



Systematic Appraisal of the American College of Rheumatology Clinical Practice Guidelines

Alí Duarte-García,¹  Milena Cavalcante,² Senada Arabelovic,² and John B. Wong² 

Objective. Because the American College of Rheumatology (ACR) practice guidelines affect the United States' and international treatment practice, we used the Appraisal of Guidelines for Research and Evaluation (AGREE) II instrument to characterize the quality of the guidelines and to identify potential areas for improvement.

Methods. Using the six quality domains in the AGREE II online tool, four reviewers assessed the practice guidelines available at the ACR website and the immediately previous version to summarize the domain scores for each guideline and examine trends over time.

Results. As of April 2016, the ACR website listed nine guidelines, four with immediate previous versions. Based on AGREE II, the minimum and maximum for each domain (with higher being better) of the current guidelines were 78–99 for Scope and Purpose, 57–99 for Stakeholder Involvement, 87–96 for Rigor of the Methodology, 83–99 for Clarity of Presentation, 49–78 for Applicability, 69–85 for Editorial Independence, and 71–96 overall. Over time, although the average domain quality of the guidelines improved for all, the Applicability and Editorial Independence domains had the least amount of improvement. For the four guidelines with previous versions, the mean (SD) absolute improvements for each domain were 18 (± 11) for Scope and Purpose, 13 (± 8) for Stakeholder Involvement, 38 (± 22) for Rigor of the Methodology, 25 (± 15) for Clarity of Presentation, 22 (± 12) for Applicability, 24 (± 17) for Editorial Independence, and 31 (± 5) overall.

Conclusion. Based on the AGREE II instrument, the ACR guidelines have achieved high quality over the past 16 years. The Applicability and Editorial Independence domains have the greatest potential for future improvement.

INTRODUCTION

Because clinical practice guidelines (CPGs) increasingly affect practice through their use as a basis for quality improvement and pay for performance measures, the American College of Rheumatology (ACR) not surprisingly places a high priority on the development of CPGs. Indeed, the ACR Rheumatology Quality Measurement Workgroup recommended in 2011 that quality measures be based on the CPGs. Over the years, the ACR has changed the approach and policies used to create guidelines, such as the incorporation of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology and a new policy regarding conflicts of interest (COI) (1,2). As the United States shifts from fee-for-service to value-based pay for performance, CPGs will only become increasingly more important.

With the proliferation of CPGs and concerns that different specialty societies may produce guidelines with conflicting recommendations, evidence-based medicine experts systematically assessed 431 guidelines developed by specialty societies. They found that only 22 (5%) met all three quality measures for reporting the professionals and stakeholders involved in guideline development, the strategy used to identify evidence, and an explicit grading of the quality of the evidence (3). Thus, tools to measure and differentiate the quality of the guideline development process and the reporting of CPGs became a precursor toward assessing the clinical validity and appropriateness of guideline recommendations (4).

Although the quality of rheumatology disease-focused CPGs (5–7) and the European League Against Rheumatism (EULAR) (8) CPGs have been evaluated, to our knowledge, a quality audit of the ACR guidelines has not been performed. We

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

¹Alí Duarte-García, MD: Mayo Clinic, Rochester, Minnesota, and Tufts Medical Center, Boston, Massachusetts; ²Milena Cavalcante, MD, Senada Arabelovic, DO, John B. Wong, MD: Tufts Medical Center, Boston, Massachusetts.

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

Address correspondence to John B. Wong, MD, Tufts Medical Center, 800 Washington St #302, Boston, MA 02111. E-mail: jwong@tuftsmedicalcenter.org.

SIGNIFICANCE & INNOVATIONS

- The quality of the clinical practice guidelines of the American College of Rheumatology has improved through the years, achieving high-quality levels since 2015.
- The Editorial Independence and Implementation domains have the greatest potential for improvement.

sought to characterize the impact of the policies and procedures for CPGs that the ACR has implemented by characterizing the evolving quality of the ACR guidelines over time with the aim to document advances and identify potential areas for future improvement.

MATERIALS AND METHODS

After obtaining and extracting CPGs available on the ACR website (9) in April 2016, we identified the most recent prior CPG versions by hand searching their reference lists and excluding

focused updates. When available, we obtained any appendices and other supportive information accompanying each guideline. Our quality assessment of the ACR CPGs used the Appraisal of Guideline Research and Evaluation II (AGREE II) instrument, which was developed to address variability in the quality of the practice guidelines. Among 40 appraisal checklists for CPGs, a systematic review identified the AGREE II tool as having been thoroughly evaluated and as being one of two comprehensive ones (10). Having construct validity, usefulness to various stakeholders, and good prediction of guideline implementation outcomes, AGREE II has been applied to CPGs across multiple medical specialties (4,11).

Four reviewers were recruited to participate in the study: a general internist with experience in evidence-based medicine and development of guidelines, an internist with rheumatology experience and interest in guideline development and guideline quality, and two clinical rheumatologists. The four reviewers completed the AGREE II online training tutorial (12), which provides trainees with a sample test practice guideline and, after submitting their assessment, provides feedback and compares the responses to those provided by experts. The online tutorial takes approximately one hour to complete (13).

Table 1. Appraisal of Guidelines for Research and Evaluation (AGREE) II instrument

Domain	Item
1. Scope and purpose	1. The overall objective(s) of the guideline is (are) specifically described.
	2. The health question(s) covered by the guideline is (are) specifically described.
	3. The population (patients, public, etc) to whom the guideline is meant to apply is specifically described.
2. Stakeholder involvement	4. The guideline development group includes individuals from all relevant professional groups.
	5. The views and preferences of the target population (patients, public, etc) have been sought.
	6. The target users of the guideline are clearly defined.
3. Rigor of development	7. Systematic methods were used to search for evidence.
	8. The criteria for selecting the evidence are clearly described.
	9. The strengths and limitations of the body of evidence are clearly described.
	10. The methods for formulating the recommendations are clearly described.
	11. The health benefits, side effects, and risks have been considered in formulating the recommendations.
	12. There is an explicit link between the recommendations and the supporting evidence.
4. Clarity of presentation	13. The guideline has been externally reviewed by experts prior to its publication.
	14. A procedure for updating the guideline is provided.
	15. The recommendations are specific and unambiguous.
	16. The different options for management of the condition or health issue are clearly presented.
	17. Key recommendations are easily identifiable.
5. Applicability	18. The guideline describes facilitators and barriers to its application.
	19. The guideline provides advice and/or tools on how the recommendations can be put into practice.
	20. The potential resource implications of applying the recommendations have been considered.
	21. The guideline presents monitoring and/or auditing criteria.
6. Editorial independence	22. The views of the funding body have not influenced the content of the guideline.
	23. Competing interests of guideline development group members have been recorded and addressed.

Table 2. Appraisal of Guidelines for Research and Evaluation (AGREE) II scores and intraclass correlation per guideline (%)

Domain	Scope and Purpose	Stakeholder Involvement	Rigor of Development	Clarity of Presentation	Applicability	Editorial Independence	Overall Score	Intraclass Correlation
SLE 1999	76	51	43	63	36	25	50	0.85
OA 2000	53	47	34	50	28	50	42	0.34
GIOP 2001	60	50	31	63	26	48	46	0.79
RA 2008	76	79	86	92	52	73	71	0.64
GIOP 2010	83	60	82	83	61	69	75	0.64
JIA 2011	82	82	83	86	53	75	71	0.65
LN 2012	78	57	84	92	60	71	83	0.71
OA 2012	81	71	88	93	49	73	79	0.81
Gout 1 2012	82	86	82	90	53	75	79	0.74
Gout 2 2012	78	89	81	90	49	73	83	0.66
PMR 2015	97	99	96	99	70	83	96	0.85
SpA 2015	99	93	95	99	76	85	96	0.62
RA 2015	93	90	93	99	58	79	96	0.69

Abbreviation: GIOP, glucocorticoid-induced osteoporosis; JIA, juvenile idiopathic arthritis; LN, lupus nephritis; OA, osteoarthritis; PMR, polymyalgia rheumatica; RA, rheumatoid arthritis; SpA, spondyloarthritis; SLE: systemic lupus erythematosus.

The four reviewers independently appraised the ACR CPGs with the AGREE II online tool. To avoid sequencing effects, we randomly assigned the order of CPG assignments across reviewers. None of them had participated in the writing or development of any of the appraised CPGs.

Having both construct validity (the items measure variability in the quality of guidelines) and face validity (from the perspective of various stakeholders) (11), AGREE II consists of 23 items organized into six quality domains: i) Scope and Purpose, ii) Stakeholder Involvement, iii) Rigor of Development, iv) Clarity of Presentation, v) Applicability, and vi) Editorial Independence (Table 1). The 23 items target various aspects of practice guideline quality (4). Each item is scored on a 7-point Likert scale with 7 (strongly agree) representing a well-reported item meeting all criteria and 1 (strongly disagree) indicating poorly reported items not meeting any criteria. Dividing the sum of the individual item scores in a domain by the maximum possible score for that domain yields the quality score for each domain, so 100 is perfect.

$$= \left(\frac{\text{obtained score} - \text{minimum possible score}}{\text{maximum possible score} - \text{minimum possible score}} \right) \times 100 \quad (1)$$

We report the domain scores and the sum of the scores for each guideline and examined trends over time.

We calculated intraclass correlation coefficients (ICCs) to estimate inter-rater reliability. ICC estimates for each guideline were calculated based on a mean rating ($k = 4$) for consistency in a 2-way mixed-effects model using Stata Statistical Software 15.1 (StataCorp LLC).

RESULTS

In April 2016, the nine guidelines available on the ACR website (see Appendix 1) included Glucocorticoid Induced Osteoporosis (GIOP) 2010, Juvenile Idiopathic Arthritis (JIA) 2011, Gout Part I and II 2012, Lupus Nephritis (LN) 2012, Osteoarthritis (OA) 2012, Axial Spondyloarthritis (SpA) 2015, Polymyalgia Rheumatica (PMR) 2015, and Rheumatoid Arthritis (RA) 2015. Four had previous versions: Systemic Lupus Erythematosus (SLE) 1999, OA 2000, GIOP 2001, and RA 2008. The results per domain of the nine contemporary guidelines at the time of the study are described below. Table 2 summarizes the AGREE II scores for each of the 13 guidelines published between 1999 and 2015, and Figure 1 displays the temporal improvement in quality scores over time. For intra-rater reliability, the mean ICC for all guidelines assessed was 0.69 (Table 2), suggesting good inter-rater reliability.

Scope and purpose. This domain relates to the overall aim of the guideline, the specific health questions, and the target population. All of the contemporary guidelines at the time of the study scored well on this domain, with only two of them scoring below 80%. The three guidelines published in 2015 (PMR, SpA, and RA) consistently scored above 90.

Stockholder involvement. This domain focuses on the extent to which guideline development included appropriate stakeholders, such as patients, and represented the views of its intended users. Scores of the current CPGs ranged from 60 for GIOP 2010 to 99 in the PMR 2015. All 2015 guidelines scored above 90.

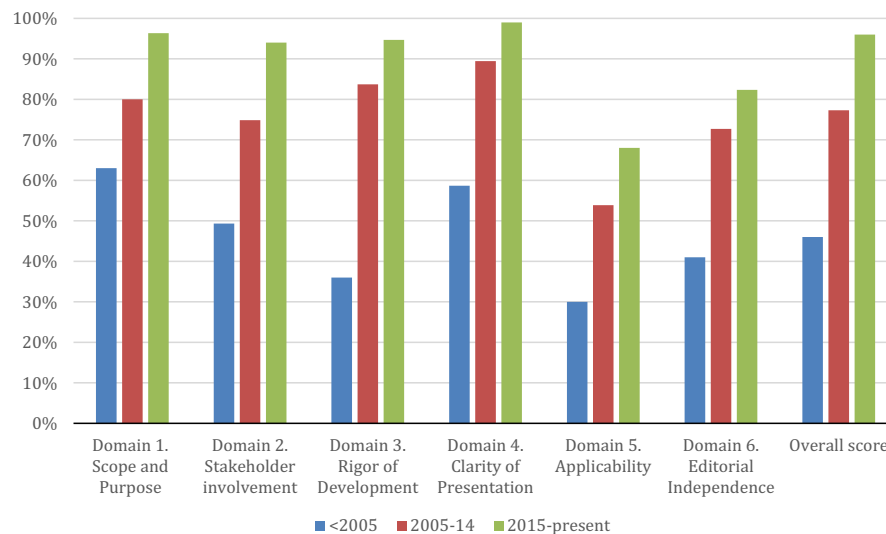


Figure 1. Appraisal of Guidelines for Research and Evaluation (AGREE) II quality scores by time periods (<2005, 2005-2014, 2015-present).

Rigor of development. This domain assesses the process used to gather and synthesize the evidence and the methods to formulate the recommendations and to update them. Reflecting this domain's importance, 8 of the 23 AGREE II items are devoted to its assessment. The ACR CPGs score highly in this domain, with all guidelines having a score that exceeds 80% and all guidelines developed after 2015 scoring above 90%.

Clarity of presentation. This domain evaluates the language, structure, and format of the guideline. All contemporary CPGs scored at or above 90% on this domain except the JIA 2011 and GIOP 2010 CPGs.

Applicability. The applicability domain includes four questions that assess existing facilitators and barriers that impact the application of guideline recommendations, the use of additional materials (eg, executive summary), the potential resource implications, and presentation of auditing criteria, among others (11). Most CPGs had lower scores on this domain, with only two CPGs scoring above 60%. Across all guidelines, items 20 and 21, which relate to costs and criteria for monitoring and audit purposes, respectively, had the lowest scores.

Editorial independence. This domain judges whether the formulation of recommendations could be unduly biased from competing interests. This was the second lowest scoring domain after Applicability. When examining the editorial independence domain, item 22 was the main cause for lower scores in the recently published ACR CPGs because it requires the presence of an explicit statement that the interests of the funding body did not influence the recommendations.

Overall score, time trends, and opportunity for improvement. When examining the overall score of the contemporary guidelines, the JIA 2011 had the lowest score of 71, and the three 2015 CPGs (PMR, SpA, and RA) scored the highest.

Figure 1 displays the temporal improvement in quality scores over each interval, especially notable for Scope and Purpose, Rigor of the Methodology, Stakeholder Involvement, and Clarity of Presentation. In contrast, the Applicability and Editorial Independence domains improved but not to the extent of the other domains. For the four guidelines with previous versions, the mean (SD) absolute improvements for each domain were 18 (± 11) for Scope and Purpose, 13 (± 8) for Stakeholder Involvement, 38 (± 22) for Rigor of the Methodology, 25 (± 15) for Clarity of Presentation, 22 (± 12) for Applicability, 24 (± 17) for Editorial Independence, and 31 (± 5) for overall.

DISCUSSION

Our study shows that the quality of ACR guidelines has improved over the past 16 years. The guidelines published after the release of the most recent ACR policy and procedure manual for CPGs in 2015 have achieved very high levels of quality with all scoring above 90% for domains 1 to 4. The ACR policy procedure manual has made GRADE the new guideline development methodology (14), encouraging the inclusion of patients, patient representatives, and other stakeholders and maintaining a strict approach to managing COI (2). Our analysis also found opportunities for improvement in the Applicability and Editorial Independence domains.

For comparison, the EULAR guideline's AGREE II audit (8) identified Stakeholder Involvement, Applicability, and Editorial Independence as areas for improvement, which encompass

our findings. In audits of CPGs for specific diseases, gout (15), rheumatoid arthritis guideless (5), JIA (16), SLE (6), JIA, and gout audits, found the applicability domain to have the lowest score, which was similar to our observations.

The applicability domain evaluates the advice and tools for dissemination and implementation (ie, recommendations for translating guidelines into practice, such as a quick reference, education tools, or computer support). It also assesses the identification of barriers for implementation and assesses whether a criteria for monitoring the guideline implementation impact is specified. Although all of the domains are relevant, the implementation and use of guidelines in daily practice is essential to attain the goal of translating evidence-based CPG recommendations into routine care and reducing unwarranted variation.

CPG recommendations seek to optimize patient care based on a systematic review of the evidence and a weighing of the pros and cons of the alternative options (17). With predicted decreases in the rheumatology workforce, care for patients with rheumatologic conditions will require innovative approaches to make care delivery more efficient, such as incorporating physician assistants, nurse practitioners, and other mid-level providers. Having high-quality CPGs clarify the evidence base, and the benefits and harms of the options can help clinicians personalize care decisions to unique patients based on the patient's preferences and individual context.

Our study includes some limitations. The appraisals were subjective, so familiarity with a particular topic, evidence-based medicine, and guideline development methodology may have influenced the assessment. Nonetheless, our assessments involved four raters, as ideally recommended by AGREE, with varying levels of clinical experience and with different expertise (two rheumatologists and two internists), the mean ICC was 0.69, which suggests a good inter-rater reliability, and was similar to that obtained by the AGREE developers during the validation of the tool (11). Our review included ACR guidelines but did not examine other rheumatology guidelines from other professional societies, so we could not compare their relative quality. Thus, our conclusions may not be generalizable to guidelines from other organizations.

In summary, the quality of ACR CPGs has improved over the past 16 years, and has achieved high quality since the implementation of the ACR policy and procedures manual for CPGs (2). Residual improvement in the quality of the ACR CPGs should address the Editorial Independence and Applicability domains. We hope our findings will contribute toward improving the quality of the ACR guidelines and their implementation in routine rheumatologic care.

ACKNOWLEDGMENTS

The authors thank Mr. Jonathan Inselman for reviewing the intraclass correlation coefficient models and estimates.

AUTHOR CONTRIBUTIONS

All authors contributed to drafting and critically revising the article for important intellectual content, and all authors approved the final version of the article to be published.

Study conception and design Duarte-Garcia, Wong.

Acquisition of data Duarte-Garcia, Cavalcante, Arabelovic, Wong

Analysis and interpretation of data Duarte-Garcia, Wong

REFERENCES

1. Saag KG, Yazdany J, Alexander C, Caplan L, Coblyn J, Desai SP, et al. Defining quality of care in rheumatology: The American College of Rheumatology white paper on quality measurement. *Arthritis Care Res (Hoboken)* 2011;63:2–9.
2. American College of Rheumatology. Policy and procedure manual for clinical practice guidelines 2015. URL: http://www.rheumatology.org/Portals/0/Files/ACR%20Guideline%20Manual_Appendices_updated%202015.pdf.
3. Grilli R, Magrini N, Penna A, Mura G, Liberati A. Practice guidelines developed by specialty societies: the need for a critical appraisal. *Lancet* 2000;355:103–6.
4. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. Agree II: advancing guideline development, reporting and evaluation in health care. *CMAJ* 2010;182:E839–42.
5. Lopez-Olivo MA, Kallen MA, Ortiz Z, Skidmore B, Suarez-Almazor ME. Quality appraisal of clinical practice guidelines and consensus statements on the use of biologic agents in rheumatoid arthritis: a systematic review. *Arthritis Rheum* 2008;59:1625–38.
6. Tunnicliffe DJ, Singh-Grewal D, Kim S, Craig JC, Tong A. Diagnosis, monitoring, and treatment of systemic lupus erythematosus: a systematic review of clinical practice guidelines. *Arthritis Care Res (Hoboken)* 2015;67:1440–52.
7. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, et al. OARSI recommendations for the management of hip and knee osteoarthritis, part I: critical appraisal of existing treatment guidelines and systematic review of current research evidence. *Osteoarthritis Cartilage* 2007;15:981–1000.
8. Colebatch-Bourn AN, Conaghan PG, Arden NK, Cooper C, Dougados M, Edwards CJ. Raising the quality of rheumatology management recommendations: lessons from the EULAR process 10 years after provision of standard operating procedures. *Rheumatology (Oxford)* 2015;54:1392–6.
9. American College of Rheumatology. Clinical practice guidelines. 2016. URL: <http://www.rheumatology.org/Practice-Quality/Clinical-Support/Clinical-Practice-Guidelines>.
10. Siering U, Eikermann M, Hausner E, Hoffmann-Eßer W, Neugebauer EA. Appraisal tools for clinical practice guidelines: a systematic review. *PLoS One* 2013;8:e82915.
11. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. Development of the AGREE II, part 2: assessment of validity of items and tools to support application. *CMAJ* 2010;182:E472–8.
12. Brouwers MC, Makarski J, Durocher LD, Levinson AJ. E-learning interventions are comparable to user's manual in a randomized trial of training strategies for the AGREE II. *Implement Sci* 2011;6:81.
13. AGREE II training tools. URL: <https://www.agreetrust.org/resource-centre/agree-ii/agree-ii-training-tools/>.
14. GRADE Working Group. 2016. URL: <http://www.gradeworkinggroup.org>.
15. Nuki G. An appraisal of the 2012 American College of Rheumatology Guidelines for the Management of Gout. *Curr Opin Rheumatol* 2014;26:152–61.

16. Smith CA, Toupin-April K, Jutai JW, Duffy CM, Rahman P, Cavallo S, et al. A systematic critical appraisal of clinical practice guidelines in juvenile idiopathic arthritis using the Appraisal of Guidelines for Research and Evaluation II (AGREE II) instrument. *PLoS One* 2015;10:e0137180.
17. Institute of Medicine (US) Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. *Clinical practice guidelines we can trust*. Washington, DC: The National Academies Press; 2011.

APPENDIX

1. Grossman JM, Gordon R, Ranganath VK, Deal C, Caplan L, Chen W, et al. American College of Rheumatology 2010 recommendations for the prevention and treatment of glucocorticoid-induced osteoporosis. *Arthritis Care Res (Hoboken)* 2010;62:1515-26.
2. Beukelman T, Patkar NM, Saag KG, Tolleson-Rinehart S, Cron RQ, DeWitt EM, et al. 2011 American College of Rheumatology recommendations for the treatment of juvenile idiopathic arthritis: initiation and safety monitoring of therapeutic agents for the treatment of arthritis and systemic features. *Arthritis Care Res (Hoboken)* 2011;63:465-82.
3. Khanna D, FitzGerald JD, Khanna PP, Bae S, Singh MK, Neogi T, et al. 2012 American College of Rheumatology guidelines for management of gout part 1: systematic nonpharmacologic and pharmacologic therapeutic approaches to hyperuricemia. *Arthritis Care Res (Hoboken)* 2012;64:1431-46.
4. Khanna D, Khanna PP, Fitzgerald JD, Singh MK, Bae S, Neogi T, et al. 2012 American College of Rheumatology guidelines for management of gout part 2: therapy and anti-inflammatory prophylaxis of acute gouty arthritis. *Arthritis Care Res (Hoboken)* 2012;64:1447-61.
5. Hahn BH, McMahon MA, Wilkinson A, Wallace WD, Daikh DI, Fitzgerald JD, et al. American College of Rheumatology guidelines for screening, treatment, and management of lupus nephritis. *Arthritis Care Res (Hoboken)* 2012;64:797-808.
6. Hochberg MC, Altman RD, April KT, Benkhalti M, Guyatt G, McGowan J, et al. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. *Arthritis Care Res (Hoboken)* 2012;64:465-74.
7. Ward MM, Deodhar A, Akl EA, Lui A, Ermann J, Gensler LS, et al. American College of Rheumatology/Spondylitis Association of America/Spondyloarthritis Research and Treatment Network 2015 recommendations for the treatment of ankylosing spondylitis and nonradiographic axial spondyloarthritis. *Arthritis Rheumatol* 2016;68:282-98.
8. DeJaco C, Singh YP, Perel P, Hutchings A, Camellino D, Mackie S, et al. 2015 recommendations for the management of polymyalgia rheumatica: a European League Against Rheumatism/American College of Rheumatology collaborative initiative. *Arthritis Rheumatol* 2015;67:2569-80.
9. Singh JA, Saag KG, Bridges SL Jr, Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. *Arthritis Rheumatol* 2016;68:1-26.
10. American College of Rheumatology Ad Hoc Committee on Systemic Lupus Erythematosus Guidelines. Guidelines for referral and management of systemic lupus erythematosus in adults. *Arthritis Rheum* 1999;42:1785-96.
11. American College of Rheumatology Subcommittee on Osteoarthritis Guidelines. Recommendations for the medical management of osteoarthritis of the hip and knee: 2000 update. *Arthritis Rheum* 2000;43:1905-15.
12. Recommendations for the prevention and treatment of glucocorticoid-induced osteoporosis: 2001 update. American College of Rheumatology Ad Hoc Committee on Glucocorticoid-Induced Osteoporosis. *Arthritis Rheum* 2001;44:1496-503.
13. Saag KG, Teng GG, Patkar NM, Anuntiyo J, Finney C, Curtis JR, et al. American College of Rheumatology 2008 recommendations for the use of nonbiologic and biologic disease-modifying antirheumatic drugs in rheumatoid arthritis. *Arthritis Rheum* 2008;59:762-84.