AAOS
American Academy of Orthopaedic Surgeons

AAOS Clinical Practice Guideline Methodology

American Academy of Orthopaedic Surgeons

Contents

Overview of Clinical Practice Guideline (CPG)

The AAOS understands that only high-quality clinical practice CPG are credible, and we go to great lengths to ensure the integrity of our evidence analyses. The AAOS addresses bias beginning with the selection of CPG work group members. Applicants with financial conflicts of interest (COI) related to the CPG topic cannot participate if the conflict occurred within one year of the start date of the CPG's development or if an immediate family member has, or has had, a relevant financial conflict. Additionally, all CPG development group members sign an attestation form agreeing to remain free of relevant financial conflicts for one year following the publication of the CPG.

CPGs are prepared by physician CPG development groups (clinical experts) with the assistance of the AAOS Clinical Quality and Value (CQV) Department (methodologists) at the AAOS. To develop CPGs, the CPG development group meets at an introductory meeting to establish the scope of the CPG. As the physician experts, the CPG work group defines the scope of the CPG by creating PICO Questions (i.e., population, intervention, comparison, and outcome) that direct the literature search. When necessary, these clinical experts also provide content help, search terms and additional clarification for the AAOS Medical Librarian. The Medical Librarian creates and executes the search(es). The supporting group of methodologists (AAOS CQV Department) review all abstracts, recall pertinent full-text articles for review and evaluate the quality of studies meeting the inclusion criteria. They also abstract, analyze, interpret, and summarize the relevant data for each PICO question and prepare the initial draft for the final work group meeting. Upon completion of the systematic reviews, physician CPG work groups participate in a one-day final meeting. To complete their charges, the physician experts and methodologists evaluate and integrate all material to develop the final recommendations. The final recommendations and rationales are edited, written and voted on.

Additional edits to the rationales are approved by the CPG work group after the meeting. The draft CPG recommendations and rationales receive final review by the methodologists to ensure that these recommendations and rationales were consistent with the data. The draft is then completed and submitted for a review period.

After the review period, the CPG draft may be edited in response to the review submissions. Thereafter, the draft CPG is sequentially approved by the AAOS Committee on Evidence-Based Quality and Value, AAOS Research and Quality Council, and the AAOS Board of Directors. All AAOS CPGs are reviewed for update or retirement every five years.

The process of AAOS CPG development incorporates the benefits from clinical physician expertise as well as the statistical knowledge and interpretation of non-conflicted methodologists. The process also includes an extensive review process offering the opportunity for over 200 clinical physician experts to provide input into the draft prior to publication. This process provides a sound basis for minimizing bias, enhancing transparency and ensuring the highest level of accuracy for interpretation of the evidence.

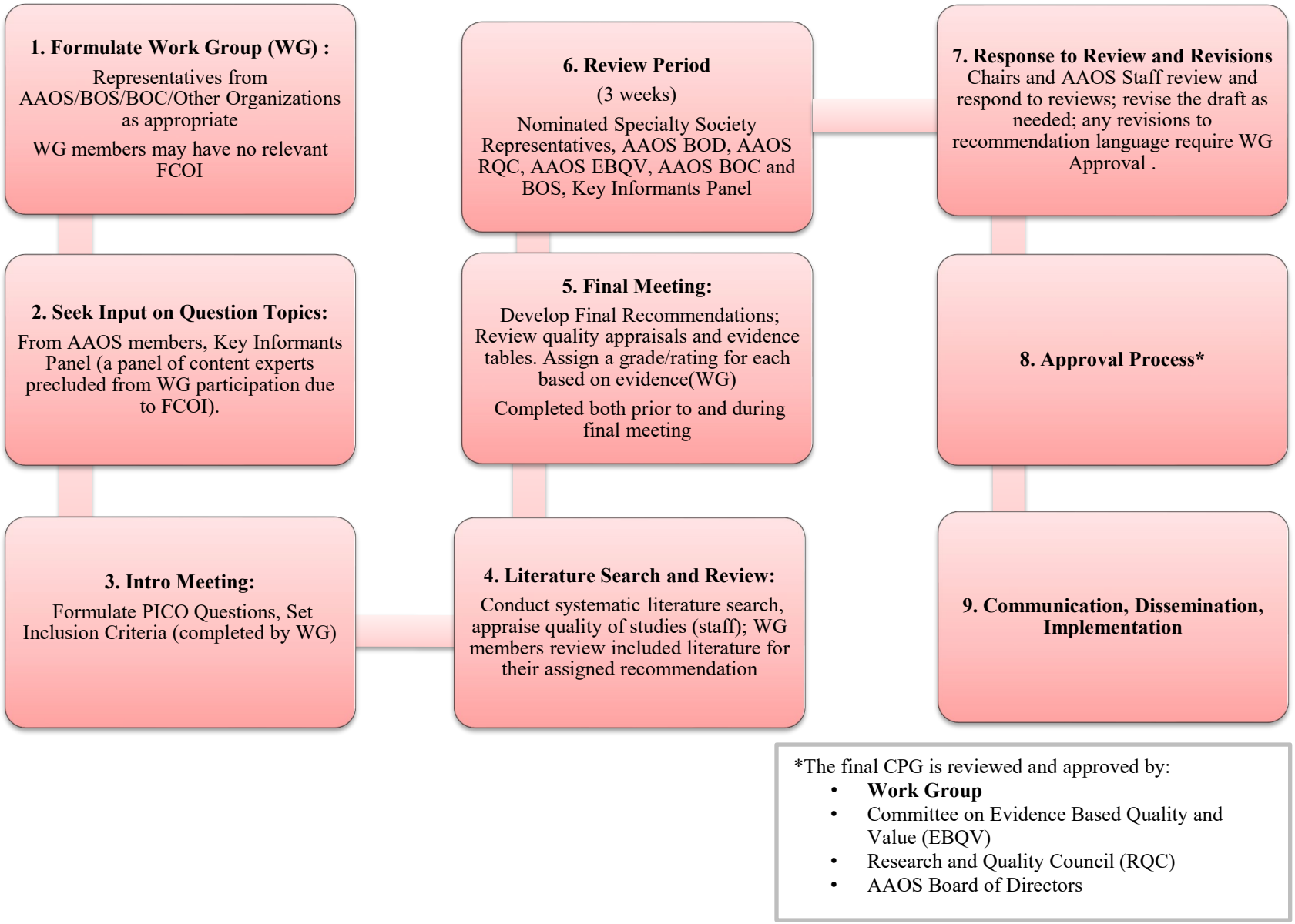
First Steps to Constructing a CPG

1a. Nominate Clinical Practice Guideline (CPG)– Open to all via electronic survey.

1b. Select a topic – The AAOS Committee on Evidence-Based Quality and Value (EBQV) prioritizes the nominated topics alongside guidelines due for an update via an electronic topic ranking form.

1c. The EBQV Committee decides which of the high priority topics should move forward as a guideline (follow CPG Process listed on page 4)

Clinical Practice Guideline Process Flowchart



*The final CPG is reviewed and approved by:

- **Work Group**
- Committee on Evidence Based Quality and Value (EBQV)
- Research and Quality Council (RQC)
- AAOS Board of Directors

Detailed Methodology

Formulating PICO Questions

The clinician work group begins their work on CPGs by constructing a set of PICO questions. These questions specify the patient population of interest (P), the intervention of interest (I), the comparisons of interest (C), and the patient-oriented outcomes of interest (O). They function as questions for the systematic review, not as final recommendations or conclusions. Once established, these *a priori* PICO questions cannot be modified until the final guideline work group meeting.

Study Selection Criteria

A priori article inclusion criteria are constructed for all CPGs. These criteria are our "rules of evidence" and articles that did not meet them are, for the purposes of this guideline, not evidence.

To be included in our CPGs an article had to meet the following criteria:

Work Group Defined Criteria

1. Study must be of an <*enter disease topic of interest*> injury or prevention thereof.
2. Study must be published in or after <work group selects date, not to precede 1966> for surgical treatment, rehabilitation, bracing, prevention and MRI
3. Study must be published in or after <*work group selects date, not to precede 1966> for* x rays and nonoperative treatment
4. Study must be published in or after <*work group selects date, not to precede 1966*> for all others non specified
5. Study should have 30 <*work group may choose to increase/decrease the sample size if justified*> or more patients per group
6. For surgical treatment a minimum of *N* days/months/year (refer to PICO questions for detailed follow up duration)
7. For nonoperative treatment a minimum of *N* days/months/year (refer to PICO questions for detailed follow up duration)
8. For prevention studies a minimum of *N* days/months/year (refer to PICO questions for detailed follow up duration)

Standard Criteria for all CPGs

- Article must be a full article report of a clinical study.
- Retrospective non-comparative case series, medical records review, meeting abstracts, meta-analyses, systematic reviews, historical articles, editorials, letters, and commentaries are ***excluded***. Bibliographies of meta-analyses and systematic reviews will be examined to ensure inclusion of all relevant literature.
- Confounded studies (i.e., studies that give patients the treatment of interest AND another treatment) are ***excluded***.
- Case series studies that have non-consecutive enrollment of patients are ***excluded***.
- Controlled trials in which patients were not stochastically assigned to groups AND in which there was either a difference in patient characteristics or outcomes at baseline AND where the authors did not statistically adjust for these differences when analyzing the results are ***excluded***.

- All studies evaluated as “very low quality” will be ***excluded***.
- Composite measures or outcomes are ***excluded*** even if they are patient-oriented.
- Study must appear in a peer-reviewed publication
- For any included study that uses “paper-and-pencil” outcome measures (e.g., SF- 36), only those outcome measures that have been validated will be included
- For any given follow-up time point in any included study, there must be ≥ 50% patient follow-up (if the follow-up is >50% but <80%, the study quality will be downgraded by one Level)
- Study must be of humans
- Study must be published in English
- Study results must be quantitatively presented
- Study must not be an in vitro study
- Study must not be a biomechanical study
- Study must not have been performed on cadavers
- We will only evaluate surrogate outcomes when no patient-oriented outcomes are available.

Best Evidence Synthesis

AAOS CPGs include only the best available evidence for any given patient- oriented outcome addressing a PICO question. Accordingly, we first include the highest quality evidence for any given outcome if it was available (see Methods for Evaluating Evidence for more information). In the absence of two or more occurrences of an outcome at this quality, we consider outcomes of the next lowest quality until at least two or more occurrences of an outcome has been acquired. For example, if there were two ‘moderate’ quality occurrences of an outcome that addressed a recommendation, we do not include ‘low’ quality occurrences of this outcome. A summary of the evidence that met the inclusion criteria but was not best available evidence is created for each CPG and can be viewed by recommendation within each document’s appendix.

Recommending for or Against a Procedure

The guideline work group considers the procedure of interest and comparison procedure when recommending or not recommending a procedure for clinical use. If the procedure of interest results in outcomes that are similar to the comparison procedure, the work group may recommend both procedures due to no statistical difference in outcomes. If the procedure of interest results in outcomes that are not statistically different than a placebo or no procedure, the work group may recommend against the procedure of interest, because it adds no measurable benefit to a patient’s outcomes.

Minimally Clinically Important Improvement

Wherever possible, we consider the effects of treatments in terms of the minimally clinically important difference (MCID) in addition to whether their effects are statistically significant. The MCID is the smallest clinical change that is important to patients and recognizes the fact that there are some treatment-induced statistically significant improvements that are too small to matter to patients. However, there were no occurrences of validated MCID outcomes in the studies included in this clinical practice guideline.

When MCID values from the specific guideline patient population are not available, we use the following measures listed in order of priority:

MCID/MID
PASS or Impact
Another validated measure Statistical Significance

Literature Searches

We begin the systematic review with a comprehensive search of the literature. Articles we consider were published prior to the start date of the search in a minimum of three electronic databases; PubMed, EMBASE, and the Cochrane Central Register of Controlled Trials. The medical librarian conducts the search using key terms determined from the guideline development group's PICO questions.

A CQV methodologist will review/include only primary literature but will supplement the electronic search with a manual search of the bibliographies of secondary literature sources, such as systematic reviews, as available. The methodologist will then evaluate all recalled articles for possible inclusion based on the study selection criteria and will summarize the evidence for the guideline work group who assist with reconciling possible errors and omissions.

A study attrition diagram is provided in the appendix of each document that details the numbers of identified abstracts, recalled and selected studies, and excluded studies that were evaluated in the CPG. The search strategies used to identify the abstracts is also included in the appendix of each CPG document.

Methods for Evaluating Evidence

All articles included from our systematic literature search are appraised by a CQV methodologist for quality. Depending on the type of study encountered, different quality forms are utilized to determine the quality rating of a study. The quality forms used by staff are described below.

Randomized Study Appraisal Form

Resources used to develop the Randomized Quality Appraisal System:

- Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org. The following domains are evaluated to determine the study quality of randomized study designs.

- Guyatt, G. H., Oxman, A. D., Sultan, S., et al. (2011). GRADE guidelines: 9. Rating up the quality of evidence. Journal of Clinical Epidemiology, 64(12), 1311–1316.

Randomized Study Quality Appraisal Questions

- Random Sequence Generation
- Allocation Concealment
- Blinding of Participants and Personnel
- Incomplete Outcome Data
- Selective Reporting
- Other Bias

Upgrading Randomized Study Quality Questions

- Is there a large magnitude of effect?
- Influence of All Plausible Residual Confounding
- Dose-Response Gradient

Randomized Study Design Quality Key

High Quality Study	<2 Flaw
Moderate Quality Study	≥2 and <4 Flaws
Low Quality Study	≥4 and <6 Flaws
Very Low Quality Study	≥6 Flaws

Combined Prognostic/Observational Study Appraisal Form

Resources used to develop the Observational Intervention Study Quality Appraisal System:

- Sterne JAC, Higgins JPT, Elbers RG, Reeves BC and the Development group for ROBINS-I. Risk of Bias In Non-randomized Studies of Interventions (ROBINS-I): detailed guidance, updated 12 October 2016. Available from http://www.riskofbias.info [accessed july 2018

- Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence–study limitations (risk of bias). J Clin Epidemiol 2011;64:407–15.

- Guyatt, G. H., Oxman, A. D., Sultan, S, et al. (2011). GRADE guidelines: 9. Rating up the quality of evidence. Journal of Clinical Epidemiology, 64(12), 1311–1316.

Combined Prognostic/Observational Appraisal Form

The following questions are used to evaluate the study quality of prognostic/observational study designs. Note that all non-randomized intervention studies begin the appraisal process at “low quality” due to design flaws inherent in observational studies. The quality is downgraded to very low if there are 4 full flaws out of the 7 domains. Bulleted items are all or nothing for scoring; if any bullet point flaw is met, the article gets the associated flaw.

1. Patient Spectrum: Did the study exclude a subset of patients that would make the sample less representative of the full patient population of the PICO question?
    - Low Risk of Bias (0 pt. reduction)
        - No important patient exclusions that would make the sample unrepresentative
    - Unclear (0.5 pt. reduction)
        - The inclusion criteria are not adequately described, or it is unclear if exclusion of patient subsets would bias study results
    - High Risk of Bias (1 pt. reduction)
        - Study excluded a subset of patients relevant to the PICO question

2. Intervention/Variable measurements: Was the treatment/prognostic factor adequately measured to limit the risk of misclassification bias
    - Low Risk of Bias (0 pt. reduction)
        - Treatment/prognostic factor status was recorded prospectively
    - Unclear (0.5 pt. reduction)
        - Treatment/prognostic factor status was obtained retrospectively through a hospital database, registry, or medical records
    - High Risk of Bias (1 pt. reduction)
        - The treatment/prognostic factor was measured by asking patients to retrospectively recall and report their treatment or prognostic status to investigators.

3. Outcome Measurement: Is there a high risk that outcomes were measured inaccurately?
    - Low Risk of Bias (0 pt. reduction)
        - Outcomes were evaluated prospectively, and evaluators were blind to treatment/prognostic status
    - Unclear (0.5 pt. reduction)
        - The study had prospective outcome data collection, but it is unclear if evaluators were blind to treatment/prognostic status
    - High Risk of Bias (1 pt. reduction)
        - Outcomes were measured retrospectively, such as through a hospital database, registry or medical records, or a prospective study explicitly states evaluators were unblinded. (Prospective data collected from normal course of patient care, would be considered unblinded)

4. Confounding: Were all relevant confounders either similar between groups at baseline or were they controlled for with multivariate modeling or matching, and were the confounding variables adequately measured?
    - Low Risk of Bias (0 pt. reduction)
        - Study is prospective and controls for all relevant confounders or all confounders were similar between groups at baseline
    - Unclear (0.5 pt. reduction)
        - Study controls for confounders or all confounders were similar at baseline, but confounding variables are measured retrospectively, such as from a registry, hospital database or patient records
    - High Risk of Bias (1 pt. reduction)
        - Study does not control for all relevant confounders, regardless of if data collection is retrospective or prospective.

5. Statistical Analysis - variables, assumptions, and models: Was statistical analysis performed and reported adequately (i.e., assumptions, variable recording and interpretation, model structure, etc.)?
    - Low Risk of Bias (0 pt. reduction)
        - All interventions/variables discussed in the methods section are reported in the results section of the paper
        - Model variables were established/reported a priori (statistical model was not built in a stepwise fashion nor after univariate screening, where a series of univariate tests were used to select variables for the final statistical model)

- Study explicitly states that statistical assumptions were tested, or a non-parametric test was used.
- No continuous variables (e.g., age) were converted to categorical variables by splitting them by their means, medians, quartiles, or other data dependent cut points (refers to author designated cut-offs based on study data, and not accepted medical cutoffs e.g., BMI)

- Unclear Risk of Bias (0.5 pt. reduction)
  - Unclear reporting; not enough information provided to determine
- High Risk of Bias (1.0 pt. reduction)
  - Some interventions/variables discussed in the methods section are not reported in the results section The study does not explicitly state that they considered statistical assumptions and a parametric test was used.
  - Stepwise models or univariate screening was used to select variables into the final statistical model
  - Data dependent cut points were used to split one or more continuous variables into categories
  - There are <10 patients per variable in the statistical model (eg 40 patients analyzed, but multivariate model contained 5+ variables)

6. Missing Data: Are there low rates of loss to follow-up, missing outcomes, treatment status, and confounder variable data?
   - Low Risk of Bias (0 pt. reduction)
     - No more than 20% of eligible patients were excluded due to loss to follow up, missing outcome, treatment status or confounder data or the missing data was accounted for using some form of imputation (e.g., ITT analysis; Last outcome carried forward; Imputation)
   - Unclear (0.5 pt. reduction)
     - Unclear if more than 20% of eligible patients were excluded due to loss to follow up, missing outcome, treatment status or confounder data and unclear if missing data was accounted for using some form of imputation (e.g., ITT analysis; Last outcome carried forward; Imputation)
   - High Risk of Bias (1 pt. reduction)
     - More than 20% of eligible patients were excluded due to loss to follow up, missing outcome, treatment status or confounder data and missing data was not accounted for using some form of imputation (e.g., ITT analysis; Last outcome carried forward; Imputation)

7. Reporting -Outcomes: Were results for all outcomes specified in the methods section also reported in the results section?
   - Low Risk of Bias (0 pt. reduction)
     - Results for all outcomes discussed in the methods section are reported in the results section of the paper
   - Unclear (0.5 pt. reduction)
     - It is unclear from the methods section what outcomes the study intended to evaluate
   - High Risk of Bias (1 pt. reduction)
     - Some outcomes listed in the methods section of the study are not reported in the results section

Upgrading Prognostic/Observational Study Quality Questions

- Is there a large magnitude of effect?
- Influence of All Plausible Residual Confounding
- Dose-Response Gradient

Prognostic/Observational Study Design Quality Key

Moderate Quality Study	Only if upgrade criteria met
Low Quality Study	< 4 flaws
Very Low Quality Study	≥4 flaws

Diagnostic Study Appraisal Form

Resources used to develop the Diagnostic Quality Appraisal System:

- Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. Ann Intern Med. 2011 ;155:529–536

Diagnostic Study Quality Appraisal Questions

*The following types of bias are considered when evaluating **study quality** for diagnostic studies*

- Patient selection/spectrum bias
    - Consecutive or random sample of patients were enrolled, and inappropriate exclusions were avoided

- Index test bias
    - Index test was interpreted without knowledge of reference test results
    - Test positivity thresholds were prespecified, instead of using the optimal threshold that was determined after the start of the study.

- Reference standard bias
    - Reference standard is likely to correctly classify the target condition
    - Reference standard is interpreted without knowledge of index test results

- Flow and timing
    - Disease status is unlikely to have changed between when the index and reference tests were performed
    - All patients received verification with the same reference standard
    - All patients recruited into the study were included in the final analysis

*The following questions are asked to determine the **applicability/generalizability** of the diagnostic study*

- Are there concerns that patients in study or clinical settings are not generalizable to the full population or clinical settings relevant to the review question?

- Are there concerns that variations in test technology, execution, or interpretation in different clinical settings may affect diagnostic accuracy?

- Is there concern that the target condition as defined by the reference standard does not match the condition asked about in the PICO question?

High Quality Study	<1 Flaw
Moderate Quality Study	≥1 and <2 Flaws
Low Quality Study	≥2 and <3 Flaws
Very Low Quality Study	≥3 Flaws

Quality of Evidence

The process for determining quality of evidence also considers the following domains:

1. **Consistency/heterogeneity** of results between studies. Do the results vary widely between studies in terms of strength of effect and direction of effect?
2. **Indirectness/generalizability**
    a. Indirectness of patient population. Is the population of the studies applicable to general clinical practice?
    b. Indirectness of interventions. That is, are the interventions in the studies applied in the same way as they would be in general clinical settings, and are they available in all clinical settings?
    c. Indirectness of outcomes. Are all relevant outcomes and follow up times evaluated in the included studies? Or does the evidence only consist of surrogate or intermediate outcomes?
3. **Imprecision of results**. Are effect estimates from the studies, or the pooled effect in a meta-analysis, highly imprecise, with very wide confidence intervals? For example, if confidence intervals include what might be considered a strong effect, even though the outcome is not statistically significant, the quality of evidence would be downgraded.
4. **Tradeoff between benefits and harms.** A moderate or strong recommendation can only be made if the benefits of implementing the recommendation clearly outweigh the harms. For example, if multiple high quality RCTs showed that a treatment improves patient reported outcomes, but also greatly increased the risk of serious adverse events, the quality of evidence would be downgraded to limited.

The physician work group also applies GRADE's Evidence to Decision (EtD) framework to determine the final strength of recommendation (appendix 1). The EtD form (appendix 1) is filled out as applicable by the work group member(s) assigned to the PICO question before the meeting, and is used to facilitate discussion about the following issues that may warrant a lower or higher recommendation grade:

1. Certainty of evidence
2. Is there uncertainty over how people value the main outcomes?
3. Are the desirable effects large?
4. Are the undesirable effects small?
5. Are the desirable effects large relative to the undesirable effects?
6. Are resources required to implement the recommendation small?
7. Are the incremental costs small relative to the net benefits?
8. Is the recommendation likely to be acceptable to key stakeholders?
9. Is the option feasible to implement?

The EtD allows the workgroup to apply their clinical experience to determine the feasibility and appropriateness of CPG recommendations in real world health care settings. The EtD is a balance between the rigid evidence rules of the systematic review and the real-world clinical expertise of the work group, which allows for a richer perspective, and results in recommendations that are more appropriate. The EtD allows the workgroup to consider possible harms of implementation that may not be well studied in RCTs. It also provides a structured and transparent way to describe how they arrived at the final strength of recommendation and allows readers to be better able to determine how the recommendation applies to their own clinical setting. For example, if high quality studies show that

a new imaging modality is good at diagnosing joint infection, but the technology is very expensive and is unlikely to be available at most community medical centers. After filling out the EtD form, the work group may decide that the recommendation should be downgraded from high to moderate because it is not feasible to implement in smaller hospitals due to cost. A reader from a small community hospital is now better able to decide if the recommendation can be implemented at his/her own institution. Conversely, a reader from a high-volume academic medical center that has the imaging technology may decide to apply the recommendation in his/her clinical practice. Furthermore, if low quality studies show that not performing a certain intervention, yields exponentially higher mortality in patients, the work group may decide that the recommendation should be upgraded from limited to moderate because of the potential to prevent loss of life.

Defining the Strength of the Recommendations

Judging the quality of evidence is only a steppingstone towards arriving at the strength of a CPG recommendation. The strength of recommendation also takes into account the quality, quantity, and the trade-off between the benefits and harms of a treatment, the magnitude of a treatment's effect, and whether data exists on critical outcomes.

Strength of recommendation expresses the degree of confidence one can have in a recommendation. As such, the strength expresses how possible it is that a recommendation will be overturned by future evidence. It is very difficult for future evidence to overturn a recommendation that is based on many high quality randomized controlled trials that show a large effect. It is much more likely that future evidence will overturn recommendations derived from a few small retrospective comparative studies. Consequently, recommendations based on the former kind of evidence are given a "strong" strength of recommendation and recommendations based on the latter kind of evidence are given a "limited" strength.

To develop the strength of a recommendation, AAOS staff first assigned a preliminary strength for each recommendation that took only the final quality and the quantity of evidence (see Table 1). The recommendations can be further downgraded or upgraded based on the GRADE and Evidence to Decision framework criteria described above.

Defining Quality of Evidence and Strength of the Recommendations

Each CPG recommendation will have two grades assigned to it. The first is the quality of evidence (QOE) of the studies supporting a recommendation. The QOE reflects the degree of confidence in the evidence after applying the GRADE methodological criteria. It also expresses how possible it is that a recommendation will be overturned by future evidence. It is very difficult for future evidence to overturn a recommendation that is based on many high quality randomized controlled trials that show a large effect. It is much more likely that future evidence will overturn recommendations derived from a few small retrospective comparative studies.

Strength of Recommendation (SOR) incorporates quality of evidence along with the domains of the Evidence to Decision Framework to arrive at a final grade. For example, if there is consistent high-quality evidence from multiple RCTs showing a beneficial effect of a treatment, but the treatment was associated with serious risks or would likely have serious barriers to implementation is some health care settings, then the overall strength of recommendation would be downgraded. In such a case, QOE would be high due to the high methodological quality of the studies, but the strength of recommendation would be moderate or limited due to concerns addressed in the evidence to decision framework. The reverse is also possible, where there are only low-quality studies available, and it is found that not performing the treatment would result in higher mortality. This set of circumstances could thereby result in the QOE being graded as limited but the strength of recommendation being upgraded to moderate.

Table 1. Strength and Quality Descriptions

Statement Strength	Evidence Quality	Statement Description	Strength Visual
Strong	High*	Evidence from two or more "High" quality studies with consistent findings recommending for or against the intervention. Or Rec is upgraded using the EtD framework.	
Moderate	Moderate*	Evidence from two or more "Moderate" quality studies with consistent findings or evidence from a single "High" quality study recommending for or against the intervention. Or Rec is upgraded or downgraded using the EtD framework.	
Limited	Low*	Evidence from two or more "Low" quality studies with consistent findings or evidence from a single "Moderate" quality study recommending for or against the intervention. Or Rec is downgraded using the EtD framework.	
Consensus*	Very Low, or Consensus*	Evidence from one "Low" quality study, no supporting evidence, or Rec is downgraded using the EtD framework. In the absence of sufficient evidence, the guideline work group is making a statement based on their clinical opinion.	

*Unless statement was upgraded or downgraded in strength, using the EtD Framework.

Applying the Recommendations to Clinical Practice

To increase the practicality and applicability of the guideline recommendations in this document, the information listed in Table 3 provides assistance in interpreting the correlation between the strength of a recommendation and patient counseling time, use of decision aids, and the impact of future research

Table 3. Clinical Applicability: Interpreting the Strength of a Recommendation

Strength of Recommendation	Patient Counseling (Time)	Decision Aids	Impact of Future Research
Strong	Least	Least Important, unless the evidence supports no difference between two alternative interventions	Not likely to change
Moderate	Less	Less Important	Less likely to change
Limited	More	Important	Change possible/anticipated
Consensus	Most	Most Important	Impact unknown

Voting on the Recommendations

The recommendations and their strength are voted on by the CPG work group members during the final meeting. If disagreement between the guideline work group occurs, there was further discussion to see whether the disagreement(s) could be resolved. Recommendations were approved and adopted in instances where a majority (60%) of the guideline work group voted to approve. Any recommendation strength upgrade or downgrade based on the Evidence to Decision framework requires a super majority (75%) approval of the work group.

Statistical Methods

Analysis of Intervention/Prevention Data
When possible, the AAOS CQV Unit recalculates the results reported in individual studies and compiles them to answer the recommendations. The results of all statistical analysis by the AAOS CQV Unit are conducted using SAS 9.4. SAS is used to determine the magnitude, direction, and/or 95% confidence intervals of the treatment effect. For data reported as means (and associated measures of dispersion) the mean difference between groups and the 95% confidence interval is calculated and a two-tailed t-test of independent groups is used to determine statistical significance. When published studies report measures of dispersion other than the standard deviation the value is estimated to facilitate calculation of the treatment effect. In studies that report standard errors or confidence intervals, the standard deviation is back-calculated. In some circumstances statistical testing is conducted by the authors and measures of dispersion is not reported. In the absence of measures of dispersion, the results of the statistical analyses conducted by the authors (i.e. the p-value) are considered as evidence. For proportions, we report both the proportion and percentage of patients that experienced an outcome. The variance of the arcsine difference is used to determine statistical significance. P-values < 0.05 are considered statistically significant.

When the data are available, meta-analyses using the random effects method of DerSimonian and Laird are performed. A minimum of three studies are required for an outcome to be considered for meta-analysis. Heterogeneity is assessed with the I-squared statistic. Meta-analyses with I-squared values less than 50% are considered as evidence. Those with I-squared larger than 50% are not considered as evidence for inclusion in guidelines. All meta-analyses are performed using SAS 9.4. The arcsine difference is used in meta-analysis of proportions. In order to overcome the difficulty of interpreting the magnitude of the arcsine difference, a summary odds ratio is calculated based on random effects meta-analysis of proportions and the number needed to treat (or harm) is calculated. The standardized mean difference is used for meta-analysis of means, and magnitude is interpreted using Cohen's definitions of small, medium, and large effect.

Review Period

Following the final meeting, the CPG draft undergoes a 3-week review period for additional input from external content experts. Written comments are provided on the structured review form. All reviewers are required to disclose their conflicts of interest. Relevant specialty societies are solicited for nominations of individual reviewers approximately six weeks before the final meeting. The review period is announced as it approaches, and others interested are able to volunteer to review the draft. The guideline work group approves the final draft of the guideline prior to dissemination. Some specialty societies (both orthopaedic and non-orthopaedic) ask their evidence-based practice (EBP) committee to provide review of the guideline. The organization is responsible for coordinating the distribution of the materials and consolidating their comments onto one form. The chair of the external EBP committees provides disclosure of committee members' conflicts of interest (COI) and manages the potential conflicts.

Again, the AAOS asks for comments to be assembled into a single response form by the specialty society and for the individual submitting the review to provide disclosure of potentially conflicting

interests. The review stage gives external stakeholders an opportunity to provide evidence-based direction for modifications that they believe have been overlooked. Since the draft is subject to revisions until its approval by the AAOS Board of Directors as the final step in the guideline development process, confidentiality of all working drafts is essential.

The CPG is also provided to members of the AAOS Board of Directors (BOD), members of the Research and Quality Council (RQC), members of the Board of Councilors (BOC), and members of the Board of Specialty Societies (BOS) and members of the Committee on Evidence-Based Quality and Value (EBQV) for review and comment. The CPG is automatically forwarded to the AAOS BOD, RQC and EBQV so that they may review it and provide comment prior to being asked to approve the document. Members of the BOC and BOS are solicited for interest and provided with the confidential draft upon request. Based on these bodies, over 200 commentators have the opportunity to provide input into each CPG.

The guideline project manager and CQV director review and draft the initial responses to comments that address methodology and/or process. These responses are then reviewed by the chair and co-chair, who respond to questions concerning clinical practice and techniques. All proposed changes to recommendation language as a result of the review period are based on the evidence and require full work group approval. Final revisions are summarized in a report that is provided alongside the guideline document throughout the remainder of the approval processes and final publication.

The AAOS believes in the importance of demonstrating responsiveness to input received during the review process and welcomes the critiques of external specialty societies. Following final approval of the guideline, all individual responses are posted on our website https://www.aaos.org/quality/quality-programs/ with a point-by-point reply to each non-editorial comment. Reviewers who wish to remain anonymous notify the AAOS to have their names de-identified; their comments, our responses, and their COI disclosures are still posted.

Structured Review Electronic Form

Reviewers are asked to read and review the draft of the CPG with a particular focus on their area of expertise. Their responses to the answers below are used to assess the validity, clarity, and accuracy of the interpretation of the evidence.

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
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 guideline's target audience is clearly described.	○	○	○	○	○
4. There is an explicit link between the recommendations and the supporting evidence.	○	○	○	○	○
5. Given the nature of the topic and the data, all clinically important outcomes are considered.	○	○	○	○	○
6. The patients to whom this guideline is meant to apply are specifically described.	○	○	○	○	○
7. The criteria used to select articles for inclusion are appropriate.	○	○	○	○	○
8. The reasons why some studies were excluded are clearly described.	○	○	○	○	○
9. All important studies that met the article inclusion criteria are included.	○	○	○	○	○
10. The validity of the studies is appropriately appraised.	○	○	○	○	○
11. The methods are described in such a way as to be reproducible.	○	○	○	○	○
12. The statistical methods are appropriate to the material and the objectives of this guideline.	○	○	○	○	○
13. Important parameters (e.g., setting, study population, study design) that could affect study results are systematically addressed.	○	○	○	○	○
14. Health benefits, side effects, and risks are adequately addressed.	○	○	○	○	○
15. The writing style is appropriate for health care professionals.	○	○	○	○	○
16. The grades assigned to each recommendation are appropriate.	○	○	○	○	○

Please provide a brief explanation of both your positive and negative answers in the preceding section. If applicable, please specify the draft page and line numbers in your comments. Please feel free to also comment on the overall structure and content of the Guideline.

Would you recommend these guidelines for use in clinical practice?*

- Strongly Recommend
- Recommend
- Would Not Recommend
- Unsure

Additional Comments:

The AAOS CPG Approval Process

This final CPG draft must be approved by the AAOS Committee on Evidence Based Quality and Value, the AAOS Research and Quality Council, and the AAOS Board of Directors. These decision-making bodies are described in the Appendix of each guideline. Their charge is to approve or reject its publication by majority vote, not suggest modifications to the content of the documents.

Revision Plans

CPGs represent a cross-sectional view of current treatment and may become outdated as new evidence becomes available. They will be revised in accordance with new evidence, changing practice, rapidly emerging treatment options, and new technology. Additionally, they will be updated or withdrawn in five years.

CPG Dissemination Plans

The primary purpose of CPGs is to provide interested readers with full documentation about not only our recommendations, but also about how we arrived at those recommendations.

To view all AAOS published CPG recommendations in a user-friendly website, please visit www.orthoguidelines.org

Or download the OrthoGuidelines app from Google Play or Apple Stores.

Shorter versions of the CPGs are available in other venues. Publication of most CPGs is announced by an Academy press release, articles authored by the CPG work group and published in the Journal of the American Academy of Orthopaedic Surgeons, and articles published in *AAOS Now*. Most CPGs are also distributed at the AAOS Annual Meeting in various venues such as on Academy Row.

Selected CPGs are disseminated by webinar, an Online Module for the Orthopaedic Knowledge Online website, Radio Media Tours, Media Briefings, and by distributing them at relevant Continuing Medical Education (CME) courses and at the AAOS Resource Center.

Other dissemination efforts outside of the AAOS will include submitting the CPGs to the ECRI Guidelines Trust, Guidelines International Network Library, and distributing the guideline at other medical specialty societies' meetings.

Appendix 1-Evidence to Decision (EtD) Framework Form

Recommendations:

Table 1 Evidence to Decision Framework

Criteria	Detailed considerations	Judgements	Research Evidence	Additional considerations/physician input
What is the overall certainty of the evidence?	Study quality, imprecision, indirectness, inconsistency, risks/harms balance	☐Low ☐Moderate ☐High		
Is there uncertainty in how much people value the main outcomes?	-	☐ Important uncertainty or variability ☐ Possibly important uncertainty or variability ☐ Probably no important uncertainty or variability ☐ No important uncertainty or variability		
Are the desirable anticipated effects large?	-	☐ No ☐ Probably no ☐ uncertain ☐ Probably yes ☐ Yes ☐ Varies		

Criteria	Detailed considerations	Judgements	Research Evidence	Additional considerations/physician input
Are the undesirable anticipated effects small?	-	☐ No ☐ Probably no ☐ uncertain ☐ Probably yes ☐ Yes ☐ Varies		
Are the desirable effects large relative to undesirable effects?	Do the benefits clearly outweigh the risks or is there a balance of benefits and harms	☐ No ☐ Probably no ☐ uncertain ☐ Probably yes ☐ Yes ☐ Varies		
Are the resources required small?	-	☐ No ☐ Probably no ☐ uncertain ☐ Probably yes ☐ Yes ☐ Varies		
Is the incremental cost small relative to the net benefits?	-	☐ No ☐ Probably no ☐ uncertain ☐ Probably yes ☐ Yes ☐ Varies		

Criteria	Detailed considerations	Judgements	Research Evidence	Additional considerations/physician input
Is the option acceptable to key stakeholders?	-are there any stakeholders who wouldn't accept risk to benefit ratio, the costs, the importance of outcomes. -would anyone morally object to intervention (in regard to ethical principles such as no maleficence, beneficence, or justice)? -would intervention effect people's autonomy	☐ No ☐ Probably no ☐ uncertain ☐ Probably yes ☐ Yes ☐ Varies		
Is the option feasible to implement?	-is intervention sustainable -any barriers limiting the feasibility of implementing recommendation	☐ No ☐ Probably no ☐ uncertain ☐ Probably yes ☐ Yes ☐ Varies		