

# American College of Rheumatology White Paper on Performance Outcome Measures in Rheumatology

LISA G. SUTER,<sup>1</sup> CLAIRE E. BARBER,<sup>2</sup> JEPH HERRIN,<sup>3</sup> AMYE LEONG,<sup>4</sup> ELENA LOSINA,<sup>5</sup> AMY MILLER,<sup>6</sup> ERIC NEWMAN,<sup>7</sup> MARK ROBBINS,<sup>8</sup> HEATHER TORY,<sup>9</sup> AND JINOOS YAZDANY<sup>10</sup>

**Objective.** To highlight the opportunities and challenges of developing and implementing performance outcome measures in rheumatology for accountability purposes.

**Methods.** We constructed a hypothetical performance outcome measure to demonstrate the benefits and challenges of designing quality measures that assess patient outcomes. We defined the data source, measure cohort, reporting period, period at risk, measure outcome, outcome attribution, risk adjustment, reliability and validity, and reporting approach. We discussed outcome measure challenges specific to rheumatology and to fields where patients have predominantly chronic, complex, ambulatory care-sensitive conditions.

**Results.** Our hypothetical outcome measure was a measure of rheumatoid arthritis disease activity intended for evaluating Accountable Care Organization performance. We summarized the components, benefits, challenges, and trade-offs between feasibility and usability. We highlighted how different measure applications, such as for rapid cycle quality improvement efforts versus pay for performance programs, require different approaches to measure development and testing. We provided a summary table of key take-home points for clinicians and policymakers.

**Conclusion.** Performance outcome measures are coming to rheumatology, and the most effective and meaningful measures can only be created through the close collaboration of patients, providers, measure developers, and policymakers. This study provides an overview of key issues and is intended to stimulate a productive dialogue between patients, practitioners, insurers, and government agencies regarding optimal performance outcome measure development.

## Introduction

Standardized assessment of health care outcomes for accountability purposes is a national priority. In contrast to assessments of health care structure or processes of

care, outcome measures evaluate the results of care and are therefore considered the most valid metrics for measuring and comparing clinical care, driving quality and outcome improvement, and potentially increasing

Dr. Suter is supported by the Centers for Medicare & Medicaid Services, an agency of the US Department of Health and Human Services (contract HHSM-500-2013-13018I, Task Order HHSM-500-T0001), the VA Connecticut Healthcare System, and has provided clinical consultation for an NIH-funded grant led by Dr. Losina. Ms Leong is supported by the NIH, the National Institute of Arthritis and Musculoskeletal and Skin Diseases, and the Archstone Foundation. Dr. Yazdany is supported by the Robert L. Kroc Chair in Rheumatic and Connective Tissue Diseases, the Agency for Healthcare Research and Quality (grant R01-HS024412), and the Russell/Engleman Medical Research Center for Arthritis.

<sup>1</sup>Lisa G. Suter, MD: Yale School of Medicine and YNHSC Center for Outcome Research and Evaluation, New Haven, Connecticut, and VA Connecticut Healthcare System, West Haven; <sup>2</sup>Claire E. Barber, MD, PhD, FRCPC: University of Calgary, Calgary, Alberta, Canada; <sup>3</sup>Jeph Herrin, PhD: Yale School of Medicine, New Haven, Connecticut, and Health Research & Educational Trust, Chicago, Illinois; <sup>4</sup>Amye

Leong, MBA: Healthy Motivation, and Bone and Joint Decade, the Global Alliance for Musculoskeletal Health, Santa Barbara, California; <sup>5</sup>Elena Losina, PhD, MSc: Orthopedics and Arthritis Center for Outcomes Research, Brigham and Women's Hospital, Harvard Medical School, and Boston University School of Public Health, Boston, Massachusetts; <sup>6</sup>Amy Miller: American College of Rheumatology, Atlanta, Georgia; <sup>7</sup>Eric Newman, MD: Geisinger Health System, Danville, Pennsylvania; <sup>8</sup>Mark Robbins, MD, MPH: Harvard Vanguard Medical Associates/Atrius Health, Somerville, Massachusetts; <sup>9</sup>Heather Tory, MD, MPH: Connecticut Children's Medical Center, Hartford, Connecticut, and University of Connecticut School of Medicine, Farmington, Connecticut; <sup>10</sup>Jinoos Yazdany, MD, MPH: University of California, San Francisco.

Ms Leong has received consulting fees and speaking fees from Horizon, GlaxoSmithKline, and Zimmer (less than \$10,000 each). Dr. Losina has received consulting fees and speaking fees from the University of North Carolina (more than \$10,000). Dr. Newman is in the process of

## Significance & Innovations

- This study highlights the current focus on outcomes measurement in medicine.
- We discuss the challenges and opportunities associated with developing and implementing outcome quality measures intended for assessing provider performance in rheumatology.
- The differences between outcome measures intended for use in clinical trials and those specified and suitable for accountability purposes are explored.

provider and health system accountability (Figure 1). Despite controversies and challenges, public reporting of performance outcome measures (POMs; measures used to assess performance of the health care system or its constituents) is associated with improvements in clinical outcomes. Examples include declines in mortality after coronary artery bypass surgery (1), central line-associated bloodstream infections (2), and hospital mortality and readmissions following acute myocardial infarction, congestive heart failure, and pneumonia (3–5). These successes, along with process measure limitations, such as their lack of concordance with patient outcomes (6), have reinforced a national shift toward outcome measurement (7). As rheumatologists are increasingly impacted by outcome measures (8,9), it is important to demystify these complex metrics for practicing clinicians.

Few validated POMs are integrated into routine clinical practice, and most evaluate either acute episodic care or surgical procedures and are not applicable to rheumatologists. The application of outcome measures to chronic disease care, which is characteristic of rheumatology, involves greater methodologic complexity. The goal of this study was to highlight the opportunities and challenges of developing and implementing POMs in rheumatology in order to improve outcomes of patients with rheumatic diseases. It is our hope that we will stimulate a dialogue between practitioners, insurers, and government agencies.

### What are POMs?

POMs are metrics used to evaluate the quality of health care (“performance”) delivered by an individual

commercializing proprietary software to improve care quality through the electronic health record. Dr. Yazdany has received an independent research award from Pfizer and speaking fees and honoraria from Genentech (less than \$10,000 each).

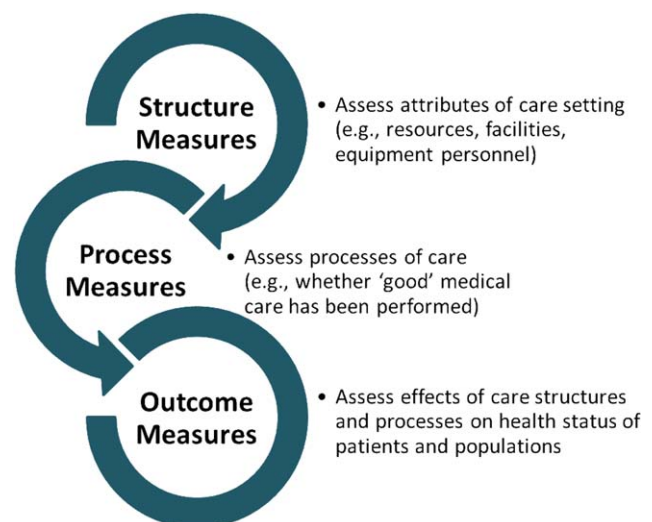
Address correspondence to Lisa G. Suter, MD, Section of Rheumatology, Department of Internal Medicine, Yale University School of Medicine, 300 Cedar Street, Room TAC S541, PO Box 208031, New Haven, CT 06520-8031. E-mail: [lisa.suter@yale.edu](mailto:lisa.suter@yale.edu).

Submitted for publication October 30, 2015; accepted in revised form April 26, 2016.

physician, group practice, hospital, or other provider. They encompass patient outcomes attributable to that provider and enable valid comparisons across providers or with an established benchmark. They include assessments of patient experience, symptoms, function, clinical events, or even costs. An example of an outcome after hospital discharge is mortality; a corresponding outcome measure might be hospital-level mortality rates within 30 days of admission. An example of an outcome following outpatient knee osteoarthritis care might be a patient’s pain level; a corresponding outcome measure might assess patients’ average pain rating during a defined period. Below we “build” a hypothetical POM in order to define each measure component and review the associated benefits and challenges; key points are shown in Table 1.

### Why measure outcomes?

The greatest advantage of measuring outcomes is they capture what matters most to patients and clinicians, including patients’ health status and experiences within the health care system. They also capture the downstream effects of care processes, some of which are difficult to directly measure (10). Measuring outcomes can often make transparent those aspects of the patient experience that may be less visible to providers. For example, asking a patient with rheumatoid arthritis (RA) how they are doing may under- or overestimate disease activity; hence the need for standardized disease activity assessments to inform treat-to-target strategies. Unlike process measures, POMs do not offer a roadmap for improving care or list the actions required to improve outcomes. Therefore, they do not supplant process measures. POMs are primarily



**Figure 1.** Review of structure, process, and outcome quality measures. Structure measures define the presence or absence of specific care resources or qualifications; process measures evaluate whether guideline-concordant or best practice care has been provided. Outcome measures assess the downstream effect of care structures and processes on the health status of patients and populations. All 3 kinds of quality measures can be used for accountability purposes (e.g., in public reporting or pay for performance programs).

**Table 1. Key take-home points regarding performance outcome measures (POMs) in rheumatology**

What are POMs?	POMs are formal tools allowing scientifically valid comparisons of quality of care across providers or to an established benchmark.
Why measure outcomes?	POMs aim to assess what matters most to patients and clinicians and capture the downstream effects of health care processes.
What are the key requirements and tradeoffs in developing rheumatology POMs?	<p>Data source: data used to create and report POMs must balance the benefits of detailed, reproducible information with the burden of data collection.</p> <p>Measure cohort (denominator): the measure cohort for POMs should accurately and reliably capture the population of interest.</p> <p>Reporting period: the reporting period for POMs should consider the measurement goal (e.g., short-term quality improvements or pay for performance).</p> <p>Period at risk: the period at risk for POMs should reflect a standard timeframe during which it is reasonable to attribute outcomes to the measured provider or group.</p> <p>Measure outcome: the measure outcome for POMs should be unambiguous, be feasible to collect and report, provide meaningful information to providers and patients, and represent an outcome influenced by the health care system or providers being assessed.</p> <p>Outcome attribution: the results of POMs should be attributed to the entity most responsible for patient care, while simultaneously recognizing minimum sample sizes needed for stable performance estimates and the multidisciplinary care required for complex chronic rheumatic diseases.</p> <p>Risk adjustment: risk adjustment of POMs is critical to ensuring that providers are not penalized for caring for patients at greater risk.</p> <p>Reliability and validity testing: POMs should be created from valid, reproducible data and tested to ensure they produce reliable and valid results.</p> <p>Implementation and results reporting: reporting of POMs should consider their primary purpose and audience.</p>
What POMs exist in rheumatology?	There are no existing National Quality Forum–endorsed national POMs suitable for the majority of rheumatologists’ patients.
What lies ahead?	POMs are here to stay, and the most effective and meaningful measures can only be created through the close collaboration of patients, providers, measure developers, and policymakers.

developed for public reporting and accountability purposes, to inform and drive quality improvement.

### What are the key requirements and tradeoffs in developing rheumatology POMs?

In order to illustrate the methods, benefits, and challenges of measuring outcomes among patients with chronic illness, we created a hypothetical POM of RA disease activity using electronic health record (EHR) data. We considered each component of this example, i.e., data source, measure cohort, reporting period, period at risk, measure outcome, outcome attribution, risk adjustment, reliability and validity, and reporting (Table 2). We used terminology consistent with the American Heart Association’s published outcome measure guidance (11), tailored for rheumatology.

**Data source.** Existing POMs use many data sources, including administrative claims, clinical registries, patient surveys, or EHR data. Each data source offers its own balance between the detail of information captured and the cost and burden of both initial and ongoing data collection. POMs bring an additional challenge of centralizing data to allow for risk adjustment of patient case mix. Our EHR measure of RA disease activity would require that all measured providers capture and export sufficient information to identify patients with RA, assess their disease activity level, and adjust for disease severity, as explained below.

**Measure cohort.** The measure cohort for any outcome measure consists of patients for whom the outcome will be measured; some measures label this the denominator. To ensure measurement is comprehensive and representative, the cohort should be clearly defined using reliably captured data and include all eligible patients with the relevant condition within a specified time period. For our hypothetical RA disease activity POM, the measure cohort might include all patients with RA in an Accountable Care Organization (ACO) as defined by the presence of a single rheumatologist visit coded for RA within a specified timeframe. This approach offers the ease of using claims data to identify the cohort and the specificity of a rheumatologist’s (versus non-rheumatologist’s) diagnosis of RA. However, this approach might identify patients with suspected RA who are subsequently determined to have another diagnosis or miss patients with RA who are not seeing a rheumatologist. Administrative claims codes minimize additional data collection burden, but are a limited reflection of patients’ health status.

**Reporting period.** To identify the measure cohort, the time period for cohort eligibility (reporting period) must be specified. The key tradeoff in choosing the reporting period is that of timeliness (i.e., how current the data are) versus precision. Shorter reporting periods (e.g., 3 months) capture more recent outcomes, allowing for more rapid detection of short-term improvements, but include fewer outcome events and therefore less precision. Longer

Table 2. POM components and key issues for rheumatology\*

Measure component	Definition	Key issues	Hypothetical RA disease activity POM example	Additional considerations
Data source	Data used to calculate the measure results	Must balance ease of collection with sufficient detail and reproducibility	EHR and administrative billing data	Data collected need to accurately identify patients and their outcomes to be measured as well as capture key risk factors, such as disease duration and severity, to allow for adequate risk adjustment (see below). Criteria used to identify patients with RA for research purposes may under-/over-identify cases appropriate for quality measurement. There is a tradeoff between the detail and amount of clinical data available for identifying eligible patients and the costs and burden of collecting this data. EHRs may offer long-term mechanisms for collecting clinical data, but require significant upfront costs, and not all clinical data in the EHR are equally appropriate for use in measurement. Longer measurement periods offer more cases and therefore more precise and stable POM results, but will not be as sensitive to rapid cycle quality improvements (or decrements) compared with a shorter measurement period.
Measure cohort	Patients included in measure and assessed for the outcome of interest	Must balance inclusivity and comprehensive measurement with capturing patients with similar risk profiles	All patients with RA in provider's practice with a rheumatologist-coded visit for RA within 36-month measurement period	Shorter reporting periods allow for more rapid detection of short-term improvements, but fewer outcome events; longer periods allow inclusion of more patients and more outcome events, but include data that are often much less current.
Reporting period	Time period for cohort eligibility	Must balance how current the data are with how many events are captured (which impacts the precision of the POM)	A 36-month period starting January 1 and updated each year, producing a rolling 3-year measure	Shorter reporting periods allow for more rapid detection of short-term improvements, but fewer outcome events; longer periods allow inclusion of more patients and more outcome events, but include data that are often much less current.

(continued)

Table 2. (Cont'd)

Measure component	Definition	Key issues	Hypothetical RA disease activity POM example	Additional considerations
Period at risk	Followup period during which outcome of interest (in our example, RA disease activity) is assessed	Must be standard timeframe to avoid bias, but challenging to define appropriate start and end dates for chronic diseases	180 days from first rheumatologist-coded visit for RA within 36-month measurement period	Should be long enough to allow opportunity for both treatment to take effect and patient to return for reassessment. May be impacted by frequency of assessments; disease activity scores for patients seen only once will be based on only one assessment and may not reflect changes that occur as a result of treatment decisions at the initial encounter. In contrast, a patient seen at baseline and 4 months later will have 2 disease activity assessments that could be averaged to calculate their score for the 180-day period, but this average may over- or underestimate their actual disease activity level during this period.
Measure outcome	Clinical outcome that the measure captures	Ideally captures meaningful variation in health outcomes or patient experiences of care influenced by provider actions	ACO-level risk-adjusted average number of days in remission	Acknowledges fluctuations in disease activity in patients with RA, but subject to variability based on numbers of assessments used to calculate measure outcome as noted above.
Outcome attribution	Indicates entity held responsible for measure result	Should emphasize shared responsibility as appropriate and avoid encouraging harmful behavior or unintended consequences	Outcome attributed to ACOs	Entity held responsible should be able to meaningfully impact outcome. Attribution of outcomes to individual physicians is often limited by smaller sample sizes that make it difficult to distinguish physician or provider performance; aggregating data to the level of a larger responsible entity can both provide greater case volume (and thus measurement precision) as well as incentivize collaborative, coordinated care.

(continued)

Table 2. (Cont'd)

Measure component	Definition	Key issues	Hypothetical RA disease activity POM example	Additional considerations
Risk adjustment	Statistical correction for differences in patient case mix	Should ideally account for patient-level risk variables associated with the measure outcome, but not provider performance	Adjust for relevant clinical comorbidities and other patient characteristics that are associated with RA disease activity	Avoid adjusting for provider characteristics and/or characteristics that reflect provider choice; for example, adjusting for biologic versus non-biologic DMARD medication.
Reliability testing	Degree to which components and results of measure are reproducible	Measure components must be reproducible and indicate the same information across patients and providers in order for the measure to provide meaningful performance assessment	Use standard approaches for defining and extracting EHR data elements; examine test-retest reliability of measure results using accepted methods	Use of standardized disease activity assessments will limit variation in measure results unrelated to provider care quality.
Validity testing	Degree to which components and results of measure truly reflect their intended meaning	Measure components and results must indicate the intended content in order for the measure to provide meaningful performance assessment	Use validated instrument for measuring disease activity and expert consensus-based approach to assess measure face validity	Use of expert consensus to assess measure face validity will provide an important safeguard to ensure results are both meaningful to clinicians and patients and actionable.
Implementation and results reporting	How measure is calculated and results are provided to target audience	Must balance feasibility of data collection, interpretability, and usability of measure results, and statistical precision	Report ACO-level risk-adjusted observed average number of days in remission in comparison to the national average	EHR data theoretically provide a rich source of clinically detailed information, but uniform data extraction from EHRs and data centralization for risk adjustment can be challenging and costly. The average number of remission days is a point estimate of ACO performance; additionally, it may be important to provide an assessment of the uncertainty surrounding that point estimate using a confidence interval or other approach. If the national average is very low, demonstrating poor performance overall, it may not be advisable to benchmark against this standard. Alternatively, clinical experts could create a consensus-defined benchmark that reflects a reasonable care goal.

\* POM = performance outcome measure; RA = rheumatoid arthritis; EHR = electronic health record; ACO = Accountable Care Organization; DMARD = disease-modifying antirheumatic drug.

Table 3. Examples and potential benefits and pitfalls of outcomes relevant to rheumatology POMs*		
Outcome categories	Examples	Potential benefits and Potential pitfalls
Intermediate outcomes	C-reactive protein level	Influenced by many factors
	Erythrocyte sedimentation rate	May not reflect meaningful outcomes to patients or providers
Physician-reported outcomes	Active urinary sediment	Often more easily captured/standardized than other assessments
	Complement studies	Better scores on assessments are often associated with improved long-term clinical outcomes
	Serum urate level	Represent patient-centered outcomes
	Swollen joint count	Potential for variability in measurement
	Tender joint count	Variable reproducibility, potentially resulting in unreliable performance
Patient-reported outcomes	Disease activity	Multiple similar instruments requiring either consensus regarding best metric or additional testing to define consistent results across metrics
	Physician global assessment	Variable burden of data collection
Safety	Patient global assessment	Variable reproducibility, potentially resulting in unreliable performance
	Pain visual analog scale	Limited responsiveness data in rheumatic diseases available for newer instruments
Experience with care	Patient-Reported Outcomes Measurement Information System (PROMIS) 29-item health profile and functional assessment	Variable burden of data collection
	Adverse drug events	May be rare, making it difficult to accurately estimate performance
Efficiency	Opportunistic infections	May be influenced by many factors, making it challenging to attribute to individual provider
	Fragility fractures	Do not provide information about quality of clinical care
Cost	Access to care	May decrease appropriate as well as inappropriate care
	Timeliness of care	Unclear how to interpret (high cost separated from clinical outcomes is neither good nor bad)

\* POMs = performance outcome measures; MRI = magnetic resonance imaging.

reporting periods (e.g., 1 or more years) allow inclusion of more patients and more outcome events, but include data that are often much less current.

Ultimately, the reporting period depends on the number of measured patients and outcome events and the intended use of the measure. Our RA disease activity POM might capture patients seen in the outpatient setting during a 36-month period and we might update the results each year, producing a rolling 3-year measure. Such a measure might be appropriate for public reporting if it captures sufficient numbers of eligible patients to provide an accurate and precise estimate of an ACO's patients' outcomes. However, it may not be useful for assessing the impact of local quality improvement efforts because it would be difficult to detect short-term changes in performance.

**Period at risk.** Distinct from the reporting period, which helps define the denominator, the period at risk is the followup period during which the outcome of interest is expected to occur and can be detected; it defines the period of time for assessing the outcome for the measure numerator. To be reliably reproduced and compared across providers, there must be a clearly defined and consistent period at risk. Though outcome events may occur outside this period, the period is typically chosen to include the time of greatest risk and attribution. Chronic diseases present a challenge because the date of their onset is often unclear, so it is difficult to know when to start the period at risk. For our RA disease activity POM, we could anchor the period at risk to an outpatient rheumatology visit. Measurement would then start on the date of the first RA-coded outpatient visit with a rheumatologist in the 36-month reporting period. This has the advantage of linking the start of measurement to a date when the provider had an opportunity to assess and influence patient health.

To eliminate the potential for detecting differences in outcomes that do not reflect care quality, the period at risk should be a standard time period. For our measure, we define the period at risk to be 180 days from that initial visit. If the period at risk were defined as the intervening time between RA-related visits, physicians who see patients with RA more frequently would have a shorter period in which to achieve an optimal measure outcome, resulting in measure results that may not reflect care quality. The period at risk should consider the intervals at which data are collected (e.g., frequency of disease activity assessments) and the relationship between providers' actions and the outcome (e.g., it is difficult for a provider to influence patient outcomes if that patient is not seen for an extended period).

**Measure outcome.** POMs typically capture clinical outcomes. The measure outcome (numerator) should be unambiguous, be feasible to collect and report, provide meaningful information to providers and patients, and be influenced by the health care system or the providers being assessed. The measure outcome could be a desirable event, such as achievement of remission, or an adverse event, such as death or infection; it could capture a

measured health state, such as pain or functional status, or it could assess the cost of a defined episode of care. POMs should be distinguished from other "outcome measures," such as the standardized clinical and radiographic assessments developed or validated by Outcome Measures in Rheumatology or other organizations, which are used to evaluate therapeutic interventions in randomized clinical trials but do not include the specifications required for performance reporting (e.g., a defined reporting period and case mix adjustment). However, these standardized assessments are critical components of POMs because they are often used to define part of the measure outcome.

"Intermediate" clinical outcome measures refer to interim assessments instead of the ultimate outcome we are trying to achieve (or avoid) with treatment. Inflammatory marker values are examples of intermediate outcomes; abnormal results are often influenced by clinical care, but most rheumatologists and patients would agree that they lack specificity when examined in isolation and assess only a limited spectrum of an RA patient's health state (Table 3).

For our RA disease activity POM, we must define the instrument(s) used to assess disease activity and decide whether we will measure static disease activity, change in disease activity, or perhaps achievement of a predefined benchmark (e.g., remission or low disease activity state). Rheumatology has the advantage of a large number of validated RA-specific assessments of disease activity and the disadvantage that no single assessment has universal endorsement for use in clinical practice (12). Further, outcomes can be captured as dichotomous (e.g., an event occurred versus did not occur), quantitative (e.g., number of infections), or graded (e.g., Multidimensional Health Assessment Questionnaire score) variables (13,14). How the outcome is captured affects how it is analyzed, reported, and received by patients and clinicians.

Our hypothetical POM will assess the average (risk-adjusted) number of days in remission or with low disease activity as defined by the Clinical Disease Activity Index (CDAI) (15) within the 180-day period at risk among all of an ACO's eligible patients with RA. Providers delivering higher-quality care (i.e., those whose patients spend more time in remission or in a low disease activity state) will have a higher POM score. Our definition recognizes low disease activity disease as a goal of care. It acknowledges that disease activity fluctuates over time and that static assessments may not accurately reflect patients' experiences. This definition is limited by when and how often the data are collected. If a patient is seen monthly during the 180-day period at risk, they may have 6 disease activity assessments from which to calculate the number of days in remission or a low disease activity state; fewer visits and therefore fewer assessments will require additional assumptions about a patient's disease activity between assessments and may over- or underestimate their actual disease activity during that period.

Both the patient population under evaluation and the intended use of the measure impact the measure outcome definition. A measure intended solely for rapid cycle quality improvement might assess short-term improvements in



CDAI scores for newly diagnosed patients with RA. Such a measure might not be very useful for accountability because it only measures outcomes for a narrow population of recently diagnosed patients. Chronic diseases such as RA, in which the goal of care is to minimize disease activity and maintain function and quality of life, have the added complexity that patient populations often reflect a broad range of both disease severity and disease activity, with disease flares occurring in sometimes otherwise stable, well-managed patients. Further complicating RA measurement are recent changes in clinical practice that utilize more aggressive treatment regimens early in the disease course, potentially followed by sequential withdrawal or tapering of medications after achieving clinical remission in order to minimize the number and potency of agents required for longer-term maintenance. This evolution in RA treatment strategy may result in patients previously in remission experiencing clinical flares that do not necessarily represent substandard care quality.

**Outcome attribution.** Outcome attribution refers to the entity held responsible for the performance assessed by the measure. This entity must be able to meaningfully influence patient outcomes either directly through the care provided or indirectly through communication or influence on care coordination. Moreover, each patient in the measure cohort should be unambiguously associated with exactly one entity. Depending on the intended measure application, this entity could be an individual care provider, such as an eligible professional defined by the Meaningful Use EHR incentive program (16), a group of physicians, a hospital, or an ACO. The choice of entity responsible for measure performance impacts many aspects of measure development, from the cohort definition to the number of patients and outcome events required to produce stable performance estimates.

For our measure, we could attribute the outcome to the ACO in which the patient is enrolled at the time of the baseline assessment. Given the rarity and diversity of rheumatic diseases and the frequency of comorbid conditions, outcomes for patients with RA can be difficult to attribute to individual rheumatologists or other health care professionals, since small sample sizes yield less precise measure results. Attributing our measure outcome to an ACO will offer greater sample sizes, and therefore greater ability to distinguish performance among ACOs, and facilitates assessment of coordinated, multidisciplinary care. Moreover, while patients may be treated by multiple providers during the period at risk, they will only be enrolled in one ACO at a time. In addition to publicly reporting ACO-level results, the reporting entity (e.g., Medicare) could privately provide each ACO with physician-level data to inform local quality improvement efforts.

**Risk adjustment.** One of the most critical and technically challenging aspects of outcome measurement is risk adjustment, which seeks to adjust the outcomes of different measured providers according to the risk level of the patients on which they are being measured. Risk adjustment is critical because it levels the playing field, allowing

comparisons between providers to be made on the basis of outcomes for similar patients; providers caring for sicker patients are not unfairly penalized if their patients have poorer observed outcomes. To most accurately compare outcome performance across measured entities, it is important to identify which patient factors impact outcomes and, to the extent possible, adjust for variation in those factors across providers. The goal of risk adjustment is to predict the outcome expected based on the individual characteristics of the patient (e.g., their preexisting risk factors and clinical characteristics), to serve as a reference point by which a provider's actual observed performance can be evaluated. Although it is typically impossible to identify or capture all risk factors that might influence an outcome, and therefore create a perfectly level playing field, adjusting for the most important differences in patient risk factors can substantially improve comparability between providers.

To account for the differences in patient risk factors, a POM typically utilizes a risk model. Abstractly, this is a set of risk factors and their estimated effect on the outcome. The specification of a risk model entails identifying the risk factors to be included and specifying their relationship to the outcome. For our measure, an appropriate risk model will need to include patient risk factors known to influence RA disease activity, including clinical factors such as seropositivity, disease duration, and baseline disease activity, and other factors (e.g., comorbidities, demographic characteristics such as age and sex, potentially duration of care under this provider, or lifestyle factors such as smoking). As with other POM components, risk variables need to be clearly defined, be reliably measured, and represent the same risk information across providers and care settings. Furthermore, they should not be related to patient treatment, because that is the subject of assessment. For example, complications of care should not be included as risk factors because, even though they often impact outcomes, they are a result of care. Patient adherence is a controversial risk factor for POMs because there is disagreement about how much providers can influence it.

There is an active, ongoing debate about including (or not including) sociodemographic factors such as ethnicity or income in POM risk models. The National Quality Forum (NQF), the national consensus authority for quality measure endorsement, recently commissioned a panel to review its policy excluding sociodemographic factors such as race or socioeconomic status from risk adjustment models. The panel's report strongly favored inclusion of sociodemographic factors in POM risk models (17). The NQF leadership has not changed NQF policy, but is pursuing a 2-year trial period during which measures can be submitted for endorsement with sociodemographic factors included in risk adjustment (18). Data collected during this trial period will inform revisions to NQF policy.

Those in favor of risk adjusting for sociodemographic factors note that providers serving low socioeconomic or minority populations might be unfairly penalized because measures do not take important patient-level factors into consideration. Those cautioning against risk adjusting for sociodemographic factors reference evidence supporting low socioeconomic status or minority populations often

underuse higher-quality providers (19,20) or are cared for by poorer-performing providers. For example, providers caring for minority patients achieve worse patient safety outcomes on their nonminority patients compared to peer providers (21), making it difficult to determine if poor performance is due solely to patient-level factors, poor provider quality, other factors, or a combination. Further, including sociodemographic factors in risk-adjustment models will potentially remove incentives for improving care for those very populations with disparate outcomes. Despite the controversy, most agree that one goal of measurement should be to reduce disparities while maintaining resources to providers serving vulnerable populations.

Not all risk factors have the same effect on the outcome; therefore, once identified, risk factors must be used to adjust the outcome measure for each measured entity. A statistical method is chosen to incorporate the risk factors into the final risk model. Several approaches are common, with the choice depending on the outcome specification (dichotomous, quantitative, graded), number and kind of risk factors (categorical or continuous), sample size, and anticipated use of the measure. The most common approach for large numbers of patients and providers is to use statistical models that can directly estimate the specific risk effects of providers, separate from the effects of risk factors (e.g., hierarchical logistic regression), and use the results to construct provider-level metrics.

**Reliability and validity testing.** Both the measure components and the overall measure results should be assessed for reliability and validity according to standard guidance (11). Reliability refers to the degree to which the same measure produces the same results when applied to entities with the same underlying performance. Validity, in the context of POMs, refers to the degree to which the outcome being measured reflects true underlying care quality. This is more difficult to establish than reliability because the ideal way to assess validity of a measure would be to compare with a gold standard of perfect care; however, such standards are rare. Therefore, measure validity usually depends on “face validity” (the extent to which the outcome and risk model represent what most people in the field believe to be true reflections of patient experience) and/or a comparison with subjective rankings of providers. For our measure, we could require the data be tested to ensure reliability and validate the extracted EHR data against manually abstracted clinical data.

**Implementation and results reporting.** Once the POM is completed, it must be implemented and the measure results must be presented to relevant stakeholders. This might involve public reporting or private sharing of the results with the entities being measured, or both. The underlying purpose of the measure dictates the format and approach to results reporting. Measures used for accountability might compare a provider to an accepted benchmark. Our measure of the risk-adjusted mean number of days in remission for an ACO’s patients with RA produces a continuous score that could be benchmarked against the national average.

As part of its measure evaluation, the NQF assesses the feasibility of POM data collection and the ability of the POM results to be interpreted by stakeholders and meaningfully impact care (i.e., usability). POM implementation and reporting can be resource intensive for patients, providers, and the entity reporting the POM results. While the EHR offers potential avenues for minimizing data collection burden, clinical practice will need to evolve to capture patient outcomes that adequately inform clinical decision making, improve quality of care, and allow for scientifically rigorous POM reporting.

### What POMs exist in rheumatology?

There is currently a paucity of POMs for assessing rheumatologic care. A search of the NQF Quality Positioning System (22–24) for endorsed outcome measures applicable to nonsurgical treatment of musculoskeletal diseases in the ambulatory care setting yielded 9 measures: change in basic mobility as measured by the AM-PAC (Activity Measure for Post-Acute Care; NQF#0429); change in daily activity function as measured by the AM-PAC (NQF#0430); functional status change for patients with knee impairments (NQF#0422); functional status change for patients with hip impairments (NQF#0423); functional status change for patients with foot and ankle impairments (NQF#0424); functional status change for patients with lumbar impairments (NQF#0425); functional status change for patients with shoulder impairments (NQF#0426); functional status change for patients with elbow, wrist, or hand impairments (NQF#0427); and functional status change for patients with general orthopedic impairments (NQF#0428). All assess patients’ risk-adjusted function and mobility and are intended as assessments of rehabilitation following acute injury, surgery, and/or admission to a medical facility. None are intended for measuring outcomes related to chronic disease management or for common rheumatic diseases like RA, gout, or systemic lupus erythematosus.

### What lies ahead?

Performance outcome measurement is here to stay. The NQF Measures Application Partnership, which advises the Centers for Medicare & Medicaid Services (CMS) on which measures are suitable for federal measurement programs, recently conditionally supported (25) using a CMS measure still under development that examines functional status and shared decision making in patients with RA for the Physician Quality Reporting System, acknowledging that the measure concept was promising, but required further development. Therefore, even before a measure is completed, CMS and others are considering how it will be implemented. Further, the Medicare Access and Children’s Health Insurance Program (CHIP) Reauthorization Act of 2015 (MACRA) mandates a new physician payment structure focused on a Merit-Based Incentive Payment System (MIPS) or Alternative Payment Models, both of which require quality measurement (26). MIPS will calculate a composite physician performance score, incorporating quality, resource use, clinical practice improvement, and

meaningful use of the EHR (27); POMs are expected to be an increasingly important component.

To ensure rheumatologists can choose to be meaningfully measured on their care of patients with rheumatic disease, rather than using measures developed for nonrheumatic diseases, the American College of Rheumatology (ACR) has developed several process measures suitable for federal reporting. In addition, the ACR will begin development of a POM for RA in 2016, using clinical data from ACR's Rheumatology Informatics System for Effectiveness (RISE) Registry. Rheumatologists can learn about the ACR's existing RA measures and how RISE can help physicians navigate MACRA through its website (online at [www.rheumatology.org](http://www.rheumatology.org)). The most effective and meaningful POMs can only be created through the close collaboration of patients, providers, measure developers, and policymakers; we hope this article will spark readers to talk to their patients and the ACR about what outcomes are most meaningful to them.

## ACKNOWLEDGMENTS

The authors would like to acknowledge the important assistance provided by Regina Parker and Janet Joyce of the ACR during the drafting and preparation of this manuscript.

## AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Suter 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.** Suter, Barber, Herrin, Losina, Miller, Newman, Robbins, Tory, Yazdany.

**Acquisition of data.** Suter, Robbins.

**Analysis and interpretation of data.** Suter, Leong, Robbins.

## REFERENCES

1. State of California Office of Statewide Health Planning and Development. The California report on coronary artery bypass graft surgery: 2007–2008 Hospital and Surgeon data. Sacramento (CA): California Office of Statewide Health Planning and Development; 2011. URL: [http://www.oshpd.ca.gov/HID/Products/Clinical\\_data/CABG/2008/CABG0708.pdf](http://www.oshpd.ca.gov/HID/Products/Clinical_data/CABG/2008/CABG0708.pdf).
2. Malpiedi P, Peterson K, Soe M, Edwards J, Douglas Scott R II, Wise M, et al. 2011 national and state healthcare-associated infections standardized infection ratio report using data reported to the National Healthcare Safety Network as of September 4, 2012. Atlanta (GA): Centers for Disease Control and Prevention; 2013. URL: [http://www.cdc.gov/hai/pdfs/SIR/SIR-Report\\_02\\_07\\_2013.pdf](http://www.cdc.gov/hai/pdfs/SIR/SIR-Report_02_07_2013.pdf).
3. Suter LG, Li SX, Grady JN, Lin Z, Wang Y, Bhat KR, et al. National patterns of risk-standardized mortality and readmission after hospitalization for acute myocardial infarction, heart failure, and pneumonia: update on publicly reported outcome measures based on the 2013 release. *J Gen Intern Med* 2014;29:1333–40.
4. Yale New Haven Health Services Corporation Center for Outcomes Research and Evaluation. CMS Medicare hospital quality chartbook: performance report on outcome measures 2013. Baltimore (MD): Centers for Medicare and Medicaid Services; 2013. URL: <http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/-Medicare-Hospital-Quality-Chartbook-2013.pdf>.
5. Yale New Haven Health Services Corporation Center for Outcomes Research and Evaluation. CMS Medicare hospital quality chartbook: performance report on outcome measures 2014. Baltimore (MD): Centers for Medicare and Medicaid Services; 2014. URL: <http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Medicare-Hospital-Quality-Chartbook-2014.pdf>.
6. Jha AK, Orav EJ, Li Z, Epstein AM. The inverse relationship between mortality rates and performance in the Hospital Quality Alliance measures. *Health Aff (Millwood)* 2007;26:1104–10.
7. Department of Health and Human Services Public Reporting of Quality and Efficiency Measures Workgroup. Prioritizing the public in HHS public reporting: using and improving information on quality, cost, and coverage to support health care improvement. Rockville (MD): US Department of Health and Human Services. Agency for Healthcare Research and Quality; 2013. URL: <http://www.ahrq.gov/workingforquality/reports/hhsreporting.pdf>.
8. Federal Register v 76, no 88 (May 6, 2011; 76 FR 26489). Rules and regulations, pp. 26489–547. Part V. Department of Health and Human Services, Centers for Medicare and Medicaid Services. Parts 422 and 480 Medicare Program; Hospital inpatient value-based purchasing program; final rule. 2011. URL: <http://www.gpo.gov/fdsys/pkg/FR-2011-05-06/pdf/2011-10568.pdf>.
9. Federal Register v 79, no 219 (Nov 13, 2014; 79 FR 67547). Book 2 of 2, pp. 67547–68092. Part II. Department of Health and Human Services, Centers for Medicare and Medicaid Services. 42 CFR Parts 403, 405, 410, et al. Medicare Program; Revisions to payment policies under the physician fee schedule, clinical laboratory fee schedule, access to identifiable data for the Center for Medicare and Medicaid Innovation Models and other Revisions to Part B for CY 2015; final rule. 2014. URL: [http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/Downloads/2015\\_Medicare\\_Physician\\_Fee\\_Schedule\\_Final\\_Rule.pdf](http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/Downloads/2015_Medicare_Physician_Fee_Schedule_Final_Rule.pdf).
10. Mant J. Process versus outcome indicators in the assessment of quality of health care. *Int J Qual Health Care* 2001;13: 475–80.
11. Krumholz HM, Keenan PS, Brush JE Jr, Bufalino VJ, Chernew ME, Epstein AJ, et al. Standards for measures used for public reporting of efficiency in health care: a scientific statement from the American Heart Association Interdisciplinary Council on Quality of Care and Outcomes Research and the American College of Cardiology Foundation. *J Am Coll Cardiol* 2008;52:1518–26.
12. 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.
13. 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.
14. 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.
15. Aletaha D, Nell VP, Stamm T, Uffmann M, Pflugbeil S, Machold K, et al. Acute phase reactants add little to composite disease activity indices for rheumatoid arthritis: validation of a clinical activity score. *Arthritis Res Ther* 2005;7:R796–806.
16. Federal Register v 79, no 171 (Sept 4, 2014). Part III. Department of Health and Human Services, Centers for Medicare and Medicaid Services. 42 CFR Part 495, 45 CFR Part 170. Medicare and Medicaid Programs; Modifications to the Medicare and Medicaid Electronic Health Record (EHR) Incentive Program for 2015 and other changes to the

- EHR incentive program; and health information technology: revisions to the certified EHR technology definition and EHR certification changes related to standards; final rule. 2014. URL: <http://www.gpo.gov/fdsys/pkg/FR-2014-09-04/pdf/2014-21021.pdf>.
17. National Quality Forum. Risk adjustment for socioeconomic status or other sociodemographic factors. Washington (DC): National Quality Forum; August 2014. URL: [https://www.qualityforum.org/Publications/2014/08/Risk\\_Adjustment\\_for\\_Socioeconomic\\_Status\\_or\\_Other\\_Sociodemographic\\_Factors.aspx](https://www.qualityforum.org/Publications/2014/08/Risk_Adjustment_for_Socioeconomic_Status_or_Other_Sociodemographic_Factors.aspx).
  18. National Quality Forum. Finding the right adjustment. Washington (DC): National Quality Forum; 2014. URL: [http://www.qualityforum.org/Finding\\_The\\_Right\\_Adjustment.aspx](http://www.qualityforum.org/Finding_The_Right_Adjustment.aspx).
  19. Huang LC, Ma Y, Ngo JV, Rhoads KF. What factors influence minority use of National Cancer Institute-designated cancer centers? *Cancer* 2014;120:399–407.
  20. Dimick J, Ruhter J, Sarrazin MV, Birkmeyer JD. Black patients more likely than whites to undergo surgery at low-quality hospitals in segregated regions. *Health Aff (Millwood)* 2013;3296:1046–53.
  21. Ly DP, Lopez L, Isaac T, Jha AK. How do black-serving hospitals perform on patient safety indicators? Implications for national public reporting and pay-for-performance. *Med Care* 2010;48:1133–7.
  22. National Quality Forum. Quality positioning system. Washington, DC: National Quality Forum; 2015. URL: <http://www.qualityforum.org/QPS/QPSTool.aspx#qpsPageState=%7B%22TabType%22%3A1>.
  23. Jette A, Haley SM, Coster W, Ni PS. The Boston University activity measure for post-acute care (AM-PAC) website. Boston (MA): Boston University; 2015. URL: <http://am-pac.com/category/home/>.
  24. Focus on Therapeutic Outcomes. 2015. URL: <http://www.fotoinc.com/>.
  25. Measure Application Partnership. 2014 recommendations on measures for more than 20 federal programs: final report. Washington (DC): National Quality Forum; 2014. URL: [http://www.qualityforum.org/Publications/2014/01/MAP\\_Pre-Rulemaking\\_Report\\_2014\\_Recommendations\\_on\\_Measures\\_for\\_More\\_than\\_20\\_Federal\\_Programs.aspx](http://www.qualityforum.org/Publications/2014/01/MAP_Pre-Rulemaking_Report_2014_Recommendations_on_Measures_for_More_than_20_Federal_Programs.aspx).
  26. Medicare Access and CHIP Reauthorization Act of 2015. 42 USC 1305, Pub. L.114–10, 129 Stat. 87 (2015).
  27. Centers for Medicare & Medicaid Services. CMS quality measure development plan: supporting the transition to the Merit-based Incentive Payment System (MIPS) and Alternative Payment Models (APMs). Baltimore (MD): Centers for Medicare & Medicaid Services; 2015. URL: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/MACRA-MIPS-and-APMs.html>.