PEN Evidence Grading Checklist

The conclusion is supported by GOOD evidence (A)
Depending on the effect size, recommendation statements could be written as[i]:

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

1. Quality or certainty of evidence

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:

Treatment/Intervention Studies

- 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.

Etiology/Prognosis Studies

- 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).

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.

	√
• 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).	
2. Importance/Relevance[vii]	
• 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 as[i]:

'X' probably reduces / increases outcome
'X' probably reduces / increases outcome slightly
'X' does not reduce / increase outcome
'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:

Treatment/Intervention Studies

- systematic review (SR) of RCTs with some minor methodological issues as described in bullet points below.
- a single RCT with low risk of bias[i]
- two or more RCTs with a clinically significant conclusion and unclear risk of bias[i]

Etiology / Prognosis Studies

- 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.

	√
• Risk of bias or study limitations[ii] – results may have unclear risk of bias with some study limitations.	
• Inconsistency[iii] – results may have minor inconsistencies at most (moderate heterogeneity e.g. I^2 30 to 60%).	
• Imprecision[v] – effect estimates may have concerns about precision if able to be determined, otherwise indicate not applicable (NA).	
• Indirectness[v]– there is minor doubt about generalizability to the population being assessed.	
• Risk of publication bias or reporting bias[vi]– may be evident if assessed in MA, otherwise indicate not applicable (NA).	
2. **Importance/Relevance**[vii]	
• 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.	
• Acceptability/values and preferences – some patients would value the outcomes as important and likely be willing to accept the intervention.	
• 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 as[i]:

'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.

1. Quality or certainty of evidence

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:

Treatment/Intervention Studies

- 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.

Etiology / Prognosis Studies

- 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.

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.

	√
• 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).	
2. Importance/Relevance[vii]	
• 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 as[i]:

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

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** 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: • 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 *in vitro*, *ex vivo* or animal model. 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.	√
• Risk of bias or study limitations[ii] - results are at high risk of bias or with major study limitations.	
• Inconsistency[iii] - usually inconsistent	
• Imprecision[iv] – effect estimates are imprecise if able to be determined, otherwise indicate not applicable (NA).	
• Indirectness[v] – not generalizable to the population being assessed, very limited generalizability or due to use of surrogate outcomes.	
• Risk of publication bias or reporting bias[vi] – evident if assessed in MA, otherwise indicate not applicable (NA).	
2. **Importance/Relevance**[vii]	
• Clinical impact/balance between desirable and undesirable effects– the results are minimal or there is little to no gradient between benefits and risks.	
• Acceptability/values and preferences – many patients would not value the outcomes or be likely to be concerned about accepting the intervention.	
• Applicability/Costs – results are not applicable or have very limited applicability to the practice setting or have high resource implications.	

Notes:

[i]The 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).

References

1. 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/
2. 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/
3. Wells G, Shea B, O'connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses. Ottawa: Ottawa Hospital Research Institute; 2011. Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
4. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, Henry D, Altman DG, Ansari MT, Boutron I, Carpenter JR, Chan AW, Churchill R, Deeks JJ, Hróbjartsson A, Kirkham J, Jüni P, Loke YK, Pigott TD, Ramsay CR, Regidor D, Rothstein HR, Sandhu L, Santaguida PL, Schünemann HJ, Shea B, Shrier I, Tugwell P, Turner L, Valentine JC, Waddington H, Waters E, Wells GA, Whiting PF, Higgins JP. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016 Oct 12;355:i4919. doi: 10.1136/bmj.i4919. PMID: 27733354; PMCID: PMC5062054. Abstract available from: https://pubmed.ncbi.nlm.nih.gov/27733354/
5. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Glasziou P, Jaeschke R, Akl EA, Norris S, Vist G, Dahm P, Shukla VK, Higgins J, Falck-Ytter Y, Schünemann HJ; GRADE Working Group. GRADE guidelines: 7. Rating the quality of evidence--inconsistency. J Clin Epidemiol. 2011 Dec;64(12):1294-302. doi: 10.1016/j.jclinepi.2011.03.017. Epub 2011 Jul 31. PMID: 21803546. Abstract available from: https://pubmed.ncbi.nlm.nih.gov/21803546/
6. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, Devereaux PJ, Montori VM, Freyschuss B, Vist G, Jaeschke R, Williams JW Jr, Murad MH, Sinclair D, Falck-Ytter Y, Meerpohl J, Whittington C, Thorlund K, Andrews J, Schünemann HJ. GRADE guidelines 6. Rating the quality of evidence--imprecision. 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/
7. 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/
8. 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/
9. Schünemann H, Brożek J, Guyatt G, Oxman, A. GRADE Handbook. 2013. Available from: https://gdt.gradepro.org/app/handbook/handbook.html