

## Standard Methods for Quality Assessment of Evidence

The methods described herein is the standard approach used by the Drug Use Research & Management faculty to assess quality of evidence incorporated into the evidence summaries for the Oregon Health Plan (OHP) Pharmacy and Therapeutics Committee. These evidence summaries inform the recommendations for management of the preferred drug list (PDL) 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.

Recommendations in evidence summaries (ie, Drug Class Reviews, Updates, or Literature Scans) are determined primarily from high quality systematic reviews. High quality clinical practice guidelines and relevant clinical trials are also used as supplementary evidence. New Drug Evaluations are more focused and primarily rely on evidence from clinical trials.

### Quality Assessment

Internally validity is assessed after determination of risk of bias. 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. Clinical trials used to assess new drugs will be assessed for risk of bias using the methods in **Appendix A**.

Assessment of applicability, or directness, to the OHP population is an important consideration. The Patient, Intervention, Comparator, Outcome, and Setting (PICOS) framework is used to assess applicability. Clinical trials used to assess new drugs will also be assessed for applicability to the OHP population using guidance provided in **Appendix A**.

High quality systematic reviews and clinical practice guidelines with applicability to the OHP population, as assessed by the PICOS framework, are included in evidence summaries. The AMSTAR measurement tool is used to assess for methodological quality of systematic reviews and is provided in **Appendix B**. Clinical practice guidelines are considered for inclusion after assessment of methodological quality using the AGREE II global rating scale provided in **Appendix C**.

### Grading Quality of Evidence

The overall quality of evidence is graded for clinically relevant outcomes of efficacy and harm: mortality, morbidity outcomes, symptom relief, quality of life, and functioning (physical, mental, or emotional). Surrogate outcomes are considered if directly linked to a clinically relevant outcome. Clinically meaningful changes in these outcomes are emphasized. Evaluation of evidence for each outcome of interest is graded as **high**, **moderate**, **low**, or **insufficient** based on the domains listed in **Appendix D**. Evidence grades are defined in **Table 1**. Major domains in the evidence considered include risk of bias (internal validity), indirectness (applicability), inconsistency, imprecision, and publication bias.

**Table 1. Quality of Evidence Grades and Definitions.**

GRADE	DEFINITION
<b>High</b>	High confidence that the estimated effects produced in the studies reflect the true effect. Further research is very unlikely to change the estimated effect.
<b>Moderate</b>	Moderate confidence that the estimated effects produced in the studies reflect the true effect. Further research may change the estimated effect.
<b>Low</b>	Limited confidence that the estimated effects produced in the studies reflect the true effect. Further research is likely to change the estimated effect.
<b>Insufficient</b>	Evidence is not available or too limited to permit any level of confidence in the estimated effect.

Final evidence summary recommendations account for the availability and quality of evidence for relevant outcomes and perceived clinical impact on the OHP population.

## APPENDIX A. Methods to Assess Quality of Studies.

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

<b>Selection Bias</b>	<p>Selection bias refers to <b>systematic differences</b> between <b>baseline characteristics of the groups</b> that were compared. The unique strength of proper <b>randomization</b> 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 is often termed <b>allocation concealment</b>.</p>
<b>Performance Bias</b>	<p>Performance bias refers to <b>systematic differences</b> between groups in the <b>care provided</b>, or in exposure to factors other than the interventions of interest. After enrolment, <b>blinding participants and investigators/care givers</b> 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.</p>
<b>Detection Bias</b>	<p>Detection bias refers to <b>systematic differences</b> between groups in how <b>outcomes were assessed</b>. <b>Blinding of outcome assessors</b> 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).</p>
<b>Attrition Bias</b>	<p>Attrition bias refers to <b>systematic differences</b> between groups in <b>withdrawals</b> (exclusions and attrition) from a study. <b>Withdrawals</b> from the study lead to incomplete outcome data. There are two reasons for withdrawals or incomplete outcome data in clinical trials. <b>Exclusions</b> refer to situations in which some participants are omitted from reports of analyses, despite outcome data being available to assessors. <b>Attrition</b> refers to situations in which outcome data are not available.</p>
<b>Reporting Bias</b>	<p>Selection bias refers to the <b>selective reporting of pre-specified outcomes</b>, 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).</p>

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

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**).

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

<b>SELECTION BIAS</b>			
<b>Risk of Bias</b>	<b>LOW</b>	<b>HIGH</b>	<b>UNCLEAR</b>
Inadequate randomization	Sequence generated by: <ul style="list-style-type: none"> <li>• Computerized random number generator</li> <li>• Random number table</li> <li>• Coin toss</li> </ul>	Sequence generated by: <ul style="list-style-type: none"> <li>• Odd or even date of birth</li> <li>• Rule based on date or admission date</li> <li>• Hospital or clinic number</li> <li>• Alternating numbers</li> </ul>	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: <ul style="list-style-type: none"> <li>• Central allocation (telephone, web-based, pharmacy-controlled)</li> <li>• Sequentially numbered drug containers of identical appearance</li> <li>• Sequentially numbered, opaque, sealed envelopes</li> </ul>	Participants or investigators could possibly foresee assignment because: <ul style="list-style-type: none"> <li>• Open random allocation</li> <li>• Envelopes without appropriate safeguards (eg, unsealed or not opaque)</li> <li>• Allocation based on date of birth or case record number</li> <li>• Alternating allocation</li> </ul>	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.)
<b>PERFORMANCE BIAS</b>			
<b>Risk of Bias</b>	<b>LOW</b>	<b>HIGH</b>	<b>UNCLEAR</b>
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.	<ul style="list-style-type: none"> <li>• 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)</li> <li>• Protocol standardized across all sites and followed consistently</li> </ul>	<ul style="list-style-type: none"> <li>• Study participants could possibly identify study assignment because there was no blinding or incomplete blinding</li> <li>• Blinding potentially broken, which likely influenced effect estimate (eg, differences easily observed in appearance, taste/smell or adverse effects between groups)</li> <li>• Some sites had a different standard of care or varied from protocol which likely influenced effect estimate</li> </ul>	Not described or insufficient information to permit definitive judgment
<b>DETECTION BIAS</b>			
<b>Risk of Bias</b>	<b>LOW</b>	<b>HIGH</b>	<b>UNCLEAR</b>
Outcome assessors un-blinded	Outcome assessors could not identify study assignment because: <ul style="list-style-type: none"> <li>• Blinding of assessors was ensured and unlikely broken</li> <li>• No blinding or incomplete blinding, but effect estimate not likely influenced by lack of blinding (ie, objective outcomes)</li> </ul>	<ul style="list-style-type: none"> <li>• Outcome data assessors could possibly identify study assignment because no blinding or incomplete blinding, which likely influenced effect estimate</li> <li>• Blinding potentially broken, which likely influenced effect estimate (eg, large differences in efficacy or safety outcomes between groups)</li> </ul>	Not described or insufficient information to permit definitive judgment

<b>ATTRITION BIAS</b>			
<b>Risk of Bias</b>	<b>LOW</b>	<b>HIGH</b>	<b>UNCLEAR</b>
High attrition or differential	<ul style="list-style-type: none"> <li>No missing data</li> <li>Reasons for missing outcome data unlikely to influence effect estimates</li> </ul>	<ul style="list-style-type: none"> <li>High Drop-out rate or loss to follow-up (eg, &gt;10% for short-term studies; &gt;20% for longer-term studies)</li> <li>Differential drop-out or loss to follow-up &gt;10% between groups</li> </ul>	Not described or insufficient reporting of attrition/exclusions post-randomization to permit judgment
Missing data handled inappropriately	<ul style="list-style-type: none"> <li>Intention-to-treat analysis performed where appropriate (eg, superiority trials)</li> <li>Intention-to-treat and per-protocol analyses performed and compared where appropriate (eg, non-inferiority trials)</li> <li>Reasons for missing outcome data unlikely to influence effect estimates</li> <li>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.)</li> </ul>	<ul style="list-style-type: none"> <li>As-treated analyses performed with substantial departure from randomized number</li> <li>Per-protocol analyses or modified-intention-to-treat with substantial amount of missing data</li> <li>Potentially inappropriate imputation of missing data (eg, LOCF for chronic, deteriorating conditions like HF, COPD, or cancer, etc.)</li> </ul>	Not described or insufficient reporting of attrition/exclusions post-randomization to permit judgment
<b>REPORTING BIAS</b>			
<b>Risk of Bias</b>	<b>LOW</b>	<b>HIGH</b>	<b>UNCLEAR</b>
Evidence of selective outcome reporting	<ul style="list-style-type: none"> <li>Study protocol is available and was followed and all pre-specified primary and secondary outcomes are reported</li> <li>Study protocol is not available, but it is clear that all expected outcomes are reported</li> </ul>	<ul style="list-style-type: none"> <li>Not all pre-specified primary and secondary outcomes reported</li> <li>Primary outcome(s) reported using measurements, analyses, or subsets of patients that were not pre-specified (eg, post-hoc analysis; protocol change without justification)</li> <li>Primary outcome(s) not pre-specified (unless clear justification provided)</li> <li>Failure or incomplete reporting of other outcomes of interest</li> <li>Inappropriate over-emphasis of positive secondary outcomes in study with negative primary outcome</li> </ul>	Insufficient information to make 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
<b>Patient</b>	<ul style="list-style-type: none"> <li>• Narrow eligibility criteria and broad exclusion criteria of those with comorbidities</li> <li>• Large differences between the demographic characteristics between the study population and patients in the OHP</li> <li>• Narrow or unrepresentative severities in stage of illness or comorbidities (eg, only mild or moderate severity of illness included)</li> <li>• Run-in period with high exclusion rate for non-adherence or adverse effects</li> <li>• Event rates in study much lower/higher than observed in OHP population</li> </ul>
<b>Intervention</b>	<ul style="list-style-type: none"> <li>• Doses, frequency schedule, formulations or duration of intervention used in study not reflective of clinical practice</li> <li>• Intensity/delivery of behavioral interventions not feasible for routine use in clinical practice</li> <li>• Concomitant interventions likely over- or underestimate effectiveness of therapy</li> </ul>
<b>Comparator</b>	<ul style="list-style-type: none"> <li>• Inadequate dose or frequency schedule of comparator</li> <li>• Use of inferior or substandard comparator relative to alternative comparators that could be used</li> </ul>
<b>Outcomes</b>	<ul style="list-style-type: none"> <li>• Short-term or surrogate outcomes assessed</li> <li>• Composite outcomes used that mix outcomes of different significance</li> </ul>
<b>Setting</b>	<ul style="list-style-type: none"> <li>• Standards of care in study setting differ markedly from clinical practice</li> <li>• Monitoring/visit frequency not feasible for routine use in clinical practice</li> <li>• Level of care from highly trained/proficient practitioners in trial not reflective of typical clinical practice where intervention likely to be used</li> </ul>

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

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

A measurement tool for the “assessment of multiple systematic reviews” (AMSTAR) was developed and shown to be a validated and reliable measurement tool to assess the methodological quality of systematic reviews. There are 11 components addressed in the measurement tool below, and each question can be scored in one of four ways: “Yes”, “No”, “Can’t Answer”, or “Not Applicable”. The AMSTAR 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 (#3); characteristics of studies not provided (#6); quality of studies were not assessed or considered when conclusions were formulated (#7 and #8)). In general, a high quality systematic review will score a “yes” on most components presented in the AMSTAR tool.

Ref. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology*. 2007;7:10. doi: 10.1186/1471-2288-7-10 .

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); Cochrane Collaboration; National Institute for Health and Care Excellence (NICE); Institute for Clinical and Economic Review (ICER); U.S. Department of Veterans Affairs (VA); Canadian Agency for Drugs and Technologies in Health (CADTH); BMJ Clinical Evidence; and the University of York Centre for Reviews and Dissemination.

## AMSTAR Quality Scoring Template

<p><b>1) Was an ‘a priori’ design provided?</b>          Note: the research question and inclusion criteria should be established before the conduct of the review and should be available.</p>	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Can’t answer <input type="checkbox"/> Not applicable
<p><b>2) Was there duplicate study selection and data extraction?</b>          Note: there should be at least two independent persons for study selection and data extraction; a consensus process for disagreements is in place; at least one person checks the other’s work.</p>	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Can’t answer <input type="checkbox"/> Not applicable
<p><b>3) Was a comprehensive literature search performed?</b>          Note: at least 2 databases (eg, MEDLINE, CINAHL, Scopus) plus one supplementary source (ie, gray literature) are searched. The review must include years and names databases used. Key words and/or Medical Subject Headings (MeSH) are stated and, if feasible, the search strategy is provided. Current reviews, specialized registers, or experts in the field of study may also be consulted.</p>	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Can’t answer <input type="checkbox"/> Not applicable
<p><b>4) Was the status of publication (ie, gray literature) used as an inclusion criterion?</b>          Note: “gray literature” or “unpublished literature” was searched. Dissertations, conference proceedings, and trial registries are all considered “gray literature” for this purpose. If a database was used that contained both gray literature and published literature, it was specified that gray literature was specifically searched. The authors should state whether any studies were excluded from the systematic review based on publication status, language, etc.</p>	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Can’t answer <input type="checkbox"/> Not applicable
<p><b>5) Was a list of studies (included and excluded) provided?</b>          Note: a list of included and excluded studies should be provided or referenced. Alternatively, there is a live electronic link to the list.</p>	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Can’t answer <input type="checkbox"/> Not applicable
<p><b>6) Were the characteristics of the included studies provided?</b>          Note: in an aggregated form (eg, a table), data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed (eg, age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases) should be reported.</p>	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Can’t answer <input type="checkbox"/> Not applicable
<p><b>7) Was the scientific quality of the included studies assessed and documented?</b>          Note: methods of assessment were provided <i>a priori</i>. For example, a quality scoring tool or checklist was used, such as a Jadad scale, risk of bias, sensitivity analysis, etc., or a description of quality items, with some kind of result for EACH study (“low” or “high” is fine, as long as it is clear which studies scored “low” and which scored “high”; a summary score/range for all studies is NOT acceptable).</p>	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Can’t answer <input type="checkbox"/> Not applicable
<p><b>8) Was the scientific quality of the included studies used appropriately in formulating conclusions?</b>          Note: interpretation and analysis of the methodological rigor and quality of the included studies should be clear stated in the conclusions and explicitly stated in formulating recommendations. For example, “results should be interpreted with caution due to poor quality of included studies” is a reasonable interpretation. Cannot score “yes” for this question if scored “no” for question #7.</p>	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Can’t answer <input type="checkbox"/> Not applicable
<p><b>9) Were the methods used to combine the findings of studies appropriate?</b>          Note: for the pooled results, a test should be performed to test for heterogeneity (ie, Chi-squared test, I<sup>2</sup>). If heterogeneity exists, a random effects model was used, an explanation for inability to combine study results due to heterogeneity was given, or the clinical appropriateness of combining individual study results was considered (ie, is it sensible to combine?).</p>	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Can’t answer <input type="checkbox"/> Not applicable
<p><b>10) Was the likelihood of publication bias assessed?</b>          Note: an assessment of publication bias was made and a graphical aid was provided (eg, funnel plot) and/or statistical tests (eg, Egger regression test) were included. Alternatively, if few studies were included, the review mentions that publication bias could not be assessed.</p>	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Can’t answer <input type="checkbox"/> Not applicable
<p><b>11) Was the conflict of interest stated?</b>          Note: potential sources of support should be clearly acknowledged in both the systematic review AND is acknowledged for the included studies.</p>	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Can’t answer <input type="checkbox"/> Not applicable

## 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](http://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**). 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
<b>SCOPE AND PURPOSE</b>	
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. <b>[SCORE: ]</b>
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. <b>[SCORE: ]</b>
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. <b>[SCORE: ]</b>
<b>STAKEHOLDER INVOLVEMENT</b>	
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. <b>[SCORE: ]</b>
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. <b>[SCORE: ]</b>
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. <b>[SCORE: ]</b>
<b>RIGOR OF DEVELOPMENT</b>	
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. <b>[SCORE: ]</b>
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. <b>[SCORE: ]</b>
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. <b>[SCORE: ]</b>
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). <b>[SCORE: ]</b>

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. <b>[SCORE: ]</b>
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. <b>[SCORE: ]</b>
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. <b>[SCORE: ]</b>
14	A procedure for updating the guideline is provided.	A clear statement about the procedure for updating the guideline should be provided. <b>[SCORE: ]</b>
<b>CLARITY OF PRESENTATION</b>		
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. <b>[SCORE: ]</b>
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. <b>[SCORE: ]</b>
17	Key recommendations are easily identifiable	Users should be able to find the most relevant recommendations easily. <b>[SCORE: ]</b>
<b>APPLICABILITY</b>		
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. <b>[SCORE: ]</b>
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. <b>[SCORE: ]</b>
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. <b>[SCORE: ]</b>
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). <b>[SCORE: ]</b>
<b>EDITORIAL INDEPENDENCE</b>		
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. <b>[SCORE: ]</b>
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. <b>[SCORE: ]</b>

**Table 2. AGREE II Global Rating Scale.**

ITEM	DESCRIPTION
1	Rate the guideline development methods. <b>[SCORE: ]</b> <ul style="list-style-type: none"> <li>• Appropriate stakeholders were involved in the development of the guideline.</li> <li>• The evidentiary base was developed systematically.</li> <li>• Recommendations were consistent with the literature. Consideration of alternatives, health benefits, harms, risks, and costs was made.</li> </ul>
2	Rate the guideline presentation. <b>[SCORE: ]</b> <ul style="list-style-type: none"> <li>• The guideline was well organized.</li> <li>• The recommendations were easy to find.</li> </ul>
3	Rate the guideline recommendations. <b>[SCORE: ]</b> <ul style="list-style-type: none"> <li>• The recommendations are clinically sound.</li> <li>• The recommendations are appropriate for the intended patients.</li> </ul>
4	Rate the completeness of reporting, editorial independence. <b>[SCORE: ]</b> <ul style="list-style-type: none"> <li>• The information is complete to inform decision making.</li> <li>• The guideline development process is transparent and reproducible.</li> </ul>



## 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
<b>High</b>	<ul style="list-style-type: none"> <li>Evidence is based on data derived from multiple randomized controlled trials with homogeneity with regard to the direction of effect between studies AND</li> <li>Evidence is based on multiple, well-done randomized controlled trials that involved large numbers of patients.</li> </ul>
<b>Moderate</b>	<ul style="list-style-type: none"> <li>Evidence is based on data derived from randomized controlled trials with some conflicting conclusions with regard to the direction of effect between studies OR</li> <li>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</li> <li>Some evidence is based on data derived from randomized controlled trials with significant methodological flaws (eg, bias, attrition, flawed analysis, etc.)</li> </ul>
<b>Low</b>	<ul style="list-style-type: none"> <li>Most evidence is based on data derived from randomized controlled trials with significant methodological flaws (eg, bias, attrition, flawed analysis, etc.) OR</li> <li>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</li> </ul>
<b>Insufficient</b>	<ul style="list-style-type: none"> <li>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</li> <li>Evidence is based on data derived from expert opinion/panel consensus, case reports or case series OR</li> <li>Evidence is not available</li> </ul>

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:

<b>QUALITY OF EVIDENCE GRADES:</b>	
• ≥4 points	= <b>HIGH</b>
• 3 points	= <b>MODERATE</b>
• 2 points	= <b>LOW</b>
• ≤1 point	= <b>INSUFFICIENT</b>

An example evidence grade assessment of individual randomized controlled trials is available in **Table 3**, with a template provided in **Table 4**.

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

<b>DOMAIN</b>	<b>DESCRIPTION</b>	<b>SCORE DEMOTION/PROMOTION (start with 4 points)</b>
<b>Risk of Bias (internal validity)</b>	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. <ul style="list-style-type: none"> <li>• <i>Did any studies have important limitations that degrade your confidence in estimates of effectiveness or safety?</i></li> </ul>	<ul style="list-style-type: none"> <li>• No serious limitation: all studies have low risk of bias: <b>(0)</b></li> <li>• Serious limitations: ≥1 trial has high or unclear risk of bias: <b>(-1)</b></li> <li>• Very serious limitations: most studies have high risk of bias: <b>(-2)</b></li> </ul>
<b>Indirectness (applicability)</b>	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. <ul style="list-style-type: none"> <li>• <i>Do studies directly compare interventions of interest in populations of interest using outcomes of interest (use of clinically relevant outcomes)?</i></li> </ul>	<ul style="list-style-type: none"> <li>• Direct: clinically relevant outcomes of important comparisons in relevant populations studied: <b>(0)</b></li> <li>• Indirect: important comparisons missing; surrogate outcome(s) used; or population not relevant: <b>(-1)</b></li> </ul>
<b>Inconsistency</b>	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. <ul style="list-style-type: none"> <li>• <i>Did trials have similar or widely varying results? Can heterogeneity be explained by differences in trial design and execution?</i></li> </ul>	<ul style="list-style-type: none"> <li>• Large magnitude of effect consistent between studies: <b>(+1)</b></li> <li>• Dose-response observed: <b>(+1)</b></li> <li>• Small magnitude of effect consistent between studies: <b>(0)</b></li> <li>• 1 study with large magnitude of effect: <b>(0)</b></li> <li>• 1 study with small magnitude of effect: <b>(-1)</b></li> <li>• Inconsistent direction of effect across studies that cannot be explained: <b>(-1)</b></li> </ul>
<b>Imprecision</b>	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). <ul style="list-style-type: none"> <li>• <i>Are confidence intervals for treatment effect sufficiently narrow to rule out no effect?</i></li> </ul>	<ul style="list-style-type: none"> <li>• Precise: all studies have 95% confidence intervals that rule out no effect: <b>(0)</b></li> <li>• Imprecise: ≥1 study demonstrated 95% confidence interval fails to rule out no effect: <b>(-1)</b></li> </ul>
<b>Publication Bias</b>	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. <ul style="list-style-type: none"> <li>• <i>Is there evidence that important trials are not represented?</i></li> </ul>	<ul style="list-style-type: none"> <li>• No publication bias: all important trials published or represented: <b>(0)</b></li> <li>• Serious publication bias: ≥1 important trial(s) completed but not published: <b>(-1)</b></li> </ul>

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

**Table 3. Example Grade Assessment of Evidence from Randomized Controlled Trials.**

Summary of Findings							Evidence Profile					
Study (NCT)	n	Control Results	Intervention Results	Difference (95% CI)	NNT/NNH	Risk of Bias	Overall Risk of Bias	Indirectness	Inconsistency	Imprecision	Publication Bias	Quality of Evidence
<b>Outcome (1): Response (defined as achieve LDL &lt;100 mg/dL and HDL &gt;40 mg/dL)</b>												
Montgomery NCT26240434	553	Placebo 89/277 (32.1%)	Megastatin 40 mg 115/276 (41.7%)	9.6% RR 1.30 (1.04 to 1.62)	10	High	-1 (40 mg) 0 (>40 mg)	-1 (all)	+2 (all)	-1 (40 mg) 0 (>40 mg)	-1 (all)	LOW (40 mg) HIGH (>40 mg)
Asnis NCT96123239	704	Placebo 51/175 (29.1%)	Megastatin 40 mg 64/176 (36.4%)	7.3% RR 1.25 (0.92 to 1.69)	NS	Low						
			Megastatin 80 mg 79/177 (44.9%)	15.8% RR 1.54 (1.16 to 2.04)	6							
			Megastatin 160 mg 85/176 (48.3%)	19.2% RR 1.66 (1.26 to 2.19)	5							
Unpublished NCT01034462	434	NR	NR									
Unpublished NCT00225511	590	NR	NR									
<b>Outcome (2): Change in CRP from baseline</b>												
Montgomery NCT26240434	553	Placebo -0.20 mg/L	Megastatin 40 mg -0.85 mg/L	0.65 mg/L (0.25 to 1.05 mg/L)	NA	High	-1 (40 mg) 0 (>40 mg)	-1 (all)	+1 (all)	-1 (40 mg) 0 (>40 mg)	-1 (all)	INSUFFICIENT (40 mg) MODERATE (>40 mg)
Asnis NCT96123239	704	Placebo +0.15 mg/L	Megastatin 40 mg -0.20 mg/L	0.35 mg/L (-0.05 to 0.60 mg/L)	NA	Low						
			Megastatin 80 mg -0.25 mg/L	0.40 mg/L (0.05 to 0.70 mg/L)	NA							
			Megastatin 160 mg -0.45 mg/L	0.60 mg/L (0.45 to 0.70 mg/L)	NA							
Unpublished NCT01034462	434	NR	NR									

**Table 4. Grade Assessment Template for Randomized Controlled Trials.**

Summary of Findings							Evidence Profile					
Study (NCT)	n	Control Results	Intervention Results	Difference (95% CI)	NNT/NNH	Risk of Bias	Overall Risk of Bias	Indirectness	Inconsistency	Imprecision	Publication Bias	Quality of Evidence
<b>Outcome (1):</b>												
<b>Outcome (2):</b>												