4. Assess the Quality of Evidence for Each Outcome Across Studies - Create an Evidence Profile Table.

Notes:

- Evidence from the data abstraction must be summarized across studies for each important outcome for the PICO practice question. All outcomes are presented together in one Evidence Profile (EP) table. An EP table includes a detailed quality assessment in addition to reporting the summary of findings (See Appendix 7).
- EP tables will be created using GRADEpro – refer to detailed instructions for using GRADEpro (in Dropbox – PEN GRADE Process – Tools: Author_Using GRADEpro.docx)
- If your search has discovered a guideline based on a high quality systematic review with GRADE evidence tables (i.e. GRADE evidence profile or summary of findings table by outcome), this information can be copied directly into the EP table.
- The GRADE approach results in an assessment of the quality of the body of evidence into one of 4 grades:
    - High – we are very confident that the true effect lies close to that of the estimate of the effect
    - Moderate – we are moderately confident in the effect estimate
    - Low – our confidence in the effect estimate is limited
    - Very Low - we have very little confidence in the effect estimate

4.1 Assess the quality of evidence across studies for each outcome

The GRADE approach to rating the quality of evidence starts with the study design:

- Randomized controlled trials (RCTs) start as high quality evidence
- Non-randomized studies (NRS) start as low quality evidence

A. There are 5 factors that can downgrade the quality of evidence rating for both RCTs and NRS (see Table 1 below). For each outcome assess the following: limitations (risk of bias), inconsistency, indirectness, imprecision and publication bias. For explanations of criteria for downgrading, see – Appendix 8 Worksheet Table to Assess the Quality of Evidence Across Studies Using GRADE

Table 1. Factors that can reduce the quality of the evidence

Factor Across Studies	Considerations	GRADE
1. Limitations in study design or execution (risk of bias)	Most information is from studies at low or unclear risk of bias	No serious limitations, do not downgrade; ↓ 1 level if serious
	Proportion of information from studies at high risk of bias is sufficient to affect interpretation of results	↓ 1 level if serious; ↓ 2 levels if very serious
2. Inconsistency of results (unexplained heterogeneity)	Unexplained heterogeneity of importance and CI does not consistently overlap between the included studies	↓ 1 level if serious
	Substantial unexplained heterogeneity of unequivocal importance and CI does not overlap between the included studies	↓ 2 levels if very serious
3. Indirectness of evidence (indirect comparison of intervention; or indirect population, intervention, comparator or outcome),	Use of surrogate outcomes that are somewhat related to a causal pathway (e.g. bone density instead of a direct measure: fractures)	↓ 1 level if serious
	Indirectness of evidence, when there are differences in the comparison of intervention (e.g. A vs B is not available but A was compared with C and B was	↓ 1 or 2 levels if serious or very serious, respectively

	compared with C) or indirect population, intervention or outcome between the question and the available evidence	
4. Imprecision (studies with relatively few patients and few events, with wide confidence intervals (CI))	If the number of patients in a review is < the number of patients using sample size calculation for an adequately powered trial, plus a wide CI	↓ 1 or 2 levels if serious or very serious, respectively
	A CI that includes both appreciable benefits and appreciable harm, unless the sample size is very large	↓ 2 levels if very serious
5. Publication bias (selective publication of studies)	Small studies especially if industry sponsored and/or a funnel plot that suggests bias.	↓ 1 level if strongly suspected

GRADE is not a quantitative system for grading the quality of evidence (2). Grading the quality of evidence requires human judgment. Each factor reflects a continuum within each category and among categories. When the body of evidence is intermediate for a particular factor, the decision about downgrading (or upgrading – see below) a study depends on judgment. GRADE encourages authors to be explicit and transparent by including footnotes to explain their decision. The overall decision to downgrade the evidence should take into consideration all of the factors together. For example, if there was some uncertainty about 3 factors (study limitations, inconsistency and imprecision), but not serious enough to downgrade each of them, authors may decide to give the studies the benefit of the doubt and not downgrade, or authors may decide to rate down the evidence by one level. In either case, authors should explain the rationale behind their choice in a footnote that they decided not to downgrade due to uncertainty; or that they downgraded for one factor and decided not to downgrade for another factor since further lowering the quality of evidence would seem inappropriate.

B. There are 3 factors that can increase the quality of evidence rating for NRS (see Table 2 below). These criteria generally apply to well-conducted NRS that have not been downgraded for any factors shown in Table 1. For explanations of criteria for upgrading, see Appendix 8 – Worksheet Table to Assess the Quality of Evidence Across Studies Using GRADE

Table 2. Factors that can increase the quality of the evidence

Factor Across Studies	Considerations	GRADE
1. Strong Association (large and consistent estimates of effect)	Large magnitude of effect (e.g. RR>2 or <0.5) based on consistent evidence from at least 2 studies with no plausible residual confounding.	↑ 1 level
	Very large magnitude of effect (e.g. RR>5 or <0.2) based on direct evidence and no serious problems with risk of bias or precision (sufficiently narrow confidence intervals) with no plausible residual confounding.	↑ 2 levels
2. All Plausible Confounding	All plausible confounding would reduce the demonstrated effect or suggest a spurious effect if no effect was observed. For example, if only sicker patients receive an exposure, yet they still fared better, it is likely that the actual exposure effect is larger than the data suggest. This is opposite to the usual effect seen by confounding.	↑ 1 level
3. Dose-response Relation	A dose-response gradient is identified	↑ 1 level

4.2 Use GRADEpro to create an Evidence Profile Table

GRADEpro is free software created by the GRADE team to help with the process of producing recommendations using the GRADE process. GRADEpro can be downloaded here: www.gradepro.org
GRADEpro allows one to create an EP table by filling out a table generated in the software (see Appendix 7 for an example of an EP table that could be sent to reviewers that includes a detailed quality assessment). Refer to the detailed instructions for *Using GRADEpro* (in Dropbox – PEN GRADE Process – Tools: Author_Using GRADEpro.docx)

The advantage of using GRADEpro is that the table is automatically generated and the user is prompted for information. The help button on the screen provides useful advice and the table can be exported as a pdf or into MS word.

4.3 Recommended Readings / Resources for Assessing Quality of Evidence:

- Training modules: http://cebgrade.mcmaster.ca/index.html
    - Assessing Inconsistency
    - Assessing Indirectness
    - Assessing Imprecision
    - Assessing Publication Bias
    - Other Factors and upgrading