

Step 4 - Appraise

Appraisal involves taking the acquired information (identified in [Step 3: Acquire](#)) and critiquing it by systematically reviewing its strengths and weaknesses, as it relates to the reliability, validity and application to your specific question (identified in [Step 2: Ask](#)).

Table of Contents

1. Appraising Study Design
2. Risk of Bias & Quality of Evidence
3. Critical Appraisal Tools
4. PEN® Tools – PEN Evidence Grading Checklist
5. External Tools

For the development of PEN content, when summarizing evidence, it is important to evaluate the body of evidence on a topic that is found in the specific studies and guidelines available. Critical appraisal is an objective, structured approach to better understand a study's strengths and weaknesses allowing for the identification of evidence that ideally comes from rigorous, reliable, unbiased and methodologically appropriate research, while also acknowledging the limitations.

Once the appropriate, available evidence has been acquired, [Step 3: Acquire](#) you need to consider the study design using a hierarchical approach. This approach is explained in more detail below. Using higher level evidence (such as systematic reviews and meta-analyses, followed by randomized controlled trials), then prospective cohorts, along with guidelines, that address the question of interest can help reduce potential bias and increase confidence in the evidence. A hierarchy of evidence provides a frame of ranking for the general quality of evidence obtained from specific study designs. This is further described below and in the [Research Methods](#) document section on "Hierarchy of Evidence", as well as in Table 1, which provides a list of common types of study designs along with some of their main strengths and weaknesses. Determination of the study design aids in assessing the study quality and risk of bias.

1. Appraising Study Design

Study's designs come in many forms, each with its inherent strengths and limitations. A description of the characteristics, strengths and limitations of commonly used study designs is available in the section on *Types of research data & study design* in [Research Methods Background](#). While the type of study design used to answer a particular research question depends on a number of factors including the nature of the question and the availability of resources, there are some designs with methodologies that are considered more rigorous, reliable and less biased. The hierarchy of study designs shown in Figure 1 is a general way of considering how *valid* the evidence may be and it is largely based on level of bias that may be associated with the type of study design.

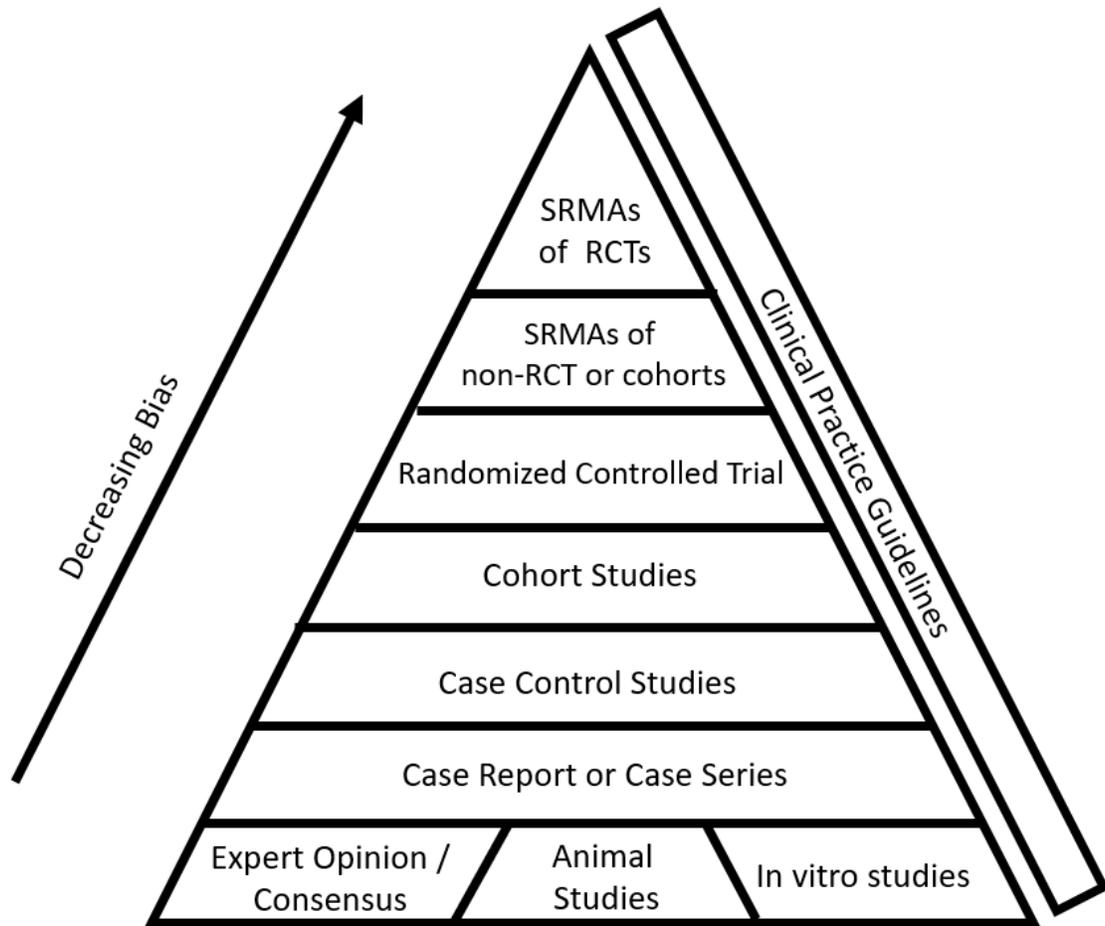
Validity refers to the degree to which a result from a study is likely to be true and free from bias (1). Interpretation of findings from a study depends on both the quality or certainty of evidence as well as the applicability of the findings as described in [Appendix 1 - PEN Evidence Grading Checklist](#).

From time-to-time there may be a situation where there is no evidence to support a well-accepted fact. In this case we refer to the fact as a truism, which is defined as "an un-doubted or self-evident truth" (2). An example may be "Boiling water coming into direct contact with human skin will burn the

skin.” Even though, the only evidence available for this may be case reports and anecdotes, the physiological rationale and basic science would support this as a truism and would warrant a higher evidence grade.

Take the following pyramid scale into consideration when doing your appraisal:

Figure 1. General Hierarchy of Study Designs (Adapted from Murad et al. 2016 (3))



Abbreviations: RCT = randomized controlled trial; SRMAs = systematic reviews and meta-analyses.

The types of study designs that appear at or near the top of the pyramid may be more valid or believable due to less bias, whereas results from study designs appearing at or near the bottom of the pyramid may be less valid or believable, due to potentially more inherent bias. This pyramid provides a general guide; however, it is also necessary to evaluate each study and guideline using your own critical thinking to assess any potential biases and ensure the evidence is reliable, valid and applicable.

Clinical practice guidelines (CPGs) are important to consider in any evidence review. However, the information presented in guidelines may or may not be based on evidence from a systematic review of your topic of interest and may not necessarily include evidence specific to nutrition elements of care. If CPGs were developed using systematic review methods and their

recommendations are generally graded, the evidence may be considered equivalent to a systematic review (i.e. top quality evidence). However, if the guidelines are based on consensus, they are equivalent to expert opinion (i.e. lower quality evidence). The possible range of quality of CPGs is the reason why CPGs are shown in the Figure 1 Hierarchy as extending from the top at low risk of bias down to the bottom at high risk of bias.

Observational studies, which include cohort, cross-sectional and case-control studies, only provide observations. These observations could be confounded by related variables, thus observational studies cannot be taken as proof of causation. This is where the phrase 'correlation, does not imply causation' comes from. For example, adults that dance are more likely to fall over may be an observation; however, the causation may be alcohol consumption. Another example is, people who eat vegetarian diets are less likely to smoke and more likely to have healthy lifestyles (e.g. more exercise, non-smoking, drink alcohol only moderately), so it is not possible to identify meat avoidance as the only causal factor in promoting health. When describing findings from observational studies, ensure you use words such as there was a 'relationship' or 'association' between dancing or vegetarian diets (the exposure variable) and falls or other health measures (the outcome variable).

If randomized controlled trials are well designed, and consistent results are seen across several RCTs, then, for the same PICO conditions, it may be possible to assume causation. For example, a trial in which individuals are randomized to eat vegetarian diets compared to those not eating vegetarian diets may conclude that meat avoidance is a causal factor for health outcomes if other factors are equal between groups (e.g. smoking status, physical activity, alcohol consumption). Well conducted randomization should create similar intervention and control groups, but this can only be assumed if the randomization is conducted in a way that the investigators are prevented from selecting which group people are allocated to (referred to as concealed allocation). The term "effect" may also be used when describing the outcome as it relates to the intervention in these trials.

Systematic reviews and meta-analyses systematically collect data about a specific topic usually from similar study designs and statistically synthesizes data. However, the quality of the evidence is dependent on the quality of the included studies and the quality of the systematic review conduct. Thus, even if there is a systematic review and meta-analysis available addressing your specific question it is still important to appraise strengths and weaknesses of the evidence by assessing any potential bias and the overall quality of the evidence. Also, the quality of the conduct of systematic reviews and/or meta-analyses varies from very strong to weak. Sometimes authors refer to their work as a systematic review when they have not followed many systematic review methods (Refer to Quality Assessment tools in Table 1 below).

Note: a narrative review differs from a systematic review in that it provides an overview of a topic area without considering or appraising all available evidence.

2. Risk of Bias & Quality of Evidence

Bias refers to a systematic error or deviation from the truth in results and inferences, meaning that multiple replications of the same study would likely lead to the wrong answer on average due to inherent errors (1). Bias can vary in magnitude, from trivial to substantial effects, and may lead to either underestimation or overestimation of the effect on the true outcome. A biased study loses validity with respect to the degree of the bias. While some study designs are more prone to bias, its presence is universal. Thus, even if a study has a systematic review and meta-analysis or randomized controlled trial design (i.e. a study design considered to be at the top of the evidence hierarchy), it can still be affected by bias.

Appraising risk of bias is part of a quality of evidence assessment. Tools often used to assess bias in randomized trials and observational studies are identified in Table 1 below.

Quality or *certainty* of evidence refers to the confidence that an estimate of effect or association is near the true value for the effect on an outcome (1). When appraising the quality of evidence, judgments about the quality of evidence are relative to the specific context in which the evidence is used. Aspects that are often considered when appraising quality of evidence include: risk of bias (study limitations), inconsistency, imprecision (size, variability, and direction of effect), indirectness (generalizability) and risk of publication bias (reporting bias) (4). These are outlined in [Appendix 1 – PEN Evidence Grading Checklist](#).

For more information about study design, bias and quality/certainty see the [Research Methods Background](#), section titled, *Types of research data & study design*.

3. Critical Appraisal Tools

Many tools are available for assessing risk of bias and quality of studies (5,6). Most tools are checklists that ask specific questions or are scales in which various components of bias and/or quality are scored and combined to give a summary score (7). It is important to be aware of these types of tools so that you can recognize when they are used in systematic reviews, and you can use them in your own appraisals of the evidence. Practice questions in the PEN System use the PEN evidence grading tool, described below to appraise the quality of each evidence statement. The PEN Evidence Grading Checklist utilizes a checklist format and provides a grade from A to D, with ‘A’ signifying “Good” evidence and ‘D’ denoting evidence where a conclusion is either not possible or extremely limited because evidence is unavailable and/or of poor quality and/or is contradictory.

4. PEN® Tools – PEN Evidence Grading Checklist

Using the [Evidence Grading Checklist – Appendix 1](#), appraise your materials to establish the quality of the evidence related to your questions. If you are feeling your critical appraisal skills are rusty or want to gain a better sense of how to effectively use the worksheets, review the relevant sections in the two Writer’s Training modules:

- [Evidence-based Process Module](#)
- [Appraising the Literature Module](#).

5. External Tools

Table 1 provides examples of risk of bias and quality of evidence tools. This is not a complete list, and these tools may also be updated by the development organization.

Note: PEN authors do not necessarily need to use these tools when appraising the evidence, but you may find it helpful to familiarize yourself with them because they are used frequently in systematic reviews.

Table 1. Risk of Bias (RoB) and Quality of Evidence Assessment Tools

Tool Name	Use	Development Organization
Cochrane RoB (8)*	Randomized controlled trials (RCTs)	The Cochrane Collaboration
RoB 2.0*	RCTs	The Cochrane Collaboration

ROBINS-1 tool*	Non-randomized studies (NRS)	Cochrane Bias Methods Group
JADAD Scale	RCTs	Jadad et al., 1996 (9)
Newcastle-Ottawa Scale	Observational studies	Wells et al., 2000 (10)
CASP Checklist	Various checklists for systematic reviews (SRs), RCTs, observational studies, qualitative studies	The Critical Appraisal Skills Programme (CASP)
JBI Critical Appraisal Checklist	Various checklists for SRs, RCTs, observational studies, qualitative studies	The Joanna Briggs Institute (JBI)
GRADE*	Assesses certainty of evidence from RCTs or NRS for each outcome	Schunemann et al., 2013 (4)
NutriGRADE	Adaptation of GRADE for nutrition research	Schwingshackl et al., 2016 (11)
AMSTAR 2*	SRs	AMSTAR Team, Bruyère Research Institute

*Indicates more rigorously developed tools

Adapted from Ma et al., 2020 (12).

References

- Higgins JPT, Altman DG, Sterne JAC (editors). Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Churchill R, Chandler J, Cumpston MS (editors), Cochrane Handbook for Systematic Reviews of Interventions version 5.2.0 (updated June 2017). 2017. Available from: www.training.cochrane.org/handbook
- "Truism." Merriam-Webster.com Dictionary, Merriam-Webster [cited 2021 July 31]. Available from: <https://www.merriam-webster.com/dictionary/truism>
- Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. Evid Based Med. 2016 Aug;21(4):125-7. doi: 10.1136/ebmed-2016-110401. Epub 2016 Jun 23. PMID: 27339128; PMCID: PMC4975798. Abstract available from: <https://pubmed.ncbi.nlm.nih.gov/27339128/>
- Schünemann H, Brożek J, Guyatt G, Oxman, A. GRADE Handbook. 2013. Available from: <https://gdt.gradeapro.org/app/handbook/handbook.html>
- Zeng X, Zhang Y, Kwong JS, Zhang C, Li S, Sun F, Niu Y, Du L. The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. J Evid Based Med. 2015 Feb;8(1):2-10. doi: 10.1111/jebm.12141. PMID: 25594108. Abstract available from: <https://pubmed.ncbi.nlm.nih.gov/25594108/>
- Ma LL, Wang YY, Yang ZH, Huang D, Weng H, Zeng XT. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? Mil Med Res. 2020 Feb 29;7(1):7. doi: 10.1186/s40779-020-00238-8. PMID: 32111253; PMCID: PMC7049186. Abstract available from: <https://pubmed.ncbi.nlm.nih.gov/32111253/>
- Higgins JPT, Savović J, Page MJ, Elbers RG, Sterne JAC. Chapter 8: Assessing risk of bias in a randomized trial. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). 2021. Available from: www.training.cochrane.org/handbook
- Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L, Sterne JA; Cochrane Bias Methods Group; Cochrane Statistical Methods Group. The

- Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011 Oct 18;343:d5928. doi: 10.1136/bmj.d5928. PMID: 22008217; PMCID: PMC3196245. Abstract available from: <https://pubmed.ncbi.nlm.nih.gov/22008217/>
9. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials*. 1996 Feb;17(1):1-12. doi: 10.1016/0197-2456(95)00134-4. PMID: 8721797. Abstract available from: <https://pubmed.ncbi.nlm.nih.gov/8721797/>
 10. Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
 11. Schwingshackl L, Knüppel S, Schwedhelm C, Hoffmann G, Missbach B, Stelmach-Mardas M, Dietrich S, Eichelmann F, Kontopantelis E, Iqbal K, Aleksandrova K, Lorkowski S, Leitzmann MF, Kroke A, Boeing H. Perspective: NutriGrade: A Scoring System to Assess and Judge the Meta-Evidence of Randomized Controlled Trials and Cohort Studies in Nutrition Research. *Adv Nutr*. 2016 Nov 15;7(6):994-1004. doi: 10.3945/an.116.013052. Erratum in: *Adv Nutr*. 2017 Mar 15;8(2):399. PMID: 28140319; PMCID: PMC5105044. Abstract available from: <https://pubmed.ncbi.nlm.nih.gov/28140319/>
 12. Ma LL, Wang YY, Yang ZH, Huang D, Weng H, Zeng XT. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? *Mil Med Res*. 2020 Feb 29;7(1):7. doi: 10.1186/s40779-020-00238-8. PMID: 32111253; PMCID: PMC7049186. Abstract available from: <https://pubmed.ncbi.nlm.nih.gov/32111253/>

PEN Evidence Grading Checklist

The conclusion is supported by GOOD evidence (A)

Depending on the effect size, recommendation statements could be written asⁱ:

- 'X' reduces / increases outcome
- 'X' reduces / increases outcome slightly
- 'X' does not reduce / increase outcome

<p>1. Quality or certainty of evidence</p> <p>The results for a specific intervention/outcome are from high quality studies for answering the practice question as described in the bullet points below. Supporting studies might consist of:</p> <p><u>Treatment/Intervention Studies</u></p> <ul style="list-style-type: none"> • good quality systematic review (SR) and meta-analysis (MA) of randomized controlled trials (RCTs) that meet most of the criteria in bullet points below • two or more high quality randomized, controlled trials that meet most of the criteria in bullet points below. <p><u>Etiology/Prognosis Studies</u></p> <ul style="list-style-type: none"> • SR and MA of cohort studies or two or more independent well-done prospective cohort studies that meet most of the criteria described in bullet points below and where treatment/exposure effects are sufficiently large (e.g., 2-fold increase or decrease in one group compared to another), where a dose-response gradient is reported or when residual confounders would be likely to decrease the effect (e.g. sicker patients receive the exposure but still fare better). <p>Note: Evidence might also be in a position statement or practice guideline from a national body or organization reporting results of research studies based on the aforementioned types of research or assessed as high certainty of evidence using GRADE process.</p>		√
• Risk of bias or study limitations ⁱⁱ - results are generally at low risk of bias with no apparent study limitations.		
• Inconsistency ⁱⁱⁱ - results are generally consistent (low heterogeneity, e.g. I ² < 40%).		
• Imprecision ^{iv} – effect estimates are generally precise if able to be determined, otherwise indicate not applicable (NA).		
• Indirectness ^v – results are generalizable to the population being assessed.		
• Risk of publication bias or reporting bias ^{vi} – not evident if assessed in MA, otherwise indicate not applicable (NA).		
<p>2. Importance/Relevance^{vii}</p>		
• Clinical impact/balance between desirable and undesirable effects – the results are clinically important with a large gradient between benefits and risks.		
• Acceptability/values and preferences – most patients would value the outcomes as important and be willing to accept the intervention.		
• Applicability/costs – results are applicable to the practice setting and resource implications are justified.		

The conclusion is supported by FAIR evidence (B)

Depending on the effect size, recommendation statements could be written asⁱ:

- 'X' probably reduces / increases outcome ro
- 'X' probably reduces / increases outcome slightly

'X' probably does not reduce / increase outcome	
1. Quality or certainty of evidence	
The results for a specific intervention/outcome are from studies of strong design with minor methodological concerns as described in the bullet points below or from studies with weaker designs for answering the practice question. Supporting studies might consist of:	
<u>Treatment/Intervention Studies</u>	
<ul style="list-style-type: none"> • systematic review (SR) of RCTs with some minor methodological issues as described in bullet points below. • a single RCT with low risk of biasⁱ • two or more RCTs with a clinically significant conclusion and unclear risk of biasⁱ 	
<u>Etiology / Prognosis Studies</u>	
<ul style="list-style-type: none"> • SR of cohort studies or two or more well-done prospective cohort studies that meet most of the criteria in bullet points below. 	
Note: Evidence might also be in a position statement or practice guideline from a national body or organization reporting results of research studies based on the aforementioned types of research or assessed as moderate certainty of evidence using GRADE process.	
	√
<ul style="list-style-type: none"> • Risk of bias or study limitationsⁱⁱ – results may have unclear risk of bias with some study limitations. 	
<ul style="list-style-type: none"> • Inconsistencyⁱⁱⁱ – results may have minor inconsistencies at most (moderate heterogeneity e.g. I² 30 to 60%). 	
<ul style="list-style-type: none"> • Imprecision^v – effect estimates may have concerns about precision if able to be determined, otherwise indicate not applicable (NA). 	
<ul style="list-style-type: none"> • Indirectness^v – there is minor doubt about generalizability to the population being assessed. 	
<ul style="list-style-type: none"> • Risk of publication bias or reporting bias^{vi} – may be evident if assessed in MA, otherwise indicate not applicable (NA). 	
2. Importance/Relevance^{vii}	
<ul style="list-style-type: none"> • Clinical impact/balance between desirable and undesirable effects – there is minor doubt about the clinical importance of results with a moderate gradient between benefits and risks. 	
<ul style="list-style-type: none"> • Acceptability/values and preferences – some patients would value the outcomes as important and likely be willing to accept the intervention. 	
<ul style="list-style-type: none"> • Applicability/costs – results are generally applicable to the practice setting and resource implications may be justified. 	

The conclusion is supported by LIMITED evidence or expert opinion (C)

Depending on the effect size, recommendation statements could be written asⁱ:

'X' may reduce or may slightly reduce / increase outcome (or may possibly be associated with reduced / increased outcome)

'X' may not reduce / increase outcome

The evidence is uncertain about the effect of 'X' on outcome or on 'Y' population. Individuals may choose to take a precautionary approach.

<p>1. Quality or certainty of evidence</p> <p>The results for a specific intervention/outcome are from studies of weak design for answering the practice question or there is substantial uncertainty attached to the conclusion as described in most of the bullet points below. Supporting studies might consist of:</p> <p><u>Treatment/Intervention Studies</u></p> <ul style="list-style-type: none"> • two or more RCTs with inconsistent results or high risk of bias • non-randomized trial or trial that used historical controls • systematic review (SR) of cohort or case-control studies (with homogeneity) or two or more well-done prospective cohort studies with consistent findings. <p><u>Etiology / Prognosis Studies</u></p> <ul style="list-style-type: none"> • SR of cohort and case-control studies (with heterogeneity) or two or more studies with some inconsistent results • results from a single cohort study or two or more case-control studies, unconfirmed by other studies • results from a number of high-quality cross-sectional studies, well described case reports or case series. <p>Note: Evidence might also be in a position statement or practice guideline from a national body or organization reporting results of research studies based on the aforementioned types of research or based on expert consensus or assessed as low or very low certainty of evidence using GRADE process.</p>		√
• Risk of bias or study limitations ⁱⁱ - results are at a high risk of bias with obvious study limitations.		
• Inconsistency ⁱⁱⁱ - results may be inconsistent (high heterogeneity e.g. I ² >60%).		
• Imprecision ^{iv} – effect estimates may be imprecise if able to be determined, otherwise indicate not applicable (NA).		
• Indirectness ^v – there may be substantial doubt about generalizability to the population being assessed or the outcome is a surrogate (e.g. markers such as blood tests).		
• Risk of publication bias or reporting bias ^{vi} – may be evident if assessed in MA, otherwise indicate not applicable (NA).		
<p>2. Importance/Relevance^{vii}</p>		
• Clinical impact/balance between desirable and undesirable effects – there is uncertainty about the clinical importance of the results with a small to moderate gradient between benefits and risks.		
• Acceptability/values and preferences – there is uncertainty about whether patients would value the outcomes as important and uncertainty about willingness to accept the intervention.		
• Costs/applicability – there is uncertainty about the applicability of the results to the practice setting and resource implications may be difficult to justify.		

A conclusion is either not possible or extremely limited because evidence is unavailable and/or of poor quality and/or is contradictory (D)

Depending on the effect size, recommendation statements could be written asⁱ:

There is no evidence about the effect of 'X' on outcome

The evidence is too uncertain to draw a conclusion about the effect of 'X' on outcome

<p>1. The evidence is very uncertain about the effect of 'X' on outcome or on 'Y' population. Until more evidence becomes available, individuals may choose to.....Quality or certainty of evidence</p> <p>The results for a specific intervention/outcome are from a single study with major design flaws or from studies with contradictory results that meet all of the criteria in bullet points below such that conclusions can't be confidently drawn. Alternatively, evidence is lacking from either authoritative sources or research involving humans. Supporting studies might consist of:</p> <ul style="list-style-type: none"> • a very poorly designed and executed trial or intervention • evidence from a single case report, case series, case-control study or ecological study unconfirmed by other studies • anecdotal reports • evidence from a small number of similar quality studies that report contradictory results (e.g. two cohort studies that report opposite associations) • research in the <i>in vitro</i>, <i>ex vivo</i> or animal model. <p>Note: Evidence might also be in a position statement or practice guideline from a national body or organization reporting results of research studies based on the aforementioned types of research or where no recommendation is able to be provided using GRADE process.</p>	√
<ul style="list-style-type: none"> • Risk of bias or study limitationsⁱ - results are at high risk of bias or with major study limitations. 	
<ul style="list-style-type: none"> • Inconsistencyⁱⁱⁱ - usually inconsistent 	
<ul style="list-style-type: none"> • Imprecision^{iv} – effect estimates are imprecise if able to be determined, otherwise indicate not applicable (NA). 	
<ul style="list-style-type: none"> • Indirectness^v – not generalizable to the population being assessed, very limited generalizability or due to use of surrogate outcomes. 	
<ul style="list-style-type: none"> • Risk of publication bias or reporting bias^{vi} – evident if assessed in MA, otherwise indicate not applicable (NA). 	
<p>2. Importance/Relevance^{vii}</p>	
<ul style="list-style-type: none"> • Clinical impact/balance between desirable and undesirable effects– the results are minimal or there is little to no gradient between benefits and risks. 	
<ul style="list-style-type: none"> • Acceptability/values and preferences – many patients would not value the outcomes or be likely to be concerned about accepting the intervention. 	
<ul style="list-style-type: none"> • Applicability/Costs – results are not applicable or have very limited applicability to the practice setting or have high resource implications. 	

Notes:

ⁱThe recommendation statements are based on GRADE guidance developed to clearly communicate findings of systematic reviews based on the certainty of evidence and size of the effect (1).

ⁱⁱ Risk of bias is an assessment of the validity of studies (i.e. the risk that they over- or underestimate the true effect of the intervention). Tools are available to assess risk of bias in RCTs (2) and observational studies (3, 4). If no quality assessment was conducted, consider study limitations. For additional information refer to PEN Writer's Training Module – [Appraising the Literature](#)

ⁱⁱⁱ Inconsistency refers to unexplained heterogeneity of results. Consistency considers whether findings are consistent across studies, considering the range of study populations and study designs, including the direction and size of the effect or degree of association, and the statistical significance. To evaluate inconsistency, authors should inspect the similarity of point estimates and the overlap of their confidence intervals or look at statistical heterogeneity (e.g. I^2 and P -value from chi-squared test) (5). If the results of the studies are in the same direction with overlapping confidence intervals, inconsistency is deemed unlikely.

^{iv} Imprecision occurs when there is uncertainty about the results and confidence intervals (CI) are wider when the estimate of the effect comes from only one or two small studies or if there are few events or considerable variability in the effects among patients (6). When a 95%CI excludes the possibility of no effect (i.e. does not cross '0'), you can be reasonably confident that there is an effect and that imprecision is unlikely.

^v Indirectness considers how the population, intervention and outcomes in the evidence match the practice question. Evidence is lower if the patients or interventions are different or when the outcome is a surrogate outcome (such as blood levels) rather than an outcome typically important to patients (7).

^{vi} Publication bias considers missing evidence due to selective publication of studies. Statistical and visual methods (e.g. funnel plot, Egger's test) can help to detect publication bias in meta-analyses (8). Publication bias is more common when studies are industry funded and with observational studies.

^{vii} The quality of the evidence is a major factor determining the evidence grade; however, consideration is given to the application of other factors that influence findings, including: impact, acceptability and applicability. In some cases, these factors can supersede the evidence base. For example, a meta-analysis with high certainty of evidence but with little clinical impact can conclude that the results do not increase or decrease the outcome (A grade) (1). However, if there is high certainty of evidence but some concern that the desirable effects do not outweigh the undesirable effects, the evidence could be downgraded (B grade) (9).

-
- J Clin Epidemiol. 2011 Dec;64(12):1283-93. doi: 10.1016/j.jclinepi.2011.01.012. Epub 2011 Aug 11. Erratum in: J Clin Epidemiol. 2021 Jun 23;137:265. PMID: 21839614. Abstract available from: <https://pubmed.ncbi.nlm.nih.gov/21839614/>
6. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Falck-Ytter Y, Jaeschke R, Vist G, Akl EA, Post PN, Norris S, Meerpohl J, Shukla VK, Nasser M, Schünemann HJ; GRADE Working Group. GRADE guidelines: 8. Rating the quality of evidence--indirectness. J Clin Epidemiol. 2011 Dec;64(12):1303-10. doi: 10.1016/j.jclinepi.2011.04.014. Epub 2011 Jul 30. PMID: 21802903. Abstract available from: <https://pubmed.ncbi.nlm.nih.gov/21802903/>
 7. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, Alonso-Coello P, Djulbegovic B, Atkins D, Falck-Ytter Y, Williams JW Jr, Meerpohl J, Norris SL, Akl EA, Schünemann HJ. GRADE guidelines: 5. Rating the quality of evidence--publication bias. J Clin Epidemiol. 2011 Dec;64(12):1277-82. doi: 10.1016/j.jclinepi.2011.01.011. Epub 2011 Jul 30. PMID: 21802904. Abstract available from: <https://pubmed.ncbi.nlm.nih.gov/21802904/>
 8. Santesso N, Glenton C, Dahm P, Garner P, Akl EA, Alper B, Brignardello-Petersen R, Carrasco-Labra A, De Beer H, Hultcrantz M, Kuijpers T, Meerpohl J, Morgan R, Mustafa R, Skoetz N, Sultan S, Wiysonge C, Guyatt G, Schünemann HJ; GRADE Working Group. GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions. J Clin Epidemiol. 2020 Mar;119:126-135. doi: 10.1016/j.jclinepi.2019.10.014. Epub 2019 Nov 9. PMID: 31711912. Abstract available from: <https://pubmed.ncbi.nlm.nih.gov/31711912/>
 9. Schünemann H, Brožek J, Guyatt G, Oxman, A. GRADE Handbook. 2013. Available from: <https://gdt.gradepro.org/app/handbook/handbook.html>