Editorial

Is the Systematic Review and Meta-Analysis the Gold Standard for Scientific Evidence?

Systematic reviews (SRs) and meta-analyses aim to summarize a substantial body of evidence regarding a specific clinical question and provide a high-quality, valuable summary for clinicians and policymakers.[1] However, SRs can be two-sided swords due to misunderstandings regarding the methodological steps and incorrect interpretations of the results.

The first concept regarding SRs is the difference between SRs and narrative reviews, minireviews, scoping reviews, and other reviews.[1] An SR follows a well-defined prespecified method to answer a PICOS question (population, intervention, comparison, outcome, study design). Moreover, these questions focus on but are not limited to interventions (drugs, devices, others), diagnostic accuracy, and prognosis.[2,3] Additionally, the protocol of each SR should be prospectively registered in an open access database, such as PROSPERO (https://www.crd.york.ac.uk/PROSPERO) or the Cochrane library.[4]

Second, the literature search and study selection process in SRs must be broad and rigorous, including at least two different databases and a combination of keywords and syntax to identify potential studies without language restriction. Several databases, including PubMed (MEDLINE), Embase, LILACS, Clinical Trials (www.clinicaltrials.gov), Cochrane Central Register of Controlled Trials, ScienceDirect, Google Scholar, and the Directory of Open Access Journals, should be explored. Additionally, a manual search of the reference lists of included studies and examinations of meeting abstracts related to the PICOS question, including both industries-funded and nonfunded research, increase the quality of and confidence in SRs.[1–4] For respiratory and sleep medicine, a search of the last five years by the American Thoracic Society (ATS), American College of Chest Physicians (Chest), and European Respiratory Society (ERS) is strongly encouraged.

Two independent reviewers perform the literature search and select the studies. The title and abstract screening process has essentially two steps: a broad database search using a syntax that aims to catch every study ever conducted; then, a precise screening process to reject all the studies that do not respond to the PICOS question. The title and abstract screening process should be done by two independent experts to systematically (using forms or spreadsheets) ascertain whether the article has original data in it and then whether it is relevant to the intended PICOS question. Then, the final study selection for qualitative and quantitative analysis should follow prespecified inclusion and exclusion criteria. Concordance between reviewers should be estimated with the Kappa (K) test.[5] This process follows the PRISMA statement for SR of

intervention,[6] PRISMA-DTA guidelines for diagnostic accuracy,[7] and MOOSE[8] guidelines for SR of observational studies. Additionally, these data should be easy to find in Fig. 1—a Prisma flowchart. Accurate knowledge of this process provides concepts about the transparency and reproducibility of the SR. An SR aims to include every study ever conducted to respond to the PICOS question and do so in a replicable and rigorous manner.

Third, the risk of bias (RoB) should also be assessed independently by two reviewers, with a third review author consulted to resolve any discrepancy between reviewers. There are different objective and subjective tools to assess RoB; for example, the Cochrane RoB tool is recommended for SRs of intervention[1]; the QUADAS-2 is recommended for a diagnostic test[9]; and the Newcastle Ottawa (NOS) and Robins-1 are recommended for observational studies nonrandomized trials.[10,11] A reasonable interpretation of the RoB of included studies is relevant because these decisions affect confidence in the results, and therefore, your Grading of Recommendations Assessment, Development and Evaluation (GRADE) recommendations.[12]

Assessing publication bias is a way to certify that all information was found and included in the SR. It will be evaluated through a visual inspection of a funnel plot and/or statistical test such as Egger's test reported in the meta-analysis. Asymmetry in the distribution of the studies raised concerns about publication bias.[1]

After a correct description of the structure of any SR, the understanding of the results includes a two-step process: (1) a detailed report of the SRs process, previously described and (2) a pooled analysis of the included studies. For this purpose, if the SR correctly identified more than two studies to combine, the meta-analysis is an acceptable method to follow.[13] Otherwise, the SR should be restricted to a qualitative report of the included studies without meta-analysis.

In brief, a meta-analysis is a statistical method that combines the individual result of each study in one effect, increasing the sample size and, therefore, the precision of the intervention. The main component of a meta-analysis includes a pooled weighted measure, precision (confidence interval).[1–3] Each included trial result is pooled in a meta-analysis using the DerSimonian and Laird method.[14] The results are shown in the Forrest plot for each comparison. Continuous data are pooled and analyzed as the mean difference (MD) or weight mean difference (WMD), and categorical data are combined as the relative ratio (RR) and odds ratio (OR). Each analysis could be done using a random effects model or a fixed

effects model. All results should include 95% confidence intervals (CIs) to show the precision of the results.[1]

Interstudy heterogeneity (percentage of total variation across studies not due to chance) is measured using visual inspection of a forest plot and an I^2 test. There is no consensus about the % of thresholds to consider significant heterogeneity. In cases of heterogeneity in the meta-analysis, subgroup analysis can be used to evaluate possible explanations of the difference between groups following a prespecified subgroup analysis considering variables such as RoB, different clinical conditions, or others as potential causes.[14]

Additionally, another form of data analysis includes meta-regression analysis. This analysis explores the potential association between the outcome and prespecified variables; however, the absence of association found on meta-regression, especially when using study-level covariates, does not warrant the assertion that the covariate of individuals is nonrelevant.[15]

The report of the main results of the SR and meta-analysis should be reported using a summary of findings (SoF) table. According to the GRADE approach, the SoF table should predefine the clinically relevant outcomes regarding the PICOS question, providing data about the relative and absolute effect, study characteristics, and quality of the evidence. The main domain from GRADE includes data about (1) RoB of included studies, (2) inconsistency or substantial heterogeneity (I^2) without explanation after subgroup analysis, (3) imprecision between results (Confidence interval of from the meta-analysis), including the minimal clinical important difference for improvement in each outcome, (4) risk of publication bias reported in the meta-analysis through the funnel plot statistical analysis and (5) indirectness related to study intervention or population.[12]

An SR that complies with the described standards should always be more informative than the individual studies that compose it. Nevertheless, the final step in the critical analysis of any SR and meta-analysis includes an assessment of its applicability, which should include the "real world" scenario, including patient values and preferences and cost effectiveness.[1–3]

After a good understanding of the methodological considerations, clinicians and researchers should be aware of the limitations of every SR. The main limitations of any SRs are related to the PICOS question (too general or too specific question); inaccurate literature search and study selection (pooling together different studies, like "apples and oranges"); the quality of the included studies (measure by the RoB); the publication bias; and the heterogeneity included in the studies. Finally, for many cases, especially for translational and basic science research (where study size is restricted to prespecified clinical and environmental conditions and experiments, and therefore, high interstudy variability regarding experiments and methods), SRs and meta-analyses do not provide the best scientific evidence.

Support

Conflict of interest

References

1. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane. 2021. Available from: www.training.cochrane.org/handbook
2. Letelier LM, Manríquez JJ, Rada G. Revisiones sistemáticas y metaanálisis: ¿son la mejor evidencia? [Systematic reviews and meta-analysis: are the best evidence?]. Rev Med Chil. 2005;133:246–9, http://dx.doi.org/10.4067/s0034-98872005000200015. Spanish.
3. Siddaway AP, Wood AM, Hedges LV. How to do a systematic review: a best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. Annu Rev Psychol. 2019;70:747–70, http://dx.doi.org/10.1146/annurev-psych-010418-102803.
4. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4:1, http://dx.doi.org/10.1186/2046-4053-4-1.
5. McGinn T, Wyer PC, Newman TB, Keitz S, Leipzig R, For GG, et al. Tips for learners of evidence-based medicine: 3. Measures of observer variability (kappa statistic). CMAJ. 2004;171:1369–73.
6. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. 2009;6, http://dx.doi.org/10.1371/journal.pmed.1000100, e1000100.
7. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, and the PRISMA-DTA Group. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. JAMA. 2018;319:388–96, http://dx.doi.org/10.1001/jama.2017.19163.
8. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA. 2000;283:2008–12, http://dx.doi.org/10.1001/jama.283.15.2008.
9. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529–36, http://dx.doi.org/10.7326/0003-4819-155-8-201110180-00009.
10. Hootman JM, Driban JB, Sitler MR, Harris KP, Cattano NM. Reliability and validity of three quality rating instruments for systematic reviews of observational studies. Res Synth Methods. 2011;2:110–8, http://dx.doi.org/10.1002/jrsm.41.
11. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355, http://dx.doi.org/10.1136/bmj.i4919, i4919.
12. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336:924–6.
13. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7:177–88, http://dx.doi.org/10.1016/0197-2456(86)90046-2.
14. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. BMJ. 2010;340:c117.
15. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? Stat Med. 2002;21:1559–73.

Gonzalo Labarca [a,*], Luz M. Letelier [b]

[a] *Division of Sleep and Circadian Disorders, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, United States*
[b] *Departamento de Medicina Interna, Escuela de Medicina, Pontificia Universidad Católica de Chile, Chile*

[*] Corresponding author.
E-mail address: glabarca@bwh.harvard.edu (G. Labarca).