GRADING OF EVIDENCE FOR META-ANALYSES
American College of Emergency Physicians Clinical Policies Committee

Citation: _____

	Rx	Dx	Px
Design 1	**MA of RCTs**	**MA of prospective cohort or cross-sectional studies**	**MA of prospect cohort studies**
Design 2	**MA of Non-RCTs**	**MA of retrospective studies**	**MA of retrospective studies**

Applicable to Clinical Question: Direct Indirect Not

Dimensions for Grading (consider both quality of execution and importance to result):
NR/NA/U: Not reported, not applicable, or unclear.

			Comments
Comprehensive article search	Y	N	NR / NA / U
Study inclusion criteria defined	Y	N	NR / NA / U
At least 2 investigators conducted search/study selection	Y	N	NR / NA / U
Quality of individual studies adequately described	Y	N	NR / NA / U
Appropriate methods used to assess heterogeneity	Y	N	NR / NA / U
Appropriate methods used to combine and report results	Y	N	NR / NA / U
Sensitivity analyses or regression model to account for study differences, including 'high-quality' studies if some studies not uniformly at low risk of bias	Y	N	NR / NA / U
Generalizability	Y	N	NR / NA / U
Data managed appropriately	Y	N	NR / NA / U
Analyses appropriate	Y	N	NR / NA / U
Conclusions supported by the results	Y	N	NR / NA / U
Industry sponsored	Y	N	NR / NA / U

Downgrading:
- ☐ No downgrading (no methodological limitations and directly applicable)
- ☐ Downgrade 1 level (only minor methodological limitations)
- ☐ Downgrade 1 level (indirectly applicable)
- ☐ Downgrade 2 levels (major methodological limitation[s])
- ☐ Fatally flawed or not applicable

Class of Evidence: **I** **II** **III** **X**

Notes:

Reviewer: _____ Date: _____

Guidelines for Use:

1. Use the top grid to assign a **Design** (1 or 2) based on the type of meta-analysis. To qualify as **Design 1**, the meta-analysis has to include results from experimental or prospective observational studies.
2. **Applicability to the clinical question** relates to whether the study being evaluated is directly, indirectly, or not applicable to the clinical question proposed as part of the clinical policy.
3. Then assess the quality of the execution of the meta-analysis using the list of important dimensions as reminders. Important dimensions to be considered when assessing the quality include:
   a. **A comprehensive search** should include and describe appropriate search strategies (including explicit search terms) and databases used (e.g., MEDLINE, Cochrane Library, EMBASE, Web of Science, CINAHL, etc.). Also, use of a medical librarian.
   b. A clear description of **how studies were included**. This description should also include a flow diagram.
   c. To avoid problems with selection **bias**, at least 2 investigators should participate in article selection process.
   d. **Quality of individual studies** should also be appropriately assessed and reported. Flaws in the design or conduct of a study can result in bias, and in some cases this can have as much influence on observed effects as that of treatment. Important intervention effects, or lack of effect, can therefore be obscured by bias. Recording the strengths and weaknesses of included studies provides an indication of whether the results have been unduly influenced by aspects of study performance.
   e. **Heterogeneity** in meta-analyses refers to the variation in study outcomes between studies. Some clinical heterogeneity is expected when studies enroll different patient populations in different settings; however, what we would like to know is whether there is more variation across study results than we would expect by chance alone. This measure of heterogeneity is typically expressed as I^2. It ranges between 0% and 100% with lower values representing less heterogeneity and values >50% indicating substantial heterogeneity.
   f. **Combining of results**: Fixed effect model: This method is only justified when minimal statistical heterogeneity is identified (e.g., I^2 < 25%). Fixed effects methods are based on the assumption that a single common (or fixed) effect underlies every study in the meta-analysis. Under this assumption, if every study were infinitely large, all would yield identical results. This approach results in narrower confidence intervals. Random effects model: A meta-analysis based on a random effects model assumes that individual studies are estimating different treatment effects. Using a random effects model is a more conservative approach since it accounts for the variation between studies and within each individual study, thus, typically yielding wider confidence intervals. A sensitivity analysis is a repeat of the meta-analysis, substituting alternative decisions or ranges of values for decisions that were arbitrary or unclear. This should describe analyses aimed at determining whether conclusions are robust. For example, if some studies in the meta-analysis were at high risk of bias, a sensitivity analysis may repeat the meta-analysis including only studies at low risk of bias to determine if the effect size changes when low quality studies are excluded.
   g. **Generalizability** refers to the ability to generalize the study's results to other patients or settings. Consider the representativeness of the patient population included in the study.
   h. **Data management** refers to whether the data were appropriately handled during collection and analyses; this may include whether authors had access to data and who performed analyses.
   i. **Analyses** should be appropriate and valid for the study design (e.g., appropriate use of multivariable methods, including meta-regression, reporting results separately for studies with large proportions of total patients, etc.).
   j. **Conclusions supported by results** refers specifically to whether the conclusions are appropriately aligned with reported results or whether the authors took liberty in over- or under-extending their conclusions.
   k. **Industry sponsored** studies often are influenced, either in their design, performance, or reporting, by the company, which may introduce bias. Who controlled and analyzed the data? Likely less important for SR/MAs.
4. At the **Downgrading** section, summarize the quality of execution and applicability to the clinical question into a decision on downgrading. The idea here is that the maximum evidence class that can be assigned is limited by the Design (i.e., Design 1 can support up to Class of Evidence I, but Design 2 can only support Class of Evidence II or lower, and so on). Essentially, the quality of execution is used to "downgrade" studies from the maximum class, as shown in the table below. Additionally, applicability to the clinical question also relates to downgrading. (e.g., not applicable studies receive a Class of Evidence "X"). Evidence Class X studies will not be used to support clinical policies. Use the downgrading results to generate a **Class of Evidence** based on the table below.

	Design		
Downgrading	1	2	3
None	I	II	III
1 level	II	III	X
2 levels	III	X	X
Fatally flawed or NA	X	X	X