{-3}$ mm^2 $seconds^{-1}$)	1.57 ± 0.07	1.54 ± 0.12	0.850	0.401
RD ($\times 10^{-3}$ mm^2 $seconds^{-1}$)	1.39 ± 0.07	1.38 ± 0.11	0.597	0.554
FA	0.21 ± 0.02	0.21 ± 0.02	1.212	0.234
Muscle condition#				
T2, msec	33.2 ± 1.5	32.6 ± 1.1	1.355	0.185
Fat infiltration**				
Fat fraction, %	3.69 ± 1.27	3.90 ± 1.81	−0.381	0.706
Size, CSA cm^2	21.8 ± 3.7	22.6 ± 5.4	−0.497	0.623

* Values are the mean ± SD unless indicated otherwise. SLE = systemic lupus erythematosus; PCr = phosphocreatine; MD = mean diffusivity; RD = radial diffusivity; FA = fractional anisotropy; CSA = cross-sectional area.
† Independent sample.
‡ One case and 1 control not analyzed due to exercise nonadherence. Two cases and 1 control not analyzed due to an artifact in the recovery curve.
§ Statistically significant.
¶ One control not analyzed due to image artifact.
# Two cases and 2 controls not analyzed due to image artifacts.
** One case and 2 controls not analyzed due to flex-M coil use. One case not analyzed due to image artifact.

Table 2. Baseline characteristics*

Characteristic	Cases (n = 19)	Controls (n = 18)	*P*†
Demographics			
Age, years	44.8 ± 14.43	42.8 ± 13.6	0.67
Female, no.	17	16	0.95‡
Symptoms			
Physical fatigue (CFS)	14.7 ± 3.6	6.9 ± 0.6	<0.0001
Anxiety (HADS)	9.3 ± 4.2	4.3 ± 2.4	0.0001
Depression (HADS)	6.7 ± 3.4	1.6 ± 1.7	<0.0001
Pain (NRS; range 0–10)	3.5 ± 2.3	0.3 ± 0.8	<0.0001
Sleep disturbance (JSS)	12.7 ± 5.3	4.8 ± 5.3	<0.0001
Physiologic measures			
Vo_2 max (ml/kg/minute)§	28.0 ± 4.4	28.4 ± 6.0	0.78
ESR (mm/hour)	18.7 ± 14.2	13.6 ± 10.4	0.28
Hemoglobin (gram/liter)	132.6 ± 11.2	131.1 ± 6.9	0.67
Creatinine (μmoles/liter)	69.7 ± 24.0	64.2 ± 13.7	0.39
Creatinine kinase (U/liter)	89.4 ± 34.2	113.8 ± 71.3	0.20

* Values are the mean ± SD unless indicated otherwise. CFS = Chalder Fatigue Scale (physical domain); HADS = Hospital Anxiety and Depression Scale; NRS = numeric rating scale; JSS = Jenkin's Sleep Scale; ESR = erythrocyte sedimentation rate.
† Derived from *t*-tests unless indicated otherwise.
‡ Derived from chi-square test.
§ Derived from Siconolfi Step Test.

fraction to generate the average value. Muscle volumes of soleus were also quantified as the cross-sectional area on the central slice of the T1-weighted anatomic image (36). Eighteen subjects per group sufficiently afforded >80% power to detect an effect size of 0.85, with a measurement error of 30% at an alpha of 0.05 (as measured by PCr recovery halftime).

Statistical analysis. Clinical parameters were expressed using simple descriptive statistics with case–control comparisons made using chi-square tests for categorical variables and *t*-tests for continuous variables. To investigate the role of skeletal muscle energetics in SLE, the case–control comparison of PCr halftime was performed using a *t*-test. To examine the role of muscle microstructure integrity and muscle volume in SLE, the case–control comparison of MD, FA, RD, T2, and fat fraction, as well as the cross-sectional area, was performed using *t*-tests. Within-case Pearson correlations were conducted using STATA software, version 12.1, to further investigate any identified group differences. Due to the small sample size, these analyses were considered exploratory.

RESULTS

Among the 37 recruited subjects (mean age 43.8 years, 89.2% female), cases (n = 19) reported significantly higher levels of physical fatigue, pain, depression, anxiety, and sleep disturbance compared to the control group, although the groups were comparable in terms of demographic and physiologic parameters (Table 2).

Overall, cases had mild SLE; only 1 patient had a history of renal involvement, and the mean ± SD SLICC score was 0.11 ± 0.3. The majority of patients experienced musculoskeletal (n = 17) and/or cutaneous (n = 12) involvement. The most commonly prescribed immunosuppressant treatment was hydroxychloroquine (n = 15), followed by methotrexate (n = 8). Other SLE-specific treatment at the time of the study included azathioprine (n = 2), mycophenolate mofetil (n = 3), and rituximab (n = 3). Only 3 patients were receiving long-term prednisolone (5–8 mg/day).

MRI analysis. In assessment of calf muscle metabolic function, there was a difference ($P = 0.045$) in the PCr halftime recovery between patients with SLE (mean ± SD 33.0 ± 9.0 seconds) and healthy controls (mean ± SD 27.1 ± 6.6 seconds). There were no significant differences in MD, RD, or FA from DTI between patients with SLE and controls (Table 1). Additionally, there were no significant differences in T2 ($P = 0.185$) or fat fraction ($P = 0.706$) between patients with SLE (mean ± SD T2 33.2 ± 1.5 msec; fat fraction 3.69 ± 1.27%) and controls (mean ± SD T2 32.6 ± 1.1 msec; fat fraction 3.90 ± 1.81%) or in muscle cross-sectional areas ($P = 0.623$) (mean ± SD SLE 21.8 ± 3.7 cm^2; controls 22.6 ± 5.4 cm^2). The MRI data from a healthy control are shown in Figure 1. There were no significant correlations identified between PCr halftime and levels of physical fatigue (r = –0.28 [95% confidence interval (95% CI) –0.60, 0.13], $P = 0.25$), or mental fatigue (r = 0.2 [95% CI –0.2, –0.54], $P = 0.41$).

DISCUSSION

To the best of our knowledge, this is the first study of a rheumatic disease to investigate the relationship between skeletal muscle and fatigue employing multimodal MRI. Among fatigued patients with SLE, calf muscle PCr recovery halftime was significantly prolonged compared to nonfatigued healthy controls.

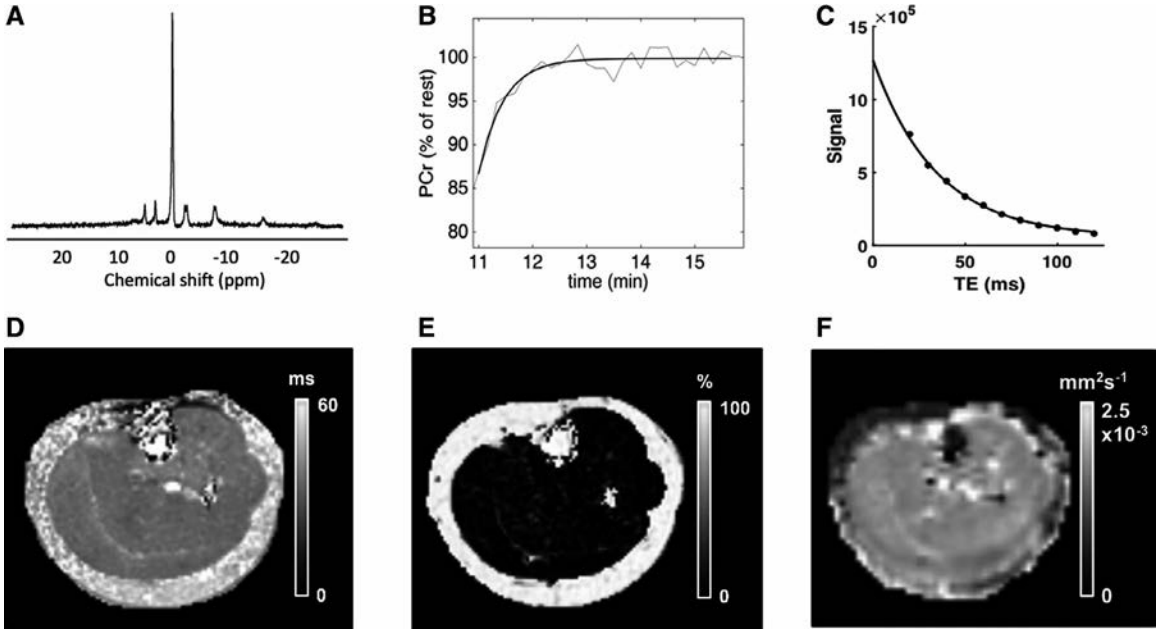

Figure 1. Magnetic resonance imaging data from a healthy control. **A**, Baseline ^{31}P spectrum; **B**, Dynamics of phosphocreatine (PCr) during the recovery period; **C**, Transverse relaxation from a single voxel within soleus muscle, shown together with fitted curve; **D**, Calculated transverse relaxation time map; **E**, Fat fraction map; and **F**, mean diffusivity map. TE = echo time; ms = milliseconds.

These differences do not appear to 0be related to physical fatigue. Further, no differences in skeletal muscle microstructure were observed between cases and controls. Taken together, skeletal muscle does not appear to serve as a major factor in SLE-related fatigue.

PCr recovery halftime reflects the muscle oxidative capacity and is used as a marker of muscle mitochondrial function (37). In SLE there is accumulating evidence to support the presence of mitochondrial abnormalities in peripheral blood cells. For example, Gergely et al observed hyperpolarized mitochondria in T cells that resulted in greater ATP depletion, oxidative stress, and ultimately cell death (38). We now provide supporting data that mitochondrial dysfunction might also exist within the skeletal muscle of patients with SLE. The same marker has previously been related to fatigue in SLE (39), although our exploratory analysis suggests that pathways other than skeletal muscle mitochondrial dysfunction may be involved in the generation of this symptom.

Microstructural MRI of skeletal muscles has been applied in only a few clinical populations and, to our knowledge, never in the investigation of fatigue. DTI has evidenced changes of muscle integrity in athletes following marathon runs, where standard sequences have failed to detect macroscopic differences (40). Furthermore, this method can distinguish disease activity in inflammatory muscle diseases with greater sensitivity than standard imaging (41). Among neuromuscular conditions, where existing clinical tests are inadequate to assess disease progression, the quantification of structural parameters such as muscle volumes and fat infiltration are providing superior biomarkers for clinical trials and practice (42). Such studies are similar in size to the present investigation, and so the absence of differences between our cases and controls in any of the sensitive microstructural metrics contradicts the hypothesis that physical fatigue is related to structural abnormalities in SLE skeletal muscle.

If not skeletal muscles, what then are the main explanations of physical fatigue among patients with SLE? A recent study of fatigue in another multisystem autoimmune disorder (antineutrophil cytoplasmic antibody–associated vasculitis) failed to detect a significant relationship between physical fatigue and skeletal muscle mass (measured using dual-energy x-ray absorptiometry) or function. Compared to healthy controls, fatigued cases evidenced reduced voluntary activation of skeletal muscle and reduced maximal voluntary contraction of skeletal muscle, and they had higher levels of perceived exertion, a finding that significantly correlated with physical fatigue (43). Together, these observations pointed toward centrally rather than peripherally driven mechanisms.

The novel application of cutting-edge MRI methods combined with a comprehensive approach to phenotyping are strengths of this study, but a number of limitations must also be considered. First, the highly selective eligibility criteria (purposely planned to enhance homogeneity by excluding known fatigue mechanisms) has resulted in a sample with generally mild disease. The results are therefore not generalizable to the wider disease spectrum. For example, patients

with a history of myositis (prevalent in 4–16% of SLE cases [44]) were excluded. Data from this study cannot be used to inform the usefulness of these methodologies in the evaluation of such manifestations (a distinct research question). Second, we recognize that patients with SLE without fatigue would have served as a more precise control group. That said, given the pervasiveness of fatigue in this disease, recruiting such patients would have been logistically challenging. Regardless, the absence of differences even with a healthy control group (as observed with almost all of the MRI metrics) indicates that these methodologies are unlikely to identify a clinically relevant fatigue-specific signal. Uncertainty also exists regarding the clinical relevance of the statistically significant PCr measure, since the 6-second difference in recovery halftime is lower in magnitude compared to other ^{31}P studies (for example, mean ± SD 18.7 ± 0.9 seconds in healthy controls versus 27.3 ± 3.5 seconds in patients with diabetes mellitus [45], or mean ± SD 35.0 ± 3.0 seconds in healthy controls versus 45.0 ± 4.0 seconds in patients with chronic obstructive pulmonary disease [46]). Third, although the sample size is equivalent to other MRI muscle studies, which have detected significant changes in other populations, we cannot be certain that larger sample sizes will not identify a significant effect. In particular, fully powered analyses of within-case correlational analysis might uncover relationships between PCr and SLE fatigue. We suspect, however, that in the absence of even a trend, any associations are unlikely to be major contributors to our understanding of physical fatigue.

This study provides evidence of feasibility for the use of multimodal MRI muscle assessment in patients with SLE. From this data, the investigation of physical fatigue would seem to be better served by examining alternatives to skeletal muscle–based pathways. Learning from other chronic diseases, the investigation of central mechanisms using advanced MRI brain techniques appears to offer greater potential (47). Such approaches have been limited in SLE and should be encouraged in an effort to better understand this considerable patient challenge.

ACKNOWLEDGMENT

The authors thank the patient community who participated in this research effort.

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. Basu 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. Erwig, Gray, Waiter, He, Basu.
Acquisition of data. Keenan, Hutcheon, Dospinescu, Basu.
Analysis and interpretation of data. Cheung, Senn, Chan, Schrepf, Waiter, He, Basu.

REFERENCES

1. Zonana-Nacach A, Roseman JM, McGwin G Jr, Friedman AW, Baethge BA, Reveille JD, et al. Systemic lupus erythematosus in three ethnic groups. VI: factors associated with fatigue within 5 years of criteria diagnosis. LUMINA Study Group. LUpus in MInority populations: NAture vs Nurture. Lupus 2000;9:101–9.
2. Overman CL, Kool MB, Da Silva JA, Geenen R. The prevalence of severe fatigue in rheumatic diseases: an international study. Clin Rheumatol 2016;35:409–15.
3. Wang B, Gladman DD, Urowitz MB. Fatigue in lupus is not correlated with disease activity. J Rheumatol 1998;25:892–5.
4. Baker K, Pope J, Silverman E, Cooper G, Fortin P, Zummer M, et al. Work disability in systemic lupus erythematosus is prevalent and predicted by socio-demographic and disease related factors: results from a multi-ethnic cohort. J Rheumatol 2009;36:2559–60.
5. Cleanthous S, Tyagi M, Isenberg DA, Newman SP. What do we know about self-reported fatigue in systemic lupus erythematosus? Lupus 2012;21:465–76.
6. Da Costa D, Dritsa M, Bernatsky S, Pineau C, Menard HA, Dasgupta K, et al. Dimensions of fatigue in systemic lupus erythematosus: relationship to disease status and behavioral and psychosocial factors. J Rheumatol 2006;33:1282–8.
7. Balsamo S, da Mota LM, de Carvalho JF, Nascimento DD, Tibana RA, de Santana FS, et al. Low dynamic muscle strength and its associations with fatigue, functional performance, and quality of life in premenopausal patients with systemic lupus erythematosus and low disease activity: a case-control study. BMC Musculoskelet Disord 2013;14:263.
8. Jones DE, Hollingsworth KG, Taylor R, Blamire AM, Newton JL. Abnormalities in pH handling by peripheral muscle and potential regulation by the autonomic nervous system in chronic fatigue syndrome. J Intern Med 2010;267:394–401.
9. Lim KL, Abdulwahab R, Lowe J, Powell RJ. Muscle biopsy abnormalities in systemic lupus-erythematosus: correlation with clinical and laboratory parameters. Ann Rheum Dis 1994;53:178–82.
10. Ai T, Yu K, Gao L, Zhang P, Goerner F, Runge VM, et al. Diffusion tensor imaging in evaluation of thigh muscles in patients with polymyositis and dermatomyositis. Br J Radiol 2014;87:20140261.
11. Galban CJ, Maderwald S, Stock F, Ladd ME. Age-related changes in skeletal muscle as detected by diffusion tensor magnetic resonance imaging. J Gerontol A Biol Sci Med Sci 2007;62:453–8.
12. Hochberg MC, for the Diagnostic and Therapeutic Criteria Committee of the American College of Rheumatology. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus [letter]. Arthritis Rheum 1997;40:1725.
13. Chalder T, Berelowitz G, Pawlikowska T, Watts L, Wessely S, Wright D, et al. Development of a fatigue scale. J Psychosom Res 1993;37:147–53.
14. Isenberg DA, Rahman A, Allen E, Farewell V, Akil M, Bruce IN, et al. BILAG 2004: development and initial validation of an updated version of the British Isles Lupus Assessment Group's disease activity index for patients with systemic lupus erythematosus. Rheumatology (Oxford) 2005;44:902–6.
15. Gladman D, Ginzler E, Goldsmith C, Fortin P, Liang M, Urowitz M, et al. The development and initial validation of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology damage index for systemic lupus erythematosus. Arthritis Rheum 1996;39:363–9.
16. Goligher EC, Pouchot J, Brant R, Kherani RB, Avina-Zubieta JA, Lacaille D, et al. Minimal clinically important difference for 7 measures of fatigue in patients with systemic lupus erythematosus. J Rheumatol 2008;35:635–42.

17. Holloway L, Humphrey L, Heron L, Pilling C, Kitchen H, Hojbjerre 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.
18. Jenkins CD, Stanton BA, Niemcryk SJ, Rose RM. A scale for the estimation of sleep problems in clinical research. J Clin Epidemiol 1988;41:313–21.
19. Marcora SM, Casanova F, Fortes MB, Maddison PJ. Validity and reliability of the Siconolfi Step Test for assessment of physical fitness in patients with systemic lupus erythematosus. Arthritis Rheum 2007;57:1007–11.
20. Siconolfi SF, Garber CE, Lasater TM, Carleton RA. A simple, valid step test for estimating maximal oxygen uptake in epidemiologic studies. Am J Epidemiol 1985;121:382–90.
21. Arpan I, Forbes SC, Lott DJ, Senesac CR, Daniels MJ, Triplett WT, et al. T_2 mapping provides multiple approaches for the characterization of muscle involvement in neuromuscular diseases: a cross-sectional study of lower leg muscles in 5-15-year-old boys with Duchenne muscular dystrophy. NMR Biomed 2013;26:320–8.
22. Schwenzer NF, Martirosian P, Machann J, Schraml C, Steidle G, Claussen CD, et al. Aging effects on human calf muscle properties assessed by MRI at 3 Tesla. J Magn Reson Imaging 2009;29:1346–54.
23. Scheel M, von Roth P, Winkler T, Arampatzis A, Prokscha T, Hamm B, et al. Fiber type characterization in skeletal muscle by diffusion tensor imaging. NMR Biomed 2013;26:1220–4.
24. Sinha U, Csapo R, Malis V, Xue YJ, Sinha S. Age-related differences in diffusion tensor indices and fiber architecture in the medial and lateral gastrocnemius. J Magn Reson Imaging 2015;41:941–53.
25. Liu CY, McKenzie CA, Yu H, Brittain JH, Reeder SB. Fat quantification with IDEAL gradient echo imaging: correction of bias from T-1 and noise. Magn Reson Med 2007;58:354–64.
26. Kuhn JP, Jahn C, Hernando D, Siegmund W, Hadlich S, Mayerle J, et al. T1 bias in chemical shift-encoded liver fat-fraction: role of the flip angle. J Magn Reson Imaging 2014;40:875–83.
27. Sprott H, Rzanny R, Reichenbach JR, Kaiser WA, Hein G, Stein G. P-31 magnetic resonance spectroscopy in fibromyalgic muscle. Rheumatology (Oxford) 2000;39:1121–5.
28. Wu JS, Buettner C, Smithline H, Ngo LH, Greenman RL. Evaluation of skeletal muscle during calf exercise by 31-phosphorus magnetic resonance spectroscopy in patients on statin medications. Muscle Nerve 2011;43:76–81.
29. Bentley R, Gray SR, Schwarzbauer C, Dawson D, Frenneaux M, He J. Dietary nitrate reduces skeletal muscle oxygenation response to physical exercise: a quantitative muscle functional MRI study. Physiol Rep 2014;2:e12089.
30. Naressi A, Couturier C, Castang I, de Beer R, Graveron-Demilly D. Java-based graphical user interface for MRUI, a software package for quantitation of in vivo/medical magnetic resonance spectroscopy signals. Comput Biol Med. 2001;31:269–86.
31. Haseler LJ, Hogan MC, Richardson RS. Skeletal muscle phosphocreatine recovery in exercise-trained humans is dependent on O-2 availability. J Appl Physiol 1999;86:2013–8.
32. Hu HH, Bornert P, Hernando D, Kellman P, Ma JF, Reeder S, et al. ISMRM workshop on fat-water separation: insights, applications and progress in MRI. Magn Reson Med 2012;68:378–88.
33. Zhong XD, Nickel MD, Kannengiesser SA, Dale BM, Kiefer B, Bashir MR. Liver fat quantification using a multi-step adaptive fitting approach with multi-echo GRE imaging. Magn Reson Med 2014;72:1353–65.
34. Ren JM, Dimitrov I, Sherry AD, Malloy CR. Composition of adipose tissue and marrow fat in humans by 1H NMR at 7 Tesla. J Lipid Res 2008;49:2055–62.
35. Yu HZ, Shimakawa A, McKenzie CA, Brodsky E, Brittain JH, Reeder SB. Multiecho water-fat separation and simultaneous R-2* estimation with multifrequency fat spectrum modeling. Magn Reson Med 2008;60:1122–34.
36. Wiewiorski M, Dopke K, Steiger C, Valderrabano V. Muscular atrophy of the lower leg in unilateral post traumatic osteoarthritis of the ankle joint. Int Orthop 2012;36:2079–85.
37. Hollingsworth KG, Newton JL, Taylor R, McDonald C, Palmer JM, Blamire AM, et al. Pilot study of peripheral muscle function in primary biliary cirrhosis: potential implications for fatigue pathogenesis. Clin Gastroenterol Hepatol 2008;6:1041–8.
38. Gergely P Jr, Grossman C, Niland B, Puskas F, Neupane H, Allam F, et al. Mitochondrial hyperpolarization and ATP depletion in patients with systemic lupus erythematosus. Arthritis Rheum 2002;46:175–90.
39. Segal BM, Thomas W, Zhu X, Diebes A, McElvain G, Baechler E, et al. Oxidative stress and fatigue in systemic lupus erythematosus. Lupus 2012;21:984–92.
40. Froeling M, Oudeman J, Strijkers GJ, Maas M, Drost MR, Nicolay K, et al. Muscle changes detected with diffusion-tensor imaging after long-distance running. Radiology 2015;274:548–62.
41. Qi J, Olsen NJ, Price RR, Winston JA, Park JH. Diffusion-weighted imaging of inflammatory myopathies: polymyositis and dermatomyositis. J Magn Reson Imaging 2008;27:212–7.
42. Morrow JM, Sinclair CD, Fischmann A, Machado PM, Reilly MM, Yousry TA, et al. MRI biomarker assessment of neuromuscular disease progression: a prospective observational cohort study. Lancet Neurol 2016;15:65–77.
43. McClean A, Morgan MD, Basu N, Bosch JA, Nightingale P, Jones D, et al. Physical fatigue, fitness, and muscle function in patients with antineutrophil cytoplasmic antibody–associated vasculitis. Arthritis Care Res (Hoboken) 2016;68:1332–9.
44. Liang Y, Leng RX, Pan HF, Ye DQ. Associated variables of myositis in systemic lupus erythematosus: a cross-sectional study. Med Sci Monit 2017;23:2543–9.
45. Schrauwen-Hinderling VB, Kooi ME, Hesselink MK, Jeneson JA, Backes WH, van Echteld CJ, et al. Impaired in vivo mitochondrial function but similar intramyocellular lipid content in patients with type 2 diabetes mellitus and BMI-matched control subjects. Diabetologia 2007;50:113–20.
46. McKeough ZJ, Alison JA, Bye PT, Trenell MI, Sachinwalla T, Thompson CH, et al. Exercise capacity and quadriceps muscle metabolism following training in subjects with COPD. Respir Med 2006;100:1817–25.
47. DeLuca J, Genova HM, Capili EJ, Wylie GR. Functional neuroimaging of fatigue. Phys Med Rehabil Clin North Am 2009;20:325–37.

Arthritis Care & Research
Vol. 71, No. 12, December 2019, pp 1647–1652
DOI 10.1002/acr.23588

American College of Rheumatology
Empowering Rheumatology Professionals

Sex Differences in Quality of Life in Patients With Systemic Lupus Erythematosus

Meenakshi Jolly,[1] Winston Sequeira,[1] Joel A. Block,[1] Sergio Toloza,[2] Ana Bertoli,[3] Ivanna Blazevic,[4] Luis M. Vila,[5] Ioana Moldovan,[6] Karina D. Torralba,[7] Davide Mazzoni,[8] Elvira Cicognani,[8] Sarfaraz Hasni,[9] Berna Goker,[10] Seminur Haznedaroglu,[10] Josiane Bourre-Tessier,[11] Sandra V. Navarra,[12] Chi Chiu Mok,[13] Michael Weisman,[14] Ann E. Clarke,[15] Daniel Wallace,[14] and Graciela Alarcón[16]

Objective. Systemic lupus erythematosus (SLE) predominantly affects women. Clinical phenotype and outcomes in SLE may vary by sex and are further complicated by unique concerns that are dependent upon sex-defined roles. We aimed to describe sex differences in disease-specific quality of life (QoL) assessment scores using the Lupus Patient-Reported Outcome (LupusPRO) tool in a large international study.

Methods. Cross-sectional data from 1,803 patients with SLE on demographics, self-identified sex status, LupusPRO, and disease activity were analyzed. The LupusPRO tool has 2 constructs: health-related QoL (HRQoL) and non-HRQoL. Disease activity and damage were evaluated using the Safety of Estrogens in Lupus Erythematosus National Assessment version of the Systemic Lupus Erythematosus Disease Activity Index and the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index, respectively. Nonparametric tests were used to compare QoL and disease activity by sex.

Results. A total of 122 men and 1,681 women with SLE participated. The mean age was similar by sex, but the damage scores were greater among men. Men fared worse on the non-HRQoL social support domain than women ($P = 0.02$). When comparing disease and QoL among men and women ages ≤45 years, men were found to have greater damage and worse social support than women. However, women fared significantly worse on lupus symptoms, cognition, and procreation domains with trends for worse functioning on physical health and pain-vitality domains.

Conclusion. In the largest study of a diverse group of SLE patients, utilizing a disease-specific QoL tool, sex differences in QoL were observed on both HRQoL and non-HRQoL constructs. Although men performed worse in the social support domain, women (especially those in the reproductive age group) fared worse in other domains. These observations may assist physicians in appropriately addressing QoL issues in a sex-focused manner.

INTRODUCTION

Systemic lupus erythematosus (SLE) is an autoimmune disease that predominantly affects young women. Quality of life (QoL) and health-related quality of life (HRQoL) among patients with SLE are poor (1–6). Among men with SLE, a higher proportion are white and a higher prevalence of serositis and nephritis has been shown. Additionally, earlier and greater damage accrual along with higher mortality has been demonstrated in men (7,8). There are only a few studies that have focused on SLE in men (7–11). QoL in men with SLE has not been previously studied, although a worse QoL among men with SLE could be expected, based on earlier and greater disease damage accrual (7–11).

Furthermore, QoL is known to vary by age and sex among healthy populations (12,13) and in diseases other than SLE (14–16). Sex-based life roles may also differentially impact QoL,

Supported by a grant from Cornell Brewer Foundation and the Intramural Research Program of the National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH.

[1]Meenakshi Jolly, MD, Winston Sequeira, MD, Joel A. Block, MD: Rush University, Chicago, Illinois; [2]Sergio Toloza, MD: Hospital San Juan Batista, Catamarca, Argentina; [3]Ana Bertoli, MD: Instituto Reumatologico Strusberg, Cordoba, Argentina; [4]Ivanna Blazevic, MD: Universidad de Buenos Aires, Buenos Aires, Argentina; [5]Luis M. Vila, MD: University of Puerto Rico, San Juan, Puerto Rico; [6]Ioana Moldovan, MD: Beaver Medical Group, Redlands, California; [7]Karina D. Torralba, MD: Loma Linda University, Loma Linda, California; [8]Davide Mazzoni, PhD, Elvira Cicognani, PhD: University of Bologna, Bologna, Italy; [9]Sarfaraz Hasni, MD: National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH, Bethesda, Maryland; [10]Berna Goker, MD, Seminur Haznedaroglu, MD: Gazi University, Ankara, Turkey; [11]Josiane Bourre-Tessier, MD: University of Montreal, Montreal, Quebec, Canada; [12]Sandra V. Navarra, MD: University of Santo Tomas, Manila, Philippines; [13]Chi Chiu Mok, MD: Tuen Mun Hospital, Hong Kong; [14]Michael Weisman, MD, Daniel Wallace, MD: Cedars Sinai Medical Center, Los Angeles, California; [15]Ann E. Clarke, MD: Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada; [16]Graciela Alarcón, MD: University of Alabama at Birmingham.

Dr. Torralba has received consulting fees, speaking fees, and/or honoraria from Pfizer, Exagen, and GlaxoSmithKline (less than $10,000 each). Rush University holds copyrights to LupusPRO. A percentage of the proceeds from its commercial use are distributed to Rush and the developers per Rush University Policy.

Address correspondence to Meenakshi Jolly, MD, Professor of Medicine and Behavioral Science, Rush University Medical Center, 1611 West Harrison Street, Suite 510, Chicago, IL 60612. E-mail: Meenakshi_Jolly@rush.edu.

Submitted for publication June 14, 2017; accepted in revised form April 17, 2018.

SIGNIFICANCE & INNOVATIONS

- Quality of life concerns in patients with systemic lupus erythematosus (SLE) can be different based on sex and age. These concerns have never previously been systematically studied.
- Awareness of sex-pertinent quality of life concerns among patients with SLE can assist physicians in providing personalized care.
- This study utilized disease-specific, patient reported outcomes to delineate sex-specific quality of life concerns in an ethnically heterogeneous, international, large group of patients with SLE.

especially because SLE predominately affects a younger age group and women. Some roles in lives vary by sex, especially in the context of age. Financial and physical independence, although relevant to all ages and both sexes, may be of greater relevance among middle-aged and older individuals than younger persons. However, completion of education, entrance into the job market, dating, and marriage are relatively more important milestones for younger individuals of either sex than older individuals. Although procreation and parenting are important roles for either sex, the role of childbearing may be more relevant among women in the reproductive age group (ages ≤45 years). Hence, it is plausible that overall QoL and specific domains of QoL that are affected may differ by sex and age in patients with SLE, especially in the younger and reproductive age groups.

We undertook this study to compare the QoL (overall and domains) among men and women with SLE, in order to better understand the sex differences in their QoL, especially in the reproductive age group. Information generated may help inform development of personalized and targeted care strategies.

PATIENTS AND METHODS

The study was approved by the institutional review board at all participating sites. The Study of Outcomes in Lupus data repository includes data obtained from SLE patients enrolled from various sites within the US, Canada, Asia, and Europe for validation of the LupusPRO tool. All patients met the American College of Rheumatology (ACR) classification criteria for SLE (17). Demographic data (age, sex, ethnicity), and QoL data were available for all patients. A subset of patients had data available on disease activity and disease damage. Disease activity was measured using the Safety of Estrogens in Lupus Erythematosus National Assessment (SELENA) version of the Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) (18). The latter index includes a physician's global assessment, which ranges from 0 to 3. Disease damage was assessed using the Systemic Lupus International Collaborating Clinics/ACR Damage Index (SDI) (19). Higher scores for the SELENA-SLEDAI and SDI represent greater disease activity and disease damage, respectively.

QoL was measured using the LupusPRO tool (20), which has been validated in several languages (21–27) and shows measurement equivalence (21–28). The LupusPRO tool has response options on a 1–5 point Likert scale, where 0 = none of the time, 1 = a little of the time, 2 = some of the time, 3 = most of the time, 4 = all of the time, 5 = not applicable (recoded as 0 for scoring). LupusPRO has 2 constructs (HRQoL and non-HRQoL) and 12 domains (8 HRQoL domains and 4 non-HRQoL domains). The HRQoL domains are lupus symptoms, cognition, lupus medications, procreation, physical health, pain-vitality, emotional health, and body image. The non-HRQoL domains are desires-goals, social support, coping, and satisfaction with treatment. Item scores are totaled for each domain item and the mean domain score is obtained by dividing the total score by the number of items in that domain. The mean raw domain score is transformed to scores ranging from 0 (worst QoL) to 100 (best QoL) by dividing by 4 (the number of responses on a 1–5 point Likert scale [5 responses] minus 1) and then multiplying by 100, as shown (mean raw domain score/4) × 100 = transformed score for the domain. Transformed domain scores are obtainable when at least 50% of the items are answered. Total HRQoL and non-HRQoL scores are obtained by averaging the transformed domain scores within each construct.

All analyses were performed using IBM SPSS software, edition 19. Descriptive summary statistics were obtained for the 2 study groups (men and women). Chi-square test and Mann-Whitney U test were used to compare discrete and continuous variables, respectively, among the 2 sex groups, because most of the data were not normative in distribution. When data were distributed normatively, Student's *t*-test was used.

For the reproductive age group subanalysis, patients ages ≤45 years were selected and matched for age. Similar comparative analyses were undertaken as for the whole group. For all tests, a 2-sided *P* value less than or equal to 0.05 was considered significant.

RESULTS

A total of 1,803 patients with SLE (mean ± SD age 40.8 ± 13.1 years) participated in the study (Table 1). Of these, 122 patients (6.7%) were men. Racial composition of the study participants was as follows: 8% African American, 37% White, 14% Hispanic, 40% Asian, and 1% other. The mean ± SD physician global assessment and SELENA-SLEDAI scores were 0.6 ± 0.7 and 3.4 ± 4.0, respectively. The mean ± SD SDI score was 0.9 ± 1.5. QoL descriptions for the entire group are shown in Table 1. The domains most affected were cognition, pain-vitality, emotional health, desires-goals, and coping.

Table 1. Characteristics and demographics of study group*

Variable	Value
Age, mean ± SD years	40.8 ± 13.1
Ethnicity, %	
African American	8.00
Caucasian	37.00
Hispanic	14.00
Asian	40.00
Other	1.00
Disease features	
PhGA	0.6 ± 0.7
SELENA-SLEDAI (n = 920)	3.4 ± 4.0
SDI (n = 1,247)	0.9 ± 1.5
SFI-Yes, %	25
LupusPRO HRQoL	
Lupus symptoms	70.3 ± 24.9
Cognition	63.8 ± 28.9
Lupus medications	68.3 ± 29.5
Procreation	77.9 ± 32.0
Physical health	76.7 ± 27.0
Pain-vitality	63.3 ± 27.8
Emotional health	58.3 ± 27.8
Body image	71.0 ± 29.1
LupusPRO non–HRQoL	
Desires-goals	65.6 ± 28.6
Social support	66.0 ± 31.4
Coping	65.5 ± 26.0
Satisfaction with treatment	66.5 ± 32.6

* Values are the mean ± SD unless indicated otherwise. PhGA = physician's global assessment; SELENA-SLEDAI = Safety of Estrogens in Lupus Erythematosus National Assessment version of the Systemic Lupus Erythematosus Disease Activity Index; SDI = Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index; SFI = SELENA-SLEDAI flare index; LupusPRO = Lupus Patient-Reported Outcome tool; HRQoL = health-related quality of life; non–HRQol = non-health–related quality of life.

There were no differences by age, ethnicity (not shown), or disease activity between men and women (Table 2). Disease damage tended to be greater in men than women (median interquartile range [IQR] 1.0–2.0 versus 0.0–1.0; $P = 0.03$). Among men, the domains with median scores ≤65 were lupus medications, emotional health, social support, and coping. Among women, the domains with median scores ≤65 were cognition, pain-vitality, and emotional health (Table 2). In the social support domain, men had worse median scores as compared to women (median 62.5 [IQR 37.5] versus median 75.0 [IQR 50.0]; $P = 0.02$). There were trends noted with worse performance among women in lupus symptoms (median 75.0 [IQR 33.3] versus median 83.3 [IQR 33.3]; $P = 0.06$). Trends were also noted for worse functioning among women in the cognition, physical health, and pain-vitality domains (Table 2). Comparison of QoL by sex is shown in Figure 1.

When men were age-matched and compared to women in the reproductive age group, greater damage was again noted among men, with a median of 1.0 (IQR 2) as compared to women who had a median of 0.0 (IQR 1) ($P < 0.001$). Women scored significantly lower than men in the lupus symptoms, cognition, and procreation domains (Table 3), with trends toward lower scores in the physical health and pain-vitality domains. For men, the domains with median scores ≤65 were the lupus medications, emotional health, coping, and social support domains (Table 3) and among women, the domains with median scores ≤65 were cognition, pain-vitality, and emotional health.

Table 2. Comparisons between men and women with SLE*

	Men (n = 122)	Women (n = 1,681)	*P*
Age, mean ± SD years	39.5 ± 14.1	40.9 ± 13.0	0.70
PhGA, median, IQR (no.)	1.0, 1.2 (55)	0.4, 0.9 (865)	0.15
SELENA-SLEDAI, median, IQR (no.)	2.0, 4.0 (77)	2.0, 4.0 (1,163)	0.83
SDI, median, IQR (no.)	1.0, 2.0 (72)†	0.0, 1.0 (1,175)†	0.03†
LupusPRO-RQoL, median, IQR			
Lupus symptoms	83.3, 33.3	75.0, 33.3	0.06
Cognition	75.0, 37.5	62.5, 37.5	0.14
Lupus medications	62.5, 50.0	75.0, 50.0	0.22
Procreation	100, 25.0	100, 37.5	0.08
Physical health	90.0, 25.0	85.0, 40.0	0.18
Pain-vitality	75.0, 45.0	65.0, 50.0	0.12
Emotional health	58.3, 45.8	58.3, 45.8	0.66
Body image	80.0, 40.0	80.0, 50.0	0.49
LupusPRO non–HRQoL, median, IQR			
Desires-goals	68.8, 46.9	68.8, 50.0	0.87
Social support	62.5, 37.5†	75.0, 50.0†	0.02†
Coping	58.3, 33.3	66.7, 33.3	0.40
Satisfaction with treatment	75.0, 56.3	75.0, 56.3	0.24

* SLE = systemic lupus erythematosus. See Table 1 for other definitions.
† Values are statistically significant ($P \leq 0.05$).

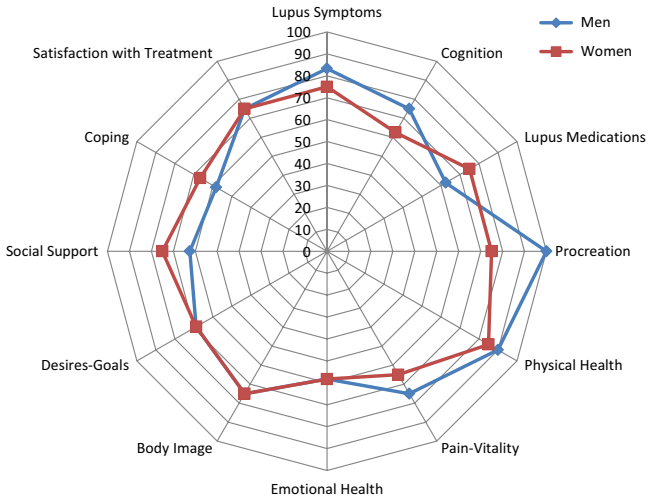

Figure 1. Quality of life spydergram for men and women with systemic lupus erythematosus.

DISCUSSION

There are sex-specific QoL issues in SLE, and it is important for physicians, and health care systems to be aware of such issues while providing individualized medical care to patients with SLE. We found poor social support to be a key QoL issue among men with SLE, irrespective of their age. This may be plausible theoretically, because SLE is traditionally known to be a "woman's disease," due to its 9:1 prevalence ratio of women to men. Men may not feel comfortable discussing their illness or matters related to SLE with their social support circle. It is possible that there are sex-based differences in both the receiver and provider of social support. Men with SLE may not be similar to women in their ability regarding communicating their need for support or reaching out for social support as readily or effectively, or in receiving the help. It is well known that men and women differ on a wide variety of behavioral, cognitive, and affective dimensions (29). Men may perceive the need for social support and/or emotional fragility as indicative of weakness or hypomasculinity, due to their traditional social role of being the provider or protector (30). One of the features of the male sex role in society is defined by strength and it includes emotional toughness, courage, self-reliance, and rationality (30). Any deviation from this expected role may subject men to harassment and discrimination.

Women, however, are more likely to seek out social support (31–34), possibly due to the paucity of sex-appropriate informational resources that are available for men with SLE because the disease is more common among women. Social support is important in SLE, and its relationship with health in SLE is recognized (35,36).

In women with SLE, the main QoL issues that differed from those in men were related to procreation, especially among the younger patients. This is plausible because the stereotypical roles assigned by society are still somewhat sex centric (37). The woman's role has traditionally included childbearing/rearing, and care of the family members and home (37). Thus, ongoing or new lupus symptoms, unpredictable disease flares, and effects on cognition may interfere with planning, scheduling, or participating in these "role performance" functions. This may explain the poorer QoL among women as compared to men (ages ≤45 years) in the lupus symptoms, cognition, and procreation domains. Notably, the emotional health domain was the most adversely affected in both sexes.

In our study, disease damage was greater among men as compared to women overall and those in the reproductive age group. Hypothetically, this finding could be attributable to disease duration and age. On further analysis (not shown) male sex

Table 3. Comparisons between men and women ages ≤45 years with SLE*

	Men (n = 117)†	Women (n = 1,080)†	*P*
Age, mean ± SD years	38.2 ± 12.9	32.9 ± 7.2	0.09
PhGA, median, IQR (no.)	1.0, 1.2 (54)	0.4, 0.8 (552)	0.60
SELENA-SLEDAI, median, IQR (no.)	2.0, 4.0 (73)	2.0, 5.0 (747)	0.9
SDI, median, IQR (no.)	1.0, 2.0 (70)‡	0.0, 1.0 (747)‡	<0.001‡
LupusPRO-HRQoL, median, IQR			
Lupus symptoms	83.3, 33.3‡	75.0, 41.7‡	0.01‡
Cognition	75.0, 37.5‡	62.5, 50.0‡	0.04‡
Lupus medications	62.5, 50.0	75.0, 50.0	0.64
Procreation	100, 25.0‡	75.0, 50.0‡	0.008‡
Physical health	90.0, 25.0	85.0, 45.0	0.10
Pain-vitality	75.0, 45.0	65.0, 45.0	0.08
Emotional health	58.3, 45.8	58.3, 45.8	0.93
Body image	75.0, 40.0	75.0, 50.0	0.59
LupusPRO non–HRQoL, median, IQR			
Desires-goals	68.8, 46.9	68.8, 43.8	0.42
Social support	62.5, 37.5‡	75.0, 50.0‡	0.01‡
Coping	58.3, 31.3	66.7, 33.3	0.47
Satisfaction with treatment	75.0, 56.3	75.0, 50.0	0.3

* See Table 1 for definitions.
† Correction added after online publication 19 November 2019: Patients' ages and sample sizes have been corrected, respectively, in the title and column headings of Table 3.
‡ Values are statistically significant.

remained an independent predictor of greater damage overall and in the reproductive age group, even after accounting for age and disease duration. Greater damage among men as compared to women with SLE has been previously demonstrated (7–11).

The findings revealed in our study have great potential application and relevance to clinical practice. Addressing procreation concerns among women of reproductive age and social support among men, in addition to an overall evaluation of QoL (especially emotional health in both sexs) is important in SLE. Discussing the side effects of medications among women with SLE within the reproductive age group (especially effects on procreation) may help allay some concerns and anxiety. In addition, physicians need to be comfortable asking for patients' immediate, intermediate, and long-term procreation plans and, accordingly, offer an individualized management plan for SLE that includes indications, relative benefits, dangers, and alternatives to various treatment options.

There are several limitations to our study. Given the relative rarity of SLE among men, we included only 122 men. The only other study of SLE to contain a comparison of QoL by sex in its analysis included 54 men and 54 women (38). Macêdo et al used the Short Form 36 tool for HRQoL and reported better functioning on vitality and mental health domains among men with SLE as compared to women with SLE. Another limitation was the lack of availability of complete data from all patients on disease activity and damage. These data were, however, available on nearly 1,250 patients, 25% of whom had an ongoing flare at the time of the study. In addition, 40% of our participants were Asian, and only 8% were African American, which may affect the generalizibility of the results. Lastly, we did not have any data on fibromyalgia. Because QoL is affected by fibromyalgia, and up to one-third of patients with SLE may have concomitant fibromyalgia, it would have been of interest to adjust for this covariate when comparing QoL domains between sexes. Despite these limitations, this study addresses HRQoL in men with SLE, an area which, to our knowledge, has not been previously addressed.

The strengths of our study include large numbers of participants from various geographic and racial backgrounds. Another advantage is the use of a disease-targeted QoL tool that includes a non-HRQoL construct and allows for comparisons, because the domains utilized in our study are not routinely represented in generic and other disease-targeted tools. Some of the domains that did show sex differences in QoL in SLE are not included in generic QoL instruments nor in some of the other disease-specific HRQoL tools. These unique domains include lupus symptoms, cognition, lupus medications, procreation, and social support. Furthermore, because the LupusPRO tool was developed using feedback from both men and women with SLE, and uses sex-neutral language, it was possible to administer the tool to patients of either sex and assess all SLE-related pertinent aspects of their QoL.

In summary, we found sex differences in damage and QoL among patients with SLE. These were especially evident in the reproductive age groups and may be partly attributable to traditional sex roles. A focused evaluation and discussion of effects of SLE on patient QoL should also include sex-specific QoL concerns, and these should be considered when developing patient management plans.

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. Jolly 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. Jolly, Sequeira, Block, Toloza, Bertoli, Blazevic, Vila, Moldovan, Torralba, Mazzoni, Cicognani, Hasni, Goker, Haznedaroglu, Bourre-Tessier, Navarra, Mok, Weisman, Clarke, Wallace, and Alarcón.

Acquisition of data. Jolly, Sequeira, Block, Toloza, Bertoli, Blazevic, Vila, Moldovan, Torralba, Mazzoni, Cicognani, Hasni, Goker, Haznedaroglu, Bourre-Tessier, Navarra, Mok, Weisman, Clarke, Wallace, and Alarcón.

Analysis and interpretation of data. Jolly, Clarke.

REFERENCES

1. 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.
2. Jolly M, Utset TO. Can disease specific measures for systemic lupus erythematosus predict patients health related quality of life? Lupus 2004;13:924–6.
3. Aggarwal R, Wilke CT, Pickard AS, Vats V, Mikolaitis R, Fogg L. Psychometric properties of the EuroQol-5D and Short Form-6D in patients with systemic lupus erythematosus. J Rheumatol. 2009;36:1209–16.
4. Thumboo J, Feng PH, Boey ML, Soh CH, Thio S, Fong KY. Validation of the Chinese SF-36 for quality of life assessment in patients with systemic lupus erythematosus. Lupus 2000;9:708–12.
5. Mok CC, Ho LY, Cheung MY, Yu KL, To CH. Effect of disease activity and damage on quality of life in patients with systemic lupus erythematosus: a 2-year prospective study. Scand J Rheumatol. 2009;38:121–7.
6. Stoll T, Gordon C, Seifert B, Richardson K, Malik J, Bacon PA, et al. Consistency and validity of patient administered assessment of quality of life by the MOS SF-36; its association with disease activity and damage in patients with systemic lupus erythematosus. J Rheumatol 1997;24:1608–14.
7. Andrade RM, Alarcón GS, Fernández M, Apte M, Vilá LM, et al, LUMINA Study Group. Accelerated damage accrual among men with systemic lupus erythematosus: XLIV. Results from a multiethnic US cohort. Arthritis Rheum 2007;56:622–30.
8. Mayor AM, Vilá LM. Gender differences in a cohort of Puerto Ricans with systemic lupus erythematosus. Cell Mol Biol 2003;49:1339–44.
9. Camilleri F, Mallia C. Male SLE patients in Malta. Adv Exp Med Biol 1999;455:173–9.
10. Mok CC, Lau CS, Chan TM, Wong RW. Clinical characteristics and outcome of southern Chinese males with systemic lupus erythematosus. Lupus 1999;8:188–96.
11. Soto ME, Vallejo M, Guillén F, Simón JA, Arena E, Reyes PA. Gender impact in systemic lupus erythematosus. Clin Exp Rheumatol 2004;22:713–21.
12. Kirchengast S, Haslinger B. Gender differences in health-related quality of life among healthy aged and old-aged Austrians: cross-sectional analysis. Gend Med 2008;5:270–8.
13. Belbeisi A, Zindah M, Walke HT, Jarrar B, Mokdad AH. Health related quality of life measures by demographics and common health risks, Jordan 2004. Int J Public Health 2009;54 Suppl 1:106–10.

14. Mujais SK, Story K, Brouillette J, Takano T, Soroka S, Franek C, et al. Health-related quality of life in CKD patients: correlates and evolution over time. Clin J Am Soc Nephrol 2009;4:1293–301.
15. Park JM, Choi MG, Kim YS, Choi CH, Choi SC, Hong SJ, et al. Quality of life of patients with irritable bowel syndrome in Korea. Qual Life Res 2009;18:435–6.
16. Nicolucci A, Cucinotta D, Squatrito S, Lapolla A, Musacchio N, Leotta S, et al; QuoLITy Study Group. Clinical and socio-economic correlates of quality of life and treatment satisfaction in patients with type 2 diabetes. Nutr Metab Cardiovasc Dis 2009;19:45–53.
17. Hochberg MC, for the Diagnostic and Therapeutic Criteria Committee of the American College of Rheumatology. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus [letter]. Arthritis Rheum 1997;40:1725.
18. Petri M, Kim MY, Kalunian KC, Grossman J, Hahn BH, Sammaritano LR, et al; OC-SELENA Trial. Combined oral contraceptives in women with systemic lupus erythematosus. N Engl J Med 2005 15;353:2550–8.
19. Stoll T, Seifert B, Isenberg DA. SLICC/ACR Damage Index is valid, and renal and pulmonary organ scores are predictors of severe outcome in patients with systemic lupus erythematosus. Br J Rheumatol 1996;35:248–54.
20. 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.
21. Navarra SV, Tanangunan RM, Mikolaitis-Preuss RA, Kosinski M, Block JA, Jolly M. Cross-cultural validation of a disease-specific patient-reported outcome measure for lupus in Philippines. Lupus 2013;22:262–7.
22. Jolly M, Toloza S, Block J, Mikolaitis R, Kosinski M, Wallace D, et al. Spanish LupusPRO: cross cultural validation study for lupus. Lupus 2013;22:431–6.
23. Bourré-Tessier J, Clarke AE, Mikolaitis-Preuss RA, Kosinski M, Bernatsky S, Block JA, et al. Cross-cultural validation of a disease-specific patient-reported outcome measure for systemic lupus erythematosus in Canada. J Rheumatol 2013;40:1327–33.
24. Kaya A, Goker B, Cura ES, Tezcan ME, Tufan A, Mercan R, et al. Turkish lupusPRO: cross-cultural validation study for lupus. Clin Rheumatol 2014;33:1079–84.
25. Bourré-Tessier J, Clarke AE, Kosinski M, Mikolaitis-Preuss RA, Bernatsky S, Block JA, et al. The French-Canadian validation of a disease-specific, patient-reported outcome measure for lupus. Lupus 2014;23:1452–9.
26. Mok CC, Kosinski M, Ho LY, Chan KL, Jolly M. Validation of the LupusPRO in Chinese patients from Hong Kong with systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2015;67:297–304.
27. Jolly M, Mazzoni D, Cicognani E. Measurement properties of web based LupusPRO: a disease targeted outcome tool among Italian patients with lupus. Value Health 2015;18:302.
28. Jolly M, Kosinski M, Toloza S, Block JA, Mikolaitis RA, Durán S, et al. Equivalence of various language versions of lupus specific patient reported outcomes measure (LupusPRO) [abstract]. Arthritis Rheum 2012;64 Suppl:S607.
29. Ptacek JT, Smith RE, Dodge KL. Gender differences in coping with stress: when stressor and appraisals do not differ. J Pers Soc Psychol 1994;20:421–30.
30. Oliffe JL, Phillips MJ. Men, depression and masculinities: a review and recommendations. J Mens Health 2008;5:194–202.
31. Billings AG, Moos RH. Coping, stress, and social resources among adults with unipolar depression. J Pers Soc Psychol 1984;46:877–91.
32. Heppner PP, Reeder BL, Larson LM. Cognitive variables associated with personal problem-solving appraisal: implications for counseling. J Counseling Psychol 1983;30:537–45.
33. Ptacek JT, Smith RE, Zanas J. Gender, appraisal, and coping: a longitudinal analysis. J Pers 1992;60:747–70.
34. Stone AA, Neale JM. New measure of daily coping: development and preliminary results. J Pers Soc Psychol 1984;46:892–906.
35. Mazzoni D, Cicognani E. Positive and problematic support, stress and quality of life in patients with systemic lupus erythematosus. Anxiety Stress Coping 2016;29:542–51.
36. Mazzoni D, Cicognani E. Problematic social support from patients' perspective: the case of systemic lupus erythematosus. Soc Work Health Care 2014;53:435–45.
37. International Labour Organization. Online Gender Learning & Information module. Unit 1. A conceptual framework for sex analysis and planning. 1998. URL: https://www.ilo.org/public/english/region/asro/mdtmanila/training/unit1/mainmenu.htm.
38. Macêdo EA, Appenzeller S, Costallat LT. Gender differences in systemic lupus erythematosus concerning anxiety, depression and quality of life. Lupus 2016;25:1315–27.

Arthritis Care & Research
Vol. 71, No. 12, December 2019, pp 1653–1659
DOI 10.1002/acr.23816

American College of Rheumatology
Empowering Rheumatology Professionals

Prevalence and Factors Associated With Bone Erosion in Patients With Gout

Mian Wu, Feng J. Liu, Jie Chen, Li Chen, Chong Wei, Zhu M. Hu, Ying Han, Jun X. Lu, Li X. Jiang, and Hai B. Chen

Objective. To determine the prevalence, distribution, and factors associated with bone erosion detectable by ultrasound in patients with gout.

Methods. Ultrasound scans were performed in 980 patients with gout, and bone erosion was detected. The prevalence and distribution of bone erosion in gout patients were calculated. Both clinical variables and ultrasound signs were entered into a multivariate logistic regression analysis to clarify the factors associated with bone erosion in patients with gout.

Results. Bone erosion was found in 431 (44.0%) of the 980 patients with gout, and in 338 (78.4%) of these patients, the bone erosion was found in the first metatarsophalangeal (MTP) joint. A multivariable logistic regression analysis showed that age, duration of gout, the existence of tophi, ultrasound-detected synovial hypertrophy, and joint effusion were independently associated with bone erosion. A tophus was the most powerful factor associated with bone erosion, with an odds ratio (OR) of 4.218 (95% confidence interval 3.092–5.731). The risk for bone erosion also increased as the number of tophi increased ($P < 0.001$). However, after stratifying the size of tophi, the ORs did not increase significantly ($P = 0.206$).

Conclusion. A high percentage of gout patients had bone erosions; the first MTP joint was the most frequently involved site. Age, duration of gout, tophi, and synovial hypertrophy were factors associated with bone erosion in gout patients. The number of tophi, but not their size, was strongly associated with bone erosion in patients with gout.

INTRODUCTION

Gout is one of the most common arthropathies, and epidemiologic data suggest that its incidence is increasing (1,2). Aggregates of monosodium urate monohydrate (MSU) crystals deposited around and in the joints and soft tissues cause bone erosion in patients with chronic gout (3). The presence of bone erosion is a central feature of the anatomic damage and is associated with disease severity and poor functional outcome, such as joint damage and deformity and, eventually, musculoskeletal disability (3,4). Evidence of bone erosions and visible tophaceous deposits are included in the current indications for considering hypouricemic therapy to ultimately prevent gout flares, ongoing joint destruction, and urate nephropathy (5). Thus, early and accurate detection of erosion is important for diagnosis, treatment, and monitoring of gout.

Few studies have focused on bone erosion in patients with gout. In a study by McQueen et al (6), 40 patients with gout of the wrist were imaged using magnetic resonance imaging (MRI). The study demonstrated that tophi were associated with bone erosion. In another study, plain radiographs and dual energy computed tomography (CT) scans of the feet were prospectively obtained from 92 people with tophaceous gout (7). The results showed that urate and soft tissue components of the tophus are strongly and independently associated with bone erosion in patients with gout. The sample sizes of these studies were quite small, so further clinical studies in a larger cohort of patients are warranted to confirm these findings and to explore the prevalence and distribution of bone erosion in patients with gout. Furthermore, a cross-sectional study compared differences in the characteristics of bone erosion in 40 patients with rheumatoid arthritis (RA) and 40 with gout by gray-scale ultrasound (8).

It is now widely accepted that ultrasound is a useful, reliable, and reproducible imaging technique compared with other imaging techniques for detecting bone erosion, with the added advantage of being an innocuous and low-cost technique (9,10). Ultrasonography may have greater sensitivity for detecting complications of the disease, such as subclinical bone erosion and tophus

Dr. Hai B. Chen's work was supported by the Xuhui District Medical Science and Technology Project (grant SHXH201649) and the National Natural Science Foundation of China (grant 81870616).

Mian Wu, MD, Feng J. Liu, MD, Jie Chen, MD, Li Chen, MD, Chong Wei, MD, Zhu M. Hu, MD, Ying Han, MS, Jun X. Lu, BS, Li X. Jiang, MD, Hai B. Chen, MD: Shanghai Jiaotong University Affiliated Sixth People's Hospital, Shanghai, China.

Drs. Wu and Liu contributed equally to this work. Drs. Jiang and Chen contributed equally to this work.

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

Address correspondence to Hai B. Chen, MD or Li X. Jiang, MD, 600 YiShan Road, XuHui District, Shanghai, China, 200233. E-mail: chenhb@sjtu.edu.cn or jinger_28@sina.com.

Submitted for publication April 25, 2018; accepted in revised form November 20, 2018.

SIGNIFICANCE & INNOVATIONS

- Bone erosions were detected in 44% of 980 gout patients, with 78.4% of defects occurring at the first metatarsophalangeal joints.
- Bone erosion was independently associated with age, duration of gout, the existence of tophi, ultrasound-detected synovial hypertrophy, and joint effusion.
- The number of tophi, but not their size, was strongly associated with bone erosion in patients with gout.

formation (11). A recent analysis of the first metatarsophalangeal (MTP) joint showed poor agreement between plain radiography and ultrasonography regarding the presence of bone erosion, as less than half of the erosions found by ultrasonography were visible by plain radiography. These findings have been replicated by other groups (12,13). The European League Against Rheumatism task force published its recommendations for the use of imaging of the joints in the clinical management of RA and gout, and the recommendations state that ultrasound is very helpful for identifying synovitis and bone erosions and, thus, for making accurate diagnoses, predicting outcomes and responses to treatment, and monitoring disease progression (14).

Therefore, the aim of this study was to measure the prevalence and distribution of bone erosion in patients with gout using ultrasound and to identify factors associated with bone erosion in gout patients.

PATIENTS AND METHODS

Data source and study design. A retrospective cohort study design was used to conduct data analysis for calculating prevalence of bone erosion and detecting factors associated with bone erosion in patients with gout. All patients were recruited from the outpatient clinic of Endocrinology in Shanghai Jiao Tong University Affiliated Sixth People's Hospital (Shanghai, China) from April 2015 to September 2017. Inclusion criteria were a history of gout based on the 2015 American College of Rheumatology/European League Against Rheumatism diagnostic criteria to undergo imaging investigations (15). This study was approved by the Ethics Committee of Shanghai Jiao Tong University Affiliated Sixth People's Hospital, and informed consent was obtained from all participants.

Clinical and laboratory assessments. Every patient had completed a uniform questionnaire at enrollment, including age, sex, disease duration, co-morbidities, medications, current and previous illnesses, flare frequency, and pain score. The clinical data of the subjects were recorded, including height, weight, waist circumference, hip circumference, systolic blood pressure, and diastolic blood pressure. The body mass index was calculated as weight (kg)/height (m^2). Venous blood samples were collected in the morning after an overnight fast to measure blood routine, alanine aminotransferase, blood urea nitrogen (BUN), cystatin C, urate, creatinine, C-reactive protein, erythrocyte sedimentation rate, blood glucose, glycosylated hemoglobin, and lipid profiles. Participants were asked to provide 24-hour urine samples to detect urinary uric acid, creatinine, and protein. The estimated glomerular filtration rate (GFR) was calculated from the 4-variable modification of diet in renal disease equation: estimated GFR (ml/min/1.73 m^2) = 186.3 × (serum creatinine [μmol/liter]/88.4) – 1.154 × (age [years]) – 0.203 × (0.742 if female) × (1.21 if black). Fractional excretion of uric acid (%) = (serum creatinine [μmol/liter] × 24-hour urine uric acid [μmol])/(serum uric acid [μmol/liter] × 24-hour urine creatinine [μmol]) × 100%. All subjects underwent standard clinical and laboratory evaluations.

Ultrasound scanning technique. The ultrasound examinations were performed by 3 experienced sonographers using Aplio 500 (Toshiba), which was equipped with a multifrequency linear transducer (12–14 MHz). On each scanner, the factory setting for superficial musculoskeletal assessment was used. Bilateral knee, ankle, foot joints as well as the first to fifth MTP joints and were routinely scanned. The dorsal, volar, and lateral aspects of all anatomical sites were explored on both longitudinal and transverse views. All joint regions were sonographically examined in a standardized manner.

Ultrasound image interpretation. Erosion was defined as an intraarticular and/or extraarticular discontinuity of the bone surface that was visible in 2 perpendicular planes in gout (16). Clearly visible erosions in both longitudinal and transverse scans were considered for the study. Vascular bone channels were distinguished from bone erosion on the basis of anatomic location, insertion of feeding vessels detected by power Doppler, and absence of adjacent synovial lesion. Tophi around the joints were identified by ultrasound, and the sizes of tophi were measured. Double contour sign, articular synovial hypertrophy, and joint effusion were also assessed at the scanned joints.

Statistical analysis. Characteristics were reported as the mean ± SD for normal distributed variables, median (25th, 75th percentiles) for non-normal quantitative variables, or number with corresponding percentage. Student's *t*-test was used to compare means for normally distributed variables, and the chi-square test was used to compare frequencies. Variables that were significantly different between the non–bone erosion group and the bone erosion group were defined as candidate factors for logistic regression analysis. A multivariable logistic regression model using the forward stepwise likelihood ratio method was fitted with candidate factors, with entry probability of 0.05 and removal of 0.10. Odds ratio (OR)

and 95% confidence intervals were calculated for a multivariate logistic model. The Hosmer-Lemeshow test was used to investigate how close the prevalence that was predicted by the multivariate model was to the observed prevalence. Then all the factors that were entered into the logistic model were categorized to compare the relative risks for bone erosion. The number and size were 2 variables on behalf of tophi; thus, the number of tophi and the size of tophi were categorized and replaced the existence of tophi in logistic regression analysis. All tests were 2-tailed, and *P* values less than 0.05 were considered statistically significant. Statistical analyses were performed using the SPSS software package, version 21.0.

RESULTS

Bone erosion prevalence and distribution. In total, 980 patients with gout were studied; 92.9% were male (924 patients), with a mean age of 50.3 years. Evidence of bone erosion on ultrasound was found in 431 of the 980 (44%) patients. First, bone erosions were classified according to the bones involved. Of the 431 patients who had bone erosion on ultrasound, 338 (78.4%) had bone erosions in the first MTP joint, 19 (4.4%) in other MTP joints, 106 (24.6%) in ankle joints, 38 (8.8%) in knee joints, 16 (3.7%) in tarsal joints, and 17 (3.9%) in the calcaneus. Next, we calculated the number of bone erosions in each patient and found that 222 of 431 (51.5%) had a single bone erosion site, 138 (32.0%) had 2 erosions, 29 (6.7%) had 3 bone erosions, and 42 (9.7%) had ≥4 bone erosions. In those patients with bone erosion disease in the first MTP joint, 242 of 338 had lesions only in the MTP joints, and the other 97 had bone erosions in the MTP and other joints.

Comparison of clinical characteristics between patients with and without bone erosion. As shown in Table 1, the patients with bone erosion were significantly older (47.0 years versus 54.6 years; $P < 0.001$) and more hyperten-

Table 1. Comparison of clinical characteristics between bone erosion patients and non–bone erosion patients*

Variable	Non–bone erosion (n = 549)	Bone erosion (n = 431)	*P*
Men, no.	518	406	0.521
Age, year	47.0 ± 16.2	54.6 ± 14.7	<0.001
BMI, kg/m^2	26.2 ± 3.5	26.0 ± 3.9	0.310
Waist hip ratio	0.93 ± 0.06	0.94 ± 0.06	0.172
Systolic BP, mmHg	125.9 ± 17.3	129.7 ± 18.8	0.001
Diastolic BP, mmHg	82.4 ± 11,4	84.1 ± 12.1	0.031
Hemoglobin, gm/liter	148.7 ± 13.8	144.6 ± 15.7	<0.001
ALT, units/liter	39.1 ± 32.5	32.5 ± 29.1	0.002
BUN, mmol/liter	4.75 ± 1.50	5.33 ± 2.21	<0.001
Serum creatine, umol/liter	90.1 ± 31.8	96.7 ± 31.9	0.003
Cystatin C, mg/liter	0.99 ± 0.33	1.15 ± 0.50	<0.001
Urinary microalbumin, mg/24 hour	29.8 ± 108.0	61.7 ± 210.6	0.018
FPG, mmol/liter	5.66 ± 0.93	5.95 ± 1.29	0.001
Glycosylated hemoglobin, %	5.64 ± 0.86	5.87 ± 1.02	0.003
ESR, mm/hour	14.5 ± 15.8	19.7 ± 20.2	<0.001
CRP, mg/liter	8.01 ± 19.0	10.4 ± 23.3	0.252
Estimated GFR, ml/min/1.73 m^2	135.4 ± 66.3	139.6 ± 64.5	0.336
Current smoker, %	31.4	32.4	0.781
Current drinker, %	38.2	35.1	0.314
Serum UA, umol/liter	499.2 ± 128.0	498.9 ± 127.0	0.969
Fraction excretion of UA, %	4.59 ± 2.54	5.01 ± 2.38	0.033
Duration of gout, years	5.2 ± 5.4	9.2 ± 7.6	<0.001
Family history of gout, %	32.7	35.7	0.433
Gout attack in the past year, no.	4.2 ± 7.5	9.0 ± 13.7	<0.001
Urate-lowering therapy, %			
Allopurinol	7.8	11.8	0.035
Febuxostat	5.5	7.2	0.288
Benzobromarone	14.9	19.0	0.089
Anodyne treatment, %			
NSAIDs	30.4	35.5	0.092
Colchicine	27.1	26.5	0.809
Glucocorticoids	2.7	2.8	0.961

* Values are the mean ± SD (continuous variables of normal distribution) unless indicated otherwise. Proportions were expressed as percentages. BMI = body mass index; BP = blood pressure; ALT = alanine aminotransferase; BUN = blood urea nitrogen; FPG = fasting plasma glucose; ESR = erythrocyte sedimentation rate; CRP = C-reactive protein; GFR = glomerular filtration rate; UA = uric acid; NSAIDs = nonsteroidal antiinflammatory drugs.

Table 2. Comparison of ultrasound findings between bone erosion patients and non–bone erosion patients*

Variable	Non–bone erosion (n = 548)	Bone erosion (n = 431)	*P*
Double contour sign	316 (57.6)	317 (73.5)	<0.001
Joint effusion	432 (78.7)	229 (53.1)	<0.001
Synovial hypertrophy	265 (48.3)	256 (59.4)	0.001
Tophi	138 (25.1)	269 (62.4)	<0.001
Size of tophi, mean ± SD mm	13.3 ± 7.7	17.5 ± 11.0	<0.001

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

sive, and they had worse kidney function (higher serum BUN, creatinine, cystatin C, and uric microalbumin), worse glycemic control (higher fasting blood glucose and HbA_{1c}), faster erythrocyte sedimentation rate, longer disease duration (5.2 years versus 9.2 years; $P < 0.001$), and more frequent gout attacks (4.2 versus 9.0; $P < 0.001$) than those without bone erosions. However, no significant difference was observed in the serum uric acid levels between the bone erosion and non–bone erosion groups of patients (499.2 μmol/liter versus 498.9 μmol/liter; $P = 0.969$).

Comparison of ultrasound findings between patients with and without bone erosion. Next, we compared the differences in the ultrasound findings between patients with and those without bone erosion. As shown in Table 2, the proportion of patients with the double contour sign was 57.6% (316 of 549) in the non–bone erosion group and 73.5% (317 of 431) in the bone erosion group ($P < 0.001$). The proportion of patients with joint effusion in the non–bone erosion group was 78.7% (432 of 549), whereas the proportion was 53.1% (229 of 431) in the bone erosion group ($P < 0.001$). The proportion of articular synovial hypertrophy in the non–bone erosion group was 48.3% (265 of 549), whereas it was 59.4% (256 of 431) in the bone erosion group ($P = 0.001$). A total of 138 of 549 patients (25.1%) in the non–bone erosion group had ultrasound-detected tophi, whereas 269 of the 431 patients in the bone erosion group (62.4%) had ultrasound-detected tophi ($P < 0.001$). Of the 431 patients with bone erosions, 238 had adjacent tophi, and another 31 had no adjacent tophi. The mean sizes of the tophi were 13.3 mm and 17.5 mm in the non–bone erosion and bone erosion groups, respectively ($P < 0.001$).

Screening for factors associated with bone erosion by multivariable regression analysis. We performed a multivariate logistic regression analysis to identify variables associated with bone erosion in patients with gout. As shown in Tables 1 and 2, there were 16 factors that were identified as being significantly different between the 2 groups ($P < 0.05$). Those variables were candidate factors for bone erosion that fit in the multivariable logistic regression analysis. The statistics from the multivariable logistic regression are shown in Table 3. Age, duration of gout, the existence of tophi, and ultrasound-detected articular synovial hypertrophy were positively independently associated with bone erosion. Joint effusion was negatively independently associated with bone erosion. Although these 5 factors were correlated with each other, one was still beyond the correlation and could not be fully predicted by the others. The modified Hosmer–Lemshow goodness-of-fit test x^2 statistic was 5.195 ($P = 0.737$), which suggested that the multivariable model was a good fit ($P > 0.05$). The proportion of variance explained by the model was 40.0–53.5% (Cox-Snell $R^2 = 0.400$, Nagelkerke's $R^2 = 0.535$) and correctly classified 81.6% of cases.

Next, all the factors that were entered in the logistic model were categorized to compare the relative risk for bone erosion. The number and size were 2 variables on behalf of tophi; thus, the number of tophi and the size of tophi were categorized and replaced the existence of tophi in logistic regression analysis. As shown in Table 4, the risk of bone erosion in patients between ages 40 and 59 years and those older than 60 years of age increased 1.746-fold ($P = 0.006$) and 2.648-fold ($P < 0.001$), respectively, compared with those younger than 40 years of age. The ORs of patients with disease durations of 1–5, 5–10, and >10 years were 1.174 ($P = 0.469$), 1.618 ($P = 0.038$), and 1.949 ($P = 0.011$), respectively, compared with those with gout duration <1 year. Finally, we stratified the patients according to the number and size of tophi adjacent to the damaged joint. The results showed that the risk of bone erosion in patients with 1–2 tophi increased 3.634-fold, 10.571-fold in those with 3–4 tophi, and 15.390-fold in those with ≥5 tophi compared with patients without tophi ($P < 0.001$). However, the risk of bone erosion in the patients with tophi size <10 mm, 10–19.9 mm, and ≥20 mm did not change significantly compared with patients without tophi ($P = 0.206$), indicating that bone erosion was related to the number of tophi but not their size.

DISCUSSION

To our knowledge, this is the first study to report the prevalence and distribution of bone erosions in a large cohort of patients with gout. In this study, we found that 44% of gout

Table 3. Multivariable logistic regression of factors associated with bone erosion*

Variable	β coefficient	OR (95% CI)	*P*
Age	0.017	1.017 (1.007–1.027)	0.001
Duration of gout	0.051	1.053 (1.027–1.079)	<0.001
Tophi	1.439	4.218 (3.092–5.753)	<0.001
Joint effusion	–1.152	0.316 (0.231–0.432)	<0.001
Synovial hypertrophy	0.626	1.870 (1.378–2.538)	<0.001

* OR = odds ratio; 95% CI = 95% confidence interval.

Table 4. Incidence of bone erosion in gout patients*

Variable	Bone erosion (%)	β coefficient	OR (95% CI)	*P*
Age, years (range)				
<40	26.4		1	<0.001 for trend
40–59	45.7	0.557	1.746 (1.175–2.595)	0.006
≥60	58.9	0.974	2.648 (1.794–3.908)	<0.001
Duration of gout, years (range)				
≤1	26.9		1	0.039 for trend
between 1 and ≤5	35.3	0.161	1.174 (0.760–1.815)	0.469
between 5 and ≤10	52.1	0.478	1.613 (1.026–2.535)	0.038
>10	65.1	0.667	1.949 (1.167–3.256)	0.011
Joint effusion				
Yes	34.6		1	
No	63.3	1.184	3.266 (2.371–4.499)	<0.001
Synovial hypertrophy				
No	38.1		1	
Yes	49.1	0.639	1.894 (1.383–2.596)	<0.001
No. of tophi				
0	28.3		1	<0.001 for trend
1–2	57.4	1.288	3.624 (1.092–12.032)	0.035
3–4	75.9	2.358	10.571 (2.884–38.744)	<0.001
≥5	80.5	2.734	15.390 (3.866–61.268)	<0.001
Size of tophi, mm				
0	29.3		1	0.206 for trend
<10	61.4	0.043	1.044 (0.300–3.628)	0.946
10–19.9	63.7	−0.345	0.708 (0.206–2.437)	0.584
≥20	72.9	−0.713	0.490 (0.136–1.767)	0.276

* Values estimated using logistic regression analysis. OR = odds ratio; 95% CI = 95% confidence interval.

patients had bone erosions; the first MTP joint was the most frequently involved site, with 78.4% of defects occurring there. Bone erosion was independently associated with age, duration of gout, tophi, synovial hypertrophy, and joint effusion. A tophus was the most powerful factor associated with bone erosion in patients with gout, and the number of tophi, but not their size, was strongly associated with bone erosion in patients with gout.

Bone erosions were detected in 44% of our patients; these results are similar to the findings demonstrated in a study by McQueen et al (6), where 25 of 40 patients scanned by 3T MRI had bone erosions. In another study by Stewart et al (17), 13 of 23 patients had bone erosions detected by ultrasound. Gout has a predilection for the first MTP joint, with as many as 50–70% of first gout attacks occurring there. Thus, it was expected that 75.4% of patients with bone erosions in our study had gout in the first MTP (18). This is different from bone erosions in patients with RA, which predominantly occurs in the metacarpophalangeal joints (19) and may be related to differences in the predilection sites of the 2 diseases. Furthermore, 51.5% of patients had a single site of bone erosion, 32.0% had 2 erosions, and 6.7% had 3 erosions, indicating that most bone erosions are independent and asymmetrical. The predilection site of bone erosion was consistent with the clinical features of a gout attack.

Among the independent factors associated with bone erosion, an ultrasound-detected tophus was the most powerful factor for bone erosion in gout patients, with an OR of 4.218. This result coincides with those from other studies. McQueen et al reported that erosions are strongly associated with tophi (OR 13.0) detected by MRI (6). Another study, by Dalbeth et al, investigated tophi by CT and found that 95% of CT-detected erosions >5 mm were associated with tophi (3). Dalbeth and colleagues scored the bone erosions by depth of the erosion and found that erosions were strongly associated with tophi, particularly with larger tophi, suggesting that very small lesions can be false positives. However, these results were somewhat different from our findings; we found that, although the presence of tophi was significantly associated with bone erosions, the risk of bone erosions did not increase with size of the tophus but with the number of tophi. As the depth of bone erosion cannot be measured by ultrasound, the relationship between the size of tophi and severity of bone erosion needs further study. The cellular mechanisms of bone erosion can be considered in the context of these results. Numerous osteoclasts are present at the interface between bone and a tophus in erosive gout, and MSU crystals promote osteoclastogenesis through interactions with stromal cells. MSU crystal–induced production of catabolic enzymes and cytokines that promote osteoclastogenesis by synovial fibroblasts, macrophages, and chondrocytes may also contribute to bone erosion (3,20). In addition, increased receptor activator of nuclear factor κB ligand and macrophage colony-stimulating factor could contribute to a pro-erosive cytokine milieu, while reduced viability and function of osteoblasts suggests disordered bone homeostasis (21).

Joint inflammation is common in patients with gout. More than 50% of gout patients have articular synovial hypertrophy or joint effusion. In patients with bone erosion, the proportions of overlying synovial hypertrophy and joint effusion were 59.4% and 53.1%, respectively. Synovitis is common in patients with chronic gout. Synovitis was detected by MRI in 26 of 40 gout patients (6), and all 8 of the patients with acute gout developed wrist synovitis (22). We also investigated the association between synovial hypertrophy and erosions and found that bone sites affected by erosion were 1.87-fold more likely to lie adjacent to regions of synovial hypertrophy than not to lie adjacent. However, joint effusion was negatively correlated with bone erosion. We speculate that, because bone erosion occurs through an "outside-in" mechanism (23), exudative inflammation may prevent direct invasion of the tophus (or granulomatous synovitis) through the articular cartilage into bone and dilute MSU-induced inflammatory cytokines. These results suggest that long-term inflammatory rather than acute inflammatory stimuli that promote progressive synovial thickening are a risk factor for bone erosion.

In the present study, we identified 2 clinical factors associated with bone erosion in gout patients that were not reported before. Average age (47.0 years versus 54.6 years) and duration of gout (5.2 years versus 9.2 years) were both significantly lower in the non–bone erosion group than the bone erosion group. In addition, compared with patients ages <40 years, the risk of bone erosion increased 1.746-fold in patients 40–59 years of age and 2.648-fold in patients ages >59 years. The duration of gout was also independently associated with bone erosion, with an OR of 1.053. These results suggest that an early diagnosis and treatment would be important for preventing bone erosion in patients with gout. Due to the cross-sectional design of the present study, however, the effects of age and duration of disease on bone erosion were not included and need to be further confirmed.

There were several limitations to this study. First, although the number of participants was substantially higher than previous studies, the sample size was small. Second, ultrasound cannot be used to evaluate the depth of bone erosion. Ultrasound allows for the detection of irregularities on the cortical bone surface and performs better in easily accessible joints, but it cannot penetrate bone like CT (although this depends on the site evaluated). Third, there were 3 sonographers for the ultrasound test in this study, and they were uniformly trained; however, not all of their reports were calibrated, and interobserver reliability was not assessed, which may have caused deviations in results. Last, because the upper limb joints were not evaluated, the prevalence of bone erosion was underestimated.

In conclusion, this study demonstrates that bone erosion is a common complication of gout. Age, duration of gout, tophi, synovial hypertrophy, and joint effusion were independently associated with bone erosion, which sheds further light on the links between crystal deposition and joint damage in patients with gout. These results suggest that an early diagnosis of gout, controlling the urate level, and decreasing local urate crystal deposition may be the most effective way to prevent bone erosion in patients with gout.

ACKNOWLEDGMENT

The authors thank all team members and participants from Shanghai Jiaotong University Affiliated Sixth People's Hospital.

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. Chen 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. H. Chen.
Acquisition of data. Wu, J. Chen, L. Chen, Wei, Jiang.
Analysis and interpretation of data. Liu, Hu, Han, Lu.

REFERENCES

1. McQueen FM, Reeves Q, Dalbeth N. New insights into an old disease: advanced imaging in the diagnosis and management of gout. Postgrad Med J 2013;89:87–93.
2. Liu R, Han C, Wu D, Xia X, Gu J, Guan H, et al. Prevalence of hyperuricemia and gout in mainland China from 2000 to 2014: a systematic review and meta-analysis. Biomed Res Int 2015;2015:762820.
3. Dalbeth N, Clark B, Gregory K, Gamble G, Sheehan T, Doyle A, et al. Mechanisms of bone erosion in gout: a quantitative analysis using plain radiography and computed tomography. Ann Rheum Dis 2009;68:1290–5.
4. Dalbeth N, Collis J, Gregory K, Clark B, Robinson E, McQueen FM. Tophaceous joint disease strongly predicts hand function in patients with gout. Rheumatology (Oxford) 2007;46:1804–7.
5. Wright SA, Filippucci E, McVeigh C, Grey A, McCarron M, Grassi W, et al. High-resolution ultrasonography of the first metatarsal phalangeal joint in gout: a controlled study. Ann Rheum Dis 2007;66:859–64.
6. McQueen FM, Doyle A, Reeves Q, Gao A, Tsai A, Gamble GD, et al. Bone erosions in patients with chronic gouty arthropathy are associated with tophi but not bone oedema or synovitis: new insights from a 3T MRI study. Rheumatology (Oxford) 2014;53:95–103.
7. Sapsford M, Gamble GD, Aati O, Knight J, Horne A, Doyle AJ, et al. Relationship of bone erosion with the urate and soft tissue components of the tophus in gout: a dual energy computed tomography study. Rheumatology (Oxford) 2017;56:129–33.
8. Ventura-Rios L, Hernandez-Diaz C, Sanchez-Bringas G, Madrigal-Santillan E, Morales-Gonzalez JA, Pineda C. A descriptive, cross-sectional study characterizing bone erosions in rheumatoid arthritis and gout by ultrasound. Clin Rheumatol 2016;35:2269–76.
9. Baillet A, Gaujoux-Viala C, Mouterde G, Pham T, Tebib J, Saraux A, et al. Comparison of the efficacy of sonography, magnetic resonance imaging and conventional radiography for the detection of bone erosions in rheumatoid arthritis patients: a systematic review and meta-analysis. Rheumatology (Oxford) 2011;50:1137–47.
10. Grassi W, Filippucci E, Farina A, Salaffi F, Cervini C. Ultrasonography in the evaluation of bone erosions. Ann Rheum Dis 2001;60:98–103.
11. Dalbeth N, McQueen FM. Use of imaging to evaluate gout and other crystal deposition disorders. Curr Opin Rheumatol 2009;21:124–31.
12. Schueller-Weidekamm C, Schueller G, Aringer M, Weber M, Kainberger F. Impact of sonography in gouty arthritis: comparison

with conventional radiography, clinical examination, and laboratory findings. Eur J Radiol 2007;62:437–43.

13. Rettenbacher T, Ennemoser S, Weirich H, Ulmer H, Hartig F, Klotz W, et al. Diagnostic imaging of gout: comparison of high-resolution us versus conventional x-ray. Eur Radiol 2008;18:621–30.
14. Richette P, Doherty M, Pascual E, Barskova V, Becce F, Castaneda-Sanabria J, et al. 2016 updated EULAR evidence-based recommendations for the management of gout. Ann Rheum Dis 2017;76:29–42.
15. Neogi T, Jansen TL, Dalbeth N, Fransen J, Schumacher HR, Berendsen D, et al. 2015 gout classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheumatol 2015;67:2557–68.
16. Gutierrez M, Schmidt WA, Thiele RG, Keen HI, Kaeley GS, Naredo E, et al. International consensus for ultrasound lesions in gout: results of delphi process and web-reliability exercise. Rheumatology (Oxford) 2015;54:1797–805.
17. Stewart S, Dalbeth N, Vandal AC, Allen B, Miranda R, Rome K. Are ultrasound features at the first metatarsophalangeal joint associated with clinically-assessed pain and function? A study of people with gout, asymptomatic hyperuricaemia and normouricaemia. J Foot Ankle Res 2017;10:22.
18. Grahame R, Scott JT. Clinical survey of 354 patients with gout. Ann Rheum Dis 1970;29:461–8.
19. Tamas MM, Filippucci E, Becciolini A, Gutierrez M, Di Geso L, Bonfiglioli K, et al. Bone erosions in rheumatoid arthritis: Ultrasound findings in the early stage of the disease. Rheumatology (Oxford) 2014;53:1100–7.
20. Harre U, Derer A, Schorn C, Schett G, Herrmann M. T cells as key players for bone destruction in gouty arthritis? Arthritis Res Ther 2011;13:135.
21. Schlesinger N, Thiele RG. The pathogenesis of bone erosions in gouty arthritis. Ann Rheum Dis 2010;69:1907–12.
22. Cimmino MA, Zampogna G, Parodi M, Andracco R, Barbieri F, Paparo F, et al. MRI synovitis and bone lesions are common in acute gouty arthritis of the wrist even during the first attack. Ann Rheum Dis 2011;70:2238–9.
23. Towiwat P, Doyle AJ, Gamble GD, Tan P, Aati O, Horne A, et al. Urate crystal deposition and bone erosion in gout: 'Inside-out' or 'outside-in'? A dual-energy computed tomography study. Arthritis Res Ther 2016;18:208.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 71, No. 12, December 2019, pp 1660–1670
DOI 10.1002/acr.23811

Course of Back Pain in the Canadian Population: Trajectories, Predictors, and Outcomes

Mayilee Canizares, Y. Raja Rampersaud, and Elizabeth M. Badley

Objective. To identify and describe back pain trajectory groups and to compare indicators of health status, medication, and health care use in these groups.

Methods. A representative sample (n = 12,782) of the Canadian population was followed-up from 1994/1995 to 2010/2011. Participants were interviewed biannually and provided data on sociodemographic (e.g., education) and behavior-related (e.g., physical activity) factors, depression, comorbidities, pain, disability, medication use (e.g., opioids), and health care use (e.g., primary care visits). We used group-based trajectory analysis to categorize participants according to patterns in the course of their back pain during the 16-year follow-up period and compared indicators of pain, disability, medication, and health care use in the trajectory groups.

Results. A total of 45.6% of the participants reported back pain at least once during follow-up. Of those, we identified 4 trajectories: persistent (18.0%), developing (28.1%), recovery (20.5%), and occasional (33.4%). The persistent and developing groups were characterized as having pain that prevented activities, disability, depression, and comorbidities. There were significant differences in the patterns of medication and health care use across the groups, with a general trend of most to least health care and medication use in the persistent, developing, recovering, and occasional groups. Those in the recovery group had an increasing trajectory reflecting opioid and antidepressant use.

Conclusion. Approximately 1 in 5 people with back pain experience a persistent pain trajectory with an associated increase in pain, disability, and health care use. Further research is needed to determine whether the groups identified represent different diagnoses, which may provide insight into the selection of stratified treatment and aid in designing early prevention and management strategies in the population.

INTRODUCTION

Back pain is among the most frequently reported health problems worldwide (1). Systematic reviews of the global burden of back pain indicate that a mean ± SD of 18.3% ± 11.7 of the adult population report back pain at any given time and that a mean ± SD of 39.8% ± 24.3 of the population are affected by back pain during their lifetime (1,2). It has also been recognized that back pain often presents as recurrent episodes (3,4), with studies showing that although many patients recover from back pain, others have symptoms persisting after 1 year (5). In addition, back pain is associated with disability, increased health care use and costs, and loss of productivity (2,6,7). Given that back pain is often recurrent, it is important to understand the course of back pain over time since this can provide additional insights into risk factors for unfavorable outcomes.

Previous studies have examined the trajectories of back pain (3). The majority of these involved patients who presented to primary care settings with back pain (8–14), had a restricted age range (8,11,15,16), and/or had short follow-up times (≤1 year) (4,8–10,12). Only a few studies have examined back pain in the general population (4,15,16) using a wide age range (4). Although these studies have methodologic differences, they have identified common trajectories of persistent, recovery, and developing back pain (3). Some of these studies have examined factors associated with the back pain trajectories identified (4,8–10,12,13); however, none have compared health-related outcomes in the back pain trajectory groups identified in the current study. Whether the patterns of back pain over time observed in the clinical samples of prior studies are also seen in the general population remains to be examined.

Mayilee Canizares, PhD, Y. Raja Rampersaud, MD, FRCSC, Elizabeth M. Badley, DPhil: Krembil Research Institute, University Health Network, Toronto, Ontario, Canada.

Dr. Rampersaud has received consulting fees from Medtronic (less than $10,000). No other disclosures relevant to this article were reported.

Address correspondence to Mayilee Canizares, PhD, the Arthritis Program, Krembil Research Institute, University Health Network, 399 Bathurst Street, 10MP-326 Toronto, Ontario M5T 2S8, Canada. E-mail: mayilee.canizares@uhnresearch.ca.

Submitted for publication March 5, 2018; accepted in revised form November 6, 2018.

SIGNIFICANCE & INNOVATIONS

- This population-based longitudinal study identified 4 back pain trajectory groups: persistent, developing, recovery, and occasional.
- Obesity, physical activity (leisurely and daily), smoking, depression, and physical chronic conditions were the main factors associated with trajectory group membership.
- Those in the persistent and developing trajectory groups had more pain and disability and increased use of health care and medication than those in the recovery and occasional trajectory groups.
- The recovery trajectory group showed increased use of opioids and antidepressants over time, suggesting that those recovering from back pain need further monitoring as they continue to receive medication.

To fill this gap in the literature, we aimed to identify back pain trajectory groups in the general population (15+ years of age) using data from a Canadian longitudinal study that spanned 16 years. Our intent was to examine whether the trajectory groups are related to sociodemographic factors (e.g., education), behavior-related factors (e.g., obesity), and comorbidities and to compare the longitudinal trajectories of health status (e.g., disability), medication use (e.g., opioids), and health care use (e.g., visits to physicians) across the previously identified pain trajectory groups.

MATERIALS AND METHODS

Data source. This study used data from the longitudinal component of the National Population Health Survey (NPHS). The NPHS surveyed a representative sample of the Canadian population from 1994/1995 to 2010/2011. The survey's target population included household residents in the 10 Canadian provinces in 1994/1995 (cycle 1), excluding persons living on Indian Reserves and Crown Lands, residents of health institutions, full-time members of the Canadian Forces bases, and persons living in some remote areas of Ontario and Quebec. The survey collected data every 2 years, thus providing 9 cycles of data. More detailed descriptions of the NPHS design and interview procedures are available from Statistics Canada (17).

Back pain. For each cycle, the NPHS collected data on chronic conditions (up to 16) that had been diagnosed by a health care professional and had lasted for at least 6 months. Back pain, excluding arthritis and fibromyalgia, was included in the list.

Covariates. We examined 4 sociodemographic variables: age, sex, education level, and household income. For each cycle, the age (in years) was calculated. The number of years of schooling was used as a measure of educational attainment and grouped as <12 years, 12–15 years, and ≥16 years. Household income was grouped in quartiles of the overall distribution, and nonresponses were kept in a separate category for analysis.

Behavior-related factors such as obesity, participation in physical activity, daily activities (e.g., usual activity or work habits), and smoking status were also included. Body mass index (BMI) was grouped as underweight/normal weight (<25.0), overweight (25–29.9), and obese (≥30.0). Information on participation in various leisurely physical activities, such as walking for exercise, running, and gardening, was collected. Based on their total energy expenditure, participants were categorized as inactive (<1.5 kcal/kg/day) and active (≥1.5 kcal/kg/day) (18). Participants were also asked to describe their usual day or work habits. Daily activities were categorized as: sedentary (those who usually sit during the day and do not walk around very much), light (those who stand or walk quite a lot during the day but do not have to carry or lift things very often), and moderate/heavy (those who usually lift or carry light loads or have to climb stairs or hills often, or those who do heavy work or carry very heavy loads). Smoking status was assessed by a variable derived from Statistics Canada that grouped participants as current smoker, former smoker, and nonsmoker (i.e., those who never smoked) (17).

In addition to back pain, participants reported the following chronic conditions: arthritis, asthma, allergies (excluding food allergies), bronchitis, emphysema, diabetes mellitus, high blood pressure, heart conditions, stroke, cancer, ulcers, urinary incontinency, migraine, glaucoma, and cataracts. Symptoms of depression were ascertained by responses (yes/no) to the question, "During the past 12 months, was there ever a time when you felt sad, blue, or depressed for 2 weeks or more in a row?"

Health-related outcomes. Indicators of health status, medication use, and health care use in the trajectory groups were compared. Indicators of health status that were examined were self-rated health (poor/fair versus good/very good/excellent), pain that prevents activities (few/some/most activities versus no pain/pain does not prevent activities), and disability measured as the number of activities for which participants needed help (at least 1 of 6 versus 0). Indicators of medication use (yes/no) were pain relievers (e.g., aspirin, acetaminophen), opioids (e.g., codeine, meperidine, and morphine), and antidepressants. We examined 6 indicators of health care use: primary care physician visits (2+ visits versus 0–1), specialist visits (1+ visits versus 0), mental health consultations (yes/no), physiotherapist visits (1+ visits versus 0), chiropractor visits (1+ visits versus 0), and visits to massage therapists (yes/no).

Statistical analysis. The first objective was to identify back pain trajectory groups. To do this, we used group-based trajectory analysis (19). Briefly, group-based trajectory analysis assumes

that there are various distinct trajectories of back pain and that participants can be allocated into 1 of many distinct groups based on patterns of back pain over time. To facilitate clinical interpretability, we first allocated respondents who did not report back pain over follow-up (54.4% of the sample) into a no back pain trajectory group. We then fit a group-based trajectory model on the remaining sample of participants, who reported back pain at least once over follow-up. Group-based trajectory models for binary outcomes were fit systematically, starting with a 1-group model (where it was assumed that all participants have the same course of back pain) and then adding another group for each successive model. The optimal number of back pain trajectory groups was chosen when Bayesian information criterion statistics remained stable. These were models in which the shape of the trajectory did not change substantially when higher order polynomials were included and each group had at least 5% of the analytical sample. We also examined the average posterior probabilities of group membership to assess model fit. Participants were allocated to a group for which they had the greatest posterior probability. Average posterior probabilities in each trajectory group were computed, and values >0.70 indicated high internal reliability (19).

The second objective was to examine the association of sociodemographic and behavior-related factors and comorbidities measured at baseline with trajectory group membership, with those not reporting back pain as reference. To address this objective, we used multinomial logistic regression.

The third objective was to compare the patterns of change over time in health-related outcomes by trajectory groups. To address this objective, we fit 2-level multilevel growth models. Multilevel growth models have been widely used in longitudinal studies that examine change in outcomes and are appropriate for longitudinal data where there are multiple observations from each participant, as is the case with the NPHS data (20). These are 2-level models, where level 2 is represented by the individual and level 1 by the repeated observations. For each outcome, a 2-level logistic model was fit, including the trajectory group variable, linear and quadratic terms for the time variable, and interactions between the time variables and the trajectory groups. In this way, we could test whether the rate of change over time for each outcome differed for each back pain trajectory group. All models were adjusted for sociodemographic and behavior-related factors and comorbidities. These covariates, with the exception of sex, were included as time varying. For ease of interpretation, the results of these analyses are presented graphically as full models in Supplementary Tables 1–5, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23811/abstract.

All analyses were conducted using SAS/STAT software (21). To fit the group-based models, we used SAS Proc Traj, which is an add-on available for free download (21,22). All models were fit to include incomplete cases up to the point when subjects dropped out or died, and maximum likelihood estimators were used to adjust for nonresponse, assuming that the data were missing at random. Multinomial regression and growth models were fit using data weights provided by Statistics Canada, which accounted for sample design and included adjustments for nonresponse and poststratification (17).

Secondary analyses. Given the long follow-up of the NPHS, 32.1% of the baseline sample died or dropped out during the course of the study. In an attempt to differentiate between participants who remained throughout the study from those who died or dropped out, baseline demographic and health characteristics were examined using nominal logistic regression. The outcome was a 3-level variable: dropped out, died, respondent in the last cycle of data collection. Participants who died during the course of the study were more likely to be older, men, severely obese, and smokers, to have lower income and/or education, to have sedentary behaviors, and to have reported multiple chronic conditions. No differences were related to physical activity. Participants who dropped out of the study were more likely to be younger, men, smokers, and physically active and to have lower income and/or education. No differences were related to BMI, sedentary lifestyle, and the number of chronic conditions reported.

Because these comparisons indicated that there are differences in important variables between participants who remained in the study and those who died or dropped out, we conducted 2 secondary analyses to examine the effect of attrition on the results. We re-estimated the trajectories that included variables indicating if the participant died or dropped out during follow-up, and we re-estimated the trajectories for participants with complete data in all 9 cycles of data collection.

One limitation of the 2-stage approach to develop trajectory groups is that it does not consider uncertainty regarding trajectory group membership. However, if we had developed the trajectory groups to include the baseline risk factors (the 1-stage approach), then the models for the outcomes would be overadjusted since we were also interested in comparing changes in outcome over time across trajectory groups and since back pain and the outcomes examined share many of the risk factors studied (e.g., age, obesity). Nevertheless, we conducted a secondary analysis in which we fit the trajectory group model including the baseline risk factors studied and compared the results with the main findings.

RESULTS

Sample description. For our analyses, we included participants who were ≥15 years in 1994/1995 and had at least 3 cycles of data from baseline onwards. This resulted in a sample of 12,782 participants at baseline: 51.0% were female with a median age of 39 years (range 15–95 years), and 14.1% and 19.5% reported back pain at baseline and at the last cycle of data

collection, respectively. Furthermore, 45.6% of the participants reported back pain at least once during follow-up. No significant differences in education, income, and physical activity were found between those with and without back pain. Those with back pain had a higher percentage of being overweight, obese, and smokers, of engaging in moderate-to-heavy, physically demanding daily activities, and of having chronic conditions, particularly arthritis, depression, high blood pressure, and migraines (Table 1).

Trajectories of back pain. We first identified a group of participants who did not report back pain (no back pain) over the follow-up period (54.4% of the sample). We then fit the

Table 1. Baseline characteristics of patients reporting back pain at least once (45.6%) compared to those without back pain (54.4%). Canadian National Population Health Survey, 1994/1995–2010/2011*

	Back pain ≥1 (n = 5,828)	No back pain (n = 6,954)	*P*†
Sociodemographic			
Age groups, years			<0.0001
15–24	13.2 (12.0–14.4)‡	21.2 (19.8–22.6)	
25–34	20.3 (18.9–21.7)	21.6 (20.3–22.9)	
35–44	22.8 (21.4–24.2)	21.5 (20.1–22.9)	
45–54	17.0 (15.8–18.2)‡	14.5 (13.4–15.6)	
55–64	13.3 (12.2–14.4)‡	9.3 (8.5–10.1)	
65+	13.4 (12.3–14.5)‡	11.6 (10.7–12.5)	
Sex, female	52.8 (51.1–54.5)‡	49.6 (48.0–51.2)	<0.0001
Education, years			
<12	28.7 (27.2–30.2)	28.0 (26.6–29.4)	0.1394
12–15	57.9 (56.2–59.6)	57.1 (55.5–58.7)	
≥16	13.4 (12.2–14.6)	14.8 (13.6–16.0)	
Income quartiles			0.2462
Q1 (bottom)	19.1 (17.9–20.3)	17.6 (16.5–18.7)	
Q2	27.8 (26.3–29.3)	25.9 (24.6–27.2)	
Q3	24.1 (22.6–25.6)	25.9 (24.5–27.3)	
Q4	24.5 (23.0–26.0)	26.1 (24.7–27.5)	
Not reported	4.4 (3.7–5.1)	4.5 (3.8–5.2)	
Health-related behaviors			
BMI groups			<0.0001
Underweight/normal	49.5 (47.8–51.2)‡	56.3 (54.7–57.9)	
Overweight	36.0 (34.3–37.7)‡	33.1 (31.6–34.6)	
Obese	14.5 (13.3–15.7)‡	10.6 (9.6–11.6)	
Physical activity			0.9401
Inactive	11.5 (10.3–12.7)	11.0 (10.0–12.0)	
Light	47.8 (46.1–49.5)	48.2 (46.6–49.8)	
Moderate/vigorous	40.7 (39.0–42.4)	40.8 (39.2–42.4)	
Daily activities			0.0287
Sedentary	21.2 (19.9–22.6)	22.5 (21.2–23.8)	
Light	49.6 (48.0–51.3)	51.3 (49.7–52.9)	
Moderate/heavy	29.1 (27.5–30.7)‡	26.1 (24.6–27.6)	
Smoking status			<0.0001
Current smokers	33.6 (32.0–35.2)‡	27.8 (26.4–29.2)	
Past smokers	31.6 (30.0–33.1)‡	29.4 (28.0–30.8)	
Nonsmokers	34.8 (33.2–36.5)‡	42.8 (41.2–44.4)	
Chronic conditions			
Depression	13.2 (12.1–14.3)‡	8.9 (8.0–9.8)	<0.0001
Arthritis	18.4 (17.2–19.6)‡	8.0 (7.2–8.8)	<0.0001
Asthma	7.0 (6.1–7.9)‡	5.1 (4.4–5.8)	<0.0001
Bronchitis	3.7 (3.1–4.3)‡	2.2 (1.8–2.6)	0.0002
Heart disease	4.0 (3.4–4.6)‡	2.8 (2.3–3.3)	<0.0001
Stroke	0.9 (0.6–1.2)‡	0.5 (0.3–0.7)	0.0060
High blood pressure	11.0 (10.0–12.0)‡	6.9 (6.2–7.6)	<0.0001
Diabetes mellitus	3.0 (2.5–3.5)	2.6 (2.1–3.1)	0.1910
Cancer	1.7 (1.3–2.1)‡	1.0 (0.6–1.4)	0.0028
Migraine	9.7 (8.8–10.6)‡	5.4 (4.6–6.2)	<0.0001
Ulcers	4.8 (4.2–5.4)‡	2.1 (1.7–2.5)	<0.0001
Urinary incontinence	1.5 (1.1–1.9)‡	0.4 (0.2–0.6)	<0.0001

* Values are the percentage (95% confidence interval). BMI = body mass index.
† Based on chi-square test.
‡ $P < 0.05$ for comparing proportions by back pain status.

group-based trajectory model to identify back pain trajectory groups on the subsample of participants reporting back pain at least once over follow-up. The results of this analysis identified 4 distinct trajectory groups: persistent, developing, recovery, and occasional (Figure 1).

The average posterior probability ranged from 0.72 for the recovery group to 0.91 for the occasional group, indicative of good model fit. The persistent group accounted for 8.1% of the total sample and 18.0% of those reporting back pain. This group had a very high probability of reporting back pain throughout follow-up (the proportion of time that back pain was reported was 80% [range 67–100%]). The developing group, with a relatively low probability of reporting back pain initially, had an increasing probability of back pain throughout follow-up (the proportion of time that back pain was reported was 45% [range 33–67%]). This group accounted for 12.8% of the sample (28.1% of those with back pain). The next group, the recovery group, included participants who reported back pain in the first cycles of the study (96% reported back pain prior to cycle 4) and then stopped reporting back pain over time (the proportion of time that back pain was reported was 45% [range 33–67%]). This group accounted for 9.4% of all participants and 20.5% of those with back pain. The fourth group that we identified (the occasional group) accounted for 15.2% of the sample (33.4% of those with back pain). Members of this group reported back pain only once during follow-up. The characteristics of these trajectory groups are presented in Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23811/abstract.

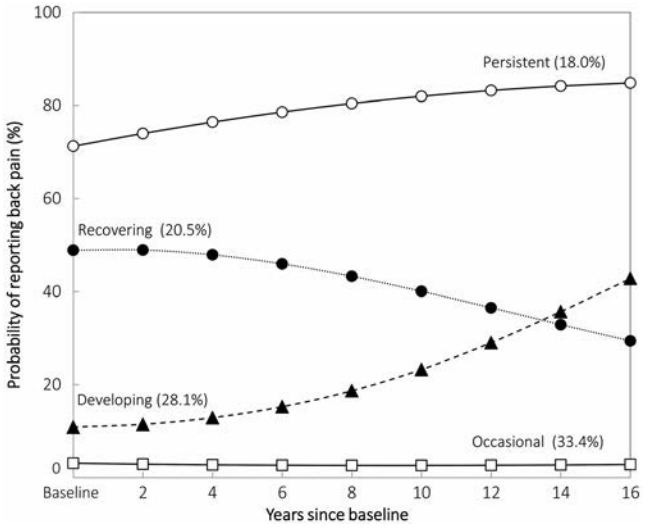

Figure 1. Results from a group-based trajectory analysis of back pain. Canadian National Population Health Survey, 1994/1995–2010/2011. The predicted probability of reporting back pain is along the y-axis, and the follow-up years since baseline are along the x-axis. Each line represents 1 distinct trajectory. The percentage of the back pain population for each trajectory is shown.

Individual characteristics of the trajectory groups. Table 2 shows the results of a multivariate, multinomial, logistic regression model that examines the factors associated with trajectory group membership, with the no back pain group used as reference. Odds ratios (ORs) for the persistent group peaked in the 45–54 year age group. For the developing and recovery groups, ORs peaked in the 55–64 year age group, while for the occasional group, ORs peaked in the 35–44 year age group. Women were more likely to be in the developing group (OR 1.28 [95% confidence interval (95% CI) 1.05–1.56]), although no significant differences were seen between men and women in the other trajectory groups. Education was only associated with the occasional group, such that those with a lower level of education were more likely to be allocated into this group. Income was not significantly associated with the trajectory groups. Of the behavior-related factors examined, obesity was significantly associated with experiencing persistent (OR 1.73 [95% CI 1.39–2.15]), developing (OR 1.25 [95% CI 1.04–1.50]), or occasional back pain (OR 1.25 [95% CI 1.06–1.47]). Participation in leisurely physical activities was significantly associated with all the trajectory groups. However, the direction of the association differed in each back pain group. Compared to inactive individuals, those who were physically active were more likely to be in the developing or occasional groups and were less likely to be allocated into the persistent or recovery groups. Those who reported more physically demanding daily activities (e.g., lifting light or heavy loads) compared to those with sedentary activities were more likely to be in any of the back pain trajectory groups except the persistent group. Furthermore, current smokers were more likely to be in any of the trajectories compared to nonsmokers. There appeared to be a gradient in the odds of reporting comorbidities from highest to lowest for those in the persistent, developing, recovery, and occasional back pain groups.

Comparing outcomes by trajectory groups. The findings from the multilevel logistic growth models are presented in Figures 2–4. These models compare changes over time in indicators of health status, medication use, and health care use in the back pain trajectory groups, adjusting for sociodemographic and behavior-related factors and comorbidities (for the full models, see Supplementary Tables 1–5, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23811/abstract). The estimates are the log odds for each covariate. In the preliminary analysis, we found no significant differences in these outcomes between the occasional and no back pain groups. Therefore, we have presented the findings for these groups as occasional/no back pain.

The trajectory for those who rated their health as poor and who had pain that prevented activities mirrored the trajectories seen in the trajectory groups (Figure 2A and 2B, respectively),

Table 2. Factors associated with trajectory group membership: results from a multinomial logistic regression model with the no back pain group as reference. Canadian National Population Health Survey, 1994/1995–2010/2011*

	Persistent	Developing	Recovery	Occasional
Age groups, years				
15–24, reference				
25–34	1.58 (1.19–2.10)†	1.14 (0.93–1.40)	1.71 (1.32–2.22)§	1.69 (1.40–2.04)§
35–44	2.00 (1.52–2.63)§	1.34 (1.10–1.63)†	1.86 (1.44–2.40)§	1.83 (1.52–2.20)§
45–54	2.11 (1.56–2.85)§	1.32 (1.06–1.64)‡	2.42 (1.86–3.15)§	1.62 (1.30–2.02)§
55–64	1.92 (1.39–2.65)§	1.65 (1.30–2.09)§	2.56 (1.93–3.40)§	1.65 (1.32–2.06)§
65+	0.88 (0.62–1.25)	1.37 (1.07–1.75)‡	1.69 (1.25–2.28)†	1.28 (1.00–1.64)‡
Sex, female	1.02 (0.87–1.20)	1.17 (1.04–1.32)‡	0.98 (0.88–1.09)	1.03 (0.92–1.15)
Education, years				
≥16, reference				
<12	0.99 (0.76–1.29)	0.86 (0.71–1.04)	1.13 (0.89–1.43)	1.25 (1.02–1.53)‡
12–15	1.18 (0.93–1.50)	0.80 (0.67–0.96)‡	1.05 (0.85–1.30)	1.16 (0.98–1.37)¶
Income quartiles				
Q4 (top), reference				
Q1 (bottom)	1.09 (0.86–1.38)	0.98 (0.81–1.19)	0.96 (0.78–1.18)	1.06 (0.89–1.26)
Q2	1.09 (0.88–1.35)	1.00 (0.85–1.18)	0.85 (0.69–1.05)	1.09 (0.90–1.32)
Q3	1.05 (0.85–1.30)	0.95 (0.86–1.05)	0.99 (0.82–1.20)	0.96 (0.82–1.12)
Not reported	0.65 (0.41–1.03)	0.95 (0.63–1.43)	1.15 (0.83–1.59)	1.16 (0.86–1.56)
BMI				
Normal, reference				
Overweight	1.26 (1.06–1.50)†	1.13 (0.99–1.29)¶	1.10 (0.95–1.27)	1.00 (0.89–1.12)
Obese	1.73 (1.39–2.15)§	1.25 (1.04–1.50)‡	0.94 (0.76–1.16)	1.25 (1.06–1.47)†
Physical activity				
Inactive, reference				
Light	0.82 (0.69–0.97)‡	1.27 (1.02–1.58)‡	0.79 (0.64–0.98)‡	1.16 (0.97–1.39)¶
Moderate/vigorous	0.87 (0.69–1.10)	1.41 (1.13–1.76)†	0.85 (0.69–1.05)	1.25 (1.04–1.50)‡
Daily activities				
Sitting, reference				
Light	0.83 (0.69–0.99)‡	1.00 (0.86–1.16)	1.03 (0.87–1.22)	1.05 (0.91–1.21)
Moderate/heavy	0.94 (0.76–1.16)	1.29 (1.09–1.53)†	1.31 (1.08–1.59)†	1.24 (1.06–1.45)‡
Smoking status				
Nonsmokers, reference				
Current smokers	1.51 (1.26–1.81)§	1.40 (1.21–1.62)§	1.38 (1.17–1.63)†	1.10 (0.97–1.25)
Past smokers	1.21 (1.01–1.45)‡	1.13 (0.98–1.30)¶	1.23 (1.05–1.44)‡	1.04 (0.91–1.19)
Chronic conditions				
Depression	2.02 (1.64–2.49)§	1.53 (1.28–1.83)§	1.54 (1.26–1.88)§	1.18 (0.99–1.41)¶
Arthritis	3.35 (2.72–4.13)§	1.78 (1.47–2.16)§	2.74 (2.28–3.29)§	1.94 (1.62–2.32)§
Asthma	1.51 (1.13–2.02)†	1.67 (1.32–2.11)§	1.52 (1.17–1.97)†	1.20 (0.95–1.52)
Bronchitis	1.00 (0.66–1.52)	0.91 (0.63–1.31)	1.46 (1.04–2.05)‡	1.16 (0.84–1.60)
Heart disease	1.96 (1.39–2.76)†	0.81 (0.57–1.15)	0.98 (0.68–1.41)	0.80 (0.57–1.12)
Stroke	1.08 (0.52–2.24)	1.63 (0.87–3.05)	0.72 (0.31–1.67)	0.73 (0.34–1.57)
High blood pressure	1.04 (0.79–1.37)	1.29 (1.04–1.60)‡	1.25 (0.99–1.58)¶	1.18 (0.96–1.45)
Diabetes mellitus	0.92 (0.61–1.39)	0.70 (0.47–1.04)	0.75 (0.49–1.15)	1.11 (0.80–1.54)
Cancer	1.73 (1.05–2.85)‡	1.07 (0.66–1.73)	1.89 (1.18–3.03)†	0.87 (0.50–1.51)
Migraine	2.15 (1.70–2.72)§	1.29 (1.03–1.62)§	2.05 (1.64–2.56)§	1.34 (1.09–1.65)†
Ulcers	2.23 (1.61–3.09)§	1.92 (1.42–2.60)§	1.46 (1.30–1.64)†	1.14 (0.86–1.51)
Urinary incontinence	3.37 (1.81–6.27)§	2.69 (1.49–4.86)†	2.53 (1.35–4.74)†	0.93 (0.43–2.01)

* Values are odds ratios (95% confidence interval). BMI = body mass index.
† $P < 0.01$.
‡ $P < 0.05$.
§ $P < 0.0001$.
¶ $P < 0.1$.

increasing for the developing group, declining for the recovery group, and remaining stable for the persistent and occasional/no back pain groups. In contrast, the trajectory of disability increased over time in all back pain trajectory groups (Figure 2C), with a general trend of greater disability for those in the persistent group, followed by the developing and recovery groups and the occasional/no back pain group.

Medication use was highest for those in the persistent group, followed by the developing and recovery groups and the occasional/no back pain group (Figure 3A and 3B). Use of opioids increased over time in all trajectory groups (Figure 3A), but the increase was more marked for those in the persistent and developing groups. Noticeably, those in the recovery group continue to receive opioids over time. Like-

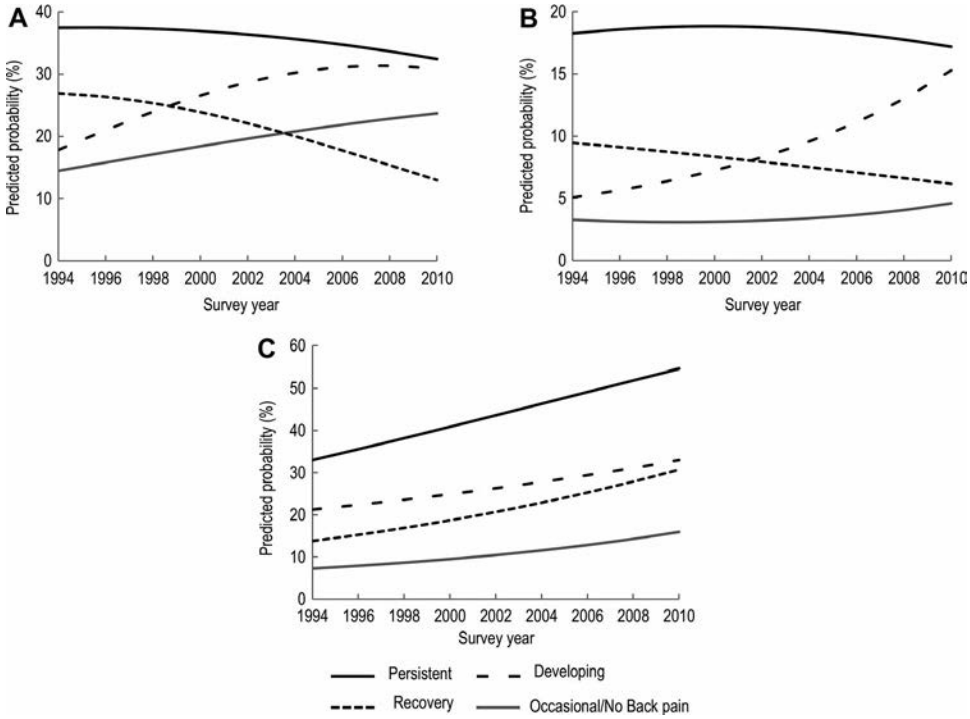

Figure 2. Results of multilevel growth models of health status in back pain trajectory groups. Canadian National Population Health Survey, 1994/1995–2010/2011. **A**, Poor/fair self-rated health. **B**, Presence of pain that prevents activities. **C**, Disability. All covariates were held at their means when calculating predicted values. Full models are presented in Supplementary Table 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23811/abstract.

wise, use of antidepressants increased over time across all groups (Figure 3B), and no significant differences were seen in the trajectories of antidepressant use between the recovery and developing groups. Similar patterns were seen in the use of pain relievers (data shown in Supplementary Tables 1–5, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23811/abstract).

The results of comparing health care use indicators in the back pain trajectory groups are shown in Figure 3C and 3D for physician visits and Figure 4 for physiotherapy, chiropractic, massage therapy,

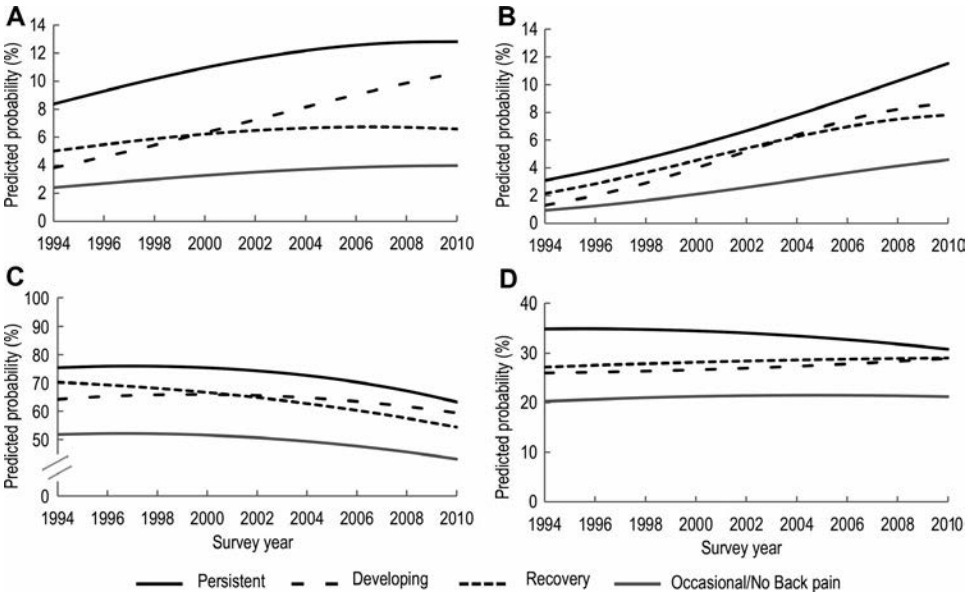

Figure 3. Results of multilevel growth models of medication use and visits to physicians in back pain trajectory groups. Canadian National Population Health Survey, 1994/1995–2010/2011. **A**, Opioids. **B**, Antidepressants. **C**, Primary care physician visits (2+). **D**, Specialist visits (1+). All covariates were held at their means when calculating predicted values. Full models are presented in Supplementary Tables 3 and 4.

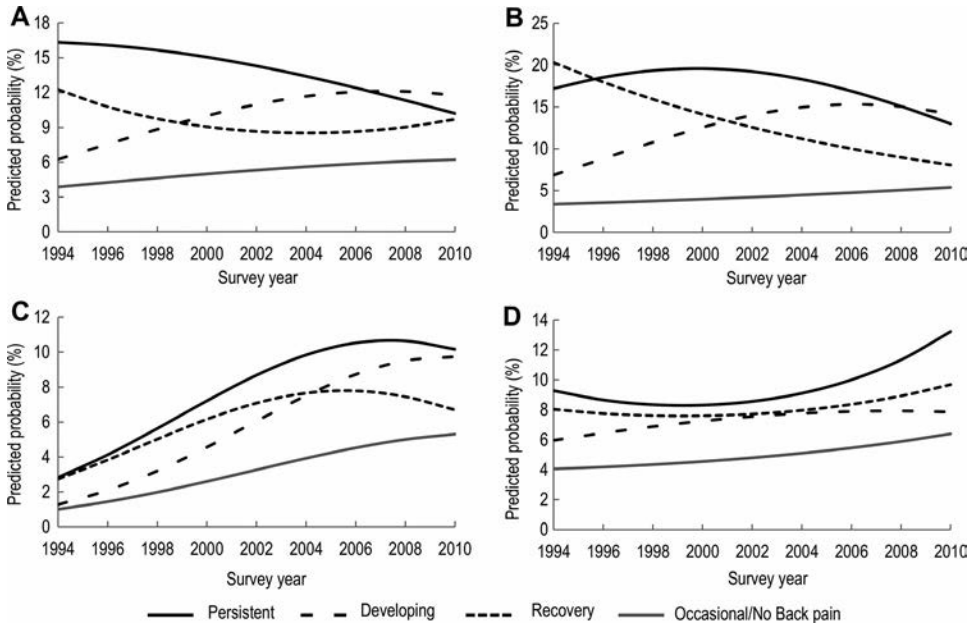

Figure 4. Results of multilevel growth models of health care use in back pain trajectory groups. Canadian National Population Health Survey, 1994/1995–2010/2011. **A**, Physiotherapy visits (1+). **B**, Chiropractic visits (1+). **C**, Consulted with massage therapists. **D**, Consulted with mental health professionals. All covariates were held at their means when calculating predicted values. Full models are presented in Supplementary Tables 3 and 5.

and mental health services. Overall, those experiencing a trajectory of persistent back pain used health care services more often compared to those with occasional/no back pain. The trajectories for primary care (Figure 3C), specialist care (Figure 3D), and mental health services (Figure 4D) were relatively stable over time across all back pain groups, and the trajectories for the developing and recovery groups were not significantly different. In addition, physiotherapy (Figure 4A) and chiropractic visits (Figure 4B) declined over time for those in the persistent and recovery groups but increased among those in the developing group. Use of massage therapy increased over time in all of the groups (Figure 4C).

Secondary analyses. In our analysis of the impact of attrition, we found that adjusting for death and dropouts resulted in findings similar to those of the main analyses. In addition, our analysis of patients with data for all 9 cycles yielded similar trajectory groups. The factors associated with the back pain groups and the differences in outcome by back pain groups showed patterns similar to those seen in our main findings. The results from a trajectory model adjusted for risk factors at baseline compared with those from an unadjusted trajectory (the main results) were not substantially different.

DISCUSSION

This study provides insights into the long-term course of back pain in the general population. Our analyses, based on data from a longitudinal survey of the Canadian population from 1994/1995 to 2010/2011, show that ~50% of responders reported back pain at least once during the 16-year follow-up. We identified 4 distinct trajectory groups: those who persistently reported back pain; those who developed back pain over time; those who recovered; and those who occasionally reported back pain over time. A major finding of this study is the negative impact of persistent back pain on a range of health-related outcomes, including health care use, after adjustments for sociodemographic and behavior-related factors and comorbidities. These findings are in accord with studies that suggest that up to one-quarter of patients with low back pain report this condition persistently and that they account for 75% of the total cost attributable to low back pain (23–25).The study findings also provide strong evidence that overall back pain trajectories in the population obscure the complexity of the course of back pain over time and counter the notion that back pain will resolve in the vast majority of cases (26).

Our study is not directly comparable to previous studies of back pain trajectories. Ours used a representative sample of the general population with a wide age range (15+ years) and a long follow-up time (16 years), whereas most of the previous studies used clinical samples with a restricted age range (3). However, our back pain trajectory groups are somewhat compatible with previous studies (4,9,13,15,16,27). For instance, our groups are similar to those in 2 population studies of adolescents transitioning to adulthood (15,16). One study (15) found trajectories of low, increasing, recovery, and persistent back pain among 17-year-old patients who were followed-up for 5 years. Whereas another study (16) found 3 trajectories of musculoskeletal pain (low,

increasing, and persistent) in a sample of 16-year-old patients who were followed-up at ages 21, 30, and 43 years old. Dunn et al (9) also found trajectories of persistent and recovery back pain in patients who visited primary care settings for back pain. A study (4) that collected daily data for 1 year on pain intensity in the general population identified 3 different persistent courses but did not find a recovery group. Another study (27), which used data on 50-year-old patients from the general population, found 3 distinct back pain trajectories: those who never reported back pain; those whose reports fluctuated; and those who reported back pain persistently.

In keeping with past research, we found that having comorbidities was linked to higher odds of reporting back pain (7,28), particularly persistent back pain (15,16). In addition to comorbidities, obesity, physical activity, and smoking were associated with trajectory group membership. We found that obese individuals were more likely to experience trajectories of persistent, developing, and occasional back pain. This is inconsistent with studies that show no association between BMI and back pain trajectories (4,11,16). Furthermore, the finding that physically demanding daily activities were associated with the developing, recovery, and occasional trajectory groups is in accord with previous observations (29). These findings are somewhat paradoxical since some of the risk factors that were associated with persistent and developing back pain were also linked to recovery from back pain. Future studies, including established psychological risk factors of back pain, such as low mood and pain somatization, are warranted (14,30,31).

New findings from our study include the negative impact of persistent back pain for a range of indicators of health services and medication use as well as for disability and pain that prevents activities. No study has compared indicators of health care and medication use in back pain trajectory groups, and few studies have examined pain and disability in relation to the course of back pain (4,5,13). One important difference between these studies and ours is that they examined the effect of health status measured at baseline in relation to trajectory group membership, while we compared the changes in these outcomes over time in the back pain trajectory groups. Nevertheless, compatible with our findings, these studies noted that having greater disability and more pain was strongly associated with experiencing persistent back pain (4,5,13).

Noticeably, the trajectories of the outcomes examined aligned in general with those seen in the trajectory groups, particularly for the persistent and occasional groups. For instance, the developing group had increasing trajectories of pain that prevents activities over time. However, this was not always true for the recovery and occasional groups, where treatment with opioids and antidepressants increased over time. A possible explanation for this is that given the long follow-up time of this study, some people in the recovery group may have been using these medications for several years after their back pain had resolved. If this was the case, then there are potential long-term consequences that highlight the importance of clinical intervention to modify or stop medication use as patients with back pain recover. On the other hand, it can be argued that continued recovery is related to ongoing drug intervention; however, it is unlikely that long-term treatment with opioids is justifiable. As we controlled for the presence of comorbidities in our analyses, ongoing treatment with opioids is not likely to be related to co-occurring conditions. Another possibility is that continued treatment with opioids in the recovery and occasional groups is related, at least partially, to the increase in opioid use (medical and nonmedical) in the general population, which has been reported previously (32–34).

It is unclear what the developing group represents. It is likely that a portion of this group may represent people who would transition to a more persistent stage, particularly since the trajectories of the outcomes for this group over time were similar to those for the persistent group. Given the high impact of back pain on both the individual and the health care system, there is a need for further research to understand the characteristics of this group. Future long-term studies that record back pain from the onset and assessing it in short time intervals (e.g., monthly) would help to fully understand the dynamics of back pain over time. It has been noted that repeated monthly pain assessments reduce recall bias while better reproducing the course of back pain over time (35).

An important limitation of the current study is that our analyses are based on self-reported data. Studies have shown that self-reported measures are subject to recall and measurement bias. Such biases should be kept in mind when interpreting our findings. We do not expect the patterns seen over time to be affected by recall or measurement bias since these biases are unlikely to vary over time. Our analyses may also be affected potentially by selection bias; participants with 1 or 2 follow-up measurements were excluded from the analyses, and the impact of these exclusions on the results is unknown. In addition, the survey that we used employs a crude standard for measuring back pain; thus, we cannot differentiate the specific nature of back pain studied, such as neck pain, low back pain, or other spinal problems. The study was subject to attrition due to dropouts and mortality. Group-based trajectory and growth models were fit using all the data available until the participants died or dropped out of the study. We conducted a series of supplementary analyses to examine the impact of attrition on the results. Analyses including indicator variables that identify those who died or dropped out during follow-up and analyses based on a sample of respondents to all 9 cycles yielded trajectory groups similar to those presented in the main analyses. These limitations notwithstanding, this study extends previous work in clinical samples by comprehensively examining the trajectories of back pain in the general population, including patients with a wide age range (15+ years) and who were followed-up for a long period of time. We

also have expanded on previous studies since we compared a variety of health outcomes in back pain trajectory groups after controlling for sociodemographic and behavior-related factors and comorbidities.

In conclusion, approximately 1 in 5 people with back pain is likely to experience a persistent trajectory, 1 in 3 develops back pain over time, and 1 in 4 recovers. The study showed that persistent back pain is associated with persistent symptoms as well as increased health care and medication use. The results have important implications for the way that we understand back pain since the different trajectory patterns potentially represent subgroups in the population that may require different interventions. In light of the trend of marked worse outcomes, particularly for the persistent and developing groups, studies are needed to determine the nature of these groups and to identify factors that may facilitate early identification and mitigation of unfavorable outcomes.

ACKNOWLEDGMENT

We thank the Statistics Canada Research Data Centres Program for providing access to the data.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Hoy D, Bain C, Williams G, March L, Brooks P, Blyth F, et al. A systematic review of the global prevalence of low back pain. Arthritis Rheum 2012;64:2028–37.
2. Hoy D, March L, Brooks P, Blyth F, Woolf A, Bain C, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis 2014;73:968–74.
3. Kongsted A, Kent P, Axen I, Downie AS, Dunn KM. What have we learned from ten years of trajectory research in low back pain? BMC Musculoskelet Disord 2016;17:220.
4. Tamcan O, Mannion AF, Eisenring C, Horisberger B, Elfering A, Müller U. The course of chronic and recurrent low back pain in the general population. Pain 2010;150:451–7.
5. Smith BH, Elliott AM, Hannaford PC, Chambers WA, Smith WC. Factors related to the onset and persistence of chronic back pain in the community: results from a general population follow-up study. Spine (Phila Pa 1976) 2004;29:1032–40.
6. Sibbritt DW, Adams J. Back pain amongst 8,910 young Australian women: a longitudinal analysis of the use of conventional providers, complementary and alternative medicine (CAM) practitioners and self-prescribed CAM. Clin Rheumatol 2010;29:25.
7. Bollegala D, Perruccio AV, Badley EM. Combined impact of concomitant arthritis and back problems on health status: results from a nationally representative health survey. Arthritis Care Res (Hoboken) 2011;63:1584–91.
8. Deyo RA, Bryan M, Comstock BA, Turner JA, Heagerty P, Friedly J, et al. Trajectories of symptoms and function in older adults with low back disorders. Spine (Phila Pa 1976) 2015;40:1352–62.
9. Dunn KM, Jordan K, Croft PR. Characterizing the course of low back pain: a latent class analysis. Am J Epidemiol 2006;163:754–61.
10. Kongsted A, Kent P, Hestbaek L, Vach W. Patients with low back pain had distinct clinical course patterns that were typically neither complete recovery nor constant pain: a latent class analysis of longitudinal data. Spine J 2015;15:885–94.
11. Enthoven WT, Koes BW, Bierma-Zeinstra SM, Bueving HJ, Bohnen AM, Peul WC, et al. Defining trajectories in older adults with back pain presenting in general practice. Age Ageing 2016;45:878–83.
12. Downie AS, Hancock MJ, Rzewuska M, Williams CM, Lin CW, Maher CG. Trajectories of acute low back pain: a latent class growth analysis. Pain 2016;157:225–34.
13. Dunn KM, Campbell P, Jordan KP. Long-term trajectories of back pain: cohort study with 7-year follow-up. BMJ Open 2013;3:e003838.
14. Chen Y, Campbell P, Strauss VY, Foster NE, Jordan KP, Dunn KM. Trajectories and predictors of the long-term course of low back pain: cohort study with 5-year follow-up. Pain 2018;159:252–60.
15. Coenen P, Smith A, Paananen M, O'Sullivan P, Beales D, Straker L. Trajectories of low back pain from adolescence to young adulthood. Arthritis Care Research (Hoboken) 2017;69:403–12.
16. Leino-Arjas P, Rajaleid K, Mekuria G, Nummi T, Virtanen P, Hammarstrom A. Trajectories of musculoskeletal pain from adolescence to middle age: the role of early depressive symptoms, a 27-year follow-up of the Northern Swedish Cohort. Pain 2018;159:67–74.
17. Statistics Canada. Information about the national population health survey. Ottawa: Statistics Canada; 2011. URL: https://www150.statcan.gc.ca/n1/pub/82f0068x/82f0068x1997001-eng.htm.
18. Katzmarzyk PT, Church TS, Craig CL, Bouchard C. Sitting time and mortality from all causes, cardiovascular disease, and cancer. Med Sci Sports Exerc 2009;41:998–1005.
19. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. Annu Rev Clin Psychol 2010;6:109–38.
20. Raudenbush SW, Bryk AS. Hierarchical linear models: applications and data analysis methods. 2nd ed. Thousand Oaks (CA): Sage; 2002.
21. SAS Institute Inc. SAS/STAT 9.3 User's Guide. Cary (NC): SAS Institute Inc.; 2013.
22. Jones BL, Nagin DS. Advances in group-based trajectory modeling and an SAS procedure for estimating them. Sociol Methods Res 2007;35:542–71.
23. Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. Spine J 2008;8:8–20.
24. Luo X, Pietrobon R, Sun SX, Liu GG, Hey L. Estimates and patterns of direct health care expenditures among individuals with back pain in the United States. Spine (Phila Pa 1976) 2004;29:79–86.
25. Hong J, Reed C, Novick D, Happich M. Costs associated with treatment of chronic low back pain: an analysis of the UK General Practice Research Database. Spine (Phila Pa 1976) 2013;38:75–82.
26. Carey TS, Garrett J, Jackman A, McLaughlin C, Fryer J, Smucker DR. The outcomes and costs of care for acute low back pain among patients seen by primary care practitioners, chiropractors, and orthopedic surgeons: the North Carolina Back Pain Project. N Engl J Med 1995;333:913–7.
27. Leboeuf-Yde C, Fejer R, Nielsen J, Kyvik KO, Hartvigsen J. Pain in the three spinal regions: the same disorder? Data from a population-based sample of 34,902 Danish adults. Chiropr Man Therap 2012;20:11.

28. Badley EM, Millstone DB, Perruccio AV. Back pain and co-occurring conditions: findings from a nationally representative sample. Spine (Phila Pa 1976) 2018;43:E935–41.
29. Heneweer H, Staes F, Aufdemkampe G, van Rijn M, Vanhees L. Physical activity and low back pain: a systematic review of recent literature. Eur Spine J 2011;20:826–45.
30. Neupane S, Leino-Arjas P, Nygård CH, Oakman J, Virtanen P. Developmental pathways of multisite musculoskeletal pain: what is the influence of physical and psychosocial working conditions? Occup Environ Med 2017;74:468–75.
31. Coggon D, Ntani G, Palmer KT, Felli VE, Harari R, Barrero LH, et al. Patterns of multisite pain and associations with risk factors. Pain 2013;154:1769–77.
32. Boudreau D, Von Korff M, Rutter CM, Saunders K, Ray GT, Sullivan MD, et al. Trends in long-term opioid therapy for chronic non-cancer pain. Pharmacoepidemiol Drug Saf 2009;18:1166–75.
33. Paulozzi LJ, Jones CM, Mack KA, Rudd RA, for the Centers for Disease Control and Prevention. Vital signs: overdoses of prescription opioid pain relievers: United States, 1999–2008. MMWR Morb Mortal Wkly Rep 2011;60:1487–92.
34. Von Korff M, Saunders K, Thomas Ray G, Boudreau D, Campbell C, Merrill J, et al. De facto long-term opioid therapy for noncancer pain. Clin J Pain 2008;24:521–7.
35. Axén I, Bergström G, Bodin L. Using few and scattered time points for analysis of a variable course of pain can be misleading: an example using weekly text message data. Spine J 2014;14:1454–9.

Arthritis Care & Research
Vol. 71, No. 12, December 2019, pp 1671–1677
DOI 10.1002/acr.23814

American College of Rheumatology
Empowering Rheumatology Professionals

Hip Involvement in Patients With Calcium Pyrophosphate Deposition Disease: Potential and Limits of Musculoskeletal Ultrasound

Andrea Di Matteo,[1] Emilio Filippucci,[1] Edoardo Cipolletta,[1] Alice Musca,[2] Marina Carotti,[1] Riccardo Mashadi Mirza,[3] Diogo Jesus,[4] Victoria Martire,[5] Daniele Pierucci,[1] Marco Di Carlo,[1] Fausto Salaffi,[1] and Walter Grassi[1]

Objective. To preliminarily explore the diagnostic potential of ultrasound (US) in detecting calcium pyrophosphate (CPP) crystal deposits at the hip joint in a cohort of patients with CPP deposition disease (CPPD) who were previously evaluated by conventional radiography (CR) and to assess the sensitivity and specificity as well as the agreement between US and CR in the evaluation of hip CPP crystal deposits.

Methods. Fifty consecutive patients with definite CPPD and 40 age/sex/body mass index–matched disease control subjects who had undergone hip CR within the previous 6 months were enrolled. Bilateral hip US examination was carried out to assess the presence of CCP crystal deposits at the acetabular labrum fibrocartilage and at the femoral head's hyaline cartilage. Two independent radiologists evaluated the presence of hip CPP crystal deposits on CR in both groups.

Results. US findings indicative of CPP crystal deposits were found in at least 1 hip in 45 of 50 patients with CPPD (90.0%) and in 73 of 100 hips (73.0%). CPP crystal deposits were more frequently found at the acetabular labrum fibrocartilage than at the femoral head's hyaline cartilage (72% and 17% of the hips in patients with CPPD, respectively). US and CR sensitivity was 90% and 86%, whereas US and CR specificity was 85% and 90%, respectively. Total agreement between the US and CR findings was 77.8%.

Conclusion. Our results provide new evidence supporting US as a first-line, sensitive, safe, and reliable imaging technique in detecting CPP crystal deposits at the hip level.

INTRODUCTION

Calcium pyrophosphate deposition disease (CPPD) is a microcrystalline arthropathy caused by the deposition of calcium pyrophosphate (CPP) crystals within articular and periarticular tissues (1). CPPD can be regarded as a great mimicker because of the high variability of its clinical manifestations, ranging from an asymptomatic form to clinical scenarios that may present symptoms that overlap with different rheumatic diseases (i.e., gout, osteoarthritis [OA], rheumatoid arthritis, and fever of unknown origin) (2). According to the diagnostic criteria proposed by Ryan and McCarty, a definite diagnosis of CPPD can be established if CPP dihydrate crystals are identified in the synovial fluid analysis, and hyaline cartilage and/or fibrocartilage calcifications are detected on conventional radiography (CR) (3).

In daily clinical practice, the application of these criteria may not always be obvious since the synovial fluid analysis is not practical in many cases and CR has the disadvantage of false-negative results, especially if this investigation is limited only to the anatomic sites that are regarded as the most frequently involved in the disease, such as the knee and wrist (4). On the other hand, a

[1]Andrea Di Matteo, MD, Emilio Filippucci, MD, PhD, Edoardo Cipolletta, MD, Marina Carotti, MD, PhD, Daniele Pierucci, MD, Marco Di Carlo, MD, Fausto Salaffi, MD, PhD, Walter Grassi, MD, PhD: Università Politecnica Delle Marche, Ancona, Italy; [2]Alice Musca, MD: Dr. I. Cantacuzino Clinical Hospital, Bucharest, Romania; [3]Riccardo Mashadi Mirza, MD: Ospedale di Cittadella, Padua, Italy; [4]Diogo Jesus, MD: Centro Hospitalar e Universitario de Coimbra, Coimbra, Portugal; [5]Victoria Martire, MD: Instituto Médico Platense, Buenos Aires, Argentina.

Dr. Di Matteo has received speaking fees from Grünenthal (less than $10,000). Dr. Filippucci has received speaking fees from AbbVie, Bristol-Myers Squibb, Celgene, Roche, and Union Chimique Belge Pharma (less than $10,000 each). Dr. Carotti has received speaking fees from Roche, Novartis, AbbVie, Pfizer, and Bristol-Myers Squibb (less than $10,000 each). Dr. Di Carlo has received speaking fees from AbbVie and Novartis (less than $10,000 each). Dr. Salaffi has received speaking fees from Roche, AbbVie, Novartis, Bristol-Myers Squibb, Pfizer, and Merck Sharp & Dohme Italia (less than $10,000 each). Dr. Grassi has received speaking fees from AbbVie, Celgene, Grünenthal, Pfizer, and Union Chimique Belge Pharma. No other disclosures relevant to this article were reported.

Address correspondence to Andrea Di Matteo, MD, Clinica Reumatologica, Università Politecnica delle Marche, Ospedale Carlo Urbani, Via Aldo Moro 25, 60035 Jesi, Ancona, Italy. E-mail: andrea.dimatteo@hotmail.com.

Submitted for publication August 21, 2018; accepted in revised form November 16, 2018.

SIGNIFICANCE & INNOVATIONS

- This study provides new evidence supporting ultrasound (US) as a sensitive and specific imaging tool in detecting calcium pyrophosphate (CPP) crystal deposits at hip level.
- The diagnostic accuracy of US and conventional radiography (CR) in the evaluation of CPP crystal deposits at hip level was comparable, and the agreement between these 2 imaging techniques was very good.
- US should be regarded as a first-line imaging technique in the assessment of the hip joint in patients with suspected CPP deposition disease (CPPD) because of its reliability and safety compared to CR.

systematic CR evaluation of the potential targets of CPPD is not acceptable from an ethical point of view.

In recent years, ultrasound (US) has emerged as a useful imaging technique for the detection of CPP crystals, and its use is encouraged in the latest European League Against Rheumatism recommendations for CPPD diagnosis (5). Hip involvement is traditionally considered uncommon and mainly observed in patients with long-lasting disease (6,7). Of interest, a recent study carried out by Abhishek et al demonstrated that hip chondrocalcinosis is anything but infrequent, with the hip joint representing the third most commonly involved anatomic site in a cohort of patients with and without hip and knee OA (4). This study has also shown that hip chondrocalcinosis is frequently seen in the absence of knee involvement, with almost half of the hips examined presenting with CR findings indicative of chondrocalcinosis without evidence of CPP crystals at knee level.

The opportunity to consider the hip as a sentinel region for the detection of CPP crystal deposits even in the absence of knee involvement is clinically relevant, especially in patients with joint pain of unknown origin. US has gained a well-defined role in the diagnosis of CPPD, having shown a greater or equal sensitivity and comparable specificity to CR in the detection of CPP crystal aggregates (8–13). However, to the best of our knowledge, there is no study that has compared the diagnostic accuracy of US and CR in the evaluation of CPP crystal deposits at the hip joint in patients with CPPD.

The primary objective of this study was to preliminarily explore the diagnostic potential of US in detecting CPP crystal deposits in a cohort of patients with CPPD who were previously evaluated by CR. The secondary objective was to assess the sensitivity and specificity of, as well as the agreement between, US and CR in the evaluation of CPP crystal deposits in the hip.

PATIENTS AND METHODS

Patients. Consecutive patients with definite CPPD according to the Ryan and McCarty diagnostic criteria (3) (both CR and synovial fluid analysis positive for the presence of CPP crystals) and age/sex/body mass index–matched disease control subjects who were diagnosed with other rheumatic diseases according to various international diagnostic/classification criteria (14–22) were enrolled. All the patients were recruited from the outpatient clinic of the Clinica Reumatologica of the "Carlo Urbani" Hospital, Rheumatology Clinic, Jesi, Ancona, Italy. Inclusion criteria were the CR examination of hips and knees performed within the previous 6 months and synovial fluid analysis (performed within the previous 3 months in disease control subjects). Exclusion criteria were prior remarkable hip injuries or surgery procedures (including hip prosthesis and severe hip OA) and inflammatory hip involvement at the time of the evaluation. Disease control subjects with radiographic evidence of knee chondrocalcinosis or with a synovial fluid analysis that was positive for CPP crystals were excluded.

Ethics committee approval was not required since all the patients underwent clinical and US evaluation according to our local protocols. All the patients gave their informed consent. No specific funding was received from any bodies in the public, commercial, or not-for-profit sectors to carry out the work described in this study.

Clinical examination. A rheumatologist (EC) gathered the demographic (age, sex, and BMI) and clinical data from all the subjects. The same rheumatologist also collected the CR images and the synovial fluid analysis results.

Ultrasound examination. A rheumatologist (ADM), blinded to the clinical and CR data, performed a bilateral US examination of the hip joint both in patients with CPPD and in the disease control subjects. The US examination was carried out using a MyLab Class C US machine (Esaote, SpA) while working with a linear (3–13 MHz) and, when necessary, a convex probe (2–7 MHz). The lower frequency probe (2–7 MHz) was used to assess overweight and obese patients, in whom relatively high frequency probes were not appropriate to examine deep structures, such as the hip.

The patients were examined while lying in a supine position on the examination bed with heels together and hips slightly externally rotated. The anterior acetabular labrum fibrocartilage of the hip and the femoral head's hyaline cartilage were scanned using both longitudinal and transverse views (23). Parameters of gray-scale gain were adapted to enhance recognition of CPP crystals. As is commonly known, crystal deposits, similarly to bone, maintain a high reflectivity even when the gain setting is minimized (24,25).

CPP crystal deposits at the acetabular labrum fibrocartilage were identified as hyperechoic spots of variable shapes (i.e., rounded or amorphous-shaped). CPP crystal deposits at the femoral head's hyaline cartilage were defined as a hyperechoic band (either focal or diffuse) within the cartilage layer (26–29). At the acetabular labrum fibrocartilage, only grades higher than 2 were considered positive according to the semiquantitative 0–3

Table 1. Demographic, clinical, and serologic data of patients with CPPD*

	Values
Male/female, no.	23/27
Age, mean ± SD years	71 ± 10.2
Disease duration, mean ± SD months†	57.3 ± 90.8
BMI, kg/m^2	25.3 ± 3.8
ESR, mean ± SD mm/hour	27.1 ± 12.6
CRP, mean ± SD mg/dl	0.9 ± 1.4
Anatomic sites‡	
Knee	47 (94.0)
Hip joint§	14 (28.0)
Radiocarpal joint	6 (12.0)
Ankle	5 (10.0)
Shoulder	4 (8.0)
Elbow	2 (4.0)
Therapy	
Colchicine	31 (62.0)
NSAIDs	9 (18.0)
MTX	4 (8.0)
HCQ	3 (6.0)
HCQ+MTX	1 (2.0)
None	4 (8.0)

* Values are the number (%) unless indicated otherwise. Percentages refer to the total number of patients with CPPD included. CPPD = calcium pyrophosphate deposition disease; BMI = body mass index; ESR = erythrocyte sedimentation rate; CRP = C-reactive protein level; NSAIDs = nonsteroidal antiinflammatory drugs; MTX = methotrexate; HCQ = hydroxychloroquine.
† Disease duration indicates the time since the first diagnosis of CPPD.
‡ Sites showing calcium pyrophosphate crystals at the synovial fluid analysis.
§ Hip joint synovial fluid analysis was performed in 16 patients. In the 2 patients with negative hip synovial fluid analysis for calcium pyrophosphate crystals, neither ultrasound nor conventional radiography showed calcifications at hip level.

scoring system proposed by Filippou et al (where 0 = absent, 1 = 1 to 2 spots, 2 = >2 spots covering <50% of the volume of the structure, and 3 = deposits covering >50% of the volume of the structure) (30). This scoring system has been used recently in a study evaluating CPP crystal deposits at the wrist level (12).

Conventional radiography examination. Two radiologists (MC, RMM) with experience in imaging assessment of musculoskeletal diseases, blind to the clinical and US data, independently evaluated the images of the hip CR examinations (anteroposterior views) of all the patients to determine the presence or absence of cartilage calcifications. Radiographic features described by Resnick et al and Martel et al were used to identify the pathologic findings (31,32).

Statistical analysis. Results are reported as mean ± SDs for the quantitative variables. Data for qualitative variables are expressed as absolute frequency and as corresponding percentage. The Mann-Whitney test was used for quantitative variables that were not normally distributed, a Student's *t*-test for the quantitative variables that had a normal distribution, and the chi-square test for the qualitative variables. The agreement between the 2 radiologists who evaluated the CR images and the agreement between CR and US were assessed using Cohen's kappa and total agreement. The kappa values were interpreted according to methods proposed by Landis and Koch (33), where <0.0 = poor, 0.00 to 0.20 = slight, 0.21 to 0.40 = fair, 0.41 to 0.60 = moderate, 0.61 to 0.80 = substantial, 0.81 to 1.00 = almost perfect. The sensitivity, specificity, and diagnostic odds ratio of US and CR in the evaluation of CPP crystal deposits were calculated. McNemar's test was used to assess whether there was a significant difference between US and CR with respect to diagnostic accuracy. Statistical analysis was performed using Statistical Package for the Social Sciences software, version 24.0.

RESULTS

Patients. Fifty patients with CPPD and 40 disease control subjects were included in the study. A total of 180 hips (100 hips in patients with CPPD and 80 hips in control subjects) were evaluated using both US and CR. The main clinical and serologic data for patients with CPPD are shown in Table 1.

In the disease control group, 13 patients had OA, 10 psoriatic arthritis, 7 rheumatoid arthritis, 5 ankylosing spondylitis, 2 polymyalgia rheumatica, 2 gout, and 1 systemic lupus erythematosus. CPP crystals were not found in the synovial fluid analysis in any of the disease control subjects.

Table 2. Prevalence and distribution of US findings indicative of CPP crystal deposits at the hip joint in patients with CPPD and disease control subjects*

	No. of patients		No. of hips	
US findings	CPPD (n = 50)	Controls (n = 40)	CPPD (n = 100)	Controls (n = 80)
In at least 1 hip	45 (90.0)	6 (15.0)†	73 (73.0)	8 (10.0)†
Acetabular labrum fibrocartilage	45 (90.0)	6 (15.0)†	72 (72.0)	8 (10.0)†
Femoral head's hyaline cartilage	13 (26.0)	0‡	17 (17.0)	0‡
In both hips	28 (56.0)	2 (5.0)†	–	–

* Values are the number (%). Percentages refer to the total number of patients. Significance was determined by chi-square test. CPP = calcium pyrophosphate; CPPD = calcium pyrophosphate deposition disease; US = ultrasound.
† $P < 0.0001$.
‡ $P < 0.001$.

Table 3. Sensitivity, specificity, and diagnostic odds ratio of US and CR in the evaluation of CPP crystal deposits at the hip joint both in patients with CPPD and disease control subjects*

Imaging technique	Sensitivity (95% CI)	Specificity (95% CI)	Diagnostic OR (95% CI)
US	0.90 (0.78–0.97)	0.85 (0.70–0.94)	51.0 (14.0–181.0)
CR	0.86 (0.73–0.94)	0.90 (0.76–0.97)	55.0 (15.0–204.0)

* Significance was determined by chi-square test ($P < 0.0001$ for all). US = ultrasound; CR = conventional radiography; CPP = calcium pyrophosphate; CPPD = calcium pyrophosphate deposition disease; 95% CI = 95% confidence interval; OR = odds ratio.

Ultrasound. The prevalence and distribution of the US findings indicative of CPP crystal deposits are reported in Table 2. US findings indicative of CPP crystal deposits were found at the acetabular labrum fibrocartilage in 72 hips of 45 patients with CPPD (in 56 hips as an isolated finding, in 16 hips associated with CPP crystal deposits at the femoral head's hyaline cartilage) and in 8 hips of 8 disease control subjects, always as an isolated finding.

US findings indicative of CPP crystal deposits were found at the femoral head's hyaline cartilage in 17 hips of 13 patients with CPPD (in 1 hip as an isolated finding, in 16 hips associated with CPP crystal deposits at the acetabular labrum fibrocartilage) and in none of the hips of disease control subjects. A total of 76 patients were assessed using a linear probe (42 patients with CPPD, 34 disease control subjects), whereas the remaining 14 patients were studied using a convex probe (9 patients with CPPD, 5 disease control subjects).

Conventional radiography. CR findings indicative of CPP crystal deposits were detected in at least 1 hip in 43 of 50 patients with CPPD (86.0%) and in 4 of 40 disease controls (10.0%) ($P < 0.0001$), in 72 of 100 hips in patients with CPPD (72.0%), and in 5 of 80 hips in disease control subjects (6.3%) ($P < 0.0001$). Calcifications were found in both hips in 29 of 50 patients with CPPD (58.0%) and in 1 of 40 disease control subjects (2.5%).

The Cohen's kappa value for the interreader agreement between the 2 radiologists was 0.77 (95% confidence interval 0.67–0.87). CR findings indicative of CPP crystal deposits were detected in at least 1 knee in 48 of 50 patients with CPPD (96.0%) in 86 of 100 knees (86.0%).

Table 4. Agreement between ultrasound and conventional radiography findings in the evaluation of calcium pyrophosphate crystal deposits at the hip joint*

	CR positive (n = 180)	CR negative (n = 180)	Total
US positive	59 (32.8)	22 (12.2)	81 (45.0)
US negative	18 (10.0)	81 (45.0)	99 (55.0)
Total	77 (42.8)	103 (57.2)	

* Values are the number (%). Total agreement was 77.8%; kappa 0.55 (95% confidence interval 0.43–0.67). CR = conventional radiography; US = ultrasound.

Ultrasound and conventional radiography. The sensitivity, specificity, and diagnostic odds ratio of US and CR in the evaluation of CPP crystal deposits at the hip are shown in Table 3. There was no significant difference between US and CR with respect to sensitivity ($P = 0.5$), specificity ($P = 0.5$), and diagnostic odds ratio ($P = 0.25$). The agreement between the US and CR findings is reported in Table 4. Representative US and CR scenarios of hip involvement in patients with CPPD are shown in Figure 1.

DISCUSSION

The results of the present study indicate that US may represent a valid alternative to CR in demonstrating the presence of CPP crystal aggregates both at the hip acetabular labrum fibrocartilage and at the femoral head's hyaline cartilage. This study provides new evidence supporting US as a sensitive and reliable imaging technique in the assessment of the disease.

To date, only a few imaging studies have evaluated hip involvement in patients with CPPD, and a variable prevalence has been reported (34,35). So far, hip involvement in patients with CPPD has been poorly investigated essentially because systematic CR evaluation of the hip in asymptomatic patients is unethical. In a previous retrospective CR study, calcifications were reported in 48% of patients with CPPD (44.7% of examined hips) (34). Only 1 study of US evaluated the involvement of the hip in patients with CPPD, reporting findings indicative of CPP crystal deposits in 6.7% of patients in 5% of scanned hips (35). The higher prevalence of hip involvement reported in our results could be related to several factors, such as longer disease duration, different inclusion criteria (only patients with definite CPPD and with hip CR performed within the previous 6 months were included in our study), and US equipment used.

In the present study, US findings indicative of CPP crystal deposits were more frequently detected at the acetabular labrum fibrocartilage than at the femoral head's hyaline cartilage. This result is consistent with what had been shown at knee level in previous US studies (9,11,29,36,37) where CPP crystal deposits were more frequently found at the meniscal fibrocartilage than at the femoral condyles' hyaline cartilage, with the exception of 1 study in which US calcifications were more frequently detected at

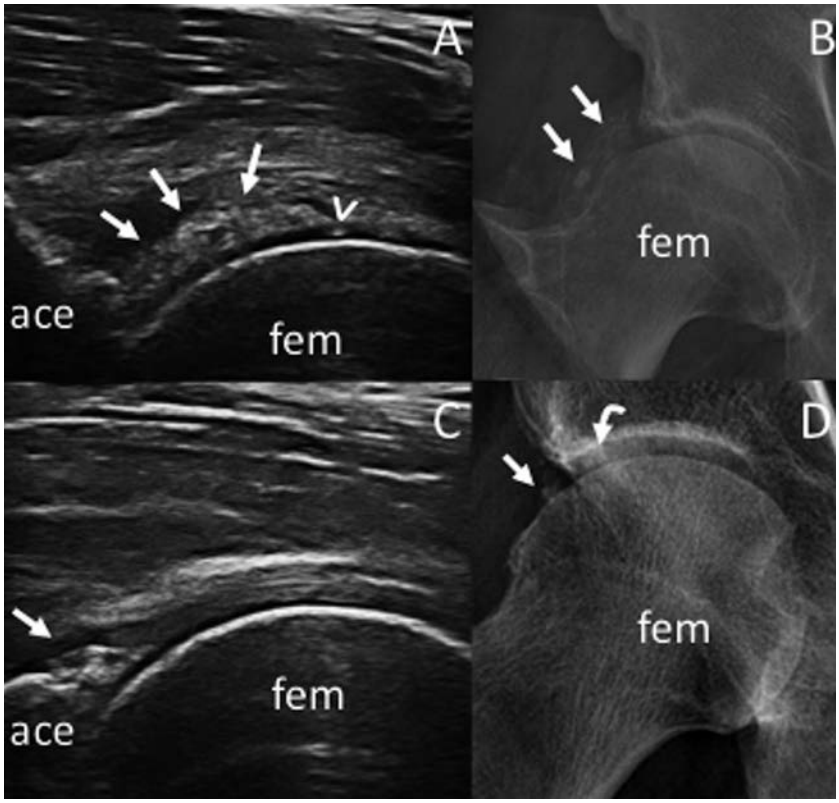

Figure 1. Ultrasound (US) and conventional radiography (CR) findings indicative of calcium pyrophosphate (CPP) crystal deposits at the hip joint in 2 patients with CPP deposition disease. In the first patient, US longitudinal scan of the hip joint (**A**) shows a diffuse and large calcification at the acetabular labrum fibrocartilage (**arrows**) and a hyperechoic spot within the femoral head's hyaline cartilage (**arrowhead**). In the same patient, CR anteroposterior view of the hip joint (**B**) reveals a large calcification at the acetabular labrum fibrocartilage (**arrows**). In the second patient, US longitudinal scan of the hip joint (**C**) shows an evident calcification at the acetabular labrum fibrocartilage (**arrow**). In this patient, the CR anteroposterior view of the hip (**D**) exhibits calcifications at the acetabular labrum fibrocartilage (**arrow**) and at the femoral head's hyaline cartilage (**curved arrow**). ace = acetabulum; fem = femoral head.

the femoral condyles' hyaline cartilage than at the meniscal fibrocartilage (38).

To the best of our knowledge, this is the first study that has explored the diagnostic value of US and CR and the agreement between these 2 imaging tools in the detection of CPP crystal deposits at the hip in a cohort of patients with CPPD. US showed a higher sensitivity than CR (90% versus 86%) but a lower specificity (85% versus 90%). The agreement between US and CR was very good (total agreement 77.8%), although in some cases the results of US and CR differed.

In 18 cases, calcifications were detected by CR but not by US. This could be mainly related to the fact that CR allows a more comprehensive evaluation of the hip joint, which also includes the cartilage areas covered by the bone. In fact, in 14 of the 18 cases, CR calcifications were located at the cartilage area covered by the iliac bone, an area that is not accessible by US (Figure 1D). On the contrary, in 22 cases CPP crystal deposits were detected by US but not by CR. As known, US has a higher resolution power than CR and allows the detection of even submillimeter microcrystal aggregates (8). This could explain why these deposits were visible by US but not by CR.

This study has the following main limitations. First is the absence of reproducibility analysis across operators of US, as examinations were performed by a single sonographer in a single center. Second, the selection criteria for a patient's inclusion, such as having a definite diagnosis of CPPD and knee and hip CR performed within the previous 6 months, may have influenced the sensitivity of US and CR in detecting CPP crystals at the hip, which resulted in higher results as compared to clinical practice. Third, the absence of a systematic fluid analysis from the hip joint may have influenced the correct classification rate for patients with CPPD. The presence of hyperechoic spots, which may be US pitfalls (39), as well as findings detectable in patients with other pathologic conditions different from CPPD (i.e., degenerative or traumatic injuries), may lead to misinterpretation of such findings. On the other hand, CR findings indicative of CPP crystal deposition are not necessary nor sufficient to make a definite diagnosis of CPPD. Finally, we included disease control subjects rather than healthy control subjects since we decided to enroll only patients who for clinical purposes underwent hip and knee CR in the previous 6 months. This could have affected the specificity of both US and CR.

In conclusion, the hip joint should be included among the anatomic regions to explore with US in patients with suspected CPPD. This study provides new evidence supporting US as a sensitive and specific imaging tool in detecting CPP crystal deposits at hip level. US should be regarded as a first-line imaging technique in the assessment of the hip joint in patients with suspected CPPD because of its reliability and safety compared to CR.

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. Di Matteo 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. Di Matteo, Filippucci, Cipolletta, Musca, Jesus, Martire, Pierucci, Di Carlo, Salaffi, Grassi.

Acquisition of data. Di Matteo, Filippucci, Cipolletta, Musca, Mashadi Mirza, Martire, Salaffi, Grassi.

Analysis and interpretation of data. Di Matteo, Filippucci, Cipolletta, Carotti, Mashadi Mirza, Jesus, Martire, Pierucci, Di Carlo.

REFERENCES

1. Ferrone C, Andracco R, Cimmino MA. Calcium pyrophosphate deposition disease: clinical manifestations. Reumatismo 2012;63:246–52.
2. Rosenthal AK, Ryan LM. Calcium pyrophosphate deposition disease. N Engl J Med 2016;374:2575–84.
3. McCarty D. Calcium pyrophosphate dihydrate crystal deposition disease: nomenclature and diagnostic criteria. Ann Intern Med 1977;87:240.
4. Abhishek A, Doherty S, Maciewicz R, Muir K, Zhang W, Doherty M. Chondrocalcinosis is common in the absence of knee involvement. Arthritis Res Ther 2012;14:R205.
5. Zhang W, Doherty M, Bardin T, Barskova V, Guerne PA, Jansen TL, et al. European League Against Rheumatism recommendations for calcium pyrophosphate deposition. Part I: terminology and diagnosis. Ann Rheum Dis 2011;70:563–70.
6. Abhishek A, Doherty M. Calcium pyrophosphate crystal–associated arthropathy. In: Hochberg M, Silman AJ, Smolen J, Weinblatt M, Weisman M, editors. Rheumatology. 6th ed. Maryland Heights: Mosby; 2015. p. 1583–95.
7. Abhishek A, Doherty M. Epidemiology and risk factors for calcium pyrophosphate deposition. In: Doherty M, Bijlsma J, Arden N, Hunter DJ, Dalberth N, editors. Oxford textbook of osteoarthritis and crystal arthropathy. 3rd ed. Oxford: Oxford University Press; 2016. p. 461–4.
8. Gutierrez M, Di Geso L, Filippucci E, Grassi W. Calcium pyrophosphate crystals detected by ultrasound in patients without radiographic evidence of cartilage calcifications. J Rheumatol 2010;37:2602–3.
9. Ottaviani S, Juge PA, Aubrun A, Palazzo E, Dieudé P. Sensitivity and reproducibility of ultrasonography in calcium pyrophosphate crystal deposition in knee cartilage: a cross-sectional study. J Rheumatol 2015;42:1511–3.
10. Filippou G, Adinolfi A, Cimmino MA, Scirè CA, Carta S, Lorenzini S, et al. Diagnostic accuracy of ultrasound, conventional radiography and synovial fluid analysis in the diagnosis of calcium pyrophosphate dihydrate crystal deposition disease. Clin Exp Rheumatol 2016;34:254–60.
11. Gutierrez M, Di Geso L, Salaffi F, Carotti M, Girolimetti R, De Angelis R, et al. Ultrasound detection of cartilage calcification at knee level in calcium pyrophosphate deposition disease. Arthritis Care Res (Hoboken) 2014;66:69–73.
12. Di Matteo A, Filippucci E, Salaffi F, Carotti M, Carboni D, Di Donato E, et al. Diagnostic accuracy of musculoskeletal ultrasound and conventional radiography in the assessment of the wrist triangular fibrocartilage complex in patients with definite diagnosis of calcium pyrophosphate dihydrate deposition disease. Clin Exp Rheumatol 2017;35:647–52.
13. Forien M, Combier A, Gardette A, Palazzo E, Dieudé P, Ottaviani S. Comparison of ultrasonography and radiography of the wrist for diagnosis of calcium pyrophosphate deposition. Joint Bone Spine 2018;85:615–8.
14. 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.
15. Aletaha D, Neogi T, Silman AJ, Funovitis J, Felson DT, Bingham CO III, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2010;62:2569–81.
16. Dasgupta B, Cimmino MA, Kremers HM, Schmidt WA, Schirmer M, Salvarani C, et al. 2012 provisional classification criteria for polymyalgia rheumatica: a European League Against Rheumatism/American College of Rheumatology collaborative initiative. Arthritis Rheum 2012;64:943–54.
17. Rudwaleit M, van der Heijde D, Landewé R, Akkoc N, Brandt J, Chou CT, et al. The Assessment of SpondyloArthritis international Society classification criteria for peripheral spondyloarthritis and for spondyloarthritis in general. Ann Rheum Dis 2011;70:25–31.
18. Taylor W, Gladman D, Helliwell P, Marchesoni A, Mease P, Mielants H, et al, and the CASPAR Study Group. Classification criteria for psoriatic arthritis: development of new criteria from a large international study. Arthritis Rheum 2006;54:2665–73.
19. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. Arthritis Rheum 1986;29:1039–49.
20. Altman R, Alarcón G, Appelrouth D, Bloch D, Borenstein D, Brandt K, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum 1991;34:505–14.
21. Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1982;25:1271–7.
22. Neogi T, Jansen TL, Dalbeth N, Fransen J, Schumacher HR, Berendsen D, et al. 2015 gout classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheumatol 2015;67:2557–68.
23. Möller I, Janta I, Backhaus M, Ohrndorf S, Bong DA, Martinoli C, et al. The 2017 EULAR standardised procedures for ultrasound imaging in rheumatology. Ann Rheum Dis 2017;76:1974–9.
24. Frediani B, Filippou G, Falsetti P, Lorenzini S, Baldi F, Acciai C, et al. Diagnosis of calcium pyrophosphate dihydrate crystal deposition disease: ultrasonographic criteria proposed. Ann Rheum Dis 2005;64:638–40.
25. Grassi W, Meenagh G, Pascual E, Filippucci E. "Crystal clear": sonographic assessment of gout and calcium pyrophosphate deposition disease. Semin Arthritis Rheum 2006;36:197–202.
26. Filippou G, Scirè CA, Adinolfi A, Damjanov NS, Carrara G, Bruyn GA, et al. Identification of calcium pyrophosphate deposition disease (CPPD) by ultrasound: reliability of the OMERACT definitions in an extended set of joints: an international multiobserver study by the OMERACT Calcium Pyrophosphate Deposition Disease Ultrasound Subtask Force. Ann Rheum Dis 2018;77:1194–9.

27. Delle Sedie A, Riente L, Iagnocco A, Filippucci E, Meenagh G, Grassi W, et al. Imaging Ultrasound imaging for the rheumatologist X. Ultrasound imaging in crystal-related arthropathies. Clin Exp Rheumatol 2007;25:513–7.
28. Filippucci E, Scirè CA, Delle Sedie A, Iagnocco A, Riente L, Meenagh G, et al. Ultrasound imaging for the rheumatologist. XXV. Sonographic assessment of the knee in patients with gout and calcium pyrophosphate deposition disease. Clin Exp Rheumatol 2010;28:2–5.
29. Filippucci E, Gutierrez M, Georgescu D, Salaffi F, Grassi W. Hyaline cartilage involvement in patients with gout and calcium pyrophosphate deposition disease: an ultrasound study. Osteoarthritis Cartilage 2009;17:178–81.
30. Filippou G, Filippucci E, Tardella M, Bertoldi I, Di Carlo M, Adinolfi A, et al. Extent and distribution of CPP deposits in patients affected by calcium pyrophosphate dihydrate deposition disease: an ultrasonographic study. Ann Rheum Dis 2013;72:1836–9.
31. Resnik CS, Resnick D. Calcium pyrophosphate dihydrate crystal deposition disease. Curr Probl Diagn Radiol 1982;11:6–40.
32. Martel W, McCarter DK, Solsky MA, Good AE, Hart WR, Braunstein EM, et al. Further observations on the arthropathy of calcium pyrophosphate crystal deposition disease. Radiology 1981;141:1–15.
33. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
34. Resnick D, Niwayama G, Goergen TG, Utsinger PD, Shapiro RF, Haselwood DH, et al. Clinical, radiographic and pathologic abnormalities in calcium pyrophosphate dihydrate deposition disease (CPPD): pseudogout. Radiology 1977;122:1–15.
35. Vele P, Simon SP, Damian L, Felea I, Muntean L, Filipescu I, et al. Clinical and ultrasound findings in patients with calcium pyrophosphate dihydrate deposition disease. Med Ultrason 2018;20:159–63.
36. Ruta S, Catay E, Marin J, Rosa J, García-Monaco R, Soriano ER. Knee effusion: ultrasound as a useful tool for the detection of calcium pyrophosphate crystals. Clin Rheumatol 2016;35:1087–91.
37. Filippou G, Frediani B, Gallo A, Menza L, Falsetti P, Baldi F, et al. Technique for the diagnosis of chondrocalcinosis of the knee: sensitivity and specificity of high-frequency ultrasonography. Ann Rheum Dis 2007;66:1126–8.
38. Coari G, Iagnocco A, Zoppini A. Chondrocalcinosis: sonographic study of the knee. Clin Rheumatol. 1995;14:511–4.
39. Filippucci E, Di Geso L, Grassi W. Tips and tricks to recognize microcrystalline arthritis. Rheumatology (Oxford) 2012;51:Suppl 7:vii18–21.

Arthritis Care & Research
Vol. 71, No. 12, December 2019, pp 1678–1684
DOI 10.1002/acr.23796

AMERICAN COLLEGE of RHEUMATOLOGY
Empowering Rheumatology Professionals

BRIEF REPORT

Spinal Radiographic Progression in Early Axial Spondyloarthritis: Five-Year Results From the DESIR Cohort

Sofia Ramiro,[1] Désirée van der Heijde,[2] Alexandre Sepriano,[3] Miranda van Lunteren,[2] Anna Moltó,[4] Antoine Feydy,[5] Maria Antonietta d'Agostino,[6] Damien Loeuille,[7] Maxime Dougados,[4] Monique Reijnierse,[2] and Pascal Claudepierre[8]

Objective. To analyze the progression of spinal radiographic damage in patients with early axial spondyloarthritis (SpA).

Methods. Axial SpA patients from the DESIR (Devenir des Spondylarthropathies Indifférenciées Récentes) cohort with 5-year spinal (cervical and lumbar) radiographs available (n = 549) were included. Two- and 5-year modified Stoke Ankylosing Spondylitis Spine Score (mSASSS) progression and development of new syndesmophytes (net change: the number of patients with positive change minus the number of patients with negative change divided by the total number of patients) were assessed in subgroups defined at baseline according to the Assessment of SpondyloArthritis international Society axial SpA criteria and its arms, modified New York criteria (mNYC) and the presence of syndesmophytes.

Results. Mean ± SD mSASSS progression was 0.2 ± 0.9 at 2 years and 0.4 ± 1.8 at 5 years. Five-year progression was higher in the imaging arm (mean ± SD 0.6 ± 2.3), magnetic resonance imaging (MRI)+/mNYC+ (mean ± SD 1.3 ± 4.0), than in the clinical arm only (mean ± SD 0.1 ± 0.7), and highest in patients with syndesmophytes (mean ± SD 2.7 ± 5.0). At 5 years, 7% of all patients had a net change of any new syndesmophyte; this value was 10% for the imaging arm (mNYC+/MRI+ with 18%), 17% for mNYC+ patients, and 42% for patients with syndesmophytes.

Conclusion. Spinal radiographic progression, although limited in early axial SpA, can be captured after 2 years. Inflammation and damage in the sacroiliac joint are associated with higher radiographic progression. The presence of baseline syndesmophytes already strongly predicts the development of further structural damage early in the disease.

INTRODUCTION

The development and evolution of spinal structural damage over time has been investigated in patients with radiographic axial spondyloarthritis (SpA). At a group level, an average progression of 2 modified Stoke Ankylosing Spondylitis Spine Score (mSASSS) units per 2 years (i.e., at the level of the cervical and lumbar spine) is seen when radiographs are scored with known time order, or 1 mSASSS unit per 2 years when scoring is blinded for chronologic order (1,2). The presence of syndesmophytes is known to be the strongest predictor for the development of further damage in radiographic axial SpA (1).

So far, no studies have focused on the development of structural damage over time in patients with early axial SpA. The

Supported by Pfizer. The DESIR study is conducted as a Programme Hospitalier de Recherche Clinique with Assistance Publique Hopitaux de Paris as the sponsor. The DESIR study is also under the umbrella of the French Society of Rheumatology, which financially supports the cohort.

[1]Sofia Ramiro, MD, PhD: Leiden University Medical Center, Leiden, and Zuyderland Medical Center, Heerlen, The Netherlands; [2]Désirée van der Heijde, MD, PhD, Miranda van Lunteren, Msc, Monique Reijnierse, MD, PhD: Leiden University Medical Center, Leiden, The Netherlands; [3]Alexandre Sepriano, MD: Leiden University Medical Center, Leiden, The Netherlands, and NOVA Medical School, Universidade Nova de Lisboa, Lisbon, Portugal; [4]Anna Moltó, MD, Msc, PhD, Maxime Dougados, MD: Paris Descartes University, Hôpital Cochin, Hôpitaux de Paris, INSERM U1153, and PRES Sorbonne Paris-City, Paris, France; [5]Antoine Feydy, MD, PhD: Paris Descartes University, Paris, France; [6]Maria Antonietta d'Agostino, MD, PhD: Ambroise Paré Hospital, AP-HP, Boulogne-Billancourt, and INSERM U1173, UFR Simone Veil, Université Versailles-Saint Quentin en Yvelines, Saint-Quentin en Yvelines, France; [7]Damien Loeuille, MD, PhD: University of Nancy, Nancy, France; [8]Pascal Claudepierre, MD: Henri Mondor Hospital, AP-HP, and Université Paris-Est Créteil, Créteil, France.

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

Address correspondence to Sofia Ramiro, MD, PhD, Department of Rheumatology, Leiden University Medical Center, Albinusdreef 2, 2333 GA Leiden, The Netherlands. E-mail: sofiaramiro@gmail.com.

Submitted for publication June 27, 2018; accepted in revised form October 16, 2018.

SIGNIFICANCE & INNOVATIONS

- Spinal radiographic progression, though limited in early axial spondyloarthritis, can be captured after 2 years.
- Inflammation and damage in the sacroiliac joints, i.e., the imaging arm of the Assessment of SpondyloArthritis international Society criteria, particularly magnetic resonance imaging positive/modified New York criteria positive, are associated with higher spinal radiographic progression.
- Syndesmophytes, which can already be present early in the axial disease, strongly predispose patients for the development of further structural damage.

development of the Assessment of SpondyloArthritis international Society (ASAS) classification criteria, although not meant for diagnosis, has raised awareness for early forms of the disease (3). However, radiographic progression in these patients without radiographic sacroiliitis, and also in patients fulfilling the different arms of the classification criteria, has not yet been investigated.

Recently we have shown that the mSASSS is also the most valid, feasible, and sensitive-to-change scoring method in patients with early axial SpA (4). The aim of the current study was to analyze the development and progression of spinal radiographic damage in patients with early axial SpA.

MATERIALS AND METHODS

Study population. Patients from the DESIR (Devenir des Spondylarthropathies Indifférenciées Récentes) cohort were included (5). Briefly, the DESIR cohort includes 708 patients with a high suspicion of recent axial SpA (<3 years of symptoms). Radiographs of cervical and lumbar spine were performed at baseline, at 2 years, and at 5 years and were read in 1 reading campaign. Patients were included in this analysis provided they had ≥2 observations with available radiographs and with an mSASSS progression score calculated. The database used for this analysis was locked in June 2016. DESIR received ethics approval from the Comité de Protections des Personnes Ile de France III.

Scoring methods. Radiographs were scored using the mSASSS (6). The anterior vertebral corners of the cervical and lumbar segments (total of 24 vertebral corners) were scored in the lateral view for the presence of erosion and/or sclerosis and/or squaring (1 point), syndesmophyte (2 points), and bridging syndesmophyte (3 points). The total score range was 0–72.

The radiographs were independently scored by 3 trained readers (scores were averaged) blinded to chronologic order, clinical characteristics, and other imaging data. For the mSASSS, only scores of radiographs with ≤3 missing vertebral corners per segment (cervical or lumbar) were used (7,8). Individual missing vertebral corners were imputed following a previous method (8). Reliability of the mSASSS readings of this study was good (4).

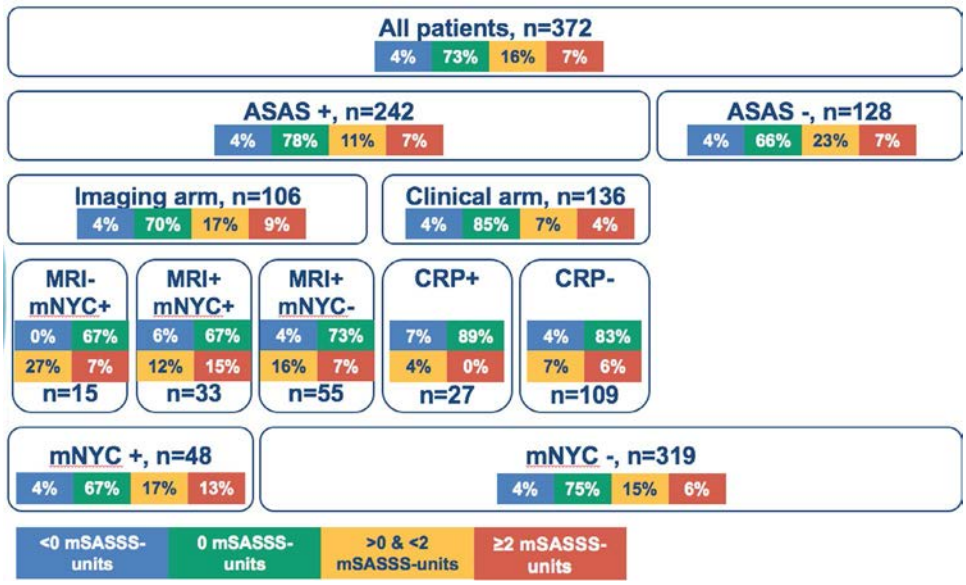

Figure 1. Categories of 5-year modified Stoke Ankylosing Spondylitis Spine Score (mSASSS) progression for the different subgroups according to the Assessment of SpondyloArthritis international Society (ASAS) and modified New York criteria (mNYC) at baseline. The total number of patients included in this flowchart is lower than the number of included patients, due to a missing radiograph at 5 years. MRI = magnetic resonance imaging; CRP = C-reactive protein.

To classify patients into different subgroups, baseline radiographs of the sacroiliac (SI) joints were also scored for the fulfillment of the modified New York criteria (mNYC) (9), and magnetic resonance imaging (MRI) for the presence of inflammation (i.e., axial SpA suggestive of bone marrow edema lesions) according to the ASAS definition (MRI+) (10).

Radiographic progression. Two- and 5-year progression scores (from baseline) were analyzed (mean ± SD) in subgroups of patients defined at baseline according to the following: ASAS axial SpA classification criteria (3) (ASAS+ or ASAS–) and its arms (imaging and clinical); fulfilment of the mNYC (mNYC+ or mNYC–), regardless of the MRI SI joint assessment (9); and the presence or absence of syndesmophytes (Figure 1). To get more insight into the progression scores, these scores were also analyzed in categories: <0, 0, >0 and <2, and ≥2 mSASSS units.

Syndesmophytes, assessed in all available vertebral corners independently of the missing vertebral corners per segment, were considered present when at least 2 of 3 readers identified them at exactly the same vertebral corner and time point. The number of syndesmophytes was analyzed at baseline and then new syndesmophytes (from baseline) were analyzed at 2 and 5 years. Both cutoffs of >0 and >1 new syndesmophytes were considered. The proportion of change is shown as the change above the cutoff (positive change), change below the cutoff (negative change), and net change. Net change corresponds to the number of patients with a positive change (e.g., ≥1) minus the number of patients with a negative change (e.g., ≥ –1) (numerator) divided by the total number of patients included in the analysis (denominator) (10).

A sensitivity analysis was conducted in a subpopulation, excluding patients who throughout follow-up obtained a different diagnosis than axial SpA (n = 25) and only including patients with an averaged (from all visits) level of confidence in the diagnosis of axial SpA of ≥5 (range 0–10). Additionally, the same analysis was conducted but restricted to patients with an averaged level of confidence in the diagnosis of axial SpA of ≥7. Stata SE software, version 12, was used.

RESULTS

In total, 549 patients were included, mean ± SD age was 34 ± 9 years, mean ± SD symptom duration was 1.5 ± 0.9 years, 46% were males, and 61% were HLA–B27 positive. In all, 63% of patients fulfilled the ASAS classification criteria (ASAS+), 13% fulfilled the mNYC criteria (mNYC+), and 7% had ≥1 baseline syndesmophyte (42% of these patients did not fulfill the ASAS classification criteria [ASAS–]). At baseline, no patients were treated with tumor necrosis factor inhibitors (TNFi), while at 2 years 31% of the included patients and at 5 years 43% were treated with a TNFi. Included patients were somewhat older, were more frequently HLA–B27 positive and ASAS+, and had a slightly higher baseline mSASSS score than those patients with missing radiographs, but differences were small (see Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23796/abstract).

Radiographic progression results. At baseline, the mean ± SD mSASSS score was 0.5 ± 1.5 for all patients, 0.6 ± 1.8 for ASAS– patients, and 0.4 ± 1.4 for ASAS+ patients, with

Table 1. Mean baseline damage and 2- and 5-year radiographic progression for the different subgroups according to the ASAS criteria, mNYC, and baseline syndesmophytes*

	Baseline mSASSS†	2-year mSASSS progression‡	5-year mSASSS progression‡
All patients	0.5 ± 1.5 (527)	0.2 ± 0.9 (488)	0.4 ± 1.8 (372)
ASAS–	0.6 ± 1.8 (196)	0.2 ± 1.0 (186)	0.6 ± 2.0 (128)
ASAS+	0.4 ± 1.4 (328)	0.1 ± 0.9 (299)	0.3 ± 1.6 (242)
Imaging arm	0.6 ± 1. 9 (151)	0.2 ± 1.2 (141)	0.6 ± 2.3 (106)
MRI+/mNYC–	0.3 ± 0.8 (81)	0.04 ± 0.3 (77)	0.3 ± 0.8 (55)
MRI–/mNYC+	0.7 ± 1.2 (23)	0.5 ± 1.6 (22)	0.3 ± 0.6 (15)
MRI+/mNYC+	1.2 ± 3.2 (43)	0.5 ± 1.9 (38)	1.3 ± 4.0 (33)
Clinical arm (only)	0.2 ± 0.7 (177)	0.02 ± 0.5 (158)	0.1 ± 0.7 (136)
CRP+	0.2 ± 0.5 (33)	0.01 ± 0.3 (29)	–0.02 ± 0.2 (27)
CRP–	0.2 ± 0.7 (144)	0.02 ± 0.5 (129)	0.2 ± 0.7 (109)
mNYC+	1.0 ± 2.7 (66)	0.5 ± 1.8 (60)	1.0 ± 3.3 (48)
mNYC–	0.4 ± 1.3 (454)	0.1 ± 0.7 (421)	0.3 ± 1.4 (319)
Baseline syndesmophytes+	4.4 ± 3.9 (36)	1.1 ± 2.9 (35)	2.7 ± 5.0 (31)
Baseline syndesmophytes–	0.2 ± 0.5 (491)	0.1 ± 0.5 (453)	0.2 ± 0.8 (341)

* Values are the mean ± SD (number of patients). Progression is measured compared to baseline. ASAS = Assessment of SpondyloArthritis international Society; mNYC = modified New York criteria; mSASSS = modified Stoke Ankylosing Spondylitis Spine Score; MRI = magnetic resonance imaging; CRP = C-reactive protein.
† In 22 of the included patients, the baseline mSASSS was missing, but at least 1 mSASSS progression interval was available and therefore the patient could be included in the analysis.
‡ The 2- and 5-year progression scores cannot be directly compared, because patients are not exactly the same in both groups (due to missing radiographs).

Table 2. Net change for the development of new syndesmophytes at 2 and 5 years compared to baseline*

		>0 at 2 years			>1 at 2 years				>0 at 5 years			>1 at 5 years		
	No. at 2 years	Positive change	Negative change	Net change	Positive change	Negative change	Net change	No. at 5 years	Positive change	Negative change	Net change	Positive change	Negative change	Net change
All patients	518	29 (6)	3 (0.6)	26 (5)	11 (2)	2 (0.4)	9 (2)	413	36 (9)	6 (1)	30 (7)	13 (3)	1 (0.2)	12 (3)
ASAS–	195	13 (7)	1 (0.5)	12 (6)	3 (2)	1 (1)	2 (1)	143	15 (10)	2 (1)	13 (9)	5 (3)	1 (1)	4 (3)
ASAS+	317	16 (5)	2 (0.6)	14 (4)	8 (3)	1 (0.3)	7 (2)	265	21 (8)	4 (2)	17 (6)	8 (3)	0 (0)	8 (3)
Imaging arm	149	10 (7)	1 (0.7)	9 (6)	5 (3)	0 (0)	5 (3)	120	13 (11)	1 (1)	12 (10)	5 (4)	0 (0)	5 (4)
MRI+/mNYC–	83	2 (2)	0 (0)	2 (2)	1 (1)	0 (0)	1 (1)	62	4 (6)	1 (2)	3 (5)	0 (0)	0 (0)	0 (0)
MRI–/mNYC+	22	3 (14)	0 (0)	3 (14)	1 (5)	0 (0)	1 (5)	16	2 (16)	0 (0)	2 (16)	0 (0)	0 (0)	0 (0)
MRI+/mNYC+	39	5 (13)	1 (3)	4 (10)	3 (8)	0 (0)	3 (8)	38	7 (18)	0 (0)	7 (18)	5 (13)	0 (0)	5 (13)
Clinical arm (only)	168	6 (4)	1 (0.6)	5 (3)	3 (2)	1 (1)	2 (1)	145	8 (6)	3 (2)	5 (3)	3 (2)	0 (0)	3 (2)
CRP+	31	2 (7)	0 (0)	2 (7)	1 (3)	0 (0)	1 (3)	29	2 (7)	1 (3)	1 (3)	1 (3)	0 (0)	1 (3)
CRP–	137	4 (3)	1 (1)	3 (2)	2 (1)	1 (1)	1 (1)	116	6 (5)	2 (2)	4 (3)	2 (2)	0 (0)	2 (2)
mNYC+	61	8 (13)	1 (2)	7 (12)	4 (7)	0 (0)	4 (7)	54	9 (17)	0 (0)	9 (17)	5 (9)	0 (0)	5 (9)
mNYC–	446	21 (5)	2 (0.4)	19 (4)	7 (2)	2 (0.4)	5 (1)	350	27 (8)	6 (2)	21 (6)	8 (2)	1 (0.3)	7 (2)
Baseline syndesmophytes+	37	15 (41)	3 (8)	12 (32)	7 (19)	2 (5)	5 (14)	33	17 (52)	3 (9)	14 (42)	8 (24)	1 (3)	7 (21)
Baseline syndesmophytes–	470	13 (3)	0 (0)	13 (3)	3 (1)	0 (0)	3 (1)	369	17 (5)	2 (1)	15 (4)	4 (1)	0 (0)	4 (1)

* Values are the number (%) unless indicated otherwise. Net change means that first the patients with a positive change were calculated, i.e., patients with a new syndesmophyte (according to 2 of 3 readers); subsequently the negative change was calculated, i.e., the number of patients in which an existing syndesmophyte disappeared, corresponding to measurement error. Net change is the number of patients with a positive change minus the number of patients with a negative change and divided by the total number of patients. ASAS = Assessment of SpondyloArthritis international Society; MRI = magnetic resonance imaging; mNYC = modified New York criteria; CRP = C-reactive protein.

an increasing gradient for patients who were MRI+/mNYC–, followed by MRI–/mNYC+, and then MRI+/mNYC+ (Table 1). Mean ± SD 2-year mSASSS progression for all patients was 0.2 ± 0.9 and 5-year progression was 0.4 ± 1.8. Following the baseline difference, 5-year progression was slightly higher in patients who were ASAS– (mean ± SD 0.6 ± 2.0) than ASAS+ patients (mean ± SD 0.3 ± 1.6). mSASSS progression was higher in the imaging arm than in the clinical arm only. Within the imaging arm, there was a gradient starting in the MRI+/mNYC– patients, with a 5-year progression of mean ± SD 0.3 ± 0.8, followed by MRI–/mNYC+ (mean ± SD 0.3 ± 0.6) and then MRI+/mNYC+ (mean ± SD 1.3 ± 4.0). mNYC+ patients had higher progression (mean ± SD 1.0 ± 3.3) than mNYC– patients, just like patients with baseline syndesmophytes, the group with the highest progression (mean ± SD 2.7 ± 5.0), compared to those patients without syndesmophytes (mean ± SD 0.2 ± 0.8). At 5 years, 23% of the patients showed mSASSS progression (16% progression >0 and <2 units, 7% progression ≥2). These percentages were higher in patients fulfilling the imaging arm criteria (26% progression >0), mNYC+ (30%, with 13% progression ≥2) (Figure 1), and were the highest in patients with baseline syndesmophytes, with a total of 74% showing a progression >0 and 39% a progression ≥2. At a group level, mean mSASSS values per time point increased from 0.5 at baseline to 1.1 at 5 years (see Supplementary Figure 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23796/abstract).

New syndesmophytes. At 5 years, 91% of the patients did not show any new syndesmophytes, 6% had 1 new syndesmophyte, 1% had 2 new syndesmophytes, and 2% had >2 new syndesmophytes (see Supplementary Figure 2, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23796/abstract). Table 2 shows the proportion of patients with a new syndesmophyte. The proportion of patients showing any new syndesmophytes (net change >0) at 5 years was 7% for all patients, 10% for the patients in the imaging arm (ranging from 5% MRI+/mNYC– to 18% MRI+/mNYC+), 17% for mNYC+, and 42% for patients with baseline syndesmophytes. Using a cutoff of >1 new syndesmophyte, the percentages dropped importantly, and the presence of new syndesmophytes was mostly captured in the patients who were mNYC+ (9%) and especially in patients with baseline syndesmophytes (21%).

Sensitivity analysis. Radiographic progression in the selected population of patients with a level of confidence in the diagnosis of axial SpA ≥5, and excluding patients with another diagnosis than axial SpA during follow-up, was very similar to the main results (see Supplementary Tables 2–4, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23796/abstract). Radiographic progression in both continuous and categorical forms, as well as net change of new syndesmophytes, was similar across all subgroups of patients analyzed. Only in the subgroup of patients with baseline syndesmophytes was progression slightly higher in the population of patients from the sensitivity analysis. Similar results are for the patients with a level of confidence of the diagnosis of axial SpA ≥7 (data not shown).

DISCUSSION

Spinal radiographic damage progression can already be captured after 2 and 5 years of follow-up in early axial SpA. Interestingly, 7% of the patients already presented with syndesmophytes at baseline, which means that the process of structural damage starts, at least for some, early in the disease.

Radiographic damage and progression were slightly higher in ASAS– patients compared to ASAS+ patients, which was surprising. However, the difference was small and it can likely be explained by the fact that almost half of the patients with baseline syndesmophytes were ASAS– and 8% of the ASAS– patients had baseline syndesmophytes, a strong predictor of further damage, also confirmed in this study (1,11,12). The presence of syndesmophytes is already associated with a bad prognosis in early axial SpA; 74% of the patients with baseline syndesmophytes showed radiographic progression (any mSASSS progression) at 5 years, and almost half developed new syndesmophytes. For the first time, syndesmophytes have been analyzed as a net change. Net change considers a negative change, i.e., a situation in which a syndesmophyte disappeared from baseline to a subsequent time point (according to at least 2 of 3 readers), in principle meaning measurement error. Despite being a conservative approach, net change still captured new syndesmophytes in an early axial SpA population and with varying frequencies in different subgroups, as expected. Net change represents a method that should be further encouraged, also when analyzing the development of new syndesmophytes (10).

As expected, the imaging arm showed more progression than the clinical arm. Within the imaging arm, a gradient was seen, with the lowest progression in the MRI+/mNYC– patients, followed by MRI–/mNYC+ patients, and by MRI+/mNYC+ patients. These findings suggest that the presence of bone marrow edema on the MRI SI joint assessment (i.e., MRI+) is associated with more structural damage in the spine in comparison to patients with a negative MRI SI joint result. Additionally, the findings show that radiographic sacroiliitis (i.e., mNYC+), and particularly the combination of both SI joint inflammation and damage (MRI+/mNYC+), seems to predispose patients to more spinal radiographic progression (10,13). New syndesmophytes at 5 years beyond measurement error are mainly seen in the subgroup of patients who are

both MRI+/mNYC+. Structural damage in axial SpA seems, at the group level, to start in the SI joint and expand cranially in the spine. Having a positive C-reactive protein level did not make any difference in the patients in the clinical arm, while clinical inflammation, i.e., disease activity, is known to be associated with spinal radiographic progression (14). Possibly a relationship between inflammation and structural progression only happens in patients who are already prone for progression, i.e., who have SI joint damage, because such a relationship has only been demonstrated in radiographic axial SpA. Sensitivity analyses in a population of patients with a high level of confidence in the diagnosis and excluding patients with other diagnoses during follow-up provided similar results, which adds to the robustness of the findings.

In this study we did not consider the influence of other important factors in radiographic progression, such as patient characteristics (e.g., sex, HLA–B27 positivity, or smoking) or medication (the observed progression was under treatment of TNFi in 40–50% of the patients during any period of the follow-up). Neither did we score degenerative changes. We aimed at evaluating the progression in the different subgroups according to classification criteria or presence of syndesmophytes. Particularly the influence of medication requires specific analyses and handling potential confounding by indication, which requires a dedicated study.

In conclusion, spinal radiographic progression in early axial SpA is low but measurable beyond measurement error. Inflammation and damage in the SI joint, i.e., the imaging arm of the ASAS criteria, particularly MRI+/mNYC+, are associated with higher spinal radiographic progression. Syndesmophytes, which can be present early in the axial disease, seem to strongly predispose patients for the development of further structural damage.

ACKNOWLEDGMENTS

The DESIR cohort is conducted under the control of Assistance Publique Hopitaux de Paris via the Clinical Research Unit Paris Center and under the umbrella of the French Society of Rheumatology and Institut National de la Santé et de la Recherche Médicale (INSERM). Database management is performed within the Department of Epidemiology and Biostatistics (Jean-Pierre Daures, Nimes, France). We also thank the different regional participating centers: Maxime Dougados (Paris-Cochin B), Andre Kahan (Paris-Cochin A), Philippe Dieudé (Paris-Bichat), Bruno Fautrel (Paris-La Pitie-Salpetriere), Francis Berenbaum (Paris-Saint-Antoine), Pascal Claudepierre (Creteil), Maxime Breban (Boulogne-Billancourt), Bernadette Saint-Marcoux (Aulnay-sous-Bois), Philippe Goupille (Tours), Jean Francis Maillefert (Dijon), Emmanuelle Dernis (Le Mans), Daniel Wendling (Besancon), Bernard Combe (Montpellier), Liana Euller-Ziegler (Nice), Pascal Richette (Paris Lariboisière), Pierre Lafforgue (Marseille), Patrick Boumier (Amiens), Martin Soubrier (Clermont-Ferrand), Nadia Mehsen (Bordeaux), Damien Loeuille (Nancy), Rene-Marc Flipo (Lille), Alain Saraux (Brest), Xavier Mariette (Le Kremlin-Bicêtre), Alain Cantagrel (Toulouse), and Olivier Vittecoq (Rouen). We also thank the research nurses, the staff members of the Clinical Research Unit of Paris Center, the staff members of the Biological Resource Center of Bichat Hospital, the staff members of the Department of Statistics of Nimes, and all the investigators, and in particular Jerome Allain, Emmanuelle Dernis, Salah Ferkal, Clement Prati, Marie-Agnes Timsit, and Eric Toussirot for active patient recruitment and monitoring.

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. Ramiro 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. Ramiro, van der Heijde, Sepriano, van Lunteren, Moltó, Feydy, d'Agostino, Loeuille, Dougados, Reijnierse, Claudepierre.

Acquisition of data. Ramiro, van der Heijde, Sepriano, van Lunteren, Moltó, Feydy, d'Agostino, Loeuille, Dougados, Reijnierse, Claudepierre.

Analysis and interpretation of data. Ramiro, van der Heijde, Sepriano, van Lunteren, Moltó, Feydy, d'Agostino, Loeuille, Dougados, Reijnierse, Claudepierre.

ROLE OF THE STUDY SPONSOR

Pfizer 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 Pfizer.

REFERENCES

1. Ramiro S, Stolwijk C, van Tubergen A, van der Heijde D, Dougados M, van den Bosch F, et al. Evolution of radiographic damage in ankylosing spondylitis: a 12 year prospective follow-up of the OASIS study. Ann Rheum Dis 2015;74:52–9.
2. Van der Heijde D, Landewé R, Baraliakos X, Houben H, van Tubergen A, Williamson P, et al. Radiographic findings following two years of infliximab therapy in patients with ankylosing spondylitis. Arthritis Rheum 2008;58:3063–70.
3. Rudwaleit M, van der Heijde D, Landewe R, Listing J, Akkoc N, Brandt J, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. Ann Rheum Dis 2009;68:777–83.
4. Ramiro S, Claudepierre P, Sepriano A, van Lunteren M, Molto A, Feydy A, et al. Which scoring method depicts spinal radiographic damage in early axial spondyloarthritis best? Five-year results from the DESIR cohort. Rheumatology (Oxford) 2018;57:1991–2000.
5. Dougados M, Etcheto A, Molto A, Alonso S, Bouvet S, Daures JP, et al. Clinical presentation of patients suffering from recent onset chronic inflammatory back pain suggestive of spondyloarthritis: the DESIR cohort. Joint Bone Spine 2015;82:345–51.
6. Creemers MC, Franssen MJ, van 't Hof MA, Gribnau FW, van de Putte LB, van Riel PL. Assessment of outcome in ankylosing spondylitis: an extended radiographic scoring system. Ann Rheum Dis 2005;64:127–9.
7. Wanders AJ, Landewé RB, Spoorenberg A, Dougados M, van der Linden S, Mielants H, et al. What is the most appropriate radiologic scoring method for ankylosing spondylitis? A comparison of the

available methods based on the Outcome Measures in Rheumatology Clinical Trials filter. Arthritis Rheum 2004;50:2622–32.
8. Ramiro S, van Tubergen A, Stolwijk C, Landewe R, van de Bosch F, Dougados M, et al. Scoring radiographic progression in ankylosing spondylitis: should we use the modified Stoke Ankylosing Spondylitis Spine Score (mSASSS) or the Radiographic Ankylosing Spondylitis Spinal Score (RASSS)? Arthritis Res Ther 2013;15:R14.
9. Van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis: a proposal for modification of the New York criteria. Arthritis Rheum 1984;27:361–8.
10. Dougados M, Sepriano A, Molto A, van Lunteren M, Ramiro S, de Hooge M, et al. Sacroiliac radiographic progression in recent onset axial spondyloarthritis: the 5-year data of the DESIR cohort. Ann Rheum Dis 2017;76:1823–8.
11. Van Tubergen A, Ramiro S, van der Heijde D, Dougados M, Mielants H, Landewe R. Development of new syndesmophytes and bridges in ankylosing spondylitis and their predictors: a longitudinal study. Ann Rheum Dis 2012;71:518–23.
12. Baraliakos X, Listing J, Rudwaleit M, Haibel H, Brandt J, Sieper J, et al. Progression of radiographic damage in patients with ankylosing spondylitis: defining the central role of syndesmophytes. Ann Rheum Dis 2007;66:910–5.
13. Van der Heijde D, Machado P, Braun J, Hermann KG, Baraliakos X, Hsu B, et al. MRI inflammation at the vertebral unit only marginally predicts new syndesmophyte formation: a multilevel analysis in patients with ankylosing spondylitis. Ann Rheum Dis 2012;71:369–73.
14. Ramiro S, van der Heijde D, van Tubergen A, Stolwijk C, Dougados M, van den Bosch F, et al. Higher disease activity leads to more structural damage in the spine in ankylosing spondylitis: 12-year longitudinal data from the OASIS cohort. Ann Rheum Dis 2014;73:1455–61.

Arthritis Care & Research
Vol. 71, No. 12, December 2019, pp 1685–1689
DOI 10.1002/acr.24105

American College of Rheumatology
Empowering Rheumatology Professionals

Reviewers

I would like to thank the following individuals for their time and effort in reviewing articles for *Arthritis Care & Research* in the last year. The continued high quality of the journal depends on the dedicated service of these individuals.

Marian T. Hannan, DSc, MPH, Editor

Abasolo, Lydia
Abbott, J.
Aberle, Teresa
Abreu, Mirhelen
Ackerman, Ilana
Adams, Jo
Adebajo, Ade
Agacayak, Elif
Aggarwal, Amita
Aggarwal, Rohit
Agricola, Rintje
Ahedi, Harbeer
Aitken, Dawn
Akamatsu, Yasushi
Akiyama, Keisuke
Akram, Qasim
Al-Mayouf, Sulaiman
Alarcón, Graciela
Albrecht, Katinka
AlHeresh, Rawan
Allanore, Yannick
Allen, Kelli
Allison, Kim
Almeida, Gustavo
Alten, Rieke
Amris, Kirstine
Anderson, Ronald J.
Andreoli, Laura
Andrews, James
Angeles-Han, Sheila
Angst, Felix
Annapureddy, Narender
Antony, Benny
Antovic, Aleksandra
Approbato Selistre, Luiz Fernando
Arai, Yuji
Aringer, Martin
Arkema, Elizabeth
Armbrust, Wineke
Arnold, John
Askling, Johan
Assassi, Shervin
Atkinson, Jane C.
Atsumi, Tatsuya
Avenatti, Eleonora
Avina-Zubieta, J. Antonio
Avouac, Jerome
Bachen, Elizabeth
Backman, Catherine
Bacon, Kathryn
Bader-Meunier, Brigitte
Badley, Elizabeth
Baer, Alan
Baker, Joshua F.
Baker, Kenneth
Baker, Nancy
Bannuru, Raveendhara
Bansback, Nick
Baraliakos, Xenofon
Baran, Anna
Barber, Claire
Barbour, Kamil
Baron, Murray
Bartels, Christie M.
Bartlett, Susan
Bartley, Emily
Barton, Jennifer
Bass, Anne
Basu, Neil
Bathon, Joan
Battafarano, Daniel
Beck, Charlotte
Becker, Ingrid
Becker, Michael
Benveniste, Olivier
Bergman, Martin
Bergstra, Sytske Anne
Bermas, Bonnie
Berner, Jon
Berthelot, Jean-Marie
Berti, Alvise
Beukelman, Timothy
Beylot-Barry, Marie
Bhutani, Gita
Bieleman, André
Bieler, Theresa
Bierma-Zeinstra, Sita M.A.
Bili, Androniki
Binder-Finnema, Pauline
Bingham III, Clifton
Birmingham, Trevor
Birrell, Fraser
Bjurstrom, Martin
Blakley, Michael
Blalock, Susan
Blanc, Paul
Blanco, Irene
Boer, Aleid
Bolster, Marcy
Boonen, Annelies
Bove, Allyn
Bovis, Francesca
Bowen, Catherine
Bowes, Michael
Bowman, Simon J.
Brady, Teresa
Braun, Juergen
Bremander, Ann
Bresee, Catherine
Bricca, Alessio
Brisson, Nicholas
Broder, Anna
Brodin, Nina
Broughton, Nigel
Brunner, Hermine
Bruno, Michael
Burdorf, Alex
Burgos Vargas, Ruben
Buttgereit, Frank
Caban-Martinez, Alberto
Callaghan, Michael
Callahan, Leigh
Callen, Jeffrey
Callhoff, Johanna
Campbell Jr., Robert
Canizares, Mayilee
Cannella, Amy
Cannon, Grant
Cao, Hua
Capin, Jacob
Cappelli, Laura
Cardiel, Mario
Carmona, Loreto
Carpenter, Lewis
Carruthers, Mollie
Carteron, Nancy
Cattalini, Marco
Cattano, Nicole
Cavallo, Sabrina
Cazenave, Tomas
Ceccarelli, Fulvia
Cedraschi, Christine
Cervera, Ricard
Chakravarty, Eliza
Chandran, Vinod
Chang, Alison
Chang, Joyce
Chang, Rowland
Changhai, zang
Chatham, Winn
Chehab, Gamal
Chen, Hsin-Hua
Chen, Wenjia
Chiba, Daisuke
Chighizola, Cecilia
Chimera, Nicole
Chinoy, Hector
Chira, Peter
Choi, Hyon

Christiansen, Meredith
Chu, Cong-Qiu
Chung, Cecilia
Chung, Lorinda
Ciani, Oriana
Cibere, Jolanda
Cimmino, Marco
Ciurea, Adrian
Claudepierre, Pascal
Coates, Laura
Coburn, Brian
Cole, Iain
Collins, Jamie
Collins, Natalie
Consolaro, Alessandro
Cook, Chad
Cook, Michael
Copsey, Bethan
Cornec, Divi
Coronado, Rogelio
Costedoat-Chalumeau, Nathalie
Costello, Kerry
Costenbader, Karen
Cott, Cheryl
Cotton, Seonaidh
Cottrell, Elizabeth
Courvoisier, Delphine
Coyle, Pete
Cozier, Yvette
Cramp, Fiona
Criscione-Schreiber, Lisa
Croft, Peter
Crowson, Cynthia
Cui, Kangping
Curtis, Jeffrey
Cuttica, Ruben
Dagfinrud, Hanne
Dai, Zhaoli (Joy)
Dalal, Deepan
Dalbeth, Nicola
Dalton, Diane
Damon, Bruce M.
Danieli, MariaGiovanna
Danila, Maria
Davis, Aileen
Davis, Dawn
Davis III, John
Davis, Jennifer
de Thurah, Annette
De Vera, Mary
Deane, Kevin D.
Dejaco, Christian
Dekker, Joost
Deligny, Christophe
Dellaripa, Paul
Denton, Christopher
Deodhar, Atul
Desai, Rishi
Desai, Sonali
Descatha, Alexis
DeVellis, Robert
Dhir, Varun
Di Stasi, Stephanie
Direskeneli, Haner
Dixon, Padraig
Dixon, William
Doria, Andrea
Dougados, Maxime
Dougherty, Paul
Drenkard, Cristina
Driban, Jeffrey
Druce, Katie
Dubreuil, Maureen
Dudler, Jean
Dufour, Alyssa
Dumitru, Raluca
Dunlop, Dorothy
Eberhard, Barbara
Eder, Lihi
Ehrmann Feldman, Debbie
Emery, Paul
England, Bryant
Englund, Martin
Erhardt, Daniel
Erkan, Doruk
Esbensen, Bente Appel
Esdaile, John
Eshed, Iris
Espinosa, Gerard
Espinoza, Luis
Fain, Olivier
Fairchild, Robert
Falahee, Marie
Falzer, Paul
Faurschou, Mikkel
Favier, Leslie
Feist, Eugen
Feldman, Brian
Feldman, Candace
Fenton, Sally
Ferraz-Amaro, Ivan
Ferreira, Ricardo
Ferucci, Elizabeth
Figueiredo Gomes, João
Fisher, Judith
Fisher, Nadine
Fitzgerald, G. Kelley
Fleischmann, Roy M.
Flowers, Portia
Foeldvari, Ivan
Foocharoen, Chingching
Forslind, Kristina
Fraenkel, Liana
Franklin, Patricia
Frech, Tracy
Frick, Kevin
Fritzler, Marvin
Fukui, Naoshi
Furuta, Shunsuke
Gademan, Maaike
Gaffo, Angelo
Gall, Eric
Galloway, James
Gandek, Barbara
Garg, Neha
Garg, Vishvas
Garriga, Cesar
Gelber, Allan
Genovese, Mark
Gensler, Lianne
George, Michael
Geraldino, Laura
Gerli, Roberto
Gheita, Tamer
Gignac, Monique
Giles, Jon
Gill, Tiffany
Girschick, Hermann
Gissel, C.
Gladman, Dafna
Glauser, Terry Ann
Golder, Vera
Golightly, Yvonne
Gonzalez-Gay, Miguel
Goodman, Susan
Gordon, Caroline
Gordon, Jessica
Gossec, Laure
Goswami, Rudra
Graber, Judith M.
Graudal, Niels
Grayson, Peter
Greenwood, Kenneth
Griffin, Marie R.
Gudbrandsson, Birgir
Gudu, Tania
Guimaraes, Maria Fernanda
Gulati, Gaurav
Guzman, Jaime
Gwynne-Jones, David
Hackett, Katie
Halbert, R.
Hall, Michelle
Hamilton, David
Hanaoka, Beatriz
Hanly, John
Haraoui, Boulos
Harper, Daniel
Harris, Ian
Harrison, Mark
Harry, Onengiya
Hart, Jaime
Hashimoto, Motomu
Hassett, Afton
Haugen, Ida
Hawkins, Melanie
Hay, M. Cameron
Hazlewood, Glen
Helfgott, Simon
Helliwell, Philip
Helliwell, Toby
Helmick, Charles
Henrickson, Michael
Henriksen, Marius
Herbolsheimer, Florian
Hernández-Díaz, Cristina
Hernandez-Molina, Gabriela
Hersh, Aimee
Hill, Catherine
Hinman, Rana
Hirata, Shintaro
Hitchon, Carol
Hoi, Alberta
Holden, Melanie
Holmqvist, Marie
Horneff, Gerd
Hoshino, Junichi
Houghton, Kristin
Howren, Alyssa
Hsu, Vivien

Petterson, Stephanie
Pettersson, Susanne
Pietrosimone, Brian G.
Pillinger, Michael
Pincus, Theodore
Pisaniello, Huai Leng
Pisetsky, David
Piva, Sara
Poon, Samuel
Pope, Janet
Power, J. Denise
Pozzi, Federico
Primdahl, Jette
Prothero, Louise
Purc-Stephenson, Rebecca
Quicke, Jonathan
Rahman, Anisur
Rahme, Elham
Ramanan, Athimalaipet
Ramiro, Sofia
Ramsey-Goldman, Rosalind
Ranganath, Veena
Rasmussen, Astrid
Rat, Anne-Christine
Rathbun, Alan
Rees, Frances
Reimold, Andreas
Rhon, Daniel
Richards, John
Riddle, Daniel
Robbins, Shawn
Robinson, Philip
Rome, Keith
Rosenbaum, James
Rubinstein, Tamar
Ruiz-Irastorza, Guillermo
Runhaar, Jos
Russell, Linda
Rutherford, Andrew
Sabio, Jose Mario
Sale, Joanna
Salvarani, Carlo
Sampaio-Barros, Percival
Saraux, Alain
Sarmanova, Aliya
Sarsour, Khaled
Sato, Emilia
Scheepers, Lieke
Scheinberg, Morton
Scheuermann, Suzette
Schikler, Kenneth
Schlenk, Elizabeth
Schmajuk, Gabriela
Schmeling, Heinrike
Schmitt, Laura
Schoenfeld, Sara
Scielzo, Shannon
Scirè, Carlo
Sebastiani, Marco
Sebro, Ronnie
Segelmark, Mårten
Selva-O'Callaghan, Albert
Seo, Philip
Shakoor, Najia
Sharif, Behnam
Sharma, Leena
Shaw, Yomei
Shea, M. Kyla
Shea, Steven
Shiboski, Caroline
Shmerling, Robert H.
Shoenfeld, Yehuda
Shupak, Rachel
Sim, Julius
Simard, Julia
Simkin, Peter
Simon, Corey
Singh, Namrata
Sirois, Fuschia
Sivaraman, Vidya
Slade, Susan
Slusher, Barbara
Smarr, Karen
Smitherman, Emily
Smolen, Josef
So, Alexander
Sokka, Tuulikki
Sokolove, Jeremy
Solomon, Daniel
Soriano, Enrique
Soubrier, Martin
Souza, Richard
Sparks, Jeffrey
Squire, Sandra
Stamm, Tanja
Stamp, Lisa
Starz, Terence
Staud, Roland
Stavropoulos-Kalinoglou, Antonios
Steen, Virginia
Stefanik, Joshua
Steiner, Jennifer
Steultjens, Martijn
Stoll, Matthew
Stone, John H.
Strand, Vibeke
Suehiro, Ricardo
Sumpton, Daniel
Sundström, Björn
Svärd, Anna
Svenungsson, Elisabet
Takeuchi, Tsutomu
Tamhane, Ashutosh
Tanaka, Yoshiya
Taylor, Janalee
Taylor, William
Tedeschi, Sara
ten Klooster, Peter
Terhorst, Lauren
Theander, Elke
Thiele, Geoffrey
Thoma, Louise
Thorne, Jennifer
Thyberg, Ingrid
Tillett, William
Timmermans, Erik
Tiong, Benedict
Toes, Rene
Touma, Zahi
Tournadre, Anne
Toyama, Shogo
Tran-Duy, An
Treharne, Gareth
Triantafyllias, Konstantinos
Trupin, Laura
Tsang-A-Sjoe, Michel
Tsao, Nicole
Tsokos, George
Tunnicliffe, David
Tupper, Susan
Turer, Christy Boling
Twilt, Marinka
Tzaribachev, Nikolay
Underwood, Martin
Ungprasert, Patompong
van den Hout, Wilbert
Van der Elst, Kristien
van der Esch, Martin
van der Heijde, Désirée
van der Horst-Bruinsma, Irene
van der Leeden, Marike
van Riel, Piet L.
Verstappen, Suzanne
Vetrano, Davide
Vilá, Luis
Vina, Ernest
Vitali, Claudio
Voinier, Dana
von Scheven, Emily
Voskuyl, Alexandre
Wahezi, Dawn
Wahl, Elizabeth
Wahrens, Eva
Walitt, Brian
Walker-Bone, Karen
Wallace, Daniel
Wallace, Zachary
Wang, Runsheng
Wang, Weiqi
Wanigatunga, Amal
Ward, Michael
Waugh, Esther
Webster, Fiona
Weisman, Michael
Weiss, Pamela
Wellsandt, Elizabeth
Welsing, Paco
Wendling, Daniel
Wener, Mark
Werth, Victoria
Westby, Marie
Westerlind, Helga
Westhovens, René
White, Daniel
Williamson, Esther
Wimmer, Markus
Windschall, Daniel
Wirth, Wolfgang
Wise, Barton
Wittink, Harriet
Wluka, Anita
Wolff, Aviva
Woo, Patricia

VOLUME 71 TABLE OF CONTENTS

VOLUME 71 • January 2019 • NO.1

VOLUME 71 • February 2019 • NO. 2

VOLUME 71 • March 2019 • NO. 3

VOLUME 71 • April 2019 • NO. 4

VOLUME 71 • May 2019 • NO. 5

Rheumatoid Arthritis

Osteoarthritis

Systemic Lupus Erythematosus

Fibromyalgia

Rheumatology Practice

Letters

VOLUME 71 • June 2019 • NO. 6

Special Articles

Osteoarthritis

VOLUME 71 • July 2019 • NO. 7

VOLUME 71 • August 2019 • NO. 8

VOLUME 71 • September 2019 • NO. 9

Rheumatoid Arthritis

Fibromyalgia

Pediatrics

Ankylosing Spondylitis

Erratum

VOLUME 71 • October 2019 • NO. 10

Special Articles

Rheumatoid Arthritis

Osteoarthritis

Psoriatic Arthritis

Letters

VOLUME 71 • November 2019 • NO. 11

Osteoarthritis

Psoriatic Arthritis

VOLUME 71 • December 2019 • NO. 12

Special Articles

Rheumatoid Arthritis

Pediatrics

Osteoarthritis

Systemic Lupus Erythematosus

Gout

Back Pain

Calcium Pyrophosphate Deposition Disease

Spondyloarthritis

ARP Announcements

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

Download the New ACR Publications Mobile App

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

ARP Membership

The Association of Rheumatology Professionals (ARP), a division of the American College of Rheumatology, appreciates your continued membership and looks forward to serving you another year. Membership costs range from $30 to $140. ARP welcomes nurse practitioners, nurses, physician assistants, office staff, researchers, physical therapists, occupational therapists, assistants, and students. Student membership is complimentary; the Annual Meeting registration fee is waived for students who submit the required student verification letter. For information, go to www.rheumatology.org and select "Membership" or call 404-633-3777 and ask for an ARP staff member.

New ACR Journal Twitter Account (@ACR_Journals) and Social Media Editor

The ACR journals are heightening our focus on social media, to benefit authors and readers. Among our first activities is the introduction of an official ACR Journals Twitter account: @ACR_Journals. Followers will enjoy special features and the opportunity to engage with authors and other fellow professionals about studies published in *Arthritis Care & Research, Arthritis & Rheumatology,* and *ACR Open Rheumatology*. Authors of published articles will have the opportunity to use @ACR_Journals to share their work and engage in dialogue with others interested in the research. The journals welcome Dr. Paul Sufka of Minneapolis as our first Social Media Editor.