

## **SUPPLEMENTARY APPENDIX 1: Methods**

### ***Methodology Overview***

We developed this guideline following the [ACR guideline development process](http://www.rheumatology.org/Practice-Quality/Clinical-Support/Clinical-Practice-Guidelines) (<http://www.rheumatology.org/Practice-Quality/Clinical-Support/Clinical-Practice-Guidelines>). This process includes the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) methodology ([www.gradeworkinggroup.org](http://www.gradeworkinggroup.org)) (1-3).

### ***Teams Involved***

A Core Leadership Team (4 members) supervised the project and was responsible for defining the scope, drafting the clinical (Patient/Intervention/Comparator/Outcomes - PICO) questions, coordinating with the Literature Review Team, overseeing the voting process, and drafting the manuscript. The Core Team, together with the Literature Review Team, was comprised of individuals with content and methodological expertise, and included a GRADE methodologist who advised on the process of developing and presenting the evidence and provided input on the quality assessment of evidence and summary of findings (SoF) tables (provided in Supplementary Appendix 3). The Literature Review Team (5 members) conducted a systematic search, assessed study quality, extracted data, computed pooled estimates of outcomes, graded the quality of overall evidence, generated the SoF tables, and compiled the resulting evidence into an evidence report. The role of the Expert Panel, composed of 9 content experts, was to provide consultation and feedback on the PICO questions. The Voting Panel (16 members) included adult and pediatric rheumatologists, internists, a nephrologist, a pulmonologist, a gastroenterologist, medical specialists with expertise and clinical experience in treating glucocorticoid induced OP (GIOP), and a patient representative. Their role was to participate in the development of the PICO questions, including

making judgments regarding the relative importance of the outcomes, and vote on the PICO questions based on the evidence report and their expertise. The ACR provided training for everyone involved in the development of this guideline, which included sessions on the ACR guideline process and GRADE methodology. See Supplementary Appendix 2 for team/panel rosters.

### ***Disclosures and Management of Conflicts of Interest***

Per ACR policy, everyone who was intellectually involved in the project (i.e., considered for guideline authorship) disclosed all relationships (<https://www.rheumatology.org/Portals/0/Files/GIOP-Guidelines-Disclosure-Summary.pdf>). Disclosures were compared 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. 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.

The project’s principal investigator (PI) and literature review leader had no relevant conflicts of interest for the full 12 months before this project began, and all but one of the guideline development team members, including the PI and literature review leader, had no relevant conflicts of interest for the duration of the project. A participant who had any relationship with an affected company was counted as conflicted (i.e., toward the allowed threshold) regardless of the type or subject of the relationship. Intellectual conflicts, such as a prior publication or scientific presentation on a GIOP therapy, were recognized as important and were required to be disclosed, but because they were ubiquitous, participants with intellectual conflicts were not counted as conflicted (i.e., toward the allowed threshold) based on their intellectual conflict alone.

Participant disclosures were included in the project plan that was posted online for public comment (see description below). 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. Finally, author disclosures are also included in this paper.

### ***Scope and Target Audience***

The scope of this project included the assessment, prevention and treatment of OP and fractures in children and adults taking glucocorticoids (prednisone dose of > 2.5 mg of prednisone for  $\geq$  3 months), including patients with organ transplant who are treated with GCs. Clinical situations not addressed by this guideline include treatment of people with stage 4-5 chronic kidney disease and people who use inhaled GCs. The target audience for this guideline includes patients with or at risk for GIOP and their clinicians. The ACR plans to develop derivative products, including a pocket card, an app version of this guideline, and a patient education tool to facilitate implementation.

### ***Establishing Key Principles and PICO Development***

The Core Team collaborated with the Expert Panel to develop the initial set of PICO-formatted clinical questions for the guideline (4). The panel ranked fracture (hip, vertebral, non-vertebral) as the only *critically important* outcome for treatment evaluation. *Important* outcomes included adverse effects of treatments, in particular the incidence of serious and total adverse events (AEs) in all clinical scenarios. These events included upper gastrointestinal AEs for bisphosphonates; rates of transplant rejections, mortality, and hypo- or hypercalcaemia in transplant recipients; and maternal and fetal risks for women of child-bearing potential. The PICO questions were posted for 30 days on the ACR web site for public comment and revised accordingly.

## ***Framework for the GIOP Guideline Development***

At the initial scoping meeting, the group agreed that the scope of the populations to be addressed would include adult men and postmenopausal women and special populations that have unique risks, such as organ transplant recipients, women of childbearing potential, children, and people receiving very high-dose GCs (defined as one or more courses of high dose GCs (initial dose of  $\geq 30$  mg/day of prednisone or equivalent) and a cumulative dose of  $\geq 5$  grams (5,6). Chronic GC use was defined as  $\geq 3$  months in duration. Adult men and women were divided into two groups based on age ( $\geq 40$  years,  $< 40$  years) because tools to predict absolute fracture risk are available only for adults  $\geq 40$  years.

After defining population risk groups, interventions and comparators were specified for each PICO question (see list of PICO questions in Supplementary Table 1). The Core Team agreed that while bone mineral density (BMD) data would be examined, the critical outcome for the analysis and literature search was fracture – particularly vertebral fracture. Vertebral fractures are more common than femoral fractures in GC treated patients and GIOP clinical trials are not of adequate size to assess the impact of interventions on femoral fractures. When necessary to use BMD to support a recommendation (which was the case in only four PICO questions, all addressing pediatric GIOP patients), the Voting Panel rated down the quality of evidence for indirectness. PICO questions included assessment and reassessment of fracture risks, treatment comparisons, and questions about duration and reassessment of treatment.

## ***Systematic Synthesis of the Literature***

### ***Literature Searches***

To identify relevant evidence for the PICO questions, a medical librarian, in collaboration with the Literature Review Team, performed systematic searches of the

published English language literature. We searched OVID Medline, PubMed, Embase, and the Cochrane Library (including Cochrane Database of Systematic Reviews; Database of Abstracts of Reviews of Effects (DARE); Cochrane Central Register of Controlled Trials (CENTRAL); and Health Technology Assessments (HTA)) from the beginning of each database through October 6-8, 2015 (Supplementary Appendix 4), and update searches were conducted on April 23, 2016. For PICO questions for which we found no direct evidence in the GIOP field, we sought indirect evidence, in particular meta-analyses of randomized trials, in the non-GIOP populations.

### ***Study Selection***

We used DistillerSR software (<https://distillercer.com/products/distillersr-systematic-review-software/>) to screen the literature search results. We performed duplicate screening of each title and abstract using two independent reviewers from among a pool of reviewers, with a third reviewer resolving conflicts. Eligible articles underwent full-text screening by two independent reviewers. Selected manuscripts were matched to PICO questions. See Supplementary Appendix 5 for details related to the study selection process.

### ***Data Extraction and Analysis***

We extracted data from RCTs for each PICO question into RevMan software (<http://tech.cochrane.org/revman>). The quality of each study was assessed using the Cochrane risk of bias tool (<http://handbook.cochrane.org/>). The critical/important outcomes (fractures and adverse events) selected for this guideline were binary, and they were analyzed using the Mantel-Haenszel method in a random effects model and reported as relative risks with 95% confidence intervals.

In clinical scenarios not addressed by RCT data (e.g., certain special populations, such as patients on high-dose steroids), we used data from observational cohort studies to estimate relative effects. In situations in which the question had not

been tested in a sample taking glucocorticoids but had been tested in a non-GIOP population, we applied the relative risk values from that study, making the assumption that the effect was generalizable but rating down the quality of evidence for indirectness.

After calculating relative treatment effects for the pooled data for each PICO question, we quantified overall absolute risk reduction using pooled control group fracture rates. Next, in order to project absolute risk reduction within each risk strata, we constructed a risk calculator to display the absolute risk reduction in vertebral fracture rates over five years, depending on hypothetical baseline fracture risk ranging from 1% to 20%. We used the following cut-points to stratify levels of risk: low risk, viewed as < 5% incidence of vertebral fractures over 5 years; medium risk, 5 to < 10%; and high risk,  $\geq$  10%. The Voting Panel then made recommendations based on absolute fracture reduction with treatment in each of these strata. We focused on vertebral fracture rates because this outcome was more consistently reported in the literature and because of the greater effects of GCs on trabecular bone.

### ***Evidence Report Formulation***

We exported RevMan files into GRADEpro software to formulate a GRADE summary of findings (SoF) table for each PICO question (4). The quality of evidence for each outcome was evaluated in duplicate by three independent reviewers using GRADE quality assessment criteria (1) with discordance resolved by discussion. We compiled the resulting SoF tables in an evidence report (Supplementary Appendix 3). The Core Team reviewed the evidence report and addressed possible evidence gaps prior to presentation to the Voting Panel.

### ***Moving from Evidence to Recommendations***

GRADE methodology specifies that panels make recommendations 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 effect estimates), 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.

A recommendation could be either in favor or against the proposed intervention and either strong or conditional. According to GRADE, a recommendation is categorized as strong if the panel is very confident that the benefits of an intervention clearly outweigh the harms (or vice versa); a conditional recommendation denotes uncertainty over the balance of benefits and harms, such as when the evidence quality is low or very low, or when patient preferences or costs are expected to impact the decision. Thus, conditional recommendations refer to decisions where incorporation of patient preferences is an essential element of decision making.

The Voting Panel's assessment was that patients would be willing to use calcium and vitamin D with only a very small absolute risk reduction, and that all or virtually all would be willing to use bisphosphonates to achieve a 5-year absolute reduction in vertebral fracture risk of 5%, and that most would choose to use oral bisphosphonates if the fracture reduction were  $\geq 3\%$  to  $<5\%$  (leading to a conditional recommendation in favor) The 5-year time period was chosen because few clinical trials have data on fracture risk reduction past 3-5 years. Further, the panel thought that most patients would decline oral bisphosphonates with an absolute reduction in 5-year risk of vertebral fractures of 1.6% to 2.9% (leading to a conditional recommendation against) and all or virtually all would decline if the risk reduction were less than 1.5% (leading, in the presence of high or moderate quality evidence, to a strong recommendation against).

For IV bisphosphonates, denosumab, raloxifene, and teriparatide, which have greater harms or burden of treatment, the threshold was higher, though the panel did not specify a threshold value. Because raloxifene may increase the risk of death due to

stroke in postmenopausal women with documented coronary heart disease or at increased risk for major coronary events and/or increase the risk of deep vein thrombosis and pulmonary embolism (7), and there is no evidence of its benefit in fracture reduction in GC-treated patients, the Voting Panel considered the drug as a treatment option only for postmenopausal women with contraindications to all other treatments.

We are unaware of published literature exploring patient values and preferences regarding these issues. Our judgments are based on the experience of the panel members (which included a patient) in shared decision making with their patients.

### ***Consensus Building***

The Voting Panel received the evidence report for review before it met to discuss and decide on the final recommendations. During a face-to-face meeting held on May 14, 2016, and subsequent conference calls and e-mails, the Voting Panel, for each PICO question, reviewed the evidence and provided votes on the direction and strength of the related recommendation. The initial voting process was conducted using Poll Everywhere software (<http://www.polleverywhere.com/>) with follow-up conference calls to vote on unresolved questions. 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.

In some instances, the Voting Panel decided to combine recommendation statements comparing treatment options (calcium and vitamin D, lifestyle changes) for all patient groups. They then voted on a new recommendation statement that combined those treatment options instead of considering the interventions separately for each risk group. When considering post-transplant patient populations with GFR  $\geq$  30 and no evidence of metabolic bone disease, the panel decided that despite the availability of transplant-specific evidence, the recommendations should be identical to



the same-age recommendations for adults without transplants. Those with GFR < 30 should be referred to the care of an expert in osteoporosis and metabolic bone disease because of the difference in harms and benefits of treatment in this group. In addition, the Voting Panel dropped a number of PICO questions because the clinical scenario was uncommon, irrelevant, or redundant (see Supplementary Table 1), or because one of the treatment options for that scenario had been eliminated by another recommendation. Consistent with GRADE guidance, in some instances, the Voting Panel chose to provide a strong recommendation despite a low quality rating of evidence (3). In such cases, a written explanation is provided describing the reasons behind this decision.

### ***Final Review and Approval of the Manuscript by the ACR***

In addition to journal peer reviews, the manuscript was reviewed by the ACR Guideline Subcommittee, ACR Quality of Care Committee, and the ACR Board of Directors – including over 40 reviewers (<https://www.rheumatology.org/Portals/0/Files/GIOP-Guidelines-Disclosure-Summary.pdf>). These ACR oversight groups did not mandate that certain recommendations be made within the guideline, but rather, these ACR committees served as peer reviewers.

### ***Moving From Recommendations to Practice***

When applying these risk-stratified recommendations in clinical settings, adults  $\geq$  40 years of age on chronic GCs should be designated to be at high, moderate, or low risk of fracture (Table 1) based on BMD, history of fracture, and 10-year risk of major OP fracture and hip fracture using a tool that combines risk factors with GC dose. Although many tools are available, the Voting Panel suggested using FRAX (Figure 1) for fracture risk assessment. When GC use is included as a risk factor in FRAX, the fracture risk generated is the risk associated with a prednisolone dose of 2.5-7 mg of per day

(prednisolone and prednisone doses are nearly equivalent). For people taking  $\geq 7.5$  mg per day, the fracture risk generated by FRAX should be increased by a relative 15% for major osteoporotic fracture (MOF) and 20% for hip fracture risk (8). For example, if the 10-year hip fracture risk was 2.0% with glucocorticoids entered in FRAX, the risk should be increased to 2.4% if the prednisone dose is  $\geq 7.5$  mg per day.

There are no tools available to estimate absolute fracture risk in children or adults  $< 40$  years of age. These groups were considered to be at high fracture risk if they had already sustained an OP fracture. The Voting Panel designated men and women  $< 40$  years of age to be moderate to high fracture risk if they were expected to continue GC treatment at  $\geq 7.5$  mg per day for 6 months and had a BMD Z score  $< -3$  at the hip or spine or a rapid decline in BMD (equivalent to  $\geq 10\%$  at the hip or spine in 1 year) during GC treatment.

## 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. 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.
3. 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.
4. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol* 2011;64:395-400.
5. van Staa TP, Geusens P, Pols HA, de Laet C, Leufkens HG, Cooper C. A simple score for estimating the long-term risk of fracture in patients using oral glucocorticoids. *QJM* 2005;98:191-8.
6. De Vries F, Bracke M, Leufkens HG, Lammers JW, Cooper C, van Staa TP. Fracture risk with intermittent high-dose oral glucocorticoid therapy. *Arthritis Rheum*. 2007;56:208-214.
7. EVISTA (raloxifene hydrochloride) Tablet for Oral Use Initial U.S. Approval: 1997. URL: [http://www.accessdata.fda.gov/drugsatfda\\_docs/label/2007/022042lbl.pdf](http://www.accessdata.fda.gov/drugsatfda_docs/label/2007/022042lbl.pdf).
8. Guyatt GH, Schünemann HJ, Djulbegovic B, Akl EA. Guideline panels should not GRADE good practice statements. *J Clin Epidemiol*. 2015;68(5):597-600.