SUPPLEMENTARY APPENDIX 1: Methods

Methodology Overview

This guideline followed the American College of Rheumatology (ACR) guideline development process (http://www.rheumatology.org/Practice-Quality/Clinical-Support/Clinical-Practice-Guidelines), using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology to rate the quality of the available evidence and to facilitate the development of the recommendations (1).

Using the GRADE approach, the quality of evidence is rated as high, moderate, low, or very low. Randomized trials begin as high quality evidence, but may be rated down as a result of serious limitations with respect to risk of bias, imprecision, inconsistency, indirectness, or publication bias (2). Observational studies are typically rated as low or very low quality evidence, but may be rated up if the effect size is large in sufficiently large studies. (4,5)

GRADE methodology specifies that recommendations are made based on a consideration of the balance of relative benefits and harms of the treatment options under consideration, the quality of the evidence (i.e., confidence in the estimated effects of an intervention), and patients' values and preferences. Key to the recommendation is the trade-off between desirable and undesirable outcomes; recommendations require estimating the relative value patients place in the outcomes. The Voting Panel, in keeping with the views of a patient advisory panel, estimated that typical patients place a much higher value on avoiding infection and a lower value on avoiding a disease flare.

Using GRADE, a recommendation can be either in favor or against the proposed intervention and either strong or conditional (3) and a clear distinction is made between the quality of the evidence and the strength of the recommendations. A strong recommendation indicates that all or almost all physicians would make the recommendation, and all or almost all informed patients would choose the

recommended action, and that additional research would be unlikely to change the recommendation. Conditional recommendations are those in which most informed patients would choose the recommended course of action, but a minority might not (6,7). All of the recommendations in this guideline are conditional due to the quality of the evidence, as there was little high quality evidence identified directly addressing questions about when to stop or re-start rheumatic medications, and much of the evidence used came from non-surgical studies. In addition, need for additional research was identified.

Teams Involved

This project was a collaboration between the American College of Rheumatology (ACR) and the American Association of Hip and Knee Surgeons (AAHKS); all participating teams included representation from both the organizations. A Core Leadership Team was comprised of ACR and AAHKS co-principal investigators (SG, BS) who co-led the project, the ACR and AAHKS Literature Review Team leaders (JS and AY), and a methodologist who had GRADE expertise (GG). Experts with content and methodology expertise helped define the scope of the project and drafted the Patient/Intervention/Comparator/Outcomes (PICO) questions (see list of PICO questions in Appendix 3), with participation of the Literature Review Team and the Voting Panel. The Expert Panel was comprised of 2 orthopaedists, 3 rheumatologists, 1 methodologist, 1 rheumatology methodologist, 2 infectious disease experts, 1 patient representative, and an SLE expert, with the support of the ACR staff.

The Literature Review Team was comprised of 8 orthopaedists and 6 rheumatologists, who had the support of ACR staff. The literature search was performed with the assistance of a medical research librarian.

A patient panel was convened to discuss patient values and preferences relative to outcomes and PICO findings; the results of the patient meeting were used as part of the weighing of risks and benefits by the

Voting Panel, which was comprised of 6 orthopaedists, 5 rheumatologists, an infectious disease expert, an SLE expert, 2 patient representatives, 2 rheumatology methodologists, as well as an expert in GRADE methodology, and was supported by ACR staff. The Voting Panel discussed the results of the literature review and reframed the PICO questions into recommendations after reviewing the evidence synthesis presented by the Literature Review Team leaders.

Patient Panel

A patient panel was convened the day prior to the Voting Panel meeting on July 10, 2016, consisting of 11 adults with RA and JIA, all of whom had undergone THA or TKA, with a range of 1 to 8 joints replaced per patient, with only one patient reporting a prosthetic joint infection. No patients with SLE or SPA were included in the panel. The mean age of the participants was 47 years (range of 23 to 71) and the mean duration of disease was 26 years (range of 8 to 42). Two members of the Core Leadership Team and one ACR staff person facilitated the day-long discussion. The participants, all of whom had completed research and guideline methodology webinars prior to meeting, were presented the background and scope of the guideline project determined at the first face-to-face meeting. The patients were specifically queried on the relevant importance of surgical-site or non-surgical site infection, rare post-operative events linked to continued immunosuppressant DMARD and biologic use, compared to the importance of flares of disease linked to withholding the medications, which are frequent after THA and TKA. The patient panel reviewed the evidence synthesized by the Literature Review Team as each PICO question was discussed. The participants were encouraged to consider their personal experiences relevant to the questions and judge the importance of the outcomes accordingly. The values and preferences of the patient panel and the voting results for each recommendation were presented to the Voting Panel by two core team members who facilitated the patient panel meeting during their discussions the following day.

Disclosures and Management of Conflicts of Interest

Per ACR policy, everyone who was considered for intellectual involvement in the project (i.e., considered for guideline authorship), disclosed all relationships (see https://www.rheumatology.org/Portals/0/Files/Perioperative-Management-GL-Disclosure%20Summary.pdf for full details on participant disclosures). The agreed upon next step was to compare disclosures against a previously drafted list of "affected companies" (i.e., companies or organizations that were considered reasonably likely to be positively or negatively affected by care delivered in accordance with the guideline) to determine which relationships were considered conflicts of interest for purposes of this project. However, because the focus of this guideline was to temporarily stop or restart medications that were already prescribed, in situations where surgery had already been scheduled, it was decided by the ACR and AAHKS that there were no affected companies for this guideline, and therefore, no conflict of interest for any individuals involved. Even so, in keeping with ACR policies, individuals whose primary employment (> 51% of work time/effort) was with a company that manufactured or sold therapeutics or diagnostics were not eligible to participate.

Intellectual conflicts, such as a prior publication or scientific presentation on perioperative management of DMARDs and biologics in patients with rheumatic diseases undergoing THA/TKA, were recognized as important and were required to be disclosed, but because they were ubiquitous, participants with intellectual conflicts were not counted as conflicted based on their intellectual conflict alone.

Participant disclosures were included in the project plan that was posted online for public comment. In addition, disclosures of all participants were shared, in writing, with each project participant. At the face-to-face Voting Panel meeting, verbal disclosures were provided before any content discussion. Updated participant disclosures, as well as ACR committee reviewer disclosures, are included online with this manuscript. In addition, author disclosures are also included in this paper.

PICO Question Development

The Core Leadership Team initially drafted the project scope, key principles and relevant clinical (PICO) questions, which were then presented to the Expert Panel, the Voting Panel, and the Literature Review Team for their review at a face-to-face meeting where the project plan was defined. The project plan, including these elements and other project details, was sent to ACR and AAHKS members via broadcast email and electronic newsletters, and was also posted on the ACR and AAHKS websites for public comment and revised accordingly. The group initially considered a wide range of outcomes, but eventually determined that infection (both deep surgical site, reported within the first year after surgery, or superficial surgical site and non-surgical site infections within 90 days of surgery) and disease flare were the most critical, although literature on other outcomes such as hospital readmission, non-surgical site or remote infection, death, and long-term arthroplasty outcome was also sought.

The outcome with the greatest weight for this guideline was deep surgical-site infection, an uncommon event on the order of 0.5 to 2.4% (8,9). The group acknowledged that there would likely not be direct high quality RCT data available comparing the risk of infection after THA or TKA in those taking versus not taking the medications of interest, or comparing the background risk of adverse events after THA and TKA in the populations of interest, due primarily to practical reasons (the inability to provide sufficient power for a study with a rare endpoint). To address this gap, two questions were included to inform the recommendations – the first sought indirect evidence of drug-related adverse effects from studies outside of the perioperative setting, and the second sought to establish the baseline risk of adverse events in patients with inflammatory arthritis undergoing THA or TKA who were not receiving the drugs of interest:

1. Indirect evidence: What is the risk for serious adverse events, infections, or hospitalizations, associated with use of each of the candidate drugs outside of the surgical setting, limiting

the search to systematic literature reviews (SLRs) and meta-analyses (MAS) for RA, SpA, and JIA, and including observational studies in SLE, as indicated?

2. What is the background risk for adverse events associated with THA or TKA in patients with RA, SpA, JIA, or SLE independent of the use of anti-rheumatic medications of interest?

Systematic Synthesis of the Literature

1. Literature Searches

Literature search strategies based on PICO questions were developed by the principal investigators, the systematic review leaders, and a research librarian, with input from the GRADE consultant. The search strategies were reviewed by another medical librarian using the Peer Review of Electronic Search Strategies (PRESS). Searches were performed in Embase (1974+), the Cochrane Library and PubMed (mid-1960s+) from January 1, 1980, through March 6, 2016.

The search strategies were developed using the controlled vocabulary or thesauri language for each database: Medical Subject Headings (MeSH) for PubMed and Cochrane Library; and Emtree terms for Embase (Supplementary Appendix 1). Text words were used in PubMed and Embase, and keyword/title/abstract words in the Cochrane Library. Searches resulted in 2,230 total references. After title and abstract and full manuscript screening, 19 papers were included as relevant for PICO 1, 9 for PICO 2, 31 for PICO 3, 20 for PICO 4, and 69 for background questions 5 and 6 (Supplementary Appendix 2). A final search was performed for the time period of January 1 to September 8, 2016, using the inclusive search terms of the disease states (RA, SpA including AS and PsA, JIA, and SLE) coupled separately with “arthroplasty;” no randomized trials were identified that were relevant to the guideline.

2. Study Selection

DistillerSR software (available at: http://systematic-review.net/) was used to screen the literature search results grouped by their match with the pertinent PICO questions. Duplicate screening of each title and abstract was performed by two independent reviewers from among a pool (BJ, AY, MT, SS, LM, MG, SL, JG, LS, MM, PS, VD), with a third reviewer (AY or JS) resolving conflicts. The second screen was done with the full text of the papers available by two independent reviewers from the same pool. Selected manuscripts were then reviewed in their entirety.

3. Evidence Report Formulation

The Literature Review Team analyzed and synthesized data from included studies that addressed the PICO questions. An evidence summary was prepared as a PowerPoint presentation for each PICO question; due to the lack of RCTs, we were unable to prepare GRADE Summary of Findings (SoF) tables for most PICO questions as planned using GRADEprofiler (GRADEpro) software. Microsoft Excel was used for abstracting data from observational studies. When available, the evidence summaries contained the benefits and harms for outcomes of interest across studies, the relative effect (95% CI), the number of participants, and number needed to treat. We rated the quality of evidence for each critical and important outcome (i.e., high, moderate, low, or very low), taking into account limitations of study design, inconsistency, indirectness, imprecision, and other considerations. The Core Leadership Team reviewed the evidence summary and discussed possible evidence gaps prior to the presentation to the patient panel on July 10, 2016, and the Voting Panel the following day.

Moving from Evidence to Recommendations

The patient panel weighed the evidence first and analyzed it in the context of their experiences. The panel participants recognized that post-operative flares were very common and very difficult for them, and infection was rare. However, the importance they attached to infection at the time of surgery was far greater than the importance attached to flares. They were unable to precisely quantify the difference

in value, noting that it was greater than 10:1 or 20:1. They felt that flares represented a "known risk." The patients viewed endurance during the perioperative period as a "job" in which their task is to focus on the eventual positive outcomes of better mobility and less pain, while minimizing major risks as much as possible. From the perspective of the patients, there was no "average" infection – as all infections had potential to develop into significantly worse possible outcomes than flares (e.g., permanent loss of joint, amputation, death). While flares were perceived as difficult, infection could postpone recovery and/or introduce other health issues, which patients felt was unacceptable because it would delay achievement of the positive outcomes they sought.

Patients agreed that close coordination between the rheumatologist and the orthopaedist was essential, including timing surgery at the end of a patient's drug dosing cycle to minimize infection *and* flare risks. The presence of a coordinated approach was important to them and would influence their perspectives about which risks they were willing to take if they were confident that their individual needs were considered.

In regards to the recommendation for glucocorticoid dosing, patients agreed that there was little support for use of supra-physiologic "stress-dose steroids," but they wondered whether flares were prevented as an unexpected benefit related to the use of "stress-dose steroids." Finally, the patients noted that they were uncomfortable providing important input into the recommendations for management of patients with SLE, as there were no lupus patients in the group.

The next day, the Voting Panel met to decide the final guideline recommendations. PICO questions had been reformulated as drafted recommendation statements, for the panel's consideration. The panel, chaired by the co-PIs, discussed the evidence in the context of their clinical experience and expertise, as well as the input from the patient panel, which was summarized and presented during the Voting Panel meeting. The panel voted anonymously and an 80% consensus was used as the threshold for a

recommendation; if 80% consensus was not achieved during an initial vote, the panel members held additional discussions before re-voting. The Voting Panel meeting discussions were supported by the systematic review leaders, the GRADE expert, and selected members of the systematic review team, who attended the meeting to summarize the evidence and provide details, as requested.

Much of the evidence for this guideline was indirect, which lowered evidence quality ratings. Included studies were heterogeneous with regard to surgical procedures, including foot or spine procedures in which infection risks vary markedly from THA or TKA. Heterogeneity in baseline medication dose and duration was particularly relevant in studies addressing glucocorticoid "stress-dose" therapy. Most studies of drug-related infection risk are derived from RCTs and are not performed in patients undergoing surgery. Therefore, observational studies were used to determine baseline risk associated with THA or TKA in the patient populations addressed by this guideline, and additionally, imprecision led to rating the evidence down where studies reported on small numbers. The patient panelists, however, provided very clear guidance on their values and preferences, rating the importance of perioperative infection, a rare event, significantly higher than post-operative flares, which were frequent (10,11); this input helped inform the Voting Panel's final decisions even in the absence of high quality literature about risks.

All recommendations were supported by over 80% of the panel, and all but one were supported unanimously. In some instances, the panel combined PICO questions into one final recommendation. When we recommended that a medication be withheld, we included a recommendation for the suggested timing of surgery in relation to the drug-dosing interval.

Final Review and Approval of the Manuscript by the ACR and AAHKS

In addition to journal peer reviews, the manuscript was reviewed by the following committees and subcommittees of the ACR and AAHKS: ACR Guidelines Subcommittee; ACR Quality of Care Committee;

ACR Board of Directors; AAHKS Evidence Based Medicine Committee; and AAHKS Board of Directors. These ACR and AAHKS oversight groups did not make or mandate that specific recommendations be made within the guideline, but rather, served as peer reviewers.

REFERENCES

1. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336(7650):924-6.
2. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schunemann HJ, et al. What is "quality of evidence" and why is it important to clinicians?. BMJ 2008;336(7651):995-8.
3. Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A, et al. Going from evidence to recommendations. BMJ 2008;336(7652):1049-51.
4. Alonso-Coello P, Oxman AD, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines. BMJ 2016;353:i2089.
5. Neumann I, Santesso N, Akl EA, Rind DM, Vandvik PO, Alonso-Coello P, et al. A guide for health professionals to interpret and use recommendations in guidelines developed with the GRADE approach. J Clin Epidemiol 2016;72:45-55.
6. Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. J Clin Epidemiol 2013;66(7):719-25.
7. Andrews JC, Schunemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Coello PA, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. J Clin Epidemiol 2013;66(7):726-35.
8. Kurtz SM, Lau E, Watson H, Schmier JK, Parvizi J. Economic burden of periprosthetic joint infection in the United States. J Arthroplasty 2012;27(8 Suppl):61-5.e1.
9. Centers for Disease Control and Prevention. General Recommendations on Immunization: Recommendations of the Advisory Committee on Immunization Practices. MMWR 2011;60(RR-2):22-23.
10. Alonso-Coello P, Montori VM, Diaz MG, Devereaux PJ, Mas G, Diez AI, et al. Values and preferences for oral antithrombotic therapy in patients with atrial fibrillation: physician and patient perspectives. Health Expect 2015;18(6):2318-27.
11. Zhang Y, Tikkinen KA, Agoritsas T, Ayeni OR, Alexander P, Imam M, et al. Patients' values and preferences of the expected efficacy of hip arthroscopy for osteoarthritis: a protocol for a multinational structured interview-based study combined with a randomised survey on the optimal amount of information to elicit preferences. BMJ Open 2014;4(10):e005536.