

A Systematic Review of Quality Measures for Inflammatory Arthritis

Matthew Cooper, Azin Rouhi, and Claire E.H. Barber

ABSTRACT. Objective. To conduct a systematic review and quality appraisal of quality measures for inflammatory arthritis, including rheumatoid arthritis (RA), spondyloarthritis, psoriatic arthritis (PsA), and juvenile idiopathic arthritis (JIA).

Methods. Embase, MEDLINE, and Cumulative Index to Nursing and Allied Health Literature (CINAHL) were searched from January 1, 2000, to October 23, 2016, using Medical Subject Headings terms for inflammatory arthritis and quality measures. A “grey literature” search of international arthritis organizations and quality measure libraries was also conducted. Two reviewers independently considered the papers for inclusion, with disagreements resolved by consensus. A modified guideline appraisal tool (AGREE II) was used to appraise the measure development process, which determined final inclusion. Measures were abstracted in duplicate and categorized into themes, measure type, and domains of quality.

Results. Thirteen measurement sets were included from 4 countries (United States, Canada, United Kingdom, Netherlands) and 1 European consortium. They included 10 sets on RA and 1 each for PsA, inflammatory arthritis, and JIA. There were 161 unique individual measures (136 process, 20 structure, and 5 outcome). Major themes included assessment, medications, and comorbidities. Measure development methods were varied, including RAND/University of California, Los Angeles appropriateness methodology, prioritization exercises, or other modified-Delphi methods. Inclusion of patients occurred in 77% of development groups. Discussion of barriers to measurement was infrequent.

Conclusion. Inflammatory arthritis quality measures cover a diversity of themes encompassing process, structure, and outcomes of care across the 6 domains of quality. However, between organizations, measure development is not standardized. Local assessment of measurement feasibility before use outside the original development context is recommended. (J Rheumatol First Release November 15 2017; doi:10.3899/jrheum.170157)

Key Indexing Terms:

RHEUMATOID ARTHRITIS JUVENILE IDIOPATHIC ARTHRITIS PSORIATIC ARTHRITIS

Quality measures are tools for measuring whether care provided is concordant with evidence-based practices. They often represent minimum standards of care and are derived from quality indicators, which are statements about best practices that are associated with high-quality care. Quality indicators are usually specified in the following format: If (a

specific clinical scenario), then (a clinical action)¹. An often-cited example of a quality indicator would be if a patient has rheumatoid arthritis (RA), then they should be prescribed a disease-modifying antirheumatic drug (DMARD)¹. Quality indicators can be further specified into quality measures, which have a specific numerator, denominator, and exclusions and can be reported as a percentage representing a quality or performance measure (depending on whether the subject of measurement is associated with improved quality or just healthcare performance). Quality measures can be used for benchmarking and quality improvement efforts, and in some countries and jurisdictions, are used in pay-for-performance programs. An example of a quality measure based on the above quality indicator would be “the percentage of patients with RA who have been prescribed a DMARD.”

According to a classic Donabedian framework², quality measures are often classified into process, structure, or outcome. Process measures determine whether clinical processes are concordant with evidence-based best practice and improved patient outcomes; structure measures record whether the health system or clinic infrastructures are present

From the Department of Medicine, and Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary; Department of Medicine, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta; Arthritis Research Canada, Richmond, British Columbia, Canada.

Azin Rouhi was funded by a Canadian Rheumatology Summer Student Research Award.

M. Cooper, MD, FRCPC, Rheumatology Resident, Department of Medicine, Cumming School of Medicine, University of Calgary; A. Rouhi, BHSc, MD program student, Faculty of Medicine and Dentistry, University of Alberta; C.E. Barber, MD, PhD, FRCPC, Assistant Professor, Division of Rheumatology, Department of Medicine, Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, and Scientist, Arthritis Research Canada.

Address correspondence to Dr. C.E. Barber, Rheumatologist, Division of Rheumatology, University of Calgary, HRIC 3AA20, 3280 Hospital Dr. NW, Calgary, Alberta T2N 4Z6, Canada. E-mail: cehbarbe@ucalgary.ca
Full Release Article. For details see Reprints and Permissions at jrheum.org
Accepted for publication August 29, 2017.

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2017. All rights reserved.

to support best practices; and outcome measures determine the effect on the health status of the patient. While outcome measures may seem the most obvious to measure, patient outcomes are the result of many factors, some of which are beyond the control of physicians. Several domains of quality exist, and the most commonly cited framework is the Institute of Medicine's 6 Domains of Health Care Quality³: safety, effectiveness, patient-centeredness, timeliness, efficiency, and equity.

The present systematic review of quality measures in rheumatology was undertaken to review the landscape of existing quality measures for inflammatory arthritis and to classify existing measures according to the Donabedian and Institute of Medicine frameworks described above. The quality of the existing measurement sets was appraised and the methods of development were also reviewed.

MATERIALS AND METHODS

The systematic review was developed and reported according to Preferred Reporting Items for Systematic Reviews and MetaAnalyses (PRISMA) guidelines⁴. The search strategy was developed in consultation with a medical librarian using quality measures as well as Medical Subject Headings terms for inflammatory arthritis including rheumatoid arthritis (RA), juvenile idiopathic arthritis (JIA), and the spondyloarthropathies (SpA): ankylosing spondylitis (AS), psoriatic arthritis (PsA), reactive arthritis (ReA), and inflammatory bowel disease (IBD)-related arthritis (see Supplementary Figure 1 for the search strategy, available with the online version of this article). The study protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO, <http://www.crd.york.ac.uk/PROSPERO>, record number 33433). EMBASE, MEDLINE, and CINAHL were searched from January 1, 2000, to October 23, 2016. A "grey literature" review of Websites for international arthritis organizations and quality measure libraries was also conducted (see list of reviewed Websites in Supplementary Table 1, available with the online version of this article).

Two reviewers independently considered the papers for inclusion with disagreements resolved by consensus. Quality measures were included if they were in English and identified an element of inflammatory arthritis care, and the method of development of the quality measures was available. Checklists or quality measures with no description of how they were selected and/or developed were excluded, as were quality standards (they lacked specific numerators and denominators and were not measurable). If the development of the quality measure was not outlined on a potentially eligible measure set, attempts were made to contact the author to determine whether the development strategy was available. Studies describing the use of quality measures, and not the development of the measures, were excluded.

A modified guideline appraisal tool, the Appraisal of Guidelines for Research and Evaluation II (AGREE II), that has previously been used for quality measure development evaluation⁵, was used to evaluate the measure development process for each set. The modified AGREE II tool includes 6 domains of measure development that are rated on a 7-point Likert scale, where 1 is "strongly disagree" and 7 is "strongly agree" that the domain item was met by the measurement set. The 6 domains include scope and purpose, user involvement, rigor of development, clarity of presentation, applicability, editorial independence, and an overall quality assessment, with the latter determining final inclusion. Description of how these elements were rated is presented in the published modification⁵. For example, "applicability" was rated based on whether the measures were supported with tools for use (e.g., data collection forms, electronic databases, etc.) and whether potential barriers to applying the indicators have been discussed. Two of 3 reviewers (CEB, MC, AR) independently scored the measurement sets. Because one

of the authors of this review (CEB) had been an author on some of the included publications, she did not participate in the quality review of her own publications. Domain scores were then calculated for each of the 6 domains according to the AGREE II scoring (example available from www.agreetrust.org). In brief, the scores for each element in the domain are summed across reviewers and the minimum possible score for each domain is subtracted from this obtained score, then the result is divided by the maximum possible score minus the minimum possible score for each domain. A maximum domain score of 100% indicates that all reviewers felt that all elements of the domain were met and rated them each a 7.

Individual measures from the included sets were abstracted in duplicate and categorized by indicator type (process, structure, outcome)² and were also based on the 6 domains of quality (safety, effectiveness, patient-centeredness, timeliness, efficiency, equity) from the Institute of Medicine³. While quality measures can fall in multiple domains of quality, for this paper we elected to assign each measure to a single domain, deemed by the 2 reviewers to be the dominant domain of quality represented by the measure.

Major and minor measurement themes were determined by 2 of the reviewers (CB and MC), and each measure was categorized with any disagreements resolved by consensus. The method of measure development was abstracted including the measurement team composition.

Ethics approval for this study was not required as per the guidelines of our institution.

RESULTS

There were 4418 citations identified and 456 were included in full text review. After quality rating, a total of 13 were included (Figure 1). From the 13 sources (Table 1), a total of 161 individual quality measures were abstracted (See Supplementary Table 2, available with the online version of this article). Seven of the sets were from North America⁶⁻¹² and 6 were from Europe (3 from the United Kingdom^{13,14,15}, 2 from the Netherlands^{16,17}, and 1 was a pan-European set¹⁸). Most of the sets were for RA^{6,7,9,11,12,13,15-18} (n = 10, 77%) with 1 each (8%) for JIA¹⁰, PsA¹⁴, and inflammatory arthritis⁸. No sets were found for ReA and none accepted for AS. The vast majority of the individual 161 measures were specific to RA (n = 143, 89%), while 12 addressed JIA (7%), 4 inflammatory arthritis (in general, 2%), and 2 PsA (1%).

Quality assessment of measurement sets. Details of the quality assessment for individual measurement sets are shown in Table 2. In general, well-described domains included the scope and purpose of the measurement sets (mean domain rating across sets 95%) and description of participant involvement in the development process (mean domain rating across sets 85%). The domain relating to the rigor of development received a mean rating across sets of 64%, while clarity of presentation received a mean rating across sets of 73%. Conversely, applicability of the measurement sets was generally poor (mean rating across sets of 34%), highlighting a lack of discussion of the organizational barriers to measurement or tools to support measure use. The domain for editorial independence had a mean rating of 50%, often because the measurement set did not declare or address potential competing interests of the measure development panel.

Quality measure themes. There were 18 major themes

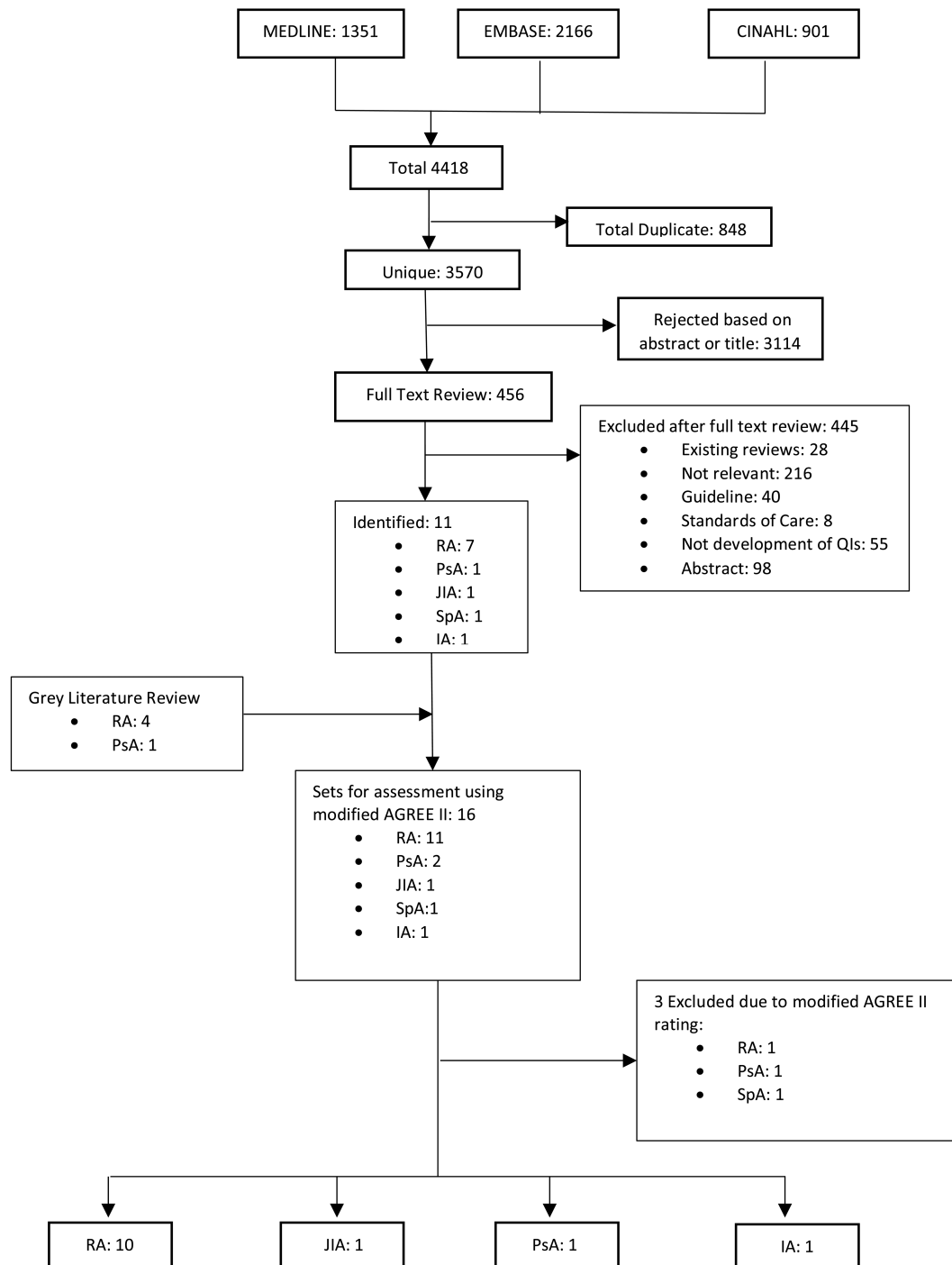


Figure 1. Flow diagram for systematic review. RA: rheumatoid arthritis; JIA: juvenile idiopathic arthritis; PsA: psoriatic arthritis; IA: inflammatory arthritis; SpA: spondyloarthropathy.

abstracted from the quality measures, and subthemes were assigned these major themes (Table 3). The most common major theme, constituting 27% of measures from all sets, was medications (including subthemes of DMARD use, steroid use, and medication counseling, and safety in pregnancy counseling). The second most common major theme was

clinical assessment, constituting 26% of all measures (including subthemes of functional status, disease activity, radiographs, and followup). The remainder of the major themes each composed < 10% of all measures, including accessibility, adjunctive therapy (with subthemes of exercise, physiotherapy, and assistive devices), comorbidity assess-

Table 1. Details of accepted quality measure sets.

Measure Set, Author	Country, Year	Organization	Target	No. Indicators		
				Process	Outcome	Structure
AF QI Set, MacLean, <i>et al</i> ¹¹	USA, 2004	AF	RA	37	0	0
Monitoring of the disease course in RA, van Hulst, <i>et al</i> ¹⁶	NLD, 2009	DRA	RA	10	3	7
Proposed set of QM for the process of care in JIA, Lovell, <i>et al</i> ¹⁰	USA, 2011	ACR, AAP, ABP, ARHP	JIA	12	0	0
Development of healthcare QI for RA in Europe, Petersson, <i>et al</i> ¹⁸	Europe, 2014	Eumusc.net	RA	12	1	1
KPI to evaluate centralized intake, Barber, <i>et al</i> ^{6*}	Canada, 2015	AAC	RA	12	0	2
Cardiovascular QI, Barber, <i>et al</i> ⁷	Canada, 2015	n/a	RA	11	0	0
System-level performance measures, Barber, <i>et al</i> ⁸	Canada, 2016	AAC	IA/RA	5	0	1
QI for physiotherapy management, Peter, <i>et al</i> ¹⁷	NLD, 2016	n/a	RA	16	1	0
ACR RA electronic QM, Yazdany, <i>et al</i> ¹²	USA, 2016	ACR	RA	4	0	0
NCQA/PCPI/ACR, Kwok, <i>et al</i> ⁹	USA, 2008	ACR	RA	5	0	0
QOF-RA, NICE-QOF ¹⁵	UK, 2011	NICE-QOF	RA	3	0	1
QS for RA, NICE ¹³	UK, 2013	NICE-QS	RA	8	0	7
QS for PsA, NICE ¹⁴	UK, 2013	NICE-QS	PsA	1	0	1

*Does not include 5 measures that were adopted from the Barber, *et al* 2015 paper⁸ into the Barber, *et al* 2016 paper⁶ (see Supplement for affected measures, available with the online version of this article). AAC: Arthritis Alliance of Canada; AAP: American Academy of Pediatrics; ABP: American Board of Pediatrics; ACR: American College of Rheumatology; AF: Arthritis Foundation; ARHP: Association of Rheumatology Health Professionals; DRA: Dutch Rheumatology Association; Eumusc.net: European Musculoskeletal Conditions Surveillance and Information Network; IA: inflammatory arthritis; JIA: juvenile idiopathic arthritis; KPI: key performance indicators; NCQA: National Committee for Quality Assurance; NICE: (UK) National Institute for Health and Care Excellence; NLD: Netherlands; PCPI: physician consortium performance improvement; PsA: psoriatic arthritis; QI: quality indicator; QM: quality measure; QOF: quality and outcomes framework; QS: quality standard; RA: rheumatoid arthritis; n/a: not applicable.

ment/management, diagnosis (of the inflammatory arthritis), documentation, education (patient self-management), experience (healthcare and patient satisfaction), clinical expertise (nurse and physician), multidisciplinary care, outcomes, surgery, triage, vaccinations, and waiting times.

The measurement themes were also categorized based on disease (RA, PsA, and JIA, data not shown). Most of the JIA measures related to the assessment theme (58%), medications (25%), and the remainder to comorbidity (ophthalmologic), experience (patient satisfaction), and wait time themes. Both PsA measures addressed the assessment major theme and the followup subtheme.

When the measures were categorized using the Institute of Medicine³ domains of quality (Table 4), 35% related to measures of healthcare effectiveness, 25% to safety, and 16% to patient-centered care. In contrast, only 11% addressed timeliness of patient care, 12% efficiency, and 1% directly measured care equity. When the measures were categorized using a Donabedian framework², according to process, structures, and outcomes, the majority related to process of care (84%), with only 20 being structure (12%), and 5 (3%) recording patient outcomes (Table 1).

Quality measure development methods. Of the 13 sets, 6 were

developed predominantly from an existing guideline or clinical care pathway, 3 were based on systematic reviews of existing guidelines and measures, 3 were based on non-systematic literature reviews, and 1 was based on discussions/questionnaires with those involved, accompanied by a nonsystematic literature review¹⁰ (Table 5). Further methods are described in Supplementary Table 3, available with the online version of this article.

Similarly, there was variability in the consensus procedures used to finalize the measurement sets: a modified Delphi in conjunction with Nominal Group Technique in 1 set¹⁰, a modified RAND/University of California, Los Angeles (UCLA) Appropriateness Method in 4 sets^{11,12,16,17}, an online modified Delphi using a platform called ExpertLens (RAND Corp.) in 3 sets^{6,7,8}, a modified Delphi in another¹⁷, and other study-specific consensus methods in 4 sets^{9,13,14,15}.

Three measure development panels included physician experts exclusively^{9,11,16}, while the other 10 included a diverse set of medical professionals such as nurses, occupational therapists, and physiotherapists. All 10 also included patients in the development process (Table 5). The measure development panels universally included rheumatologists.

Table 2. Modified AGREE II⁵ quality assessment of quality measure sets. Values are %.

Measure Set, Author	Domain 1: Scope and Purpose	Domain 2: Involvement	Domain 3: Rigor of Development	Domain 4: Clarity of Presentation	Domain 5: Applicability	Domain 6: Editorial Independence	Overall
AF QI set, MacLean, <i>et al</i> ¹¹	91.7	50	65.5	91.7	0	25	83.3
Monitoring of the disease course in rheumatoid arthritis, van Hulst, <i>et al</i> ¹⁶	79.2	50	66.7	75	16.7	75	75
Proposed set of QM for the process of care in JIA, Lovell, <i>et al</i> ¹⁰	100	91.7	35.7	66.7	0	0	75
Development of healthcare QI for RA in Europe, Petersson, <i>et al</i> ¹⁸	95.8	91.7	59.5	83.3	12.5	95.8	91.7
KPI to evaluate centralized intake, Barber, <i>et al</i> ⁶	100	100	69	0	0	16.7	83.3
Cardiovascular QI, Barber, <i>et al</i> ⁷	100	100	73.8	50	16.7	50	83.3
System-level performance measures, Barber, <i>et al</i> ⁸	100	100	73.8	66.7	41.7	25	83.3
QI for physiotherapy management, Peter, <i>et al</i> ¹⁷	100	100	77.4	66.7	83.3	100	91.7
ACR RA electronic QM, Yazdany, <i>et al</i> ¹²	91.7	75	76.2	100	100	75	100
NCQA/PCPI/ACR, Kwoh, <i>et al</i> ⁹	95.8	83.3	35.7	83.3	50	16.7	58.3
QOF-RA, NICE-QOF ¹⁵	91.7	83.3	48.8	75	62.5	45.8	75
QS for RA, NICE ¹³	91.7	91.7	73.8	91.7	29.2	62.5	83.3
QS for PsA, NICE ¹⁴	91.7	91.7	73.8	91.7	29.2	62.5	83.3
Average	94.6	85.3	63.8	72.5	34	50	82
Median	95.8	91.7	67.9	75	29.2	50	83.3

AGREE II: Appraisal of Guidelines for Research and Evaluation II; ACR: American College of Rheumatology; AF: Arthritis Foundation; JIA: juvenile idiopathic arthritis; KPI: key performance indicators; NCQA: (US) National Committee for Quality Assurance; NICE: (UK) National Institute for Health and Care Excellence; PCPI: physician consortium performance improvement; PsA: psoriatic arthritis; QI: quality indicator; QM: quality measure; QOF: quality and outcomes framework; QS: quality standard; RA: rheumatoid arthritis.

DISCUSSION

To our knowledge, this is the first systematic review of quality measures including all types of inflammatory arthritis that has rigorously assessed the quality of the methods of development and reporting of quality measures using a published modification of the AGREE II⁵. The review has identified 13 high-quality sets of quality measures in inflammatory arthritis, with 161 individual quality measures. The majority of sets included were developed for RA (n = 10), while only a few sets for inflammatory arthritis, PsA, and JIA were identified. Although 1 set was identified for AS¹⁹, it was not included because development was unclear, indicators were not measurable, and they were more consistent with quality standards because they lacked clear numerators and denominators. There were no sets identified pertaining to other SpA subtypes including ReA or IBD-related arthritis.

Methodologic review of the included sets revealed substantial heterogeneity in the methods used for development, including the evidence base, the consensus technique used, and the panel composition. For example, considering the evidence base used for the measure development, only 3 sets of quality measures were based on systematic reviews of existing guidelines and measures^{7,16,18}, with another 3 based on nonsystematic literature reviews^{6,8,11}. The rest of the measures were developed based on discussions and/or

questionnaires with those involved or on existing guidelines. Similarly, the methods for consensus varied among development groups. While most groups used a modified-Delphi method or a RAND/UCLA Appropriateness Method²⁰, others used nonspecified consensus techniques to develop their quality measures. More recently, the RAND/UCLA method has been adapted for online panels using a platform called ExpertLens²¹. This platform was used for 3 panels^{6,7,8} and has the advantage of including a larger number of participants (40–50) than traditional consensus panels²⁰, allowing for broader input. While there is no current standard methodology for quality measure development, at a minimum it is recommended to develop measures based on a high level of evidence (usually from a guideline). Measure development groups should also use transparent reporting of the consensus procedures used for measure development and ensure broad input, including patients.

For quality measures to be useful, they must reflect a number of important attributes, including face/content validity, reproducibility, acceptability, feasibility, reliability, sensitivity to change, and predictive validity²². Quality measures must also target important improvements and be precisely defined and specified²³. Some of these attributes are important to establish during the development stage (e.g., face/content validity, importance, and acceptability), while

Table 3. Condensed major themes and subthemes of each quality measurement set. Values are n.

Major theme	Minor theme	Barber, <i>et al</i> ⁶	Barber, <i>et al</i> ⁷	Barber, <i>et al</i> ⁸	Kwoh, <i>et al</i> ⁹	Lovell, <i>et al</i> ¹⁰	MacLean, <i>et al</i> ¹¹	NICE-QOF ¹⁵	NICE-QS RA ¹³	NICE-QS PsA ¹⁴	Peter, <i>et al</i> ¹⁷	Pettersson, <i>et al</i> ¹⁸	van Hulst, <i>et al</i> ¹⁶	Yazdany, <i>et al</i> ¹²	Total
Accessibility	No. physicians	—	—	1	—	—	—	—	—	—	—	—	—	—	1
Adjunctive therapy	Assistive device	—	—	—	—	—	3	—	—	—	2	1	—	—	6
	Exercise	—	1	—	—	—	1	—	—	—	2	1	—	—	5
	Physiotherapy	—	—	—	—	—	—	—	—	—	2	—	—	—	2
	Safety	—	—	—	—	—	—	—	—	—	1	—	—	—	1
	Treatment plan	—	—	—	—	—	—	—	—	—	1	—	—	—	1
Assessment	Clinical documentation	—	—	—	—	—	1	—	—	—	—	1	—	—	2
	Disease activity	—	—	—	1	—	—	—	—	—	—	1	2	1	5
	Flare	—	—	—	—	—	—	—	2	—	—	—	—	—	2
	Followup	—	—	1	—	—	1	—	2	2	—	1	1	—	8
	Frequency of followup	—	—	—	—	—	—	1	—	—	—	—	—	—	1
	Functional status	—	—	—	1	1	—	—	—	—	1	—	1	1	5
	HRQOL	—	—	—	—	1	—	—	—	—	—	—	—	—	1
	Joint count	—	—	—	—	1	—	—	—	—	—	—	—	—	1
	Pain	—	—	—	—	1	—	—	—	—	—	—	—	—	1
	Physician's global assessment	—	—	—	—	1	—	—	—	—	—	—	—	—	1
	Physiotherapy	—	—	—	—	—	—	—	—	—	2	—	—	—	2
	Prognosis	—	—	—	1	—	—	—	—	—	—	—	—	—	1
	Radiographs	—	—	—	—	—	—	2	—	—	—	1	—	3	6
	Red flags	—	—	—	—	—	—	—	—	—	—	2	—	—	2
	Self-efficacy	—	—	—	—	1	—	—	—	—	—	—	—	—	1
	Treat to target	—	—	—	—	—	—	—	—	—	—	—	1	1	2
	Treatment plan	—	—	—	—	—	—	—	—	—	—	1	—	—	1
Clinic	Capacity	1	—	—	—	—	—	—	—	—	—	—	—	—	1
	Comorbidity	—	2	—	—	—	—	1	—	—	—	—	—	—	3
Comorbidity	Osteoporosis	—	—	—	—	—	—	1	—	—	—	—	—	—	1
	Diabetes	—	1	—	—	—	—	—	—	—	—	—	—	—	1
	Dyslipidemia	—	1	—	—	—	—	—	—	—	—	—	—	—	1
	Hypertension	—	2	—	—	—	—	—	—	—	—	—	—	—	2
	Obesity	—	1	—	—	—	—	—	—	—	—	—	—	—	1
	Ophthalmologic	—	—	—	—	1	—	—	—	—	—	—	—	—	1
	Osteoporosis	—	—	—	—	—	2	—	—	—	—	—	—	—	2
	Smoking	—	1	—	—	—	—	—	—	—	—	—	—	—	1
Diagnosis	—	—	—	—	—	—	—	—	—	—	1	—	—	1	
Documentation	Disease activity	—	—	—	—	—	—	—	—	—	—	—	2	—	2
Education	Disease course	—	—	—	—	—	—	—	—	—	1	—	—	—	1
	Flare	—	—	—	—	—	—	—	—	—	—	1	—	—	1
	General	1	—	—	—	—	—	—	—	—	—	1	—	—	2
	Self-management	—	—	—	—	—	—	—	2	—	—	—	—	—	2
	Treatment plan	—	—	—	—	—	—	—	—	—	—	1	—	—	1
Experience	Healthcare worker satisfaction	3	—	—	—	—	—	—	—	—	—	—	—	—	3
	Patient satisfaction	1	—	—	—	1	—	—	—	—	—	—	—	—	2
Expertise	Nurse	—	—	—	—	—	—	—	—	—	—	—	1	—	1
	Physician	—	—	—	—	—	—	—	—	—	—	—	1	—	1
Medication	Standardization	—	—	—	—	—	—	—	—	—	—	—	1	—	1
	Contraindications	—	—	—	—	—	2	—	—	—	—	—	—	—	2
	Counseling	—	1	—	—	1	5	—	—	—	—	1	—	—	8
	Counseling — reproductive	—	—	—	—	—	1	—	—	—	—	—	—	—	1
	DMARD	—	—	2	—	—	3	—	2	—	—	—	—	1	8
	Monitoring	—	—	—	—	1	5	—	—	—	—	—	—	—	6
	Safety	—	—	—	2	1	5	—	—	—	—	1	—	1	10
	Steroids	—	1	—	—	—	1	—	—	—	—	—	—	—	2
	Treat to target	—	—	—	—	—	—	—	3	—	—	1	2	—	6
Multidisciplinary care	Nursing	—	—	—	—	—	—	—	—	—	—	—	2	—	2

Table 3. Continued.

Major theme	Minor theme	Barber, <i>et al</i> ⁶	Barber, <i>et al</i> ⁷	Barber, <i>et al</i> ⁸	Kwoh, <i>et al</i> ⁹	Lovell, <i>et al</i> ¹⁰	MacLean, <i>et al</i> ¹¹	NICE-QOF ¹⁵	NICE-QS RA ¹³	NICE-QS PsA ¹⁴	Peter, <i>et al</i> ¹⁷	Petersson, <i>et al</i> ¹⁸	van Hulst, <i>et al</i> ¹⁶	Yazdany, <i>et al</i> ¹²	Total
Outcome Register Surgery Triage	Disease activity	—	—	—	—	—	—	—	—	—	—	—	3	—	3
		—	—	—	—	—	—	1	—	—	—	—	—	—	1
		—	—	—	—	—	2	—	—	—	—	1	—	—	3
	Accuracy of diagnosis	1	—	—	—	—	—	—	—	—	—	—	—	—	1
	Provider participation	1	—	—	—	—	—	—	—	—	—	—	—	—	1
	Referral efficiency	3	—	—	—	—	—	—	—	—	—	—	—	—	3
	Scheduling efficiency	1	—	—	—	—	—	—	—	—	—	—	—	—	1
	Volume	1	—	—	—	—	—	—	—	—	—	—	—	—	1
Vaccination	Contraindications	—	—	—	—	—	1	—	—	—	—	—	—	—	1
	Indications	—	—	—	—	—	1	—	—	—	—	—	—	—	1
Wait times		1	—	2	—	1	1	—	4	—	—	1	—	—	10
Total		14	11	6	5	12	37	4	15	2	17	14	20	4	

RA: rheumatoid arthritis; PsA: psoriatic arthritis; NICE: (UK) National Institute for Health and Care Excellence; QOF: quality and outcomes framework; QS: quality standard; HRQOL: health-related quality of life; CV: cardiovascular; DMARD: disease-modifying antirheumatic drug.

Table 4. Breakdown of themes from the Institute of Medicine³. Values are n.

Study, Author	Safe	Effective	Patient-centered	Timely	Efficient	Equitable
AF QI set, MacLean, <i>et al</i> ¹¹	24	9	3	1	0	0
Monitoring of the disease course in RA, van Hulst, <i>et al</i> ¹⁶	0	12	1	1	6	0
Proposed set of QM for the process of care in JIA, Lovell, <i>et al</i> ¹⁰	3	2	6	1	0	0
Eumusc QI for RA, Petersson, <i>et al</i> ¹⁸	2	5	4	3	0	0
KPI to evaluate centralized intake, Barber, <i>et al</i> ⁶	0	0 (1)*	2	1 (3)*	10	1 (1)*
Cardiovascular QI, Barber, <i>et al</i> ⁷	3	7	1	0	0	0
AAC system-level performance measures, Barber, <i>et al</i> ⁸	0	2	0	3	0	1
QI for physiotherapy management, Peter, <i>et al</i> ¹⁷	4	4	7	0	2	0
ACR RA EQM, Yazdany, <i>et al</i> ¹²	1	2	0	1	0	0
NCQA/PCPI/ACR, Kwoh, <i>et al</i> ⁹	2	3	0	0	0	0
QOF-RA, NICE-QOF ¹⁵	0	3	0	0	1	0
Quality standards for RA, NICE ¹³	2	5	2	6	0	0
Quality standards for PsA, NICE ¹⁴	0	2	0	0	0	0
All	41	56	26	17	19	2

*The numbers in parentheses indicate measures that were adopted from the Barber 2015 paper into the Barber 2016 central triage measures and therefore are counted only in the 2016 paper. AAC: Arthritis Alliance of Canada; ACR: American College of Rheumatology; AF: Arthritis Foundation; Eumusc: European musculoskeletal conditions surveillance and information network; EQM: electronic quality measures; JIA: juvenile idiopathic arthritis; KPI: key performance indicators; NCQA: (US) National Committee for Quality Assurance; NICE: (UK) National Institute for Health and Care Excellence; PCPI: physician consortium performance improvement; PsA: psoriatic arthritis; QI: quality indicator; QM: quality measure; QOF: quality and outcomes framework; RA: rheumatoid arthritis.

others must be defined during testing of the measures (e.g., reproducibility, feasibility, reliability, etc.). Interestingly, only a few of these attributes are routinely reported in the studies describing the measure development of the inflammatory arthritis measure sets in this review. For example, to be accepted, quality measures needed variably to be valid¹¹, valid and feasible^{7,12}, valid and important⁶, valid and relevant⁸, relevant¹⁷, and applicable and feasible¹⁸. A few sets did not follow any published attributes for measure devel-

opment and used the following statements for consensus development: “adequately reflected the quality of disease course monitoring”¹⁶, “consistent with [previous] recommendations”⁹, or simply included measures with higher “priority” rankings^{10,13,14,24}. Measure developers seeking to adopt and/or develop new measures should be encouraged to assess at a minimum the validity and feasibility of the candidate measures (the latter by testing the measure in clinical practice). Tools for measure assessment in key domains have

Table 5. Simplified overview* of the development methods for the included quality measurement sets.

Measure Set, Author	Initial QM Development	Consensus Procedure	Measure Development Team
AF QI set, MacLean, <i>et al</i> ¹¹	Nonsystematic literature review	Modified RAND/UCLA AM	Physicians
Monitoring of the disease course in RA, van Hulst, <i>et al</i> ¹⁶	Systematic review	Modified RAND/UCLA AM	Physician
QM for the process of care in JIA, Lovell, <i>et al</i> ¹⁰	Discussions/questionnaires with those involved	Modified Delphi + NGT	Physicians + AHP + patients
Eumusc RA QI, Petersson, <i>et al</i> ¹⁸	Systematic review	Study-specific consensus methods	Physicians + AHP + patients
KPI to evaluate centralized intake, Barber, <i>et al</i> ⁶	Nonsystematic literature review	Modified RAND/UCLA AM	Physicians + AHP + patients
Cardiovascular QI, Barber, <i>et al</i> ⁷	Systematic review	Modified RAND/UCLA AM	Physicians + AHP + patients
AAC system-level performance measures, Barber, <i>et al</i> ⁸	Non-systematic literature review	Modified RAND/UCLA AM	Physicians + AHP + patients
QI for physiotherapy management, Peter, <i>et al</i> ¹⁷	Guideline or clinical care pathway	Modified Delphi with modified RAND/UCLA AM	Physicians + AHP + patients
ACR RA EQM, Yazdany, <i>et al</i> ¹²	Guideline or clinical care pathway	Modified RAND/UCLA AM	Physicians + AHP + patients
NCQA/PCPI/ACR, Kwoh, <i>et al</i> ⁹	Guideline or clinical care pathway	Study-specific consensus methods	Physicians + AHP + patients
QOF-RA, NICE-QOF ¹⁵	Guideline or clinical care pathway	Study-specific consensus methods	Physicians + AHP + patients
QS for RA, NICE ¹³	Guideline or clinical care pathway	Study-specific consensus methods	Physicians + AHP + patients
QS for PsA, NICE ¹⁴	Guideline or clinical care pathway	Study-specific consensus methods	Physicians + AHP + patients

*For extended version, see Supplementary material, available with the online version of this article. AAC: Arthritis Alliance of Canada; ACR: American College of Rheumatology; AF: Arthritis Foundation; AHP: allied health professionals; Eumusc: European Musculoskeletal Conditions Surveillance and Information Network; EQM: electronic quality measures; JIA: juvenile idiopathic arthritis; KPI: key performance indicators; NCQA: (US) National Committee for Quality Assurance; NICE: (UK) National Institute for Health and Care Excellence; NGT: nominal group technique; PCPI: physician consortium performance improvement; PsA: psoriatic arthritis; QI: quality indicator; QM: quality measure; QOF: quality and outcomes framework; QS: quality standard; RA: rheumatoid arthritis; RAND/UCLA AM: RAND/University of California, Los Angeles Appropriateness Method.

been published by Stelfox and Straus²³ and may be useful for measure developers, allowing them to assess candidate measures along the domains of measure importance, clarity, reliability, validity, and feasibility. In the United States, the National Quality Forum (NQF) is a not-for-profit organization that works to endorse quality measures used for federal reporting as well as other programs, including pay-for-performance programs. Measures considered for endorsement by NQF undergo a rigorous process and 4 principal criteria are used: importance, scientific acceptability, usability and relevance for intended uses, and feasibility²⁵. It was unfortunately beyond the scope of our review to individually appraise each quality measure according to these criteria. We encourage users to evaluate individual metrics presented herein based on these criteria prior to implementation or use, because it is possible the measurement set fulfilled modified AGREE II criteria but that an individual measure has poor validity or feasibility in a particular healthcare context.

Published measures for inflammatory arthritis care were almost exclusively process and structure measures, with only 3% (n = 5) found to be outcome measures, all of which came out of Europe^{16,17,18}. While Rademakers, *et al*²⁶ found that structure and process most determined patients' assessment of healthcare, there is limited evidence that measuring the process and structure of care improves patient outcomes. A study investigated the documentation of disease activity and functional status by 18 rheumatologists and found that documentation was infrequent; however, this appeared to

have no association with radiographic outcomes over 24 months²⁷. As patient outcomes are of primary importance to healthcare providers, health systems, and patients, it is likely that in the future, increasing emphasis will be on measuring outcomes²⁸. While it is tempting to suggest potential outcome measures for inflammatory conditions such as RA (e.g., the percentage of patients in remission or low disease activity state), caution should be exercised because the development of outcome measures requires careful consideration of several additional factors during development, as outlined by Suter, *et al*²⁸, including outcome attribution and appropriate risk adjustment.

Although to our knowledge this is the most comprehensive review of quality measures in inflammatory arthritis, several limitations should be recognized. First, although a systematic process was used to identify selected measurement sets, it is possible that some sets were missed, especially if they were published only on a health organization Website that was not reviewed. Additionally, non-English publications were excluded. Further, it should be noted that 1 of the authors on 4 of the included sets is the senior author of this article, and this could lead to bias in assessment. This potential bias was mitigated by adding an additional independent reviewer to determine inclusion of these sets and to rate the quality of these sets. There is also no standardized classification system for arthritis measurement themes and subthemes, and it is possible for another set of reviewers to have classified the individual measures differently.

Unfortunately, there is also no standardized tool available to assess the quality of measure development; however, a published modification of the AGREE II tool was used⁵. While this tool addresses many concepts important to quality measure development, some domains may be less applicable to quality measure development than to guideline development. For example, the AGREE II tool includes an entire domain on “applicability” that relates to tools for implementation. By their nature, quality measures often represent “tools” that can be used to monitor quality. In our interpretation of this domain, we rated measure sets that included a clear description of tools for quality measure use (e.g., online tools or electronic medical record, and clear specifications for measurement), which often led to lower rating on this domain. We also recommend that the feasibility of measures should be tested prior to implementation. Prior testing appeared to occur infrequently in the reviewed measurement sets, except for the American College of Rheumatology’s RA quality measure set, which was tested in a new electronic clinical care registry (Rheumatology Informatics System for Effectiveness) prior to publication¹², and the Eumusc.net measures, where the initial set was audited with questions involving applicability and feasibility before using that feedback to develop the final set. This concept of testing feasibility prior to implementation was not included in the AGREE II tool and may have led to a biased assessment of a measurement set’s quality.

A final limitation of our review was that because only development of measures was assessed, and not their testing or use, it was often unclear whether the measures included were in routine use for quality improvement and/or pay-for-performance programs. The intended use of the measure was also not included as a domain in the modified AGREE II instrument; however, this should be considered when evaluating measures, because measures used in pay-for-performance may be more likely to represent a “minimal standard” of care, which may influence rigor of development and/or testing, rather than those in use solely for quality improvement purposes.

This systematic review of quality measures for inflammatory arthritis reveals a major focus on process and structure measures for RA. Review of development of these measures shows heterogeneity, which may undermine their generalizability to other healthcare settings. In future measure development efforts, we recommend using clear frameworks for measure development and testing the measures for feasibility prior to implementation. Quality measures are important in ensuring that high-quality care that meets consistent standards is being provided to populations with inflammatory arthritis. Development of effective, well-designed measures for quality monitoring should be an ongoing goal. Understanding what measures are currently available and their strengths, weaknesses, and underlying assumptions will ensure that the best measures are selected for use and help direct future measure development.

ACKNOWLEDGMENT

Lorraine Toews, MLIS, Librarian, University of Calgary, for her help with the literature search.

ONLINE SUPPLEMENT

Supplementary material accompanies the online version of this article.

REFERENCES

1. Desai SP, Yazdany J. Quality measurement and improvement in rheumatology: rheumatoid arthritis as a case study. *Arthritis Rheumatol* 2011;63:3649–60.
2. Donabedian A. The quality of care: how can it be assessed? *JAMA* 1988;260:1743–8.
3. Institute of Medicine (US) Committee on Quality of Health Care in America. Crossing the quality chasm: a new health system for the 21st century. Washington(DC): National Academies Press; 2001.
4. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;151:264–9.
5. Abrahamyan L, Boom N, Donovan LR, Tu JV; Canadian Cardiovascular Society Quality Indicators Steering Committee. An international environmental scan of quality indicators for cardiovascular care. *Can J Cardiol* 2012;28:110–8.
6. Barber CE, Patel JN, Woodhouse L, Smith C, Weiss S, Homik J, et al. Development of key performance indicators to evaluate centralized intake for patients with osteoarthritis and rheumatoid arthritis. *Arthritis Res Ther* 2015;17:322.
7. Barber CE, Marshall DA, Alvarez N, Mancini GB, Lacaille D, Keeling S, et al. Development of cardiovascular quality indicators for rheumatoid arthritis: results from an international expert panel using a novel online process. *J Rheumatol* 2015;42:1548–55.
8. Barber CE, Marshall DA, Mosher DP, Akhavan P, Tucker L, Houghton K, et al. Development of system-level performance measures for evaluation of models of care for inflammatory arthritis in Canada. *J Rheumatol* 2016;43:530–40.
9. Kwok CK, Maclean C, Anderson LG, Dodge C, Fitterman N, Grunwald M, et al. Rheumatoid arthritis physician performance measurement set. National Committee for Quality Assurance/Physician Consortium for Performance Improvement/American College of Rheumatology; 2008.
10. Lovell DJ, Passo MH, Beukelman T, Bowyer SL, Gottlieb BS, Henrickson M, et al. Measuring process of arthritis care: a proposed set of quality measures for the process of care in juvenile idiopathic arthritis. *Arthritis Care Res* 2011;63:10–6.
11. 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 Rheumatol* 2004;51:193–202.
12. Yazdany J, Robbins M, Schmajuk G, Desai S, Lacaille D, Neogi T, et al. Development of the American College of Rheumatology’s rheumatoid arthritis electronic clinical quality measures. *Arthritis Care Res* 2016;68:1579–90.
13. National Institute for Health and Care Excellence. Rheumatoid arthritis in over 16s - Quality standard. [Internet. Accessed September 26, 2017.] Available from: www.nice.org.uk/guidance/qs33
14. National Institute for Health and Care Excellence. Psoriasis. 2013. [Internet. Accessed September 19, 2017.] Available from: www.nice.org.uk/guidance/qs40
15. National Institute for Health and Care Excellence. NICE quality and outcomes framework indicator. [Internet. Accessed September 19, 2017.] Available from: www.nice.org.uk/Standards-and-Indicators/QOFIndicators?categories=&page=4
16. van Hulst LT, Fransen J, den Broeder AA, Grol R, van Riel PL, Hulscher ME. Development of quality indicators for monitoring of

- the disease course in rheumatoid arthritis. *Ann Rheum Dis* 2009;68:1805-10.
17. Peter WF, Hurkmans EJ, van der Wees PJ, Hendriks EJ, van Bodegom-Vos L, Vliet Vlieland TP. Healthcare quality indicators for physiotherapy management in hip and knee osteoarthritis and rheumatoid arthritis: a Delphi study. *Musculoskeletal Care* 2016;14:219-32.
 18. Petersson IF, Strombeck B, Andersen L, Cimmino M, Greiff R, Loza E, et al. Development of healthcare quality indicators for rheumatoid arthritis in Europe: the eumusc.net project. *Ann Rheum Dis* 2014;73:906-8.
 19. Munoz-Fernandez S, Carmona L, Collantes E, Mulero J, Garcia-Yebenes MJ, de Miguel E, et al. A model for the development and implementation of a national plan for the optimal management of early spondyloarthritis: the Esperanza program. *Ann Rheum Dis* 2011;70:827-30.
 20. Brook RH. The RAND/UCLA appropriateness method. In: McCormick KA, Moore SR, Siegel RA, editors. *Methodology perspectives*, AHCPR Pub No 95-0009. Rockville, MD: Public Health Service, U.S. Department of Health and Human Services; 1994:59-70.
 21. Dalal S, Khodyakov D, Srinivasan R, Straus S, Adams J. ExpertLens: A system for eliciting opinions from a large pool of non-collocated experts with diverse knowledge. *Technol Forecast Soc Change* 2011;78:1426-44.
 22. Campbell SM, Braspenning J, Hutchinson A, Marshall MN. Research methods used in developing and applying quality indicators in primary care. *BMJ* 2003;326:816-9.
 23. Stelfox HT, Straus SE. Measuring quality of care: considering conceptual approaches to quality indicator development and evaluation. *J Clin Epidemiol* 2013;66:1328-37.
 24. National Institute for Health Care Excellence. Health and social care directorate: quality standards process guide. [Internet. Accessed September 19, 2017.] Available from: www.nice.org.uk/media/default/Standards-and-indicators/Quality-standards/Quality-standards-process-guide-April-2014.pdf
 25. National Quality Forum (NQF). What NQF endorsement means. [Internet. Accessed September 17, 2017.] Available from: www.qualityforum.org/Measuring_Performance/ABCs/What_NQF_Endorsement_Means.aspx
 26. Rademakers J, Delnoij D, de Boer D. Structure, process or outcome: which contributes most to patients' overall assessment of healthcare quality? *BMJ Qual Saf* 2011;20:326-31.
 27. 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.
 28. Suter LG, Barber CE, Herrin J, Leong A, Losina E, Miller A, et al. American College of Rheumatology white paper on performance outcome measures in rheumatology. *Arthritis Care Res* 2016;68:1390-401.