

Development of the American College of Rheumatology's Rheumatoid Arthritis Electronic Clinical Quality Measures

JINOOS YAZDANY,¹ MARK ROBBINS,² GABRIELA SCHMAJUK,¹ SONALI DESAI,³ DIANE LACAILLE,⁴ TUHINA NEOGI,⁵ JASVINDER A. SINGH,⁶ MARK GENOVESE,⁷ RACHEL MYSLINSKI,⁸ NATALIE FISK,⁸ MELISSA FRANCISCO,⁸ AND ERIC NEWMAN⁹

Objective. Electronic clinical quality measures (eQMs) rely on computer algorithms to extract data from electronic health records (EHRs). On behalf of the American College of Rheumatology (ACR), we sought to develop and test eQMs for rheumatoid arthritis (RA).

Methods. Drawing from published ACR guidelines, a working group developed candidate RA process measures and subsequently assessed face validity through an interdisciplinary panel of health care stakeholders. A public comment period followed. Measures that passed these levels of review were electronically specified using the quality data model, which provides standard nomenclature for data elements (category, datatype, and value sets) obtained through an EHR. For each eQM, 3 clinical sites using different EHR systems tested the scientific feasibility and validity of measures. Measures appropriate for accountability were presented for national endorsement.

Results. Expert panel validity ratings were high for all measures (median 8–9 of 9). Health system performance on the eQMs was 53.6% for RA disease activity assessment, 69.1% for functional status assessment, 93.1% for disease-modifying antirheumatic drug (DMARD) use, and 72.8% for tuberculosis screening. Kappa statistics, which evaluated whether the eQM validly captured data obtained from manual EHR chart review, demonstrated moderate to substantial agreement (0.54 for functional status assessment, 0.73 for tuberculosis screening, 0.84 for disease activity, and 0.85 for DMARD use).

Conclusion. Four eQMs for RA have achieved national endorsement and are recommended for use in federal quality reporting programs. Implementation and further refinement of these measures is ongoing in the ACR's registry, the Rheumatology Informatics System for Effectiveness (RISE).

Introduction

Quality measurement in rheumatoid arthritis (RA) is a national priority in health care. Stakeholders convened by the National Quality Forum (NQF) recently selected RA as one of the top 20 Medicare chronic conditions for quality measure development (1). This designation resulted from the relatively high prevalence of RA, which affects 1.3 million Americans, and from its significant morbidity and

costs (2). Moreover, previous quality measurement efforts have identified important gaps in health care for RA. For example, socioeconomic and racial/ethnic disparities exist in disease-modifying antirheumatic drug (DMARD) use, and there is significant variation in implementation of best practices for ensuring patient safety and optimizing disease control through the use of standardized outcome measures (3,4).

The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality.

Supported by the American College of Rheumatology. Dr. Yazdany's work was supported by the Robert L. Kroc Chair in Rheumatic and Connective Tissue Diseases (I), AHRQ R01-HS024412, and the Russell/Engleman Medical Research Center for Arthritis.

¹Jinoos Yazdany, MD, MPH, Gabriela Schmajuk, MD, MSc: University of California, San Francisco; ²Mark Robbins, MD, MPH: Harvard Vanguard Medical Associates, Atrius Health, Somerville, Massachusetts; ³Sonali Desai, MD, MPH: Brigham & Women's Hospital, Boston, Massachusetts; ⁴Diane Lacaille, MD, FRCP, MHS: University of British Columbia, Arthritis

Research Centre of Canada, Vancouver, British Columbia, Canada; ⁵Tuhina Neogi, MD, PhD: Boston University School of Medicine, Boston, Massachusetts; ⁶Jasvinder A. Singh, MD, MPH: University of Alabama at Birmingham; ⁷Mark Genovese, MD: Stanford University Medical Center, Palo Alto, California; ⁸Rachel Myslinski, MBA, Natalie Fisk, Melissa Francisco, MBA: American College of Rheumatology, Atlanta, Georgia; ⁹Eric Newman, MD: Geisinger Medical Center, Danville, Pennsylvania.

Dr. Yazdany has received an independent research grant from Pfizer. Dr. Singh has received research grants from Takeda and Savient, consultant fees from Takeda, Savient, Regeneron, Iroko, Merz, Bioiberica, Crealta, and Allergan

Over the last decade, quality measures in RA have largely relied on 2 data sources: administrative billing claims and chart reviews. Each of these methods has limitations, including the restricted clinical information available in claims and the resource-intensive nature of chart review. Moreover, while these approaches have enabled retrospective performance measurement in RA, they have been less conducive to providing information to clinicians in real-time to support rapid-cycle quality improvement. To address these limitations, there is increasing interest in leveraging electronic health records (EHRs) to develop electronic clinical quality measures (eCQMs), which are a new type of quality measure that relies on automated extraction of information from the EHR. Coupled with local data analytics or innovations, such as nationally Qualified Clinical Data Registries that centrally analyze and feed data back to practices, eCQMs can be used as tools to drive continuous quality improvement.

In this study, we sought to develop and test eCQMs for RA using a multistakeholder process with input from an interdisciplinary team of clinicians, patients, payers, and medical informaticists. Using practices with different EHR systems, we also sought to study the early feasibility and reliability of RA eCQMs before submission to the NQF for endorsement and implementation in the American College of Rheumatology (ACR) Informatics System for Effectiveness (RISE) registry.

Materials and methods

The ACR's overall process for developing eCQMs is outlined in Figure 1 and is also described in detail elsewhere (5). Here we describe how this process was applied to develop RA eCQMs.

Measure conceptualization. A working group (JY, GS, SD, DL, TN, JAS, MG, and EN) was assembled to draft measures for RA based on the most recent ACR guidelines (6). We reviewed guidelines referencing RA, reviewed and characterized the level of scientific evidence supporting various measure concepts, and also considered harmonization with existing measures. For this latter portion, our goal was to avoid duplication with existing measures in national reporting programs.

The working group drafted potential eCQM concepts in an iterative manner. Although both process measures (e.g., what clinicians do in providing care) and outcome measures

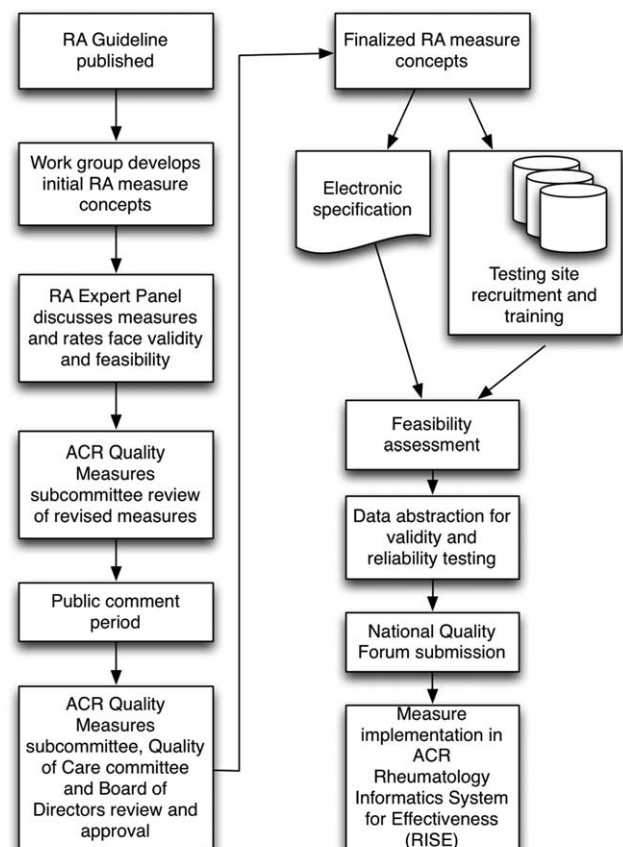


Figure 1. Overview of the American College of Rheumatology (ACR) rheumatoid arthritis (RA) electronic clinical quality measure development program.

(e.g., health outcomes that result from care) were considered, the working group decided to proceed with process measures, since research evaluating risk-adjustment models for RA was not available. We drafted eCQM concepts in an “if, then” format and presented these concepts to an expert panel for review (7).

Interdisciplinary consensus ratings. Nominations were sought for a multistakeholder panel of experts on RA. In addition to rheumatologists in both academic and community practice, we included patient and payer representatives. A member from each of the Association of Rheumatology Health Professionals, the American College of Physicians, and the American Academy of Orthopedic Surgery was also invited (see Supplementary Appendix A, available on the *Arthritis Care & Research* web site at <http://onlinelibrary.wiley.com/doi/10.1002/acr.22984/abstract>). Panel members did not receive payment for participation. The chairperson of the panel and the majority of its members ($\geq 50\%$) had no financial conflicts of interest with any product made for RA.

Expert panel meetings and ACR committee review and public comment. Expert panel members participated in a webinar introducing the project. Members received a summary of the RA measure concepts under consideration. Included were references to corresponding sections of the ACR RA guidelines and a summary of existing analogous

(more than \$10,000 each), and is the principal investigator for an investigator-initiated study funded by Horizon pharmaceuticals through a grant to DINORA, Inc., a 501c3 entity. Dr. Genovese has received consultant fees, speaking fees, and/or honoraria from AbbVie, Amgen, Astellas, Celgene, Gilead, Galapagos, Lilly, Sanofi, Vertex, UCB, Johnson & Johnson, Boehringer Ingelheim, and Roche (less than \$10,000 each).

Address correspondence to Jinoos Yazdany, MD, MPH, Division of Rheumatology, University of California, San Francisco, Box 0920, San Francisco, California 94143-0920. E-mail: jinoos.yazdany@ucsf.edu.

Submitted for publication March 30, 2016; accepted in revised form July 27, 2016.

measures in national reporting programs. For example, information on specifications and performance data on the National Committee on Quality Assurance's DMARD measure, implemented over the last decade using administrative claims to assess health plan performance, was provided (3).

We used a modification of the RAND/UCLA Appropriateness Method to have expert panel members rate the measures (8). Details about our methods for conducting this session and analyzing the results are provided elsewhere (5). Measures that were rated as valid and feasible were reviewed by the ACR Quality Measures Subcommittee and distributed for public comment. Public comments informed revisions, and the measures were sent to the ACR Quality of Care Committee and Board of Directors for final approval.

Electronic specification. To convert measure concepts to eCQM format, we used a multistep process that aligned with current national standards, including the Health Quality Measures Format. RA eCQMs were first specified using the measure-authoring tool (MAT) and quality data model (QDM) (9). We then worked with a clinical informaticist (see Supplementary Appendix A, available on the *Arthritis Care & Research* web site at <http://onlinelibrary.wiley.com/doi/10.1002/acr.22984/abstract>) who used QDM elements to elaborate all possible code sets to represent measure concepts in EHRs, including International Classification of Diseases, Ninth and Tenth Revision (ICD-9, -10), Systematized Nomenclature of Medicine (SNOMED)—Clinical Terms, Logical Observation Identifiers Names and Codes (LOINC), Current Procedural Terminology, Healthcare Common Procedure Coding System, and RxNorm.

There were 2 instances where uniform EHR nomenclature was not available in current terminology ("RA disease activity measure score," and "RA functional status measure score"); the ACR submitted requests to have these added to the Value Set Authority Center at the National Library of Medicine. Once the code lists were finalized, physicians from the working group worked in pairs to review all codes, using clinical judgment to assess their appropriateness for inclusion; any discordance was adjudicated through discussion. The MAT was then used to build the final eCQMs.

eCQM field-testing. For each eCQM, we recruited 3 sites using different EHR products to test the measures. Data elements for all eCQMs were extracted from EHRs using computer programming, and therefore by virtue of automation, this process is repeatable (reliable); however, because data algorithms must be implemented accurately, testing focused on the technical feasibility and concurrent validity of each measure (5). Each site first completed a feasibility survey and then worked with local information technology staff to build the RA eCQM extraction algorithms. This required review of the eCQMs specifications, including measure background information, required data elements, measure logic and measure calculation instructions, human-readable formats of the measure, as well as a detailed spreadsheet with value sets (i.e., code sets) for each measure.

We decided a priori to perform feasibility testing for 3 key data elements: disease activity score, functional status

score, and RA diagnosis. Sites completed a detailed survey assessing data availability and accuracy (e.g., Is information for the eCQM collected in the EHR and is that information correct?), data standards (e.g., Are standard value sets used to collect the data elements?), and operational or workflow issues (e.g., How is the data element entered in the EHR?). Both quantitative data, which included the NQF feasibility assessment scale (described previously [5]), and qualitative information, which outlined challenges to eCQM implementation, were collected.

We also assessed concurrent validity, or whether the information from the EHR data pull was similar to the information that a human abstractor obtains by manually reading data in the EHR. Rheumatology providers in each practice performed an EHR chart review. We used kappa statistics to determine whether, for each measure and site, the manual chart review and automated EHR data extracts identified the same patients as meeting the numerator of each measure. In our analyses, the manually extracted data were used as the gold standard for both the numerator and denominator of each measure. In these analyses $\kappa = 1.0$ when the automated EHR query agrees exactly with data obtained through manual chart extraction, and $\kappa = 0$ when the agreement appears entirely due to chance. For the denominator components, we calculated the percentage agreement between the chart review and automated EHR extracts.

In addition, we calculated the sensitivity and specificity for the numerator of the performance scores, again using chart review as the gold standard. In these analyses, true positives were the individuals with RA who received recommended care based on the chart review, and the sensitivity was the proportion of those true positive patients who were correctly identified as receiving recommended care in the computer extract. True negatives were those who did not receive recommended care in the chart review, and the specificity is the proportion of the true negative patients who we identified as not receiving recommended care in the EHR extract.

All data were analyzed at the individual patient level. For each validation project, a simple random sample was constructed that was powered for the analyses.

Submission for national endorsement and implementation in RISE. Because a goal of the ACR eCQM development project was to contribute toward a coherent performance measurement strategy for US rheumatologists, an important priority was to submit measures for national endorsement. RA eCQMs were therefore submitted to the NQF. Measures were also implemented in the ACR's RISE Registry.

Results

Below we present the results of each phase of the eCQM development work. Our results, including the evidence summaries, reflect the data included for the national endorsement process.

Conceptualization of measures. The working group drafted measure concepts relevant to RA (Table 1). Below, we briefly review the rationale and scientific evidence supporting each measure. In the context of this scientific

Table 1. RA quality measure concepts, numerators, and denominators*

| Measure title | Brief description of measure | Measure properties | Denominator | Numerator | Numerator or denominator detail | Exclusions | Score type |
|---------------------------------|---|---|---|--|---|--|--|
| Assessment of disease activity | Percentage of patients age ≥ 18 years with RA diagnosis, and $\geq 50\%$ of total number of outpatient RA encounters in the measurement year with assessment of disease activity, using a standardized measure | Type of measure: process Data source: electronic clinical data; EHR Level of analysis: individual clinician Time period: 12 months | Patients age ≥ 18 years with RA diagnosis seen for ≥ 2 face-to-face encounters for RA with the same clinician during the measurement period | Number of patients with $\geq 50\%$ of total number of outpatient RA encounters in the measurement year with assessment of disease activity using a standardized measure | Numerator: RA disease activity measurement tools must include 1 of the following: CDAI, DAS28-ESR or -CRP, PAS, PAS-II, RAPID-3, or SDAI A score recorded in the EHR qualifies as meeting numerator performance | None | Rate/proportion, with higher rate meaning better quality |
| Assessment of functional status | Percentage of patients age ≥ 18 years with RA diagnosis for whom a functional status assessment was performed at least once during the measurement period | Type of measure: process Data source: electronic clinical data; EHR Level of analysis: individual clinician Time period: 12 months | Patients age ≥ 18 years with RA diagnosis seen for ≥ 2 face-to-face encounters for RA with the same clinician during the measurement period | Number of patients in the measurement year with functional status assessments using a standardized measure | Numerator: functional status can be assessed by using one of a number of instruments, including HAQ-II, MDHAQ, PROMIS PF-10 or -20, or PFCAT | None | Rate/proportion, with higher rate meaning better quality |
| DMARD therapy | Percentage of patients age ≥ 18 years with RA diagnosis who are prescribed a DMARD in the measurement year | Type of measure: process Data source: electronic clinical data; EHR Level of analysis: individual clinician Time period: 12 months | Patients age ≥ 18 years with RA diagnosis seen for ≥ 2 face-to-face encounters for RA with the same clinician during the measurement period | Number of patients in the measurement year prescribed a DMARD | Numerator: all available DMARDs in the measurement period Biologic agents: abatacept, adalimumab, anakinra, certolizumab, etanercept, golimumab, infliximab, rituximab, tocilizumab Non-biologic agents: azathioprine, cyclophosphamide, cyclosporine, gold, hydroxychloroquine, leflunomide, methotrexate, minocycline, penicillamine, sulfasalazine New synthetic DMARD: tofacitinib | Patients with HIV diagnosis, pregnancy, or inactive RA | Rate/proportion, with higher rate meaning better quality |

(continued)

Table 1. (Cont'd)

| Measure title | Brief description of measure | Measure properties | Denominator | Numerator | Numerator or denominator detail | Exclusions | Score type |
|---------------|--|---|---|--|---|------------|--|
| TB screening | Percentage of patients age ≥ 18 years with RA diagnosis who have documentation of TB screening performed within 12 months prior to receiving a first course of biologic therapy or certain new synthetic DMARDs | Type of measure: process Data source: electronic clinical data; EHR Level of analysis: individual clinician Time period: 12 months | Patients age ≥ 18 years with RA diagnosis seen for ≥ 1 face-to-face encounter for RA and are newly started on biologic therapy or tofacitinib during the measurement period | Any record of TB testing documented or performed (PPD, IFN- γ -release assays, or other appropriate screening or treatment) in the medical record in the 12 months preceding the biologic DMARD or tofacitinib initiation | Denominator: for the purposes of this measure, patients who are newly started on biologic or tofacitinib therapy are those who have been prescribed these DMARDs during the measurement period and who were NOT prescribed these DMARDs in the 12 months preceding the encounter where the drug was started. Biologic DMARDs included abatacept, adalimumab, anakinra, certolizumab, etanercept, golimumab, infliximab, tocilizumab. New synthetic DMARD: tofacitinib | None | Rate/proportion, with higher rate meaning better quality |

* RA = rheumatoid arthritis; EHR = electronic health record; CDAI = Clinical Disease Activity Index; DAS28 = Disease Activity Score in 28 joints; CRP = C-reactive protein; ESR = erythrocyte sedimentation rate; PAS = Patient Activity Scale; RAPID-3 = Routine Assessment of Patient Index Data with 3 measures; SDAI = Simplified Disease Activity Index; HAQ-II = Health Assessment Questionnaire; MDHAQ = Multidimensional HAQ; PROMIS = Patient-Reported Outcomes Measurement Information System; PF-10 = physical function 10-item; PFCAT = PF computerized adaptive tests; DMARD = disease-modifying antirheumatic drug; HIV = human immunodeficiency virus; TB = tuberculosis; PPD = purified protein derivative; IFN- γ = interferon- γ .

evidence, we also discuss the reasoning of the working group in defining specific aspects of each measure.

Assessment of disease activity. The paradigm of RA management has undergone a significant transformation with the introduction of both new drugs and scientific evidence demonstrating improved outcomes when these drugs are used in conjunction with a treat-to-target strategy (10,11). The concept of treating-to-target relies on adjusting therapy until a state of remission (or low disease activity) is achieved. Despite widespread endorsement from the rheumatology community, evidence suggests a significant gap in care in this area (4).

Evidence consists of important clinical trials of different treat-to-target strategies anchored on disease activity assessments showing better RA outcomes in the treat-to-target groups (12–15). Additionally, in an observational study involving 1,297 individuals, achievement of recommended disease targets was associated with improved physical function, health-related quality-of-life, and reduced hospitalizations (16). Finally, a large observational study in the Geisinger Health System demonstrated that implementing RA disease activity assessments using health information technology tools was associated with statistically significant improvements in RA disease control over time (17). This latter study is the only one that has found a link between the process of measuring and displaying RA disease activity and improved outcomes; the remainder use disease assessments as part of a larger treat-to-target strategy.

The working group recommended a measure requiring use of a validated outcome tool, as recommended by the ACR (6,18): the Clinical Disease Activity Index (CDAI), the Disease Activity Score in 28 joints (DAS28; using the erythrocyte sedimentation rate or C-reactive protein level), the Patient Activity Scale (PAS and PAS-II), the Routine Assessment of Patient Index Data with 3 measures (RAPID-3), and the Simplified Disease Activity Index. Each measure is an accurate reflection of disease activity, is sensitive to change, has remission criteria, is feasible to perform in clinical settings, and discriminates well between low, moderate, and high disease activity states. In other words, these measures can support a treat-to-target strategy in clinical practice (18). While starting with these measures, the working group also recommends that this list be periodically updated by the ACR to incorporate the latest advances in RA outcomes measurement.

Furthermore, the working group recommended that these measures be used in a majority ($\geq 50\%$) of RA encounters. The threshold of $\geq 50\%$ was chosen for several reasons. First, patients sometimes have an encounter for RA to address an acute issue (e.g., infection, drug adverse effect), and a disease activity measure may not be relevant at all encounters. Second, the working group recognized that instituting measures in clinical practice requires complex changes in clinical workflow. Experience at leading rheumatology centers suggests that achievement of 100% performance is not attainable and may even have unintended consequences in diverting resources from other clinical activities (17). In response to these issues, the working group recommended RA disease activity measurement occur at a majority ($\geq 50\%$) of encounters.

The working group considered other evidence suggesting gaps in care that justify use of the measure. Data from the ACR's Rheumatology Clinical Registry (RCR), used by rheumatologists for the Physician Quality Reporting System (PQRS), suggests room for improvement on this measure. In 2011, participating rheumatologists had a performance rate of 43.4% on a measure requiring assessment of disease activity at least once per year; performance has increased each year (43.4% in 2011, 54.4% in 2012, and 81.0% in 2013). Other studies also suggest a gap in performance. For example, one study from an academic medical center found that RA disease activity was only recorded in 29.0% of visits (19). In addition, although information on health care disparities are limited, data from a large US registry using the CDAI also found important differences in mean disease activity level across racial/ethnic groups, with African Americans being less likely to achieve clinical remission and having higher disease activity overall (20).

Finally, the working group considered existing analogous measures. The PQRS program has included a measure recommending that RA disease activity be assessed once per year. This measure had several limitations, including that no specific instruments were recommended, there was no requirement to record an actual outcome score (making it difficult to evaluate for improvement or provide benchmarking), and the measure only required assessment once per year, which may not adequately capture the clinical course of a patient with a chronic disease. The working group recommended that the newly proposed measure replace the older PQRS measure concept.

Assessment of functional status. Patient-reported outcomes (PROs) are of strong interest nationally and are meant to capture the patient's perspective in a structured way. Among chronic conditions, RA has robust scientific evidence around the validity of functional status PROs. Functional status assessments have been important outcome measures in RA clinical trials and studies, are responsive to therapy changes, and are strong predictors of future disability and mortality (21). Measuring physical function is recommended in RA guidelines because it is a key factor in assessing prognosis and therefore the choice of DMARDs, and because assessment at regular intervals helps determine if a key treatment goal, i.e., maintaining functional capacity, is being achieved (6,22–24). Both US and international groups therefore recommend that provider treatment decisions take functional status into consideration (6,11,23).

Although there is strong evidence supporting the importance of functional status as a health outcome in RA, few studies have examined the impact of PRO implementation on health outcomes. However, there is some published experience in implementing PROs in RA, including the Swedish Quality Register for Arthritis, large US health systems such as Geisinger Health System, and in many practice settings (17,25,26). In addition, studies have demonstrated that functional status assessments impact therapy decisions. For example, in a German study of 1,467 individuals with RA who had undergone a treatment change or started a DMARD, after disease activity assessment using the DAS, functional status assessment had the highest influence on therapy decisions (27).

The working group recommended that the functional status quality measure require use of a validated tool. Members of the working group reviewed the scientific literature on available measures. In addition, a survey was created and administered to experts in RA functional status assessment (for the list of experts, see Supplementary Appendix A, available on the *Arthritis Care & Research* web site at <http://onlinelibrary.wiley.com/doi/10.1002/acr.22984/abstract>). Measures that had high quality evidence supporting their psychometric properties and were deemed by experts to be feasible for use in clinical practice were recommended. Feasibility assessment took into account time to administer the PRO, time to score the questions, availability in multiple languages, suitability for lower health literacy populations, and whether there were examples of successful use of the measures in clinical practice. Measures selected included the Multidimensional Health Assessment Questionnaire (MDHAQ), the HAQ-II, and the Patient-Reported Outcomes Measurement System Physical Function Instruments (PROMIS-PF10, PROMIS-PF20, and PROMIS-PF computer adaptive test [21,28–31]); older legacy measures such as the original HAQ and Short Form 36 (SF-36) were less preferable because of weaker psychometric properties, length (HAQ, SF-36), and licensing regulations (SF-36). The working group also recommended that scientific advances in PRO measures be incorporated into future iterations of this measure.

The working group considered whether there was opportunity for improvement for this measure. Although population-wide data are lacking, data reported through the ACR's RCR show that performance on a related measure (recording of functional status once per year using any method, including narrative assessment) was 69.6% in 2011 and improved to 86.6% in 2012. This older PQRS measure has several limitations, including that no specific instruments to assess functional status were recommended and there was no requirement to record an actual outcome score (making it difficult to evaluate for improvement or provide benchmarking). The working group recommended that the newly proposed measure replace the older PQRS measure concept.

DMARD therapy. Use of DMARDs in every patient with active RA at the earliest stage of disease, ideally within 3 months of disease onset, is recommended in guidelines (6,23). These guidelines are based on results from numerous clinical trials demonstrating that DMARDs slow the progression of RA by decreasing inflammation and reducing articular erosions. In addition, both clinical trials and observational studies demonstrate that DMARDs improve functional status and health-related quality of life (32). The working group recommended that the measure include a continuously updated list of all available DMARDs that demonstrate efficacy for inflammatory arthritis in clinical trials.

The working group also considered exclusions. Since 2005, the National Committee for Quality Assurance (NCQA) has maintained a DMARD quality measure that relies on billing data and is used for health plan quality reporting. This measure excludes individuals with human immunodeficiency virus (HIV) and pregnant women. These exclusions are justified since there is inadequate evidence regarding the use of most DMARDs in HIV and since many DMARDs are

either frankly teratogenic (e.g., methotrexate, leflunomide) or have not been adequately studied in pregnant women (33). The working group also recommended adding an exclusion for inactive RA (as indicated by coding “diagnosis, inactive: rheumatoid arthritis”) based on feedback from rheumatologists on the NCQA measure over the past decade. This exclusion was felt to be clinically justified since some studies suggest that up to 9–15% of individuals with RA may achieve a drug-free remission over the course of their disease (34).

The working group considered whether there was currently opportunity for improvement for this measure. Several studies suggest significant variation in DMARD use among individuals with RA (35). For example, research using the DMARD measure in billing data has found a relatively large difference in use based on age, with older individuals being less likely to receive DMARDs. African Americans, those with low personal incomes, and those residing in zip codes with low socioeconomic status also have significantly lower DMARD use (3,36). DMARD use is higher for patients seeing rheumatologists, with recent estimates from the RCR showing 96.8%; however, no studies have found a 100% DMARD use rate, as there may be reasonable clinical exceptions in practice. Although performance on this measure is expected to be high among rheumatologists, previous studies showing potential disparities in care justified the need for continued use of this measure.

Tuberculosis (TB) screening. Latent TB infection affects an estimated 9.6 to 14.9 million people in the US (37). Biologic and new small-molecule DMARDs increase the risk of reactivation of latent TB infection. Tumor necrosis factor (TNF) plays an important role in host responses to mycobacteria, and TNF inhibitors are therefore associated with a higher risk of TB infection. Similar associations have been discovered with other biologic DMARDs used in RA, with the possible exception of B cell-depleting agents such as rituximab.

No trials have examined the effectiveness of different screening strategies for TB prior to initiation of biologic or new small-molecule DMARDs. Instead, data on TB risk and screening is observational and has accumulated from clinical trials, postmarketing surveillance, and large registries. Early randomized clinical trials of TNF inhibitors performed before TB screening became standard of care demonstrated a 4-fold higher risk of TB infection (38,39). Similar data have been reported for the new small molecule drug tofacitinib (40). Based on this evidence, the ACR, Centers for Disease Control and Prevention (CDC), and international guidelines recommend testing patients for latent TB prior to initiating these DMARDs regardless of presence of risk factors (6,23,41). Biologic or tofacitinib DMARD therapy is contraindicated in those with either active or latent TB until appropriate antimicrobial therapy is started (6).

Consistent with ACR and CDC guidelines, the working group recommended that the eCQM capture screening for TB with either a tuberculin skin test or an interferon-release assay. In addition, we considered the scenario in which patients were treated for latent or active TB in the past. These patients have persistently positive TB screening tests, and retesting will not add new information. For this population, the working group recommended that the eCQM include evidence of prior treatment as satisfying the numerator.

Table 2. Data from the American College of Rheumatology's RA quality measures project expert panel rating process for face validity and feasibility, by measure*

| Measures | Median score for validity | Median score for feasibility | Raters with validity score ≤ 3 , no. | Raters with validity score ≥ 7 , no. | Total raters, no. | Raters with score ≤ 3 , % |
|---------------------------------|---------------------------|------------------------------|---|---|-------------------|--------------------------------|
| Assessment of disease activity | 9 | 7 | 1 | 11 | 14 | 7.14 |
| Assessment of functional status | 8 | 8 | 1 | 11 | 14 | 7.14 |
| DMARD therapy | 9 | 8 | 1 | 13 | 14 | 7.14 |
| TB screening | 9 | 8.5 | 0 | 14 | 14 | 0 |

* Panelists were provided with the following instructions: "Your validity ratings should reflect whether you believe that the measure can be used to reflect the quality of care for RA. Questions to consider in determining your validity ratings should include: Is there adequate scientific evidence or professional consensus to support the indicator? Are there identifiable health benefits to patients who receive care specified by the indicator? Based on your professional experience, would you consider providers with significantly higher rates of adherence to the indicator higher quality providers? Are the majority of factors that determine adherence to the indicator under the control of the physician or health care system?" Measure scale definitions: for validity, 1 = definitely NOT valid to 9 = definitely valid; for feasibility, 1 = definitely NOT feasible to 9 = definitely feasible. RA = rheumatoid arthritis; DMARD = disease-modifying antirheumatic drug; TB = tuberculosis.

In devising the denominator population for the measure, the working group recognized that identifying prior TB screening in prevalent biologic agent or tofacitinib DMARD users would be difficult. Challenges include that such screening might have been documented in paper records prior to the transition to EHRs or may be documented in a different prescribing physician's records. For these reasons, the working group recommended that the measure examine incident users of all biologic DMARDs (except rituximab, where no safety signal has been found) and tofacitinib, since the current prescriber of these DMARDs could reasonably be held accountable for documenting TB screening and treatment in the EHR at the time of prescription. An incident user was defined as a patient with no prescription for a biologic DMARD or tofacitinib in the year preceding the measurement year.

The working group also considered whether there was currently opportunity for improvement for this measure. Although population-based studies in the US are not available, data from the PQRS program found that performance on the TB screening measure was 73.6% in 2011, rising to 92.9% in 2012, and 90.5% in 2013.

Interdisciplinary consensus ratings and electronic specification. The measure concepts and data reviewed above were presented to an interdisciplinary consensus panel, and slight revisions were made. For example, we clarified that attribution of all measures was to the

rheumatologist (rather than other health care providers). Ratings on the revised measure concepts are provided in Table 2. Median scores for validity were high (8 or 9). Panel members rated the measures as potentially feasible, with median ratings between 7 and 8.5. Disagreement, as assessed by the number of raters with validity scores ≤ 3 was low or not present. Public comment and review by the ACR Quality of Care Committee and Board of Directors resulted in only recommendations to improve clarity but not change content. The approved measures were then converted to electronic measure format using the MAT and QDM, as detailed extensively elsewhere (5).

eCQM field testing. Characteristics of sites where testing was performed are listed in Supplementary Appendix A, available on the *Arthritis Care & Research* web site at <http://onlinelibrary.wiley.com/doi/10.1002/acr.22984/abstract>. Performance on the quality measures is described in Table 3. Five sites were involved in testing, but each site tested only 3 measures. Below we summarize the key findings of field-testing for each eCQM.

Feasibility assessment. Quantitative results of the targeted feasibility assessment are included in the Supplementary Appendix A, available on the *Arthritis Care & Research* web site at <http://onlinelibrary.wiley.com/doi/10.1002/acr.22984/abstract>. In general, sites rated current feasibility of the 3 data elements between 2 and 3 (with 3 indicating that the information is from the most authoritative source when

Table 3. Performance on RA eCQMs, as assessed by proportion of a random sample of eligible patients receiving recommended care*

| | Overall performance, no. (%) | Site 1, no. (%) | Site 2, no. (%) | Site 3, no. (%) |
|---------------------------------|------------------------------|-----------------|-----------------|-----------------|
| Assessment of disease activity | 190 (53.6) | 37 (89.1) | 34 (38.2) | 119 (59.7) |
| Assessment of functional status | 223 (69.1) | 70 (62.8) | 34 (2.9) | 119 (93.2) |
| DMARD therapy | 175 (93.1) | 81 (88.9) | 34 (94.1) | 60 (98.3) |
| TB screening | 47 (72.8) | 66 (86.4) | 40 (67.5) | 41 (56.1) |

* Values are the number in random sample of patients from the electronic health record meeting denominator for manual chart review (%). RA = rheumatoid arthritis; eCQMs = electronic clinical quality measures; DMARD = disease-modifying antirheumatic drug; TB = tuberculosis.

Table 4. Performance measure score concurrent validity for numerators of RA eQMs*

| | Assessment of disease activity | Functional status assessment | DMARD therapy | TB screening |
|--|--------------------------------------|------------------------------------|------------------|-----------------|
| Site 1† | | | | |
| Records assessed at site, no. | 70 | 81 | 81 | 62 |
| Kappa‡ | 1.00 | 1.00 | 1.00 | 1.00 |
| Sensitivity, % | 100 | 100 | 100 | 100 |
| Specificity, % | 100 | 100 | 100 | 100 |
| Site 2§ | | | | |
| Records assessed at site, no. | 34 | 34 | 34 | 35 |
| Kappa‡ | 0.17 | NA | NA | 0.85 |
| Sensitivity, % (95% CI) | 44 (25–66) | 3 (0.1–66) | 94 (80–99) | 93 (76–99) |
| Specificity, % (95% CI) | 86 (42–100) | Undef. | Undef. | 100 (63–100) |
| Site 3¶ | | | | |
| Records assessed at site, no. | 117 | 117 | 58 | 36 |
| Kappa‡ | 0.98 | 0.73 | NA | 0.33 |
| Sensitivity, % (95% CI) | 99 (92–100) | 96.4 (91–99) | 98 (91–100) | 69 (50–84) |
| Specificity, % (95% CI) | 100 (92–100) | 100.0 (54–100) | Undef. | 100 (40–100) |
| Overall | | | | |
| Records assessed at site, no. | 221 | 232 | 173 | 133 |
| Kappa‡ | 0.84 | 0.54 | 0.85 | 0.73 |
| Sensitivity, % (95% CI) | 88 (82–93) | 80.4 (74–86) | 98 (95–100) | 89 (82–94) |
| Specificity, % (95% CI) | 99 (94–100) | 100 (89–100) | 100 (66–100) | 100 (84–100) |
| * RA = rheumatoid arthritis; eQMs = electronic clinical quality measures; DMARD = disease modifying antirheumatic drug; TB = tuberculosis; 95% CI = 95% confidence interval; NA = not applicable; Undef. = undefined. | | | | |
| † Site 1 had perfect agreement on all measures. | | | | |
| ‡ As a measure of agreement, $\kappa = 1.0$ when the automated electronic health record data query agrees exactly with data obtained through manual chart extraction, and $\kappa = 0$ when the agreement appears entirely due to chance. Some kappas were undefined because there were no true negatives; this is an inherent limitation of the kappa statistic that requires a distribution to produce a result. Similarly, specificity was undefined in instances where the denominator of the specificity calculation (false positive + true negative) was zero. | | | | |
| § Site 2 was different for TB screening. | | | | |
| ¶ Site 3 was different for DMARD therapy and TB screening. | | | | |

it enters the EHR and is highly likely to be correct, 2 indicating that the information only has a moderate likelihood of being correct, and 1 indicating that the data is not accurately captured in the EHR). For the 2 quality measures of assessment of disease activity and assessment of functional status, some practices had fully operationalized workflows to enter assessments such as CDAI and RAPID-3 scores into their EHR systems, while others had not.

Concurrent validity assessment. Results for concurrent validity of the numerator of the eQMs are outlined in Table 4. As shown in the Table, there was variability between sites in our statistical measure of agreement (kappa), as well as in sensitivity and specificity.

Site 1 had an advanced EHR with well-established workflows to capture information on RA quality measures. Data elements for RA quality measures at this site were refined and tested over many years, leading to perfect agreement ($\kappa = 1.0$, 95% confidence interval 1.0–1.0) as well as a sensitivity and specificity of 100% on all measures. There was more variability at other sites. For example, while all sites were able to demonstrate good sensitivity and specificity for capturing DMARD use among patients with RA (sensitivity 98%, specificity 100% across sites), one site did not routinely capture disease activity or functional status in a structured EHR field, leading to lower sensitivity and

specificity (44% and 86% for disease activity, respectively, and 3% and undefined for functional status, respectively). Similarly, for the quality measure TB screening, sensitivity was reduced (69%) at a site serving a high-risk population in which many patients were previously treated for latent TB. Information about prior latent TB treatment was available in the text of clinical notes, but did not appear elsewhere in a structured format.

We also assessed validity for the denominator of each measure. There was excellent agreement across all measures. For example, 221 of 223 individuals tested across 3 sites for assessment of disease activity and assessment of functional status had agreement between the automated data extract and the chart review defining RA (99% accuracy). In the discordant cases, patients met the denominator definition for inclusion in the eQMs (including age ≥ 18 years, and 2 face-to-face encounters during the measurement year for RA) but did not have RA. This discrepancy resulted from the clinician incorrectly coding the patient's diagnosis as RA when they had a related condition with inflammatory arthritis (e.g., mixed connective tissue disease).

Similarly, agreement was excellent for the quality measure DMARD therapy (173 of 175 patients accurately identified, 99% accuracy). Agreement was slightly lower for the

denominator component of TB screening (133 of 147 patients, accuracy 90%) because of instances where the patient was not a new biologic DMARD user; medication reconciliation was incomplete in these cases.

Submission for national endorsement. Measures were submitted to the NQF in March 2014 for endorsement. An interdisciplinary panel of 21 national experts, the Musculoskeletal Standing Committee, convened to review the measures. This was followed by public and member comment through requests sent to NQF members and through the NQF web site. Finally, the measures were examined by the NQF Consensus Standards Approval Committee and Board of Directors, who voted to either fully or conditionally approve the RA measures for endorsement. The full report that includes all of these ratings and deliberations can be found at http://www.qualityforum.org/Publications/2015/01/NQF-Endorsed_Measures_for_Musculoskeletal_Conditions.aspx.

Implementation in RISE. The RA eCQMs were implemented in the ACR's national EHR-enabled registry, RISE. RISE passively collects data from practices, analyzes data centrally to allow benchmarking, and can be used for national quality-reporting programs. Using an iterative data mapping process, the RISE data team worked with individual practices to make sure data elements in each eCQM were adequately captured. For 2 RA eCQMs, RA: assessment of disease activity and assessment of functional status, information to satisfy the numerator (disease activity score and functional status score) was not available in a structured data field for some practices, and the team developed text-mining algorithms to capture these scores from clinical notes in the EHR. Work is ongoing to refine eCQM extraction using this methodology.

Discussion

Using a multifaceted approach that relied on scientific evidence, interdisciplinary stakeholder involvement, and electronic specification and testing of measures in different EHR systems, the ACR has developed eCQMs for RA. The 4 eCQMs cover assessment of key outcomes (disease activity and functional status), treatment (DMARD use), and patient safety (TB screening). The measures build on the foundation of quality measurement in RA over the last decade, while incorporating newer data standards such as the QDM and testing in EHRs to create a set of measures designed for rapid cycle quality improvement. The measures have now been implemented nationwide in the ACR's EHR-based registry, RISE.

Although the development of eCQMs holds significant promise in leveraging the rich resource of EHR data, we anticipate that methods for refinement and further testing of such measures will continue to evolve. Significant methodological advances in the last several years include development of the QDM, an Office of the National Coordinator-sponsored standard for representing clinical concepts for the Meaningful Use program. The QDM has created a standard for constructing quality measures. For example, in the QDM, clinical data are represented as a set of codes from a

standardized terminology system, such as the ICD, the LOINC, or the SNOMED. However, as illustrated in our project, execution of these QDM-based algorithms still requires mapping at individual sites to ensure that both measure data elements and logic are captured appropriately.

Testing of eCQMs at several clinical sites allowed us to analyze the feasibility and validity of EHR implementation before scaling efforts to the national registry. Several types of challenges were identified during the course of testing. First, 2 of the measures, assessment of disease activity and assessment of functional status, required extracting an outcome measurement score from the EHR. Because these outcomes were not captured in the QDM, the ACR submitted the measures to the Value Set Authority Center for inclusion. Next, for practices that record this information in a structured field in the EHR, mapping to identify those data required customization at each site. However, as demonstrated in our testing, some practices had not yet transitioned to collecting this information in a structured field. Using the QDM and standardized structured data queries would therefore be inadequate for capturing clinical performance in these practices in the near term. This allowed us to anticipate that implementing eCQMs in RISE would require using procedures such as text mining to capture required data elements. Further work to validate and refine text-mining algorithms is needed, and will likely continue to play a role in capturing eCQM data in the future. EHR vendors can also facilitate eCQMs by providing options for structured and standardized workflows for capture of high priority data elements.

Our work also highlights challenges of eCQM implementation and data extraction. First, the current lack of interoperability between data systems poses significant barriers to accurate data capture. For example, in the safety net hospital testing site, different EHR systems are used for the inpatient and outpatient settings, and medication histories prior to the recent outpatient EHR implementation are not available. This created an important data gap for the TB screening measure. A large number of patients in this setting have latent TB and have been treated appropriately for this condition in the past. However, record of this treatment is not available in the current EHR system and requires manual chart review of older clinical data. eCQM performance therefore looks falsely low and would require implementation of a new and targeted data collection strategy to improve. Similarly, the TB screening eCQM required identification of incident users of biologic or new small-molecule medications. Incomplete or inaccurate medication reconciliation also posed challenges for this measure. Examples include incomplete capture of infusible biologic medications, which are not e-prescribed, or difficulties ascertaining incident biologic agent users because of out-of-date medication information.

We see the methods described here as foundational and expect that both our eCQM specifications and methods to extract data from EHRs will continue to evolve. For example, as new drugs become available, our eCQMs will be continually updated and applied to RISE. In addition, we anticipate that improvements in EHR standardization and interoperability over time will lead to increasingly

accurate data capture. However, in the near term, working with practices to map individual data elements and using methods such as text mining and natural language processing will likely be required to paint an accurate picture of clinical quality in rheumatology practices. Finally, as more rheumatologists create workflows to capture RA outcomes in clinical practice, measurement and benchmarking of patient outcomes to facilitate quality improvement and population management strategies will become possible.

This effort to develop the first set of eCQMs in rheumatology has limitations, many of which are inherent to a new and developing field. First, although we tested our eCQMs in commonly used EHR systems to understand their feasibility and validity, the results presented here are not representative of all EHR systems in the US. Our testing occurred largely in health systems that had the information technology support to build the eCQMs locally. Second, some of the clinical sites have established workflows to not only collect, but also report, performance on RA quality measures. Data quality and performance at these sites likely exceeds that of many rheumatology clinics. Finally, many questions remain about the feasibility and validity of widespread eCQM implementation in clinical practice, and rheumatology is among the first specialties to embark on a national EHR-enabled registry to collect such measures. RISE is already mapping to more than 30 different EHR products, with customization and mapping at each individual clinical site to ensure data accuracy. Currently, data on more than 90,000 individuals with RA is being collected in the registry. These experiences with eCQM deployment will influence future iterations of the measures and their implementation.

In conclusion, we used a multifaceted process to build eCQMs in RA. We found that a diverse group of stakeholders, including rheumatologists, patients, and national organizations rated the content of RA measures as important and valid measures of quality. Testing revealed that eCQM deployment is feasible in most practices, but that the lack of standardization of data elements in current EHRs necessitates local mapping and customization to ensure that data is accurate. These initial results in developing and testing RA eCQMs have laid an important foundation for using EHRs as a resource for quality improvement in rheumatology.

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. Yazdany 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. Yazdany, Robbins, Schmajuk, Desai, Lacaille, Neogi, Singh, Genovese, Myslinski, Newman.

Acquisition of data. Yazdany, Myslinski, Fisk, Francisco.

Analysis and interpretation of data. Yazdany, Fisk, Francisco.

REFERENCES

1. National Quality Forum. Measure Prioritization Advisory Committee Report. Prioritization of high-impact Medicare conditions and measure gaps. 2010 URL: <http://www.qualityforum.org/projects/prioritization.aspx>.
2. Helmick CG, Felson DT, Lawrence RC, Gabriel S, Hirsch R, Kwoh CK, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I. *Arthritis Rheum* 2008;58:15–25.
3. Schmajuk G, Trivedi AN, Solomon DH, Yelin E, Trupin L, Chakravarty EF, et al. Receipt of disease-modifying antirheumatic drugs among patients with rheumatoid arthritis in Medicare managed care plans. *JAMA* 2011;305:480–6.
4. Desai SP, Yazdany J. Quality measurement and improvement in rheumatology: rheumatoid arthritis as a case study. *Arthritis Rheum* 2011;63:3649–60.
5. Yazdany J, Myslinski R, Miller A, Francisco M, Desai S, Schmajuk, et al. Methods for Developing the American College of Rheumatology's Electronic Clinical Quality Measures. *Arthritis Care Res (Hoboken)* 2016;68:1402–9.
6. Singh JA, Furst DE, Bharat A, Curtis JR, Kavanaugh AF, Kremer JM, et al. 2012 update of the 2008 American College of Rheumatology recommendations for the use of disease-modifying antirheumatic drugs and biologic agents in the treatment of rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2012;64:625–39.
7. MacLean CH, Saag KG, Solomon DH, Morton SC, Sampsel S, Klippel JH. Measuring quality in arthritis care: methods for developing the Arthritis Foundation's quality indicator set. *Arthritis Rheum* 2004;51:193–202.
8. Brook RH. The RAND/UCLA appropriateness method. In: McCormick KA, Moore SR, Siegel RA, eds. *Methodology perspectives*, AHCPR Pub No. 95-009, Rockville (MD) Public Health Service, U.S. Department of Health and Human Services; 1994. pp. 59–70.
9. National Quality Forum. Quality data model. 2012. URL: https://ecqi.healthit.gov/system/files/qdm/qdm_122012.pdf
10. Schoels M, Knevel R, Aletaha D, Bijlsma JW, Breedveld FC, Boumpas DT, et al. Evidence for treating rheumatoid arthritis to target: results of a systematic literature search. *Ann Rheum Dis* 2010;69:638–43.
11. 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.
12. 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.
13. Verstappen SM, Jacobs JW, van der Veen MJ, Heurkens AH, Schenk Y, ter Borg EJ, et al. Intensive treatment with methotrexate in early rheumatoid arthritis: aiming for remission. Computer Assisted Management in Early Rheumatoid Arthritis (CAMERA, an open-label strategy trial). *Ann Rheum Dis* 2007;66:1443–9.
14. Franssen J, Moens HB, Speyer I, van Riel PL. Effectiveness of systematic monitoring of rheumatoid arthritis disease activity in daily practice: a multicentre, cluster randomised controlled trial. *Ann Rheum Dis* 2005;64:1294–8.
15. Symmons D, Tricker K, Roberts C, Davies L, Dawes P, Scott DL. The British Rheumatoid Outcome Study Group (BROSG) randomised controlled trial to compare the effectiveness and cost-effectiveness of aggressive versus symptomatic therapy in established rheumatoid arthritis. *Health Technol Assess* 2005; 9:iii–iv, ix–x, 1–78.
16. Alemao E, Joo S, Kawabata H, Al MJ, Allison PD, Rutten-van Molken MP, et al. Effects of achieving target measures in rheumatoid arthritis on functional status, quality of life, and resource utilization: analysis of clinical practice data. *Arthritis Care Res (Hoboken)* 2016;68:308–17.
17. Newman ED, Lerch V, Billet J, Berger A, Kirchner HL. Improving the quality of care of patients with rheumatic disease using patient-centric electronic redesign software. *Arthritis Care Res (Hoboken)* 2015;67:546–53.
18. 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.
19. Desai SP, Liu CC, Tory H, Norton T, Frits M, Lillegraven S, et al. Rheumatoid arthritis quality measures and radiographic progression. *Semin Arthritis Rheum* 2014;44:9–13.
 20. Greenberg JD, Spruill TM, Shan Y, Reed G, Kremer JM, Potter J, et al. Racial and ethnic disparities in disease activity in patients with rheumatoid arthritis. *Am J Med* 2013;126:1089–98.
 21. Maska L, Anderson J, Michaud K. Measures of functional status and quality of life in rheumatoid arthritis: Health Assessment Questionnaire Disability Index (HAQ), Modified Health Assessment Questionnaire (MHAQ), Multidimensional Health Assessment Questionnaire (MDHAQ), Health Assessment Questionnaire II (HAQ-II), Improved Health Assessment Questionnaire (Improved HAQ), and Rheumatoid Arthritis Quality of Life (RAQoL). *Arthritis Care Res (Hoboken)* 2011;63 Suppl 11:S4–13.
 22. Stoffer MA, Smolen JS, Woolf A, Ambrozic A, Bosworth A, Carmona L, et al. Development of patient-centred standards of care for rheumatoid arthritis in Europe: the eumusc.net project. *Ann Rheum Dis* 2014;73:902–5.
 23. Smolen JS, Landewe R, Breedveld FC, Buch M, Burmester G, Dougados M, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2013 update. *Ann Rheum Dis* 2014;73:492–509.
 24. Bykerk VP, Akhavan P, Hazlewood GS, Schieir O, Dooley A, Haraoui B, et al. Canadian Rheumatology Association recommendations for pharmacological management of rheumatoid arthritis with traditional and biologic disease-modifying antirheumatic drugs. *J Rheumatol* 2012;39:559–82.
 25. Ovretveit J, Keller C, Hvitfeldt Forsberg H, Essen A, Lindblad S, Brommels M. Continuous innovation: developing and using a clinical database with new technology for patient-centred care: the case of the Swedish quality register for arthritis. *Int J Qual Health Care* 2013;25:118–24.
 26. Pincus T, Castrejon I. MDHAQ/RAPID3 scores: quantitative patient history data in a standardized “scientific” format for optimal assessment of patient status and quality of care in rheumatic diseases. *Bull NYU Hosp Jt Dis* 2011;69:201–14.
 27. Malysheva O, Bedrich A, Kuipers JG, Kleine H, Wolff B, Baerwald CG. Use of clinical scores to guide therapeutic decisions in patients with rheumatoid arthritis in daily care. *Clin Exp Rheumatol* 2015;33:255–8.
 28. 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.
 29. 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.
 30. Fries JF, Cella D, Rose M, Krishnan E, Bruce B. Progress in assessing physical function in arthritis: PROMIS short forms and computerized adaptive testing. *J Rheumatol* 2009;36:2061–6.
 31. 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.
 32. Gaujoux-Viala C, Smolen JS, Landewe R, Dougados M, Kvien TK, Mola EM, et al. Current evidence for the management of rheumatoid arthritis with synthetic disease-modifying antirheumatic drugs: a systematic literature review informing the EULAR recommendations for the management of rheumatoid arthritis. *Ann Rheum Dis* 2010;69:1004–9.
 33. 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.
 34. Van der Woude D, Young A, Jayakumar K, Mertens BJ, Toes RE, van der Heijde D, et al. Prevalence of and predictive factors for sustained disease-modifying antirheumatic drug-free remission in rheumatoid arthritis: results from two large early arthritis cohorts. *Arthritis Rheum* 2009;60:2262–71.
 35. Schmajuk G, Solomon DH, Yazdany J. Patterns of disease-modifying antirheumatic drug use in rheumatoid arthritis patients after 2002: a systematic review. *Arthritis Care Res (Hoboken)* 2013;65:1927–35.
 36. Yazdany J, Tonner C, Schmajuk G, Lin GA, Trivedi AN. Receipt of glucocorticoid monotherapy among Medicare beneficiaries with rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2014;66:1447–55.
 37. Taylor Z, Nolan CM, Blumberg HM; American Thoracic Society; Centers for Disease Control and Prevention; Infectious Diseases Society of America. Controlling tuberculosis in the United States. Recommendations from the American Thoracic Society, CDC, and the Infectious Diseases Society of America. *MMWR Recomm Rep* 2005;54:1–81.
 38. Maini R, St.Clair EW, Breedveld F, Furst D, Kalden J, Weisman M, et al. Infliximab (chimeric anti-tumour necrosis factor α monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised phase III trial. ATTRACT Study Group. *Lancet* 1999;354:932–9.
 39. Keane J, Gershon S, Wise RP, Mirabile-Levens E, Kasznica J, Schwiertman WD, et al. Tuberculosis associated with infliximab, a tumor necrosis factor α -neutralizing agent. *New Engl J Med* 2001;345:98–104.
 40. Winthrop KL, Park SH, Gul A, Cardiel MH, Gomez-Reino JJ, Tanaka Y, et al. Tuberculosis and other opportunistic infections in tofacitinib-treated patients with rheumatoid arthritis. *Ann Rheum Dis* 2016;75:1133–8.
 41. Mazurek GH, Jereb J, Vernon A, LoBue P, Goldberg S, Castro K, et al. Updated guidelines for using interferon gamma release assays to detect Mycobacterium tuberculosis infection: United States, 2010. *MMWR Recomm Rep* 2010;59:1–25.