MSTS Clinical Practice Guideline Methodology*

***Adopted and approved by the MSTS executive committee 12/20/2023**

Contents

To view all MSTS published clinical practice guidelines recommendations, visit MSTS.org

Overview of Clinical Practice Guideline (CPG)

The primary goal of a clinical practice guideline (CPG) is to improve quality of care with secondary goals of determining gaps in research and identifying recommendations for future directions. The Musculoskeletal Tumor Society (MSTS) understands that only high-quality CPGs are credible, and we go to great lengths to ensure the integrity of our evidence analyses and to reduce the potential for bias in creating our recommendations.

The MSTS 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 within one year of starting the guidelines. 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.

Prior to beginning a CPG, the MSTS Guideline and Evidence-Based Medicine Committee nominates Clinical Practice Guideline (CPG) topics and prioritizes topics based on burden of disease, treatment costs, practice variation, assumed research availability, public or political demand, and resources available to complete the topic. We will avoid topics where there already exists broad agreement and uniformity of care among providers. We will focus on topics where there are variations in clinical practice and adequate evidence to make recommendations. For many topics in musculoskeletal oncology, there will be no randomized clinical trials or other high-quality evidence. However, this is not an insurmountable barrier to the production of high-quality CPGs if guideline developers follow a rigorous, trustworthy, and transparent process.

The process of MSTS CPG development incorporates the benefits from clinical physician expertise as well as the statistical and methodological knowledge and interpretation of non-conflicted methodologists. The process also includes an extensive review process offering the opportunity for many 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 regarding interpretation of the evidence.

CPGs are prepared by physician guideline development groups (GDG) with the assistance of guideline methodologists. As the physician experts, the GDG defines the scope of the CPG by creating PICO Questions (i.e. population, intervention, comparison, and outcome) that direct the literature search. Once this is completed, individual GDG members are assigned one or more PICO questions of which they are responsible for assisting the methodologists when evaluating the clinical relevance of published research articles for inclusion in the guideline, reviewing the final literature report for their assigned PICO questions, preparing the preliminary recommendation language and EtD form (see Appendix I), and acting as the content experts during the final meeting discussion(s).

The MSTS will use a formal consensus methodology in producing its guidelines. The details of our consensus methodology are described below. The guiding principle of our process is to prevent, to the greatest extent possible, the introduction of GDG members' individual biases into guideline recommendations and strengths.

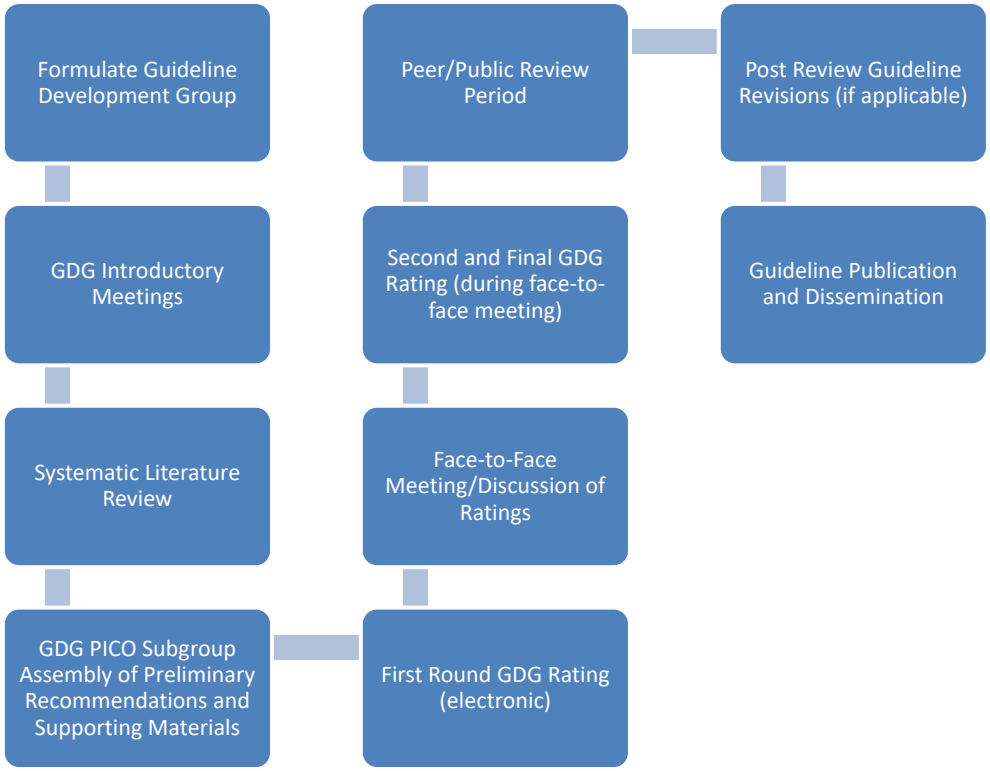
When necessary, members of the GDG also provide content help, search terms and additional clarification for the medical librarian, who creates and executes the search(es). The supporting group of methodologists (and, as necessary, clinical experts) review all abstracts, recall pertinent full-text articles for review and evaluate the quality of studies meeting the inclusion criteria. With the help of the GDG, the methodologists also abstract, analyze, interpret, and summarize the relevant data for each PICO question and prepare the initial data findings for the GDG. Upon completion of the systematic reviews, the GDG translates the data findings into action-oriented recommendations. These Action Statements will be written clearly enough that adherence can be measured. Each statement will be accompanied with a structured rationale. For additional details about how the work group formulates and votes on action statements and recommendation strengths, please see <INSERT LINK to relevant section in this document>.

After the full guideline draft is finalized by the GDG, the document is submitted for peer review to all MSTS members and specialty societies of whom have interest in the topic under study. Subsequent to peer review, the CPG draft may be edited in response to the review submissions. Thereafter, the draft CPG is sequentially approved by the MSTS Guideline and Evidence-Based Medicine (GEBM) Committee and the MSTS Executive Committee. Once approved, the CPG is submitted to an appropriate journal for publication. All MSTS CPGs are reviewed every five

years by the MSTS GEBM, and either updated or retired.

MSTS is committed to eradicating health care disparities amongst different physician and patient demographic or clinical subgroups. As part of this goal, MSTS guidelines will incorporate recommendations that address disparities and actively recruit underrepresented minority physician experts and/or laypeople into guideline workgroups.

Figure #1: Clinical Practice Guideline Process Flowchart



Detailed Methodology

Choosing Guideline Topics

Guideline topics are selected by the MSTS GEBM committee. The committee may solicit input from the Society members via electronic survey or other tools. To prioritize topics for development of CPGs, the committee will assign scores for the following domains:

1. High Disease Burden
2. Cost
3. Practice Variation
4. Sufficient Research
5. Substantial Public or Political Demand
6. Feasibility

A high average score across the above domains suggests a higher priority should be placed for that topic.

Guideline Development Group (GDG)

Each guideline development group (GDG) is led by two GDG co-chairs and one GDG oversight chair. MSTS will utilize multidisciplinary clinician workgroups when applicable. The primary responsibility of the GDG co-chairs are to lead introductory calls/meetings to ensure that the work group stays on task and within the parameters of the guideline process when developing PICO questions and guideline recommendations. They are also tasked with final writing, reviewing, and editing of the CPG manuscript to be submitted for publication. GDG co-chairs are appointed by the MSTS Guidelines and Evidence-based Medicine Committee (GEBM). Co-chair appointments are made based on GDG member's expertise with topic under study, as well as their previous experience with the development of clinical practice guidelines and/or evidence-based initiatives. The GDG oversight chair is a non-voting member of the GDG whose primary charge is to ensure that the GDG adheres to all guideline methodologies throughout the process. The oversight chair is typically a member of the MSTS GEBM Committee, although this is not mandatory.

Recruiting GDG members

GDG members will generally be recruited from MSTS members via public notification, for example, via email to the membership. If more members express interest than there are GDG positions available, initial selection will be via lottery. Those not selected will be eligible to join the GDG if, after analysis of conflicts of interest (see below), any of those initially selected are deemed ineligible. In order to maintain as broad participation as possible, each Guideline project will be preceded by a new call for volunteers. In cases where Guideline workgroup will be multidisciplinary, MSTS will request the respective societies to nominate volunteers.

Financial Disclosures and Conflict of Interest

GDG applicants have the obligation to disclose all potentially relevant financial or nonfinancial conflicts of interests. All MSTS GDG applicants are required to participate in the AAOS Orthopaedic Disclosure Program. AAOS policy requires detailed financial information for each conflict noted by an applicant or member of a CPG Work Group. This detailed information is not available to the public, is accessible to a limited number of AAOS and MSTS staff and will be used solely for purposes of resolution of conflict-of-interest issues. MSTS staff review the AAOS Orthopaedic Disclosure Program as well as data available via CMS OpenPayments. For non-MSTS GDG members, MSTS staff will review disclosures submitted to their respective societies or request disclosures directly from the members.

If a possible relevant financial or nonfinancial conflict of interest is deemed to exist, MSTS staff will contact the applicant for additional information, which may include specific details and clarification as to the relevancy to the CPG topic. If a relevant financial or nonfinancial conflict exists, the applicant is invited to divest the conflict. An applicant denied participation in a CPG Work Group due to perceived relevant financial or nonfinancial conflict may appeal the decision, after which the CPG Oversight Chair shall discuss the situation with the MSTS GEBM Committee and will make a final determination on the applicant's eligibility.

GDG Eligibility

GDG members must not have any relevant financial conflicts of interest related to their respective CPG for one year prior to development, during development of the guideline, and for one-year post-approval of the guideline. Financial conflicts of interest will be deemed

significant, and preclude participation on the GDG, if relevant financial conflicts exceed $1,000 cumulatively. GDG members will be asked to verbally disclose any new financial or nonfinancial conflicts of interest at the introductory meeting and at subsequent meetings. MSTS staff will generate a general disclosure report that will be disseminated at the meetings and published in the final CPG. 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. The MSTS disclosure and conflict of interest processes have been developed with the goal of transparent and appropriate decision-making. This Guidance Document may be modified as other MSTS policies and procedures are developed

Formulating PICO Questions

Guideline Development Group Introductory Meeting

Commented [YD1]: Do you want to specify that this may be 2 meetings? Also do we want to say that there will be other people involved? Like the librarians.

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.

PICO Workgroups

Once the PICO questions have been formulated, smaller workgroups are selected to address one or more PICO questions, preferably grouped by subject matter. Each PICO workgroup should consist of 1-3 individuals, and one GDG member can serve on more than one PICO workgroup. Each PICO workgroup will designate a leader for recording, communicating, and finalizing their recommendation and recommendation strength. As will be detailed below, PICO workgroups draft guideline recommendations (“action statements”), propose recommendation strengths, and perform Evidence-to-Decision (EtD) scoring. However, they cannot vote on their own recommendations and will recuse themselves during GDG discussions of their recommendations.

Commented [YD2]: I would specify that there will be a leader in each workgroup

Study Selection

Standard Criteria for all CPGs

1. Article must be a full article report of a clinical study.
2. Medical records review, meeting abstracts, 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.
3. Confounded studies (i.e., studies that give patients the treatment of interest AND another treatment) are ***excluded***.
4. Composite measures or outcomes are ***excluded*** even if they are patient-oriented.
5. Study must appear in a peer-reviewed publication
6. Study must be of humans
7. Study must be published in English
8. Study results must be quantitatively presented
9. Study must not be an in vitro study
10. Study must not be a biomechanical study
11. Study must not have been performed on cadavers
12. Surrogate outcomes are evaluated only when no patient-oriented outcomes are available.

Project Dependent 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.

The following criteria may be adjusted by the GDG prior to beginning the systematic literature review, depending on the topic under study:

1. Study must be published in or after date selected by GDG, not to precede 1966
2. Study should have at least 10 or more patients per group, modifiable by the GDG
3. For surgical treatment, the minimum follow up will be specified by the GDG and/or relevant PICO question
4. For nonoperative treatment the minimum follow up will be specified by the GDG and/or

relevant PICO question

5. For prevention studies the minimum follow up will be specified by the GDG and/or relevant PICO question

Literature Searches

The systematic review begins with a comprehensive search of the literature. Articles we consider must be 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 medical librarian and methodologist will search, review/include primary literature, evaluate all recalled, full-text articles for possible inclusion based on the study selection criteria, and will summarize the evidence for the guideline development group of whom 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 recommendations are accompanied by a three tier "Strength of Evidence", describing the quality and consistency of the *aggregate* evidence supporting the recommendation, and by a two tier "Strength of Recommendation", which will depend largely, but not entirely, on the Strength of Evidence. The preliminary level of evidence for any given recommendation is created by staff research analysts with the aid of GDG members who are assigned to the PICO question of interest and is based on the individual quality assessments of each included study. However, the aggregate strength of evidence determination, as well as the final strength of recommendation is determined based on the GDG's assessment of the evidence and other factors reviewed within the EtD Framework (e.g., cost, feasibility of implementation, risk/benefits assessment, etc.).

Level of evidence

All articles included in the systematic literature search are appraised by a methodologist for quality (see Appendix II). Depending on the type of study encountered, different quality forms are utilized to determine the quality rating of a study. The quality forms used during this evaluation are described below.

Along with these formal appraisal evaluations, the process for determining level 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. 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 do not include what might be considered a strong effect, even though the outcome is statistically significant, the level 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 level of evidence would be downgraded to limited.

Recommending for or Against a Procedure

The guideline work group considers the procedure of interest and comparison procedure, if

available/applicable, when recommending or not recommending a procedure for clinical use. If the procedure of interest results in outcomes that are like 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. All recommendations are crafted utilizing the best available comparative effectiveness research, supplemented by expert opinion, and considerations regarding risk/benefits, feasibility, and cost (both cost of treatment and cost of resources utilized to effectively administer treatment).

Defining the Strength of the Recommendations

Judging the quality of evidence is only a steppingstone towards arriving at the strength of a CPG Action Statement. The strength of recommendation also considers 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.

To develop the strength of a recommendation, methodologists and clinician PICO subgroup members first assign a preliminary strength for each recommendation that takes only the final quality and the quantity of evidence. The recommendations can be further downgraded or upgraded based on the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) EtD framework (EtDF).

The EtD form (Appendix 1) is filled out as applicable by the PICO work group members assigned to the question before the meeting and is used to facilitate discussion about the following issues that may warrant a lower or higher recommendation grade. In order to make scoring of EtD form more consistent and transparent, individual EtDF items have been assigned numeric scores. These scores are added and recommendations to increase or decrease the strength of evidence are offered based on pre-selected thresholds. Below are the items considered in the EtD form. Please refer to Appendices 1 and 2 for details and weighting.

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 EtDF 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 Framework 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 EtDF 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, in a situation where 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, it may not be applicable in all settings. 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, if low quality studies show that not performing a certain intervention yields 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. When recommendation strength is modified using the EtDF, the primary factors from the Framework will be identified in the Rationale accompanying the recommendation.

To improve transparency and consistency, physician workgroup members will utilize an EtD form with numerical scores associated with the individual items. The scores will be summed and predetermined score thresholds will be used to suggest whether a recommendation should be

upgraded or downgraded.

The action statements created by the PICO workgroups are then assigned a strength of recommendation by the same workgroup based on a two-tiered system. The two potential strengths are “strong” and “limited/conditional”. A strong recommendation is defined as one in which the literature is very clear as to the benefit (or detriment if there is a recommendation against) of a specific intervention, such that it suggests and models a standard of care all treating surgeons should be providing and is unlikely to be overturned with future evidence. However, a limited or conditional recommendation is one in which the literature is unclear for or against an intervention, the risks and benefits are balanced, or the intervention seems appropriate or feasible only in certain circumstances (eg. certain patient populations, certain stages of disease, or in the hands of surgeons with specific experience or skill sets). In simple terms, a strong recommendation is one in which all surgeons should be recommending to their patients in that given situation, while a limited or conditional recommendation is one that should qualify the conditions where it may be met and involves more of a full PARQ (**P**roposed treatment, **A**lternatives to it, **R**isks, and time for patient **Q**uestions) conference with the patient, such that their values and preferences are taken into consideration in shared decision-making.

Creating and Voting on Recommendations

Guideline recommendations (“Action Statements”) are developed and finalized through a two-step process. In the first round, the small PICO workgroups draft Action Statements based on the summary of included evidence. Using the strength of evidence of the included evidence, each PICO workgroup arrives at an aggregate strength of included evidence. PICO workgroups also complete the EtD form for their questions. Using the aggregate strength of evidence and EtD form scores, they propose a recommendation strength.

After completion of the literature review and the report of findings and prior to the final meeting discussions, the GDG is charged with developing action statements for their assigned PICO questions, along with a summary of any included evidence, and clinical considerations outlined in the EtD form (see Appendix I) for each of the recommendations. The evidence and clinical considerations document will describe the underlying logic or justification for a given recommendation.

Once each PICO workgroup has completed its work, all recommendations along with their supporting evidence are submitted to the entire GDG for the first-round review. As above, PICO workgroup members cannot vote on their own recommendations and will recuse themselves during GDG discussions of those recommendations. Finalizing and approving the action statements and strength of recommendation is then done via a modified Delphi consensus process as outlined below.

First Round Review and Rating of Recommendations

Commented [YD3]: I remember there being a lot of back-and-forth about "what happened to this paper? What happened to that paper? Why isn't it in the review?" where are we addressing unresolved questions about lit review? At the first round? Or earlier or later?

Prior to the final meeting, the GDG members not assigned to the PICO question of interest anonymously rate their agreement with each of the recommendation statements, based on the recommendations adherence to available literature, considering the other clinical considerations outlined in the EtD form (e.g., risk/harms, benefits, cost, feasibility of implementation, etc.). Agreement is measured via an electronic survey of which allows eligible voting GDG members the opportunity to rate the clinical accuracy of the recommendation(s) using a 5-point Likert scale (i.e. 1 = Strongly disagree with the recommendation as written, 5 = Strongly agree with the recommendation as written). Additionally, open responses are collected from eligible voting GDG members to assist the GDG members assigned to the PICO question/recommendation under consideration with refining the suggested recommendation before the final meeting discussions.

Any preliminary recommendations which lacked agreement (a priori agreement threshold defined by GDG at the introductory meeting), will be returned to the PICOT workgroups for revision or elaboration on the statement prior to the final meeting discussion(s) where the resulting, possibly revised, recommendations and round one voting results will be discussed in greater detail. All preliminary recommendations meeting the agreement threshold will be considered approved by the group for inclusion in the CPG without further need for edits/discussion.

Defining Agreement Thresholds for Voting

"Approval" of recommendations during the development of clinical practice guidelines requires, at minimum, a supermajority (i.e., two-thirds or 67% for limited/conditional recommendations and four-fifths or 80% for strong recommendations) of votes in the "4" or "5" range (i.e., "agree"

or “strongly agree with the recommendation, as written”). GDGs may choose to continue revising a recommendation even if supermajority is reached to refine the recommendation with the aim of achieving consensus of the entire GDG. All approvals will be recorded in the final guideline document to ensure transparency to the end user.

Face-to-Face Meeting Discussion(s)

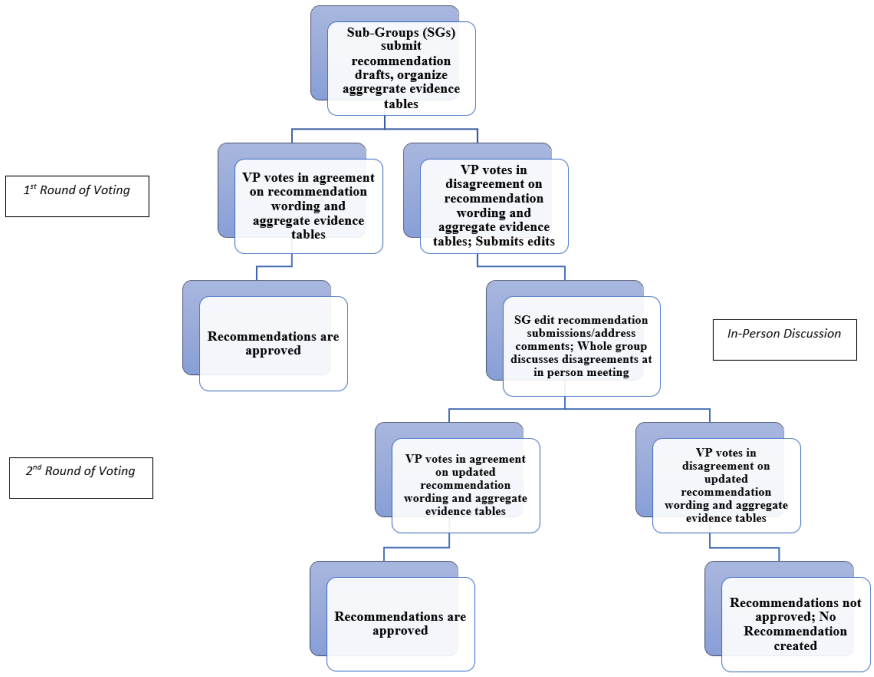
For those recommendations which did not meet the predefined agreement thresholds during the first round of rating, the GDG will discuss in greater detail the rationale behind the results of the first round ratings (i.e. what issues the group had with the preliminary recommendations and/or the supporting materials), the revised recommendation(s) or the PICO subgroup’s rationale for not revising the recommendation(s), and possible suggestions for improving the clinician and evidential accuracy of the recommendation(s).

The discussion will be jointly led by a methodologist. The PICO workgroups will recuse themselves during this discussion. However, if the GDG may ask questions of the PICO workgroup members, if needed, during the discussion. Where appropriate, these interactions will be mediated by the methodologist.

Second Round Review and Rating of Recommendations

After the face-to-face discussion(s) and subsequent revisions to the recommendations (if applicable), the eligible voting GDG members will re-rate the recommendations of which lacked agreement during the first round of rating. Any recommendations of which do not surpass the a priori definition of “agreement” due to concerns, of which are more than minor semantic/grammatical errors, will be excluded from the final clinical practice guideline.

Figure #2. Voting Flowchart



Rationale

Each Action Statement (i.e., recommendation) will be accompanied by a structured rationale. Rationales will be written by the responsible PICOT workgroups and modified as needed following review by the full GDG.

The Rationale will follow the following structure:

1. "*Included Evidence*": List of included articles/evidence and the quality score for each. Limit to top 10 items, if appropriate.
2. "*Recommendation Strength Modification*": If recommendation was upgraded or downgraded via EtD framework, the top 3 items from the EtDF that influenced the decision, otherwise "N/A".

3. "*Summary of Evidence*": Summary of findings from the included evidence that led to the recommendation.
4. "*Future Research*": Further context and research recommendations the GDG feels would be useful to readers, otherwise "N/A"

Peer and Public Review

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.

To guide who participates, the GDG identifies specialty societies at the introductory meeting. O*rganizations*, not *individuals*, are specified. The specialty societies are solicited for nominations of individual reviewers to all for approximately six weeks of review before closure of the review period. The review period is announced as it approaches, and others interested can volunteer to review the draft. The co-chairs of the guideline work group review the draft of the guideline prior to dissemination.

Some specialty societies (both orthopaedic and non-orthopaedic) ask their evidence-based practice (EBP) committee or equivalent to provide review of the guideline. The organization is responsible for coordinating the distribution of our materials and consolidating their comments onto one form. The chair of the external EBP committees provides disclosure of their conflicts of interest (COI) and manages the potential conflicts of their members.

The MSTS asks for comments to be assembled into a single response form (Appendix III) 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 MSTS Executive Committee 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 MSTS Executive Committee, relevant external medical organizations, and the broader MSTS membership for review. Based on these bodies,

over 200 commentators should have the opportunity to provide input into each CPG.

The co-chairs of the GDG and the methodologists draft the initial responses to comments that address methodology and the co-chairs also organize initial responses to questions concerning clinical practice and techniques. All comments received and the initial drafts of the responses are also reviewed by all members of the guideline development group. All proposed changes to recommendation language stemming from the review period must be based on the evidence and must be approved by the GDG. 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 MSTS 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 (http://www.MSTS.org/guidelines) with a point-by-point reply to each non-editorial comment. Reviewers who wish to remain anonymous notify the MSTS 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.

The MSTS CPG Approval Process

This final CPG draft must be approved by the MSTS Guidelines and Evidence Based Medicine Committee and the MSTS Executive Committee. 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 MSTS published CPG recommendations, please visit http://www.MSTS.org/guidelines.

Shorter versions of the CPGs are available in other venues. Publication of CPGs is announced by an MSTS press release, articles authored by the CPG work group, and published in appropriate medical journals as indicated based on topic and allied society collaboration.

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

Publication and Authorship

Principles

1. Traditional authorship principles used in high level journal writing shall apply (see below)
2. Authorship of published CPG guidelines should be as inclusive as the prospective journal will allow
3. Authorship should reflect the intellectual property, work effort, and leadership of the CPG working group

Authorship

1. Each CPG working group should have two co-chairs, A and B, with first and last authorship pre-determined
2. Co-chairs should rotate in an equitable fashion for each CPG, without repetition of A or B positions when other interested parties exist
3. Co-authors in middle authorship shall be listed by alphabetical order and shall comprise

the remainder of the working group

4. Following these will be listed the participating methodologists, in alphabetical order
5. Non-working group members such as external advisory or voting panels will not qualify for authorship
6. For journals with very limited authorship (e.g. JAAOS), the two co-chairs A and B will be the principal authors, with one middle author optional based on significant manuscript production work and intellectual contribution. All other contributing members to be listed in acknowledgements.
7. For PICO questions which have such clinical significance as to be pulled out of the umbrella CPG topic and published as stand-alone recommendations, the two specific PICO leaders will move into first/last authorship roles in lieu of the CPG co-chairs

MSTS CPG Authorship Guidelines

An "author" will be considered someone who has made substantive intellectual contributions to the work, and must meet the following criteria:

1. An author must contribute substantially to the conception or design of the work, or the acquisition, analysis, and interpretation of data, or drafted the work or substantively revised it.
2. AND have approved the submitted version and any modified version;
3. AND agreed to be personally accountable for the author's own contributions and all other aspects of the work in ensuring that questions related to the accuracy or integrity of the work are appropriately investigated and resolved.

The MSTS Guidelines and Evidence-Based Medicine Committee will not consider "gift authorship" for department heads, other chairs, professional mentors of committee members and valid authors of the work, industry representatives or individuals offering material support for the work, ghostwriters, or facilitators of technical completion of the work when all above criteria are not met. Such individuals may be credited in acknowledgement.

Lead authorship will be considered for someone who has performed the central tasks of the project, including the first complete draft of the manuscript. The lead author is responsible for

ensuring that all other authors meet the requirements for authorship as well as the integrity of the work itself. The lead author should serve as the corresponding author.

Co-authors must each review and approve the manuscript.

Relevant references

1) International Committee of Medical Journal Editors. "Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals." icmje.org. Updated December 2019.
2) Leopold, SS. Editorial: Research is a Team Sport: Updated Authorship Guidelines for CORR. *Clin. Orthop. Relat. Res.*. 2013;471:701-2.
3) Harvard University Guidelines on Authorship and Acknowledgement
4) Yale University Guidance on Authorship in Scholarly or Scientific Publications

MSTS CPG Timeline Targets

	Timeline	Targeted Task Completion
Phase I: Project Launch	6 Weeks	Introductory Meeting -Obtain disclosure of interest forms Vetting of Clinical Topic -Relevance -Absence of current evidence-based guidelines -Topic consistent with MSTS objectives -Adequacy of systematic evidence for review MSTS Board Approval Establish Workgroups and Key Questions
Phase II: Review of Evidence	4 Months	Systematic Evidence Report Screening and Selecting Relevant Studies Compilation of Studies Recruitment/Assignment of Collaborators on Voting Panels
Phase III: Consensus Process	3 Months	Delphi Survey -Review and feedback on evidence Strength of Recommendation Strength of Evidence
Phase IV: Write-Up	2 Months	Writing Assignments -Project Outline Drafting of Subsections -Key questions -Recommendations -Supporting text/evidence
Phase V: Evaluation and Submission	6 Weeks	Completed Draft Review and Approval by Workgroup Members Authorship Assignments Formatting Submission

Phase VI: Manuscript preparation and submission	6 months	Month 12-15 (3 mo): writing and manuscript assembly Month 15-18 (3 mo): submission, edits, publication

.

Appendix I - Evidence to Decision (EtD) Framework

Table 1 Evidence to Decision Framework

Criteria	Detailed considerations	Judgements (points)
What is the overall certainty of the evidence?	Study quality, precision, directness, consistency, level of evidence	☐ Low (3) ☐ Moderate (4) ☐ High (5)
What is the value and importance of the outcomes to clinical practice?		☐ None (0) ☐ Low (2) ☐ Moderate (3) ☐ High (5)
What is the magnitude of the desired effect?		☐ None (0) ☐ Low (2) ☐ Moderate (3) ☐ High (5)
What is the magnitude of undesirable effects/complications?		☐ High (0) ☐ Moderate (1) ☐ Low (2) ☐ None (3)
Do the benefits outweigh the risks?	Do the benefits clearly outweigh the risks or is there a balance of benefits and harms?	☐ No (0) ☐ Probably No (1) ☐ Uncertain (2) ☐ Probably Yes (3) ☐ Yes (5)
What amount of resources are required to produce the desired effect?	What is the estimated equipment need, space, time, and ability of any institution to provide these needs?	☐ Prohibitive (0) ☐ High (1) ☐ Moderate (2) ☐ Minimal (3) ☐ None (5)
What is the cost to produce the desired effect	What is the estimated monetary cost?	☐ Prohibitive (0) ☐ High (1) ☐ Moderate (2) ☐ Minimal (3) ☐ None (4)
Is the intervention/outcomes 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 (regarding ethical principles such as no maleficence, beneficence, or justice)? -Would intervention effect people's autonomy?	☐ No (0) ☐ Probably No (1) ☐ Uncertain (2) ☐ Probably Yes (4) ☐ Yes (5)
Is the intervention feasible to implement?	-Is intervention sustainable? -Any barriers limiting the feasibility of implementing recommendation?	☐ No (0) ☐ Probably No (1) ☐ Uncertain (2) ☐ Probably Yes (4) ☐ Yes (5)

Table 2: Quality of Observational Literature Score Thresholds

Upgrade/Downgrade Level	**Threshold**
Increase recommendation strength +2	38-42
Increase recommendation strength +1	31-37
No change in recommendation strength	18-30
Decrease recommendation strength -1	13-17
Decrease recommendation strength -2	3-12

Appendix II – Study Quality Assessment Forms

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 Flaws
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?

o Measured at the same time treatment started and did not rely on patient recall.

☐ The authors took 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).

☐ Outcomes were measured accurately

- 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, Côté 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 referencestandard
  - 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

Appendix III: Peer Review Form

	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: