Drug Use Research & Management Program
Oregon State University, 500 Summer Street NE, E35, Salem, Oregon 97301-1079
Phone 503-947-5220 | **Fax** 503-947-1119

Review Standards and Methods for Quality Assessment of Evidence

Updated: August 2025

REVIEW STANDARDS AND PREFERRED SOURCES OF EVIDENCE

1. The P&T Committee and department staff will evaluate drugs, drug classes, and select non-drug item reviews based on sound evidence-based research and processes widely accepted by the medical profession. These evidence summaries inform the recommendations for management of the preferred drug list (PDL), preferred non-drug items, and clinical prior authorization (PA) criteria. These methods support the principles of evidence-based medicine and will continue to evolve to best fit the needs of the Committee and stay current with best practices.

2. The types of reviews may include, but are not limited to, the following:

Type of Review	Rationale for Review
Abbreviated Drug Review	New drug with evidence only for non-funded condition(s)
Class Literature Scan	Used when limited literature is found which would affect clinical changes in PDL status or PA criteria based on efficacy or safety data (may include new drug formulations or expanded indications if available literature would not change PDL status or PA criteria). Provides a summary of new or available literature, and outcomes are not evaluated via the GRADE methodology listed in **Appendix D**.
New Drug Evaluation (NDE)	Single new drug identified and the PDL class was recently reviewed, or the drug is not assigned to a PDL drug class
Class Review	New PDL class
Class Update	New systematic review(s) and clinical trials identified that may inform change in PDL status or clinical PA criteria in an established PDL class
Class Update with New Drug Evaluation	New drugs(s) or indication(s) also identified (excludes new formulations, expanded indications, biosimilars, or drugs for unfunded indications)
DERP Summary Report	New DERP report which evaluates comparative evidence
Drug Use Evaluation	Analysis of utilization trends in FFS population in order to identify safety issues or inform future policy decisions
Policy Evaluation	Evaluation safety, efficacy, and utilization trends after implementation of a policy to identify areas for improvement

Prior Authorization Update	To evaluate targeted updates to PA criteria based on current policy guidance from the Health Evidence Review Commission, recommendations from the Mental Health Clinical Advisory Group, or expanded labeling from the FDA
Non-Drug Item Review (Specific non-drug items billed through pharmacy [e.g., non-durable medical equipment] as requested by Oregon Health Authority (OHA)	Assessment of products identified by OHA where pharmacy point-of-sale dispensing is considered necessary.

3. The P&T Committee will rely primarily on high quality systematic reviews and randomized controlled trials in making its evidence summary recommendations. High quality clinical practice guidelines and relevant clinical trials are also used as supplementary evidence.

4. Emphasis will be placed on the highest quality evidence available. Poor quality trials, systematic reviews or guidelines are excluded if higher quality literature is available and results offer no additional value. Unless the trial evaluates an outcome or comparison of high clinical importance, individual RCTs with the following study types will be excluded from class updates, class reviews, and literature scans:
    a. Non-comparative, placebo-controlled trials
    b. Non-inferiority trials
    c. Extension studies
    d. Poor quality studies (as assessed in **Appendix A**)

5. Individual drug evaluations rely primarily on high quality RCTs or clinical trials used for FDA approval. Evidence from poor quality RCTs may be included if there is no higher quality evidence available.

6. Phase 2 trials may be considered if there is a compelling reason to include, such as use for FDA approval. Preference will be given for inclusion of applicable phase 3 and 4 trials over earlier phase studies. If fully published, of adequate duration, and with appropriate clinical outcome measures, authors may include phase 2 studies if phase 3 or 4 trials are inadequate or when direct comparative evidence and/or dose response are reported in a comparable population to available phase 3 or 4 studies.

7. The following are preferred sources that provide high quality evidence at this time:

    a. Drug Effectiveness Review Project at Oregon Health & Science University (OHSU)
    b. U.S. Department of Veterans Affairs/Department of Defense
    c. Agency for Healthcare Research and Quality (AHRQ)
    d. Canadian Agency for Drugs and Technologies in Health (CADTH)
    e. National Institute for Clinical Excellence (NICE)
    f. Scottish Intercollegiate Guidelines Network (SIGN)
    g. Oregon Mental Health Clinical Advisory Group (MHCAG)

8. The following types of evidence are preferred and will be considered only if they are of high methodological quality as evaluated by the quality assessment criteria below:

    a. Systematic reviews of randomized controlled trials
    b. Direct comparative randomized controlled trials (RCTs) evaluating clinically relevant outcomes; placebo-controlled studies not related to initial FDA-drug approval or new indications may be considered if likely to impact current policy
    c. FDA review documents
    d. Clinical Practice Guidelines developed using explicit evidence evaluation processes

9. The following types of literature are considered unreliable sources of evidence and will rarely be reviewed by the P&T Committee:

    a. Observational studies, case reports, case series
        i. However, observational studies and systematic reviews of observational studies will be included to evaluate significant safety data beyond the FDA labeling information. Observational studies will only be included when there is not adequate data from higher quality literature.
    b. Unpublished studies (posters, abstracts, presentations, non-peer reviewed articles) that do not include sufficient methodological details for quality evaluation, with the exception of FDA review documents
    c. Individual studies that are poorly conducted, do not appear in peer-reviewed journals, are inferior in design or quality compared to other relevant literature, or duplicate information in other materials under review.
    d. Studies not designed to investigate clinically relevant outcomes
    e. Systematic reviews identified with the following characteristics:
        i. Evidence is of poor or very poor quality
        ii. Evidence is of limited applicability to a US population
        iii. Systematic review does not meet defined applicability criteria (PICOTS criteria) for the topic
        iv. Systematic review is of poor methodological quality as evaluated by AMSTAR II criteria (see **Appendix B**)
        v. Evidence is based on indirect comparisons from network meta-analyses
        vi. Conflicts of interest which are considered to be a "fatal flaw" (see quality assessment for conflicts of interest)
    f. Guidelines identified with the following characteristics:
        i. There is no systematic guideline development method described
        ii. Strength of evidence for guideline recommendations are not provided
        iii. Recommendations are largely based on expert opinion
        iv. Poor methodological quality as assessed in **Appendix C** (AGREE II score is less than 113 points OR modified AGREE II-GRS score is less than 30 points)
        v. Conflict of interest which are considered to be a "fatal flaw" (see quality assessment for conflicts of interest)

10. When assessing efficacy and safety of non-drug items (e.g., devices, digital health technologies), primary emphasis will be on studies that compare the new technology or device to currently available health and social care system technologies or the current standard of care. Included literature for medical technologies and non-drug items will focus on clinical efficacy and safety outcomes measured by relevant outcome indicators. Because the efficacy and safety of medical technologies and non-drug items may be dependent on the training and experience of the user, may be influenced by

organizational factors, and may be influenced by changes in the technology over time, pragmatic usability details (e.g., user experience) will also be included.

QUALITY ASSESSMENT

1. The standard methods used by the DURM faculty to assess quality of evidence incorporated into the evidence summaries for the OHP Pharmacy and Therapeutics Committee are described in detail in **Appendix A-C**.

2. The Cochrane Risk of Bias tool (modified) described in **Appendix A** is used to assess risk of bias (i.e., internal validity) of randomized controlled trials. The quality of non-inferiority trials will be also assessed using the additional criteria for non-inferiority trials in **Appendix A**. Internal validity of clinical trials are graded as poor, fair, or good quality.

3. The AMSTAR II measurement tool is used to assess for methodological quality of systematic reviews and is provided in **Appendix B**. Systematic reviews, meta-analyses or guidance identified from 'best sources' listed in **Appendix B** undergo methodological rigor and are considered to be high quality and are not scored for quality using the AMSTAR II tool.

4. Clinical practice guidelines are considered for inclusion after assessment of methodological quality using the AGREE II global rating scale provided in **Appendix C**. If there are concerns regarding applicability of guidelines to the Medicaid population, the AGREE-REX tool is available for use (https://www.agreetrust.org/resource-centre/agree-rex-recommendation-excellence/).

5. The Patient, Intervention, Comparator, Outcome, and Setting (PICOS) framework is used to assess applicability, or directness, of randomized controlled trials to the OHP population. Detailed guidance is provided in **Appendix A**. Only randomized controlled trials with applicability to the OHP population, as assessed by the PICOS framework, are included in evidence summaries.

6. Emphasis of the review will be on clinically relevant outcomes. The following clinically relevant outcomes are graded for quality: mortality, morbidity outcomes, symptom relief, quality of life, functioning (physical, mental, or emotional), early discontinuation due to adverse events, and severe adverse effects. Surrogate outcomes are considered if directly linked to mortality or a morbidity outcome. Clinically meaningful changes in these outcomes are emphasized.

7. The overall quality of evidence is graded for clinically relevant outcomes of efficacy and harm using the GRADE methodology listed in **Appendix D**. Evaluation of evidence for each outcome of interest is graded as **high**, **moderate**, **low**, or **insufficient**. Final evidence summary recommendations account for the availability and quality of evidence for relevant outcomes and perceived clinical impact on the OHP population.

    a. Evidence grades are defined as follows:
        i. High quality evidence: High confidence that the estimated effects produced in the studies reflect the true effect. Further research is very unlikely to change the estimated effect.
        ii. Moderate quality evidence: Moderate confidence that the estimated effects produced in the studies reflect the true effect. Further research may change the estimated effect.
        iii. Low quality evidence: Limited confidence that the estimated effects produced in the studies reflect the true effect. Further research is likely to change the estimated effect.

iv. Insufficient evidence: Evidence is not available or too limited to permit any level of confidence in the estimated effect.

8. Conflict of Interest
    a. Conflict of interest is a critical component of quality assessment. A conflict of interest is "a set of circumstances that creates a risk that professional judgement or actions regarding a primary interest will be unduly influenced by a second interest." Conflict of interest includes any relationships or activities that could be perceived to have influenced or give the appearance of potentially influencing the literature.
        i. Reference: IOM (Institute of Medicine). 2009. *Conflict of Interest in Medical Research, Education, and Practice.* Washington, DC: The National Academies Press.

    b. Conflict of interest analysis for DURM reviews:
        1. Sources will be excluded due to conflict of interest concerns if they contain one of the "fatal flaws" in **Table 1** below.
        2. If no "fatal flaws" exist, an analysis of the conflicts of interest will be completed and any limitations (examples in **Table 1** below) will be first and foremost discussed in the evidence review.
        3. Conflict of interest is also assessed through the Cochrane risk of bias, AMSTAR II, and AGREE tools (**Appendix A, B, and C**).

Table 1. DURM Conflict of Interest Analysis

Type of literature	"Fatal flaws"	If no "fatal flaws" exist, potential limitations to discuss when including the piece of literature	Other considerations- specific to the type of literature
Randomized controlled trial	• Conflict of interest not documented	• Authors or committee members have significant conflicts of interest • Concerning high dollar amounts of conflicts of interest are documented • Mitigation strategies (described in the article or journal/organization	• Higher risk of bias when the study sponsor is the pharmaceutical manufacturer and is included in data analysis and manuscript writing
Systematic review	• Conflict of interest not documented • Conflict of interest mitigation strategies not documented or are insufficient to mitigate potential bias • *Example mitigation strategies:* persons with potential conflicts of interest are excluded from the assessment or review process, independent second review of articles considered for inclusion in SR that are reviewed first by their own author who is on the SR team		• May consider funding sources or conflicts of interest for both the systematic review and the included studies

Guideline	• Conflict of interest not documented • Chair has a conflict of interest • Conflict of interest mitigation strategies not documented or are insufficient to mitigate potential bias • *Example mitigation strategies:* excluding persons with significant conflict of interest from the review process, recusing members with significant conflict of interest from voting on recommendations or having them leave the room during the discussion	policies) are documented but could be more robust	• Guidelines with "fatal flaws" which are commonly used in practice may be included for clinical context but will not be considered when creating conclusions or recommendations

APPENDIX A. Methods to Assess Quality of Studies.

Table 1. Types of Bias: Cochrane Risk of Bias (modified).

Selection Bias	Selection bias refers to **systematic differences** between **baseline characteristics of the groups** that were compared. The unique strength of proper ***randomization*** is that, if successfully accomplished, it prevents selection bias in allocating interventions to participants. Successful randomization depends on fulfilling several interrelated processes. A rule for allocating patients to groups must be specified, based on some chance (random) process. Furthermore, steps must be taken to secure strict implementation of that schedule of random assignments by preventing foreknowledge of the forthcoming allocations. This process if often termed ***allocation concealment***.
Performance Bias	Performance bias refers to **systematic differences** between groups in the **care provided**, or in exposure to factors other than the interventions of interest. After enrolment, ***blinding participants and investigators/care givers*** will reduce the risk that knowledge of which intervention was received affected the outcomes, rather than the intervention itself. Effective blinding ensures that all groups receive a similar amount of attention, ancillary treatment and diagnostic investigations. Therefore, risk of differences in intervention design and execution, care experiences, co-interventions, concomitant medication use, adherence, inappropriate exposure or migration, cross-over threats, protocol deviations and study duration between study groups are minimized.
Detection Bias	Detection bias refers to **systematic differences** between groups in how **outcomes were assessed**. ***Blinding of outcome assessors*** will reduce the risk that knowledge of which intervention was received, rather than the intervention itself, affected outcome measurement. Blinding of outcome assessors can be especially important for assessment of subjective outcomes (eg, degree of post-operative pain).
Attrition Bias	Attrition bias refers to **systematic differences** between groups in **withdrawals** (exclusions and attrition) from a study. ***Withdrawals*** from the study lead to incomplete outcome data. There are two reasons for withdrawals or incomplete outcome data in clinical trials. ***Exclusions*** refer to situations in which some participants are omitted from reports of analyses, despite outcome data being available to assessors. ***Attrition*** refers to situations in which outcome data are not available.
Reporting Bias	Reporting bias refers to the **selective reporting of pre-specified outcomes**, on the basis of the results. Of particular concern is that statistically non-significant (negative) primary endpoints might be selectively reported while select positive secondary endpoints are over-emphasized. Selective reporting of outcomes may arise in several ways: 1) there can be selective omission of pre-specified outcomes (ie, only some of the pre-specified outcomes are reported); 2) there can also be selection of choice data for an outcome that differs from what was pre-specified (eg, there may be different time points chosen to be reported for an outcome, or different methods used to measure an outcome at the same time point); and 3) there can be selective analyses of the same data that differs from what was pre-specified (eg, use of continuous vs. dichotomous outcomes for A1c lowering, selection from multiple cut-points, or analysis of between endpoint scores vs. change from baseline).
Other Bias	Other sources of bias may be present depending on conflict of interests and funding sources, trial design, or other specific circumstances not covered in the categories above. Of particular concern is how conflicts of interest and funding sources may potentially bias results. Inappropriate influence of funders (or, more generally, of people with a vested interest in the results) is often regarded as an important risk of bias. Information about vested interests should be collected and presented when relevant, with specific regard for methodology that might be been influenced by vested interests and which may lead directly to a risk of bias. Additional sources of bias may result from trial designs (e.g. carry-over in cross-over trials and recruitment bias in cluster-randomized trials); some can be found across a broad spectrum of trials, but only for specific circumstances (e.g. contamination, whereby the experimental and control interventions get 'mixed', for example if participants pool their drugs).

Ref. *Cochrane Handbook for Systematic Reviews of Interventions, v. 5.1.0 (2011). The Cochrane Collaboration. (http://handbook.cochrane.org)*

A bias is a systematic error, or deviation from the truth, in study results. It is not possible to determine the extent biases can affect results of a particular study, but flaws in study design, conduct and analysis of data are known to lead to bias. Biases vary in magnitude but can underestimate or overestimate the true effect of the intervention in clinical trials; therefore, it is important to consider the likely magnitude of bias and direction of effect. For example, if all methodological limitations of studies were expected to bias the results towards a lack of effect, and the evidence indicates that the intervention is effective, then it may be concluded that the intervention is effective even in the presence of these potential biases. Assess each domain separately to determine if risk of each bias is likely **LOW**, **HIGH** or **UNCLEAR** (**Table 2**). Unclear risk of bias will be interpreted as high risk of bias when quality of evidence is graded (**Appendix D**).

Conflicts of interest should also be assessed when determining risk of bias. This may be considered part of risk of reporting bias. Funding sources for the trial, conflicts of interest of the authors, and role the study sponsor played in the trial should be considered in this domain.

The quality of each trial will be graded as **good, fair,** or **poor** based on the following thresholds for converting the Cochrane Risk of Bias Tool to AHRQ Standards. A good quality trial will have low risk of bias for all domains. A fair quality trial will have one domain with high risk of bias or 2 domains with unclear bias, with the assessment that the one or more biases are unlikely to influence the outcome, and there are no known limitations which could invalidate results. A poor quality trial will have high risk of bias for one or more domains or have 2 criteria with unknown bias for which there may be important limitations which could invalidate the results or likely bias the outcome. Trials of poor quality will be excluded from review if higher quality sources of evidence are available.

Table 2. Methods to Assess Risk of Bias in Clinical Trials: Cochrane Risk of Bias (modified).

SELECTION BIAS			
Risk of Bias	**LOW**	**HIGH**	**UNCLEAR**
Inadequate randomization	Sequence generated by: • Computerized random number generator • Random number table • Coin toss	Sequence generated by: • Odd or even date of birth • Rule based on date or admission date • Hospital or clinic number • Alternating numbers	Method of randomization not described or sequence generation process not described in sufficient detail for definitive judgment
Inadequate allocation concealment	Participants or investigators could not foresee assignment because: • Central allocation (telephone, web-based, pharmacy-controlled) • Sequentially numbered drug containers of identical appearance • Sequentially numbered, opaque, sealed envelopes	Participants or investigators could possibly foresee assignment because: • Open random allocation • Envelopes without appropriate safeguards (eg, unsealed or not opaque) • Allocation based on date of birth or case record number • Alternating allocation	Method of concealment not described or not described in sufficient detail for definitive judgment
Unbalanced baseline characteristics	Important prognostic factors similar between groups at baseline	Important prognostic factors are not balanced, which indicates inadequate sequence generation, allocation concealment, or failed randomization. *Statistical tests of baseline imbalance are not helpful for randomized trials.	Important prognostic factors are missing from baseline characteristics (eg, co-morbidities, other medications, medical/surgical history, etc.)
PERFORMANCE BIAS			
Risk of Bias	**LOW**	**HIGH**	**UNCLEAR**
Systematic differences in how care was provided between groups due to un-blinding of participants or investigators/care providers or because of standard of care was not consistent across all sites.	• Study participants could not identify study assignment because blinding of participants was ensured and unlikely to be broken (ie, double-dummy design with matching descriptions) • Protocol standardized across all sites and followed consistently	• Study participants could possibly identify study assignment because there was no blinding or incomplete blinding • Blinding potentially broken, which likely influenced effect estimate (eg, differences easily observed in appearance, taste/smell or adverse effects between groups) • Some sites had a different standard of care or varied from protocol which likely influenced effect estimate	Not described or insufficient information to permit definitive judgment

DETECTION BIAS			
Risk of Bias	**LOW**	**HIGH**	**UNCLEAR**
Outcome assessors un-blinded	Outcome assessors could not identify study assignment because: • Blinding of assessors was ensured and unlikely broken • No blinding or incomplete blinding, but effect estimate not likely influenced by lack of blinding (ie, objective outcomes)	• Outcome data assessors could possibly identify study assignment because no blinding or incomplete blinding, which likely influenced effect estimate • Blinding potentially broken, which likely influenced effect estimate (eg, large differences in efficacy or safety outcomes between groups)	Not described or insufficient information to permit definitive judgment
ATTRITION BIAS			
Risk of Bias	**LOW**	**HIGH**	**UNCLEAR**
High attrition or differential	• No missing data • Reasons for missing outcome data unlikely to influence effect estimates	• High Drop-out rate or loss to follow-up (eg, >10% for short-term studies; >20% for longer-term studies) • Differential drop-out or loss to follow-up >10% between groups	Not described or insufficient reporting of attrition/exclusions post-randomization to permit judgment
Missing data handled inappropriately	• Intention-to-treat analysis performed where appropriate (eg, superiority trials) • Intention-to-treat and per-protocol analyses performed and compared where appropriate (eg, non-inferiority trials) • Reasons for missing outcome data unlikely to influence effect estimates • Appropriate censoring rules applied depending on nature of study (eg, last-observation-carried-forward (LOCF) for curative conditions, or for treatments that improve a condition over time like acute pain, infection, etc.)	• As-treated analyses performed with substantial departure from randomized number • Per-protocol analyses or modified-intention-to-treat with substantial amount of missing data • Potentially inappropriate imputation of missing data (eg, LOCF for chronic, deteriorating conditions like HF, COPD, or cancer, etc.)	Not described or insufficient reporting of attrition/exclusions post-randomization to permit judgment
REPORTING BIAS			
Risk of Bias	**LOW**	**HIGH**	**UNCLEAR**
Evidence of selective outcome reporting	• Study protocol is available and was followed and all pre-specified primary and secondary outcomes are reported • Study protocol is not available, but it is clear that all expected outcomes are reported	• Not all pre-specified primary and secondary outcomes reported • Primary outcome(s) reported using measurements, analyses, or subsets of patients that were not pre-specified (eg, post-hoc analysis; protocol change without justification) • Primary outcome(s) not pre-specified (unless clear justification provided) • Failure or incomplete reporting of other outcomes of interest • Inappropriate over-emphasis of positive secondary outcomes in study with negative primary outcome	Insufficient information to make determination
OTHER BIAS			
Risk of Bias	**LOW**	**HIGH**	**UNCLEAR**

Evidence of other biases not described in the categories above	• No conflicts of interest present or study sponsor was not involved in trial design, data analysis or publication • No other potential sources of bias identified	• Conflicts of interest are present based on funding source or conflicting interests of authors • Study sponsor is involved in trial design, data analysis, and publication of data • There is a run-in period with pre-randomization administration of an intervention that could enhance or diminish the effect of a subsequent, randomized, intervention • Recruitment bias in cluster-randomized trials with differential participant recruitment in clusters for different interventions • Cross-over trials in which the crossover design is not suitable, there is significant carry-over effects, or incompletely reported data (data reported only for first period) • Conduct of the study is affected by interim results ((e.g. recruiting additional participants from a subgroup showing more benefit) • Deviation from the study protocol in a way that does not reflect clinical practice (e.g. post hoc stepping-up of doses to exaggerated levels).	• Conflicts of interest for authors or funding sources are not reported or not described • Insufficient information regarding other trial methodology and design to make a determination

Ref. *Cochrane Handbook for Systematic Reviews of Interventions, v. 5.1.0 (2011). The Cochrane Collaboration. (http://handbook.cochrane.org)*

The Patient, Intervention, Comparator, Outcome, and Setting (PICOS) framework is used to assess applicability (ie, directness) of the evidence to the OHP population (**Table 3**).

Table 3. PICOS Domains that Affect Applicability.

PICOS Domain	Conditions that Limit Applicability
Patient	• Narrow eligibility criteria and broad exclusion criteria of those with comorbidities • Large differences between the demographic characteristics between the study population and patients in the OHP • Narrow or unrepresentative severities in stage of illness or comorbidities (eg, only mild or moderate severity of illness included) • Run-in period with high exclusion rate for non-adherence or adverse effects • Event rates in study much lower/higher than observed in OHP population
Intervention	• Doses, frequency schedule, formulations or duration of intervention used in study not reflective of clinical practice • Intensity/delivery of behavioral interventions not feasible for routine use in clinical practice • Concomitant interventions likely over- or underestimate effectiveness of therapy
Comparator	• Inadequate dose or frequency schedule of comparator • Use of inferior or substandard comparator relative to alternative comparators that could be used
Outcomes	• Short-term or surrogate outcomes assessed • Composite outcomes used that mix outcomes of different significance
Setting	• Standards of care in study setting differ markedly from clinical practice • Monitoring/visit frequency not feasible for routine use in clinical practice • Level of care from highly trained/proficient practitioners in trial not reflective of typical clinical practice where intervention likely to be used

Ref. *Cochrane Handbook for Systematic Reviews of Interventions, v. 5.1.0 (2011). The Cochrane Collaboration. (http://handbook.cochrane.org)*

Non-inferiority (NI) trials are designed to prove a new treatment is not worse than the control treatment by a pre-determined difference, with a given degree of confidence. The pre-determined margin of difference in non-inferiority trials is defined as delta. Correctly determining this margin is a challenge in the design and interpretation of NI trials. The greatest challenge in use of NI trials is recognizing inappropriate use.

Non-inferiority trials will only be included in evidence summaries when there is a compelling reason to include them, and higher quality evidence is not available. The compelling reason for inclusion will be clearly stated as an introduction to the reporting of the NI trial.

The following template was developed using CONSORT and FDA guidance[1,2] and will be used as a guideline to evaluate non-inferiority studies included in DURM evidence summaries. Unless the trial evaluates an outcome or comparison of high clinical importance, individual non-inferiority trials will be excluded from class updates, class reviews, and literature scans. Evidence from poor quality RCTs may be included in individual drug evaluations if there is no higher quality evidence available. Items in bold (#1-5) are essential to conducting a non-inferiority trial with good methodological rigor. In general, a non-inferiority trial with high quality methods will score a "yes" on most of the components listed below.

Table 4. Non-inferiority Trial Quality Scoring Template

Developed using CONSORT and FDA guidance[1,2] Use Template to evaluate trials supporting New Drug Evaluations and Class Update Reports A high-quality trial will meet all bolded assessments below	
1. Rationale for choosing comparator with historical study results confirming efficacy (or safety) of this comparator is provided.	□ **Yes** □ **No** □ **Can't answer**
2. Active control (or comparator) represents current standard of care.	□ **Yes** □ **No** □ **Can't answer**
3. Non-inferiority margin was specified a priori and based on statistical reasoning and clinical considerations regarding benefit, risk, and cost.	□ **Yes** □ **No** □ **Can't answer**
4. Noninferiority margin is not larger than the expected difference between active control (or comparator) and placebo.	□ **Yes** □ **No** □ **Can't answer**
5. If a superiority conclusion is drawn for outcome(s) for which noninferiority was hypothesized, the justification for switching is provided and superiority analysis was defined a priori.	□ **Yes** □ **No** □ **Can't answer**
6. Investigator reported both ITT and per-protocol analysis in detail and the results of both analyses demonstrate noninferiority. (If only one analysis is provided, per protocol is subject to less bias than ITT analysis in noninferiority trials.)	□ **Yes** □ **No** □ **Can't answer**
7. Rationale for using a noninferiority design is included (or why it would likely be unethical to conduct a placebo-controlled superiority trial of the new therapy).	□ Yes □ No □ Can't answer
8. Study hypothesis is stated in terms of noninferiority.	□ Yes □ No □ Can't answer
9.Eligibility criteria for participants and the settings in which the data were collected are similar to those in any trial(s) that established efficacy (or safety) of the reference treatment.	□ Yes □ No □ Can't answer
10. Trial is designed to be consistent with historical placebo-controlled trials.	□ Yes □ No □ Can't answer
11. The reference treatment in the noninferiority trial is identical (or very similar) to that in any trial(s) that established efficacy (or safety).	□ Yes □ No □ Can't answer
12. The outcomes in the noninferiority trial are identical (or very similar) to those in any trial(s) that established efficacy (or safety) of the reference treatment.	□ Yes □ No □ Can't answer
13. The lower bound of that CI is clinically significant.	□ Yes □ No □ Can't answer
14. For the outcome(s) for which noninferiority was hypothesized, a figure showing confidence intervals and the noninferiority margin is included.	□ Yes □ No □ Can't answer
15. Results are interpreted in relation to the noninferiority hypothesis.	□ Yes □ No □ Can't answer

References:
1. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *Jama*. 2012;308(24):2594-2604.
2. FDA Industry Guidance for Noninferiority Trials. November 2016. https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM202140.pdf.

APPENDIX B. Methods to Assess Methodological Quality of Systematic Reviews.

A measurement tool for the "assessment of multiple systematic reviews" (AMSTAR II) was developed and shown to be a validated and reliable measurement tool to assess the methodological quality of systematic reviews. There are 16 components addressed in the measurement tool below, and questions can be scored in one of four ways: "Yes", "Partial Yes", "No", or "Not Applicable". The AMSTAR II is used as a guideline to identify high quality systematic reviews eligible for inclusion in DURM evidence summaries. High quality systematic reviews do not contain a "fatal flaw" (ie, comprehensive literature search not performed (#4); characteristics of studies not provided (#8); quality of studies were not assessed or considered when conclusions were formulated (#9 and #13)). Other areas identified as important domains in the AMSTAR II criteria include registration of a protocol (#2); justification for excluding individual studies (#7); appropriateness of meta-analysis methods (#11); and assessment of publication bias (#15). In general, a high quality systematic review will score a "yes" on most components presented in the AMSTAR II tool.

Ref. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, Moher D, Tugwell P, Welch V, Kristjansson E, Henry DA. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017 Sep 21;358:j4008.

Systematic reviews or guidance identified from 'best sources' undergo methodological rigor considered to be of high quality and are not scored for quality. 'Best sources' include, but are not limited to: Drug Effectiveness Review Project (DERP) at the Pacific Northwest Evidence-based Practice Center; Agency for Healthcare Research and Quality (AHRQ); National Institute for Health and Care Excellence (NICE); U.S. Department of Veterans Affairs (VA); and Canadian Agency for Drugs and Technologies in Health (CADTH); and BMJ Clinical Evidence.

AMSTAR II Quality Scoring Template			
1)	**Did the research questions and inclusion criteria for the review include the components of PICO?** **For Yes**: □ Population □ Intervention □ Comparator group □ Outcome	Optional (recommended) □ Timeframe for follow-up	□ Yes □ No
2)	**Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?** **For Partial Yes:** The authors state that they had a written protocol or guide that included ALL the following: □ review question(s) □ a search strategy □ inclusion/exclusion criteria □ a risk of bias assessment	**For Yes:** As for partial yes, plus the protocol should be registered and should also have specified: □ a meta-analysis/synthesis plan, if appropriate, and □ a plan for investigating causes of heterogeneity □ justification for any deviations from the protocol	□ Yes □ Partial Yes □ No
3)	**Did the review authors explain their selection of the study designs for inclusion in the review?** **For Yes**, the review should satisfy ONE of the following: □ Explanation for including only RCTs □ OR Explanation for including only NRSI □ OR Explanation for including both RCTs and NRSI		□ Yes □ No

4) Did the review authors use a comprehensive literature search strategy?

For Partial Yes (all the following):
- searched at least 2 databases (relevant to research question)
- provided key word and/or search strategy
- justified publication restrictions (e.g. language)

For Yes, should also have (all the following):
- searched the reference lists / bibliographies of included studies
- searched trial/study registries
- included/consulted content experts in the field
- where relevant, searched for grey literature
- conducted search within 24 months of completion of the review

- Yes
- Partial Yes
- No

5) Did the review authors perform study selection in duplicate?

For Yes, either ONE of the following:
- at least two reviewers independently agreed on selection of eligible studies and achieved consensus on which studies to include
- OR two reviewers selected a sample of eligible studies and achieved good agreement (at least 80 percent), with the remainder selected by one reviewer.

- Yes
- No

6) Did the review authors perform data extraction in duplicate?

For Yes, either ONE of the following:
- at least two reviewers achieved consensus on which data to extract from included studies
- OR two reviewers extracted data from a sample of eligible studies and achieved good agreement (at least 80 percent), with the remainder extracted by one reviewer.

- Yes
- No

7) Did the review authors provide a list of excluded studies and justify the exclusions?

For Partial Yes:
- provided a list of all potentially relevant studies that were read in full-text form but excluded from the review

For Yes, must also have:
- Justified the exclusion from the review of each potentially relevant study

- Yes
- Partial Yes
- No

8) Did the review authors describe the included studies in adequate detail?

For Partial Yes (ALL the following):
- described populations
- described interventions
- described comparators
- described outcomes
- described research designs

For Yes, should also have ALL the following:
- described population in detail
- described intervention in detail (including doses where relevant)
- described comparator in detail (including doses where relevant)
- described study's setting
- timeframe for follow-up

- Yes
- Partial Yes
- No

9) Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?

RCTs

For Partial Yes, must have assessed RoB from:
- unconcealed allocation, and
- lack of blinding of patients and assessors when assessing outcomes (unnecessary for objective outcomes such as all-cause mortality)

For Yes, must also have assessed RoB from:
- allocation sequence that was not truly random, and
- selection of the reported result from among multiple measurements or analyses of a specified outcome

- Yes
- Partial Yes
- No
- Includes only NRSI

NRSI

For Partial Yes, must have assessed RoB:
- from confounding, and
- from selection bias

For Yes, must also have assessed RoB:
- methods used to ascertain exposures and outcomes, and
- selection of the reported result from among multiple measurements or analyses of a specified outcome

- Yes
- Partial Yes
- No
- Includes only RCTs

10) Did the review authors report on the sources of funding for the studies included in the review?

For Yes: Must have reported on the sources of funding for individual studies included in the review. Note: Reporting that the reviewers looked for this information but it was not reported by study authors also qualifies

- Yes
- No

11) If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?

RCTs

For Yes:
- The authors justified combining the data in a meta-analysis
- AND they used an appropriate weighted technique to combine study results and adjusted for heterogeneity if present.
- AND investigated the causes of any heterogeneity

- Yes
- No
- No meta-analysis conducted

NRSI	**For Yes:** □ The authors justified combining the data in a meta-analysis □ AND they used an appropriate weighted technique to combine study results, adjusting for heterogeneity if present □ AND they statistically combined effect estimates from NRSI that were adjusted for confounding, rather than combining raw data, or justified combining raw data when adjusted effect estimates were not available □ AND they reported separate summary estimates for RCTs and NRSI separately when both were included in the review	□ Yes □ No □ No meta-analysis conducted
12)	**If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?** **For Yes:** □ included only low risk of bias RCTs □ OR, if the pooled estimate was based on RCTs and/or NRSI at variable RoB, the authors performed analyses to investigate possible impact of RoB on summary estimates of effect.	□ Yes □ No □ No meta-analysis conducted
13)	**Did the review authors account for RoB in individual studies when interpreting/ discussing the results of the review?** **For Yes:** □ included only low risk of bias RCTs □ OR, if RCTs with moderate or high RoB, or NRSI were included the review provided a discussion of the likely impact of RoB on the results	□ Yes □ No
14)	**Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?** **For Yes:** □ There was no significant heterogeneity in the results □ OR if heterogeneity was present the authors performed an investigation of sources of any heterogeneity in the results and discussed the impact of this on the results of the review	□ Yes □ No
15)	**If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?** **For Yes:** □ performed graphical or statistical tests for publication bias and discussed the likelihood and magnitude of impact of publication bias	□ Yes □ No □ No meta-analysis conducted
16)	**Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?** **For Yes:** □ The authors reported no competing interests OR □ The authors described their funding sources and how they managed potential conflicts of interest	□ Yes □ No

APPENDIX C. Methods to Assess Methodological Quality of Clinical Practice Guidelines.

Clinical practice guidelines are systematically developed statements that assist clinicians in making clinical decisions. However, guidelines can vary widely in quality and utility. The Appraisal of Guidelines, Research, and Evaluation (AGREE) Instrument (www.agreetrust.org) assesses the methodologic rigor in which a guideline is developed and used. The AGREE II is an updated instrument that has been validated. It consists of 23 items in 6 domains (scope, stakeholder involvement, rigor of development, clarity, applicability, and editorial independence) to rate (**Table 1**). Because it is time-consuming to administer, a consolidated global rating scale (GRS) was developed, and is generally a reasonable alternative to AGREE II if resources are limited. The AGREE II-GRS instrument consists of only 4 items (**Table 2**). As the AGREE II-GRS does not take into account conflicts of interest, questions 22 and 23 regarding "Editorial Independence" will also be evaluated in conjunction with the AGREE II-GRS. With both instruments, each item is rated on a 7-point scale, from 0=lowest quality to 7=highest quality. High quality clinical practice guidelines are eligible for inclusion in DURM evidence summaries. These guidelines will score 6-7 points for each component on rigor of development. In general, a high quality clinical practice guideline will score 5-7 points on most components presented in the AGREE II and each component of the AGREE II-GRS.

Table 1. AGREE II Instrument.

	ITEM	DESCRIPTION
SCOPE AND PURPOSE		
1	The overall objective(s) of the guideline is (are) specifically described.	The overall objective(s) of the guideline should be described in detail and the expected health benefits from the guideline should be specific to the clinical problem or health topic. **[SCORE:]**
2	The health question(s) covered by the guideline is (are) specifically described.	A detailed description of the health questions covered by the guideline should be provided, particularly for key recommendations, although they need not be phrased as questions. **[SCORE:]**
3	The population to whom the guideline is meant to apply is specifically described.	A clear description of the population (ie, patients, public, etc.) covered by a guideline should be provided. The age range, sex, clinical description, and comorbidities may be provided. **[SCORE:]**
STAKEHOLDER INVOLVEMENT		
4	The guideline development group includes individuals from all relevant professional groups.	This may include members of the steering group, the research team involved in selection and review of the evidence and individuals involved in formulation of the final recommendations. **[SCORE:]**
5	The views and preferences of the target population have been sought.	Information about target population experiences and expectations of health care should inform the development of guidelines. There should be evidence that some process has taken place and that stakeholders' views have been considered. For example, the public was formally consulted to determine priority topics, participation of these stakeholders on the guideline development group, or external review by these stakeholders on draft documents. Alternatively, information could be obtained from interviews of these stakeholders or from literature reviews of patient/public values, preferences or experiences. **[SCORE:]**
6	The target users of the guideline are clearly defined.	The target users should be clearly defined in the guideline so the reader can immediately determine if the guideline is relevant to them. For example, the target users for a guideline on low back pain may include general practitioners, neurologists, orthopedic surgeons, rheumatologists, and physiotherapists. **[SCORE:]**
RIGOR OF DEVELOPMENT		
7	Systematic methods were used to search for evidence.	Details of the strategy used to search for evidence should be provided, which include search terms used, sources consulted, and dates of the literature covered. The search strategy should be as comprehensive as possible and executed in a manner free from potential biases and sufficiently detailed to be replicated. **[SCORE:]**
8	The criteria for selecting the evidence are clearly described.	Criteria for including/excluding evidence identified by the search should be provided. These criteria should be explicitly described and reasons for including and excluding evidence should be clearly stated. **[SCORE:]**

9	The strengths and limitations of the body of evidence are clearly described.	Statements that highlight the strengths and limitations of the evidence should be provided. This ought to include explicit descriptions, using informal or formal tools/methods, to assess and describe the risk of bias for individual studies and/or for specific outcomes and/or explicit commentary of the body of evidence aggregated across all studies. **[SCORE:]**
10	The methods for formulating the recommendations are clearly described.	A description of the methods used to formulate the recommendations and how final decisions were arrived at should be provided. For example, methods may include a voting system, informal consensus, or formal consensus techniques (eg, Delphi, Glaser techniques). **[SCORE:]**
11	The health benefits, adverse effects, and risks have been considered in formulating the recommendations.	The guideline should consider both effectiveness/efficacy and safety when recommendations are formulated. **[SCORE:]**
12	There is an explicit link between the recommendations and the supporting evidence.	An explicit link between the recommendations and the evidence on which they are based should be included in the guideline. **[SCORE:]**
13	The guideline has been externally reviewed by experts prior to its publication.	A guideline should be reviewed externally before it is published. Reviewers should not have been involved in the guideline development group. Reviewers should include both clinical and methodological experts. **[SCORE:]**
14	A procedure for updating the guideline is provided.	A clear statement about the procedure for updating the guideline should be provided. **[SCORE:]**
CLARITY OF PRESENTATION		
15	The recommendations are specific and unambiguous.	A recommendation should provide a precise description of which option is appropriate in which situation and in what population. It is important to note that in some instances, evidence is not always clear and there may be uncertainty about the best practice. In this case, the uncertainty should be stated in the guideline. **[SCORE:]**
16	The different options for management of the condition or health issue are clearly presented.	A guideline that targets the management of a disease should consider the different possible options for screening, prevention, diagnosis or treatment of the condition it covers. **[SCORE:]**
17	Key recommendations are easily identifiable	Users should be able to find the most relevant recommendations easily. **[SCORE:]**
APPLICABILITY		
18	The guideline describes facilitators and barriers to its application.	There may be existing facilitators and barriers that will impact the application of guideline recommendations. **[SCORE:]**
19	The guideline provides advice and/or tools on how the recommendations can be put into practice.	For a guideline to be effective, it needs to be disseminated and implemented with additional materials. For example, these may include: a summary document, a quick reference guide, educational tools, results from a pilot test, patient leaflets, or computer/online support. **[SCORE:]**
20	The potential resource implications of applying the recommendations have been considered.	The recommendations may require additional resources in order to be applied. For example, there may be a need for more specialized staff or expensive drug treatment. These may have cost implications on health care budgets. There should be a discussion in the guideline of the potential impact of the recommendations on resources. **[SCORE:]**
21	The guideline presents monitoring and/or auditing criteria	Measuring the application of guideline recommendations can facilitate their ongoing use. This requires clearly defined criteria that are derived from the key recommendations in the guideline (eg, HbA1c <7%, DBP <95 mm Hg). **[SCORE:]**
EDITORIAL INDEPENDENCE		
22	The views of the funding body have not influenced the content of the guideline.	Many guidelines are developed with external funding (eg, government, professional associations, charity organizations, pharmaceutical companies). Support may be in the form of financial contribution for the complete development, or for parts of it (eg, printing/dissemination of the guideline). There should be an explicit statement that the views or interests of the funding body have not influenced the final recommendations. **[SCORE:]**
23	Competing interests of guideline development group members have been recorded and addressed	There should be an explicit statement that all group members have declared whether they have any competing interests. **[SCORE:]**

Table 2. AGREE II Global Rating Scale (modified).

	ITEM	DESCRIPTION
1	Rate the guideline development methods. **[SCORE:]**	• Appropriate stakeholders were involved in the development of the guideline. • The evidentiary base was developed systematically. • Recommendations were consistent with the literature. Consideration of alternatives, health benefits, harms, risks, and costs was made.
2	Rate the guideline presentation. **[SCORE:]**	• The guideline was well organized. • The recommendations were easy to find.
3	Rate the guideline recommendations. **[SCORE:]**	• The recommendations are clinically sound. • The recommendations are appropriate for the intended patients.
4	Rate the completeness of reporting, editorial independence. **[SCORE:]**	• The information is complete to inform decision making. • The guideline development process is transparent and reproducible.
5	The views of the funding body have not influenced the content of the guideline. **[SCORE:]**	• Many guidelines are developed with external funding (eg, government, professional associations, charity organizations, pharmaceutical companies). Support may be in the form of financial contribution for the complete development, or for parts of it (eg, printing/dissemination of the guideline). There should be an explicit statement that the views or interests of the funding body have not influenced the final recommendations.
6	Competing interests of guideline development group members have been recorded and addressed. **[SCORE:]**	• There should be an explicit statement that all group members have declared whether they have any competing interests. • All competing interests should be listed • There should be no significant competing interests

APPENDIX D. GRADE Quality of Evidence.

Grading of Recommendations Assessment, Development and Evaluation (GRADE) provides a framework to assess quality of evidence for an *outcome* that emphasizes transparency of how evidence judgments are made, though it does not necessarily guarantee consistency in assessment. Quality assessment in GRADE is 'outcome-centric' and distinct from quality assessment of an individual study. Information on risk of bias (internal validity), indirectness (applicability), imprecision, inconsistency, and publication bias is necessary to assess quality of evidence and overall confidence in the estimated effect size. The GRADE framework provides an assessment for each outcome.

DURM evidence summaries, unless a single drug is evaluated, depend on the whole body of available evidence. Evidence from high quality systematic reviews is the primary basis for recommendations in the evidence summaries. High quality evidence-based clinical practice guidelines and relevant randomized controlled trials are used to supplement the whole body of evidence.

High quality systematic reviews and clinical practice guidelines often use the GRADE framework to assess overall quality of evidence for a given outcome. In such cases, the grade of evidence provided in the respective report can be directly transferred to the DURM evidence summary. When an evidence summary includes relevant clinical trials, or when high quality systematic reviews or clinical practice guidelines that did not use the GRADE framework were identified, quality of evidence will be graded based on hierarchy of available evidence, homogeneity of results for a given outcome, and methodological flaws identified in the available evidence (**Table 1**).

Table 1. Evidence Grades for Benefit and Harm Outcomes When a Body of Evidence is Evaluated.

GRADE	TYPE OF EVIDENCE
High	• Evidence is based on data derived from multiple randomized controlled trials with homogeneity with regard to the direction of effect between studies AND • Evidence is based on multiple, well-done randomized controlled trials that involved large numbers of patients.
Moderate	• Evidence is based on data derived from randomized controlled trials with some conflicting conclusions with regard to the direction of effect between studies OR • Evidence is based on data derived from randomized controlled trials that involved small numbers of patients but showed homogeneity with regard to the direction of effect between studies OR • Some evidence is based on data derived from randomized controlled trials with significant methodological flaws (eg, bias, attrition, flawed analysis, etc.)
Low	• Most evidence is based on data derived from randomized controlled trials with significant methodological flaws (eg, bias, attrition, flawed analysis, etc.) OR • Evidence is based mostly on data derived from non-randomized studies (eg, cohort studies, case-control studies, observational studies) with homogeneity with regard to the direction of effect between studies
Insufficient	• Evidence is based mostly on data derived from non-randomized studies (eg, cohort studies, case-control studies, observational studies) with some conflicting conclusions with regard to direction of effect between studies OR • Evidence is based on data derived from expert opinion/panel consensus, case reports or case series OR • Evidence is not available

New Drug Evaluations cannot depend on evidence from systematic reviews and clinical practice guidelines. A body of evidence that solely consists of one or more clinical trials is initially assigned 4 points. For every relevant limitation, points are deducted; but points are added for consistently large effect sizes between studies or for a consistent dose-response observed in the studies (**Table 2**). The quality of evidence is subsequently graded as shown:

QUALITY OF EVIDENCE GRADES:

- ≥4 points = **HIGH**
- 3 points = **MODERATE**
- 2 points = **LOW**
- ≤1 point = **INSUFFICIENT**

Table 2. Domains to Grade Evidence for Benefit and Harm Outcomes from Clinical Trials: Cochrane Evidence Grades (modified).

DOMAIN	DESCRIPTION	SCORE DEMOTION/PROMOTION (start with 4 points)
Risk of Bias (internal validity)	Risk of bias is the likelihood to which the included studies for a given comparison and outcome has an inadequate protection against bias that affects the internal validity of the study. • *Did any studies have important limitations that degrade your confidence in estimates of effectiveness or safety?*	• No serious limitation: all studies have low risk of bias: **(0)** • Serious limitations: ≥1 trial has high or unclear risk of bias: **(-1)** • Very serious limitations: most studies have high risk of bias: **(-2)**
Indirectness (applicability)	Directness (applicability) relates to evidence that adequately compares 2 or more reasonable interventions that can be directly linked to a clinically relevant outcome in a population of interest. • *Do studies directly compare interventions of interest in populations of interest using outcomes of interest (use of clinically relevant outcomes)?*	• Direct: clinically relevant outcomes of important comparisons in relevant populations studied: **(0)** • Indirect: important comparisons missing; surrogate outcome(s) used; or population not relevant: **(-1)**
Inconsistency	Inconsistency (heterogeneity) is the degree to which reported effect sizes from included studies appear to differ in direction of effect. Effect sizes have the same sign (ie, are on the same side of ''no effect'') and the range of effect sizes is narrow. • *Did trials have similar or widely varying results? Can heterogeneity be explained by differences in trial design and execution?*	• Large magnitude of effect consistent between studies: **(+1)** • Dose-response observed: **(+1)** • Small magnitude of effect consistent between studies: **(0)** • 1 study with large magnitude of effect: **(0)** • 1 study with small magnitude of effect: **(-1)** • Inconsistent direction of effect across studies that cannot be explained: **(-1)**
Imprecision	Imprecision is the degree of uncertainty surrounding an effect estimate with respect to a given outcome (ie, the confidence interval for each outcome is too wide to rule out no effect). • *Are confidence intervals for treatment effect sufficiently narrow to rule out no effect?*	• Precise: all studies have 95% confidence intervals that rule out no effect: **(0)** • Imprecise: ≥1 study demonstrated 95% confidence interval fails to rule out no effect: **(-1)**
Publication Bias	Publication bias is the degree in which completed trials are not published or represented. Unpublished studies may have negative outcomes that would otherwise change our confidence in the body of evidence for a particular comparison and outcome. • *Is there evidence that important trials are not represented?*	• No publication bias: all important trials published or represented: **(0)** • Serious publication bias: ≥1 important trial(s) completed but not published: **(-1)**

Ref. *Cochrane Handbook for Systematic Reviews of Interventions, v. 5.1.0 (2011). The Cochrane Collaboration. (http://handbook.cochrane.org)*