AAOS
American Academy of Orthopaedic Surgeons

AAOS Clinical Practice Guideline Rapid Update Methodology

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

American Academy of Orthopaedic Surgeons

Contents

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

Overview of Clinical Practice Guideline (CPG) Methodology

The AAOS understands that only high-quality clinical practice CPG are credible, and we go to great lengths to ensure the integrity of the evidence analyses. The AAOS addresses bias beginning with the selection of CPG work group members. Applicants must participate in the AAOS Orthopaedic Disclosure Program, with enhanced disclosures pertaining to financial conflicts of interest, and CMS OpenPayments data is reviewed as well. 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 view the full AAOS Clinical Practice Guideline Methodology please visit:
https://www.aaos.org/quality/research-resources/methodology/

Rapid Update Eligibility

AAOS Clinical Practice Guidelines (CPGs) are eligible for update 5 years following publication. Rapid updates will be conducted when both (1) the original scope of the CPG (i.e. PICO questions and the treatments included, thereof) have continued relevance at the time of consideration of update and (2) substantial new evidence has not been published since the date of the last literature search.

When considering a guideline for a full versus a rapid update, the EBQV committee may evaluate the following before a decision is made:

1. Breadth of new publications: The committee will be provided with descriptive statistics outlining the results of a preliminary literature search of which includes all possibly relevant articles published after the original guideline's literature search. Understanding the 10% rule of abstract return to included articles (i.e. on average, 10% of abstracts are recalled for full text review and 10% of full texts are found relevant to the PICO questions of interest and included in the final guideline), the committee will evaluate if there is precedent to proceed with a rapid review (i.e. if the amount of new literature is small enough to benefit from a rapid review in place of convening a full work group and if the state of evidence does not substantially change the recommendations from the prior CPG – saving direct and indirect resources for other projects).
2. Prior to approving a rapid update, the EBQV committee may ask staff liaisons to reach out to the former chairs of the guideline to garner input regarding:

    - The preliminary literature search (see #1)
    - Novel, important therapies/techniques/procedures/clinical advances of which were absent from the previous guideline which would warrant important new PICO questions
    - Relevancy of prior CPG to current standards of care specifically with regards to any substantial change in clinical treatment or knowledge which would warrant consideration of a full update or recommendation to sunset the original CPG
    - Recommendations on moving forward with a rapid review in lieu of convening a full work group

Rapid Update Procedural Steps

The AAOS Committee on Evidence-Based Quality and Value oversees the Rapid Update Process; any topic eligible for a rapid update must be approved by the Committee. Following approval, AAOS staff methodologists work under the guidance of Committee content experts via the following procedural steps:

1. The AAOS Medical Librarian runs the original search strategy to identify novel literature beginning from the end date of the last search up to the current date.
2. Committee content experts complete a review of the abstracts discovered via the updated literature search and highlight the abstracts for which appear particularly relevant to one or more of the PICO questions. The AAOS Medical Librarian subsequently recalls the full text of the identified articles/abstracts for review.
3. AAOS Staff Methodologists review the full text articles identified by the committee experts (as outlined in #2) and performs a secondary review of the abstracts to ensure all relevant literature is evaluated.
    a. Committee content experts provide clinical input as necessary for study interpretation and relevancy to the PICO question as full-text articles are reviewed.
4. After full text article review has been completed (i.e. articles have either been identified for inclusion or exclusion), AAOS staff methodologists will perform quality appraisal and data extraction for all included articles.
5. Committee content experts, along with AAOS staff methodologists, will compare the updated literature search data with the original guideline recommendations and provide feedback regarding possible recommendation language and or strength of recommendation changes to the larger EBQV committee.
    - The recommendations are upgraded or downgraded as warranted when sufficient evidence is found. Recommendations without any new evidence and/or with insufficient evidence to warrant an upgraded strength remain the same. Recommendations are eligible for update only if the quality of additional literature warrants a strength change; the GRADE Evidence-to-Decision Framework will not be applied when determining updated recommendation strength.
    - Recommendation language is only updated to reflect the new strength of recommendation utilizing standard language stems that denote the strength of the recommendation. Recommendation language is not adjusted or editorialized apart from the language stem and remains as written in the original guideline.
    - The rationales as written by the original work group will remain unedited in most cases, save for inclusion of one to two sentences regarding the updated literature. If newly discovered evidence requires major revisions, AAOS staff will reach out to former chairs to ensure accuracy of revision language.
6. Guidelines published as rapid updates will be clearly marked with a disclaimer explaining the update process. A summary of changes made will be provided and the original recommendations will be listed in an appendix. Original work group member names will be included in the appendix, but clearly delineated as authors of the original guideline, not the rapid update. Authorship credit will be listed as AAOS Committee on Evidence-based Quality and Value, along with contributors, as warranted (e.g. former guideline chairs).
7. The Review Period reports from the original guideline will remain live on www.aaos.org and clearly marked as pertaining to the original guideline. As no substantive edits are made during a rapid update, a second external review will not be performed.
8. Guidelines published as rapid updates adhere to the same approval process as all AAOS Quality Products and require sequential approval from the AAOS Committee on Evidence-Based Quality and Value, AAOS Council on Research and Quality, and AAOS Board of Directors.
9. If a guideline is published as a rapid update, the accompanying Appropriate Use Criteria (AUC) may also be renewed and republished without reconvening writing and voting panels if the Committee content experts determine it is still appropriate as written.
10. Upon Board of Directors Approval, the updated guideline will be disseminated on OrthoGuidelines,

aaos.org, in the GIN Library, the ECRI Guidelines Trust, Headline News Now, AAOS Social Media, and AAOS Now (if warranted).

Financial Conflict of Interest (FCOI)

Clinical Practice Guidelines undergoing a Rapid Update are eligible only for recommendation strength upgrades based on any new supporting evidence that increases the strength of evidence as defined by AAOS methodology (e.g., studies published after the date of the original guideline's last literature search). If new evidence is discovered during the rapid update process of which the reviewers believe may change any of the original recommendations or prior decisions by the workgroup to upgrade or downgrade the recommendation based upon the Evidence to Decision framework, the rapid update process will cease and a full update with a newly formed clinician work group will commence. The overall scope (e.g., PICO questions and inclusion criteria) and final language of the entire document (e.g., recommendations, rationale, and introduction) are determined by the original work group and are not modifiable via the Rapid Updates methodology; content expert members of the Committee on Evidence-Based Quality and Value are asked to validate that the original recommendations are still supported by any newly discovered evidence. The original work group is fully vetted for any FCOI via the Clinical Practice Guidelines methodology and the Rapid Update subgroup, despite their lack of editorial freedom, is required to disclose via the AAOS disclosure process. The final decision regarding adoption of guidelines through the Rapid Update process will be discussed and voted on by the Evidence Based Quality and Value Committee, the Research and Quality Council, and the AAOS Board of Directors.

Detailed Methodology for CPG Literature Review

Study Selection Criteria

A priori article inclusion criteria are constructed for all CPGs. To be included in AAOS CPGs an article had to meet the following criteria:

<u>Work Group Defined Criteria</u>

1. Study must be of an <*disease topic of interest in original CPG*> injury or prevention thereof.
2. Study must be published after date of prior search.
3. Study should have <*number of patients as specified by the original work group*> or more patients per group
4. Study should have a minimum of <*as specified by the original work group*> days/weeks/months/years follow-up time

<u>Standard Criteria for all CPGs</u>

- 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.

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 the 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

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.

Observational Study Design Quality Appraisal Questions

The following questions are used to evaluate the study quality of 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. They can only be upgraded to moderate quality in rare cases if they meet one of the criteria for upgrading listed below.

- Does the strategy for recruiting participants into the study differ across groups?
    - Enrolled new users of a treatment rather than current users of a treatment
    - Patients were not excluded for outcomes that occurred after the start of the study.

- Is treatment status measured/recorded accurately?
    - measured at the same time treatment started and did not rely on patient recall.
- Did the authors fail to take important confounding variables into account in the design and/or analysis (e.g., through matching, stratification, interaction terms, multivariate analysis, or other statistical adjustment such as instrumental variables)?
- Is there a high risk that outcomes were measured inaccurately?
    - Measured the same way in all patients
    - Blinded outcome evaluation or outcome was objective and couldn't be influenced by lack of blinding
- Are there low rates of missing outcome, treatment status, and confounder variable data OR were the rates and/or reasons for missing data similar between groups?
- Were results for all outcomes, statistical analyses and patient populations specified in the methods section, also reported in the results section?
    - No selective reporting of outcomes
    - Results from all statistical models described in methods section are reported
    - Study was not a subgroup analysis of a previously published study
    - No conflict of interest

Upgrading Observational Study Quality Questions

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

Observational Study Design Quality Key

Moderate Quality Study	Only if upgrade criteria met
Low Quality Study	< 3 flaws
Very Low Quality Study	≥3 flaws

Prognostic Study Appraisal Form

Resources used to develop the prognostic quality appraisal form

- Hayden JA, Coˆte´ P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Ann Intern Med. 2006;144:427-37.
- Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing Bias in Studies of Prognostic Factors. Ann Intern Med. 2013 ;158:280–286.

- Hayden JA, Côté P, Steenstra IA, Bombardier C. QUIPS-LBP Working Group. Identifying phases of investigation helps planning, appraising, and applying the results of explanatory prognosis studies. J Clin Epidemiol. 2008;2:552–560. doi: 10.1016/j.jclinepi.2007.08.005.

Prognostic Study Quality Appraisal Questions

Univariate studies that do not control for confounding factors automatically start at low quality and can be further downgraded to very low quality if there are additional study limitations. Only confirmatory studies can start out at high quality. Confirmatory studies are designed to determine if a prognostic factor is independently associated with outcomes after controlling for known confounding factors. If a study uses a univariate analysis to screen statistically significant variables into the final multivariate model, or uses stepwise regression modeling techniques, then the study will be rated no higher than moderate quality due to the exploratory nature of these analyses. According to Hayden (2008), exploratory studies constitute weaker prognostic evidence because "it should be recognized that results from multiple studies in this exploratory phase of investigation often have widely varying results, as spurious associations are common, and real effects are sometimes missed, and some associations are present in one population but not in another."

Prognostic questions:

- What study design was used
    - Univariate with no matching or multivariate modeling to control for confounding factors
    - Multivariate or matched study design to account for confounding factors
- Was the spectrum of patients studied for this prognostic variable representative of the patient spectrum seen in actual clinical practice?
- Was loss to follow up unrelated to key characteristics?
- Was the prognostic factor of interest adequately measured in the study to limit potential bias?
- Was the outcome of interest adequately measured in study participants to sufficiently limit bias?
- Were all important confounders adequately measured in study participants to sufficiently limit potential bias?
- Was the statistical analysis appropriate for the design of the study, limiting potential for presentation of invalid results?
    - Adequate number of patients and events per variable in the model
    - Avoidance of exploratory design (no use of stepwise models or univariate screening)
    - Statistical assumptions tested

Prognostic Study Design Quality Key

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

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?

Diagnostic Study Design Quality Key

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

Strength of Evidence

The process for determining strength 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 strength 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 strength of evidence would be downgraded to limited.

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.

Table 1. Strength of Recommendation Descriptions

Strength of Recommendation	Overall Strength Of Evidence	Description of Evidence quality	Strength Visual
Strong	**Strong or Moderate**	Evidence from two or more "High" quality studies with consistent findings for recommending for or against the intervention. Or Rec is upgrade from Moderate using the EtD framework	
Moderate	**Strong, Moderate or Limited**	Evidence from two or more "Moderate" quality studies with consistent findings, or evidence from a single "High" quality study for recommending for or against the intervention. Or Rec is upgraded or downgraded from Limited or Strong using the EtD framework.	
Limited	**Limited or Moderate**	Evidence from one 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 from Moderate using the EtD Framework.	
Consensus	**No reliable evidence**	There is no supporting evidence, or higher quality evidence was downgraded due to major concerns addressed in the EtD framework. In the absence of reliable evidence, the guideline work group is making a recommendation based on their clinical opinion.	

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 2. 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

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.