Tools for the Critical Appraisal of Systematic Reviews

Estefanía Quarteroni, Juan P. Rodríguez and Luis Garegnani
Centro Cochrane Asociado, Universidad Hospital Italiano Buenos Aires, Argentina.

ABSTRACT
Systematic reviews aim to collect, analyze, and synthesize all available scientific evidence on a specific research question. Although reviews are a fundamental tool for evidence-based decision-making, their correct interpretation and analysis can present certain challenges. This manuscript addresses different tools to guide the critical appraisal of systematic reviews and to determine their usefulness in supporting decision-making, based on two instruments: A Measurement Tool to Assess Systematic Reviews Version 2 (AMSTAR 2) and the checklist of the United Kingdom Critical Appraisal Skills Programme (CASP). Critical appraisal of systematic reviews is crucial to ensure the validity, quality, and applicability of their results.

Keywords: systematic reviews, critical appraisal, analysis, evidence-based medicine.

Herramientas para la lectura crítica de revisiones sistemáticas

RESUMEN
Las revisiones sistemáticas tienen como objetivo recopilar, analizar y sintetizar toda la evidencia científica disponible sobre una pregunta de investigación específica. A pesar de que las revisiones son una herramienta fundamental para la toma de decisiones basadas en la evidencia, su correcta interpretación y análisis pueden plantear ciertos desafíos. Este manuscrito aborda diferentes herramientas para guiar la lectura crítica de revisiones sistemáticas y determinar su utilidad con el fin de asistir a la toma de decisiones, basada en dos instrumentos como *A Measurement Tool to Assess Systematic Reviews Version 2* (AMSTAR 2) y la lista de verificación del *Critical Appraisal Skills Programme* (CASP) del Reino Unido. La lectura crítica de las revisiones sistemáticas es crucial para garantizar la validez, la calidad y la aplicabilidad de sus resultados.

Palabras clave: revisiones sistemáticas, lectura crítica, análisis, medicina basada en la evidencia.

INTRODUCTION

A systematic review (SR) is a type of secondary research aimed at collecting, analyzing, and synthesizing all available scientific evidence (published or unpublished) that meets predefined eligibility criteria in order to answer a specific research question. SRs apply a transparent and reproducible methodology, assessing the methodological quality of the included studies to minimize bias[1]. Moreover, when the studies included in an SR are sufficiently similar in clinical, statistical, and methodological terms in the reporting of their results, they may include a meta-analysis, a statistical technique that allows the mathematical combination of results from multiple studies. This approach also improves the precision of effect estimates and enables the assessment of their impact across different patient subgroups[2].

Author for correspondence: juanp.rodriguez@hospitalitaliano.org.ar, Rodríguez JP.

Received: 2025/05/21 | Accepted: 2026/05/01 | Published: 2026/03/26

DOI: http://doi.org/10.51987/rev.hosp.ital.b.aires.v46i1.1274

Quarteroni E, Rodríguez JP, Garegnani L. Tools for the Critical Appraisal of Systematic Reviews. Rev Hosp Ital B.Aires. 2026;46(1):e0001274

ISSN (en línea) 2314-3312

SRs are an essential tool for patients, clinicians, researchers, and health policymakers, as they provide an objective synthesis of the available evidence, facilitating evidence-based decision-making. In this way, they constitute a fundamental pillar of the evidence-based medicine (EBM) paradigm, which integrates patients' values and preferences with clinical expertise and the most up-to-date scientific evidence[3].

Despite their importance for evidence-based decision-making, the proper interpretation and analysis of SRs may present certain challenges. In clinical practice settings, limited time for in-depth and reflective reading of an SR is a common barrier[4], compounded by difficulties related to their methodological complexity. The use of technical language and methodological concepts may render SRs less accessible to readers who are not familiar with them. In addition, the results of an SR are not always conclusive, easy to interpret, or readily applicable to clinical practice due to variability among studies, publication bias, or the methodological quality of the included trials, all of which may affect the validity of the findings[5]. To address these challenges, EBM promotes the critical appraisal of scientific manuscripts in general, and of SRs in particular, by systematically examining an article to assess its validity, reliability, and relevance within a given context, thereby guiding professionals in the effective interpretation and use of scientific evidence[6].

There are multiple tools available to assess different aspects of SRs, such as the PRISMA checklist (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)[7], used to evaluate reporting, and the ROBIS tool (Risk of Bias in Systematic Reviews)[8], primarily intended for guideline developers to assess the risk of bias in SRs. In this manuscript, we will focus on two additional tools developed to facilitate the critical appraisal of SRs.

One of the tools used to guide the critical appraisal of intervention SRs is *A Measurement Tool to Assess Systematic Reviews Version 2 (AMSTAR 2)*[9]. AMSTAR 2 allows users to determine whether an SR provides an accurate and comprehensive summary of the results of available studies addressing a specific research question. Its primary purpose is to help clinicians, researchers, and decision-makers assess the level of confidence in an SR when informing important clinical or health-related decisions. The application of AMSTAR 2 requires answering a series of questions that address key methodological aspects of an SR. It consists of a 16-item questionnaire, including 6 critical domains, with simple response options ("yes," "no," or "partial yes"). This questionnaire, which can be completed in less than 30 minutes[10], yields an overall rating of confidence in an SR (Table 1) to support decision-making. The impact and importance of these domains are such that the absence of prospective protocol registration alone may result in

Table 1. Critical domains and overall confidence rating of an SR according to AMSTAR 2

Critical domains of the A Measurement Tool to Assess Systematic Reviews Version 2 (AMSTAR 2)	
1	Protocol registered before the review
2	Adequate literature search
3	Justification for excluded studies
4	Risk of bias in the individual studies included
5	Appropriate meta-analytic methods
6	Consideration of risk of bias when interpreting the results of the review
7	Assessment of the presence and likely impact of publication bias
Overall confidence rating in the results of the review	
Confidence leve	Justification
High	No critical weaknesses and up to one non-critical weakness: the SR* provides an accurate and comprehensive summary of the results of the available studies
Moderate	No critical weaknesses and more than one non-critical weakness (although, if numerous, low confidence may be justified): the SR* has weaknesses but no critical flaws and may provide an accurate summary of the results of the available studies
Low	Up to one critical weakness, with non-critical weaknesses or none at all: the SR* may not provide an accurate and complete summary of the available studies
Critically low	More than one critical weakness, with or without non-critical weaknesses: the SR* is not reliable

SR: Systematic review

an SR being rated as low confidence. For example, the SR by von Luckner et al.[11] on the effect of magnesium supplementation for migraine prophylaxis does not report, in its Methods section, a registration number or protocol publication. In contrast, the review by Talandashti et al.[12] addressing a similar research question does provide this information. Protocol registration in an SR enhances transparency and reproducibility in research.

It is recommended that this quality assessment methodology be applied to each SR, as it can inform clinical decision-making or the development of health policies[13,14]. Depending on the characteristics, type, and methodology of each review, specialized input may be required to adequately assess certain domains –for example, to determine whether the authors appropriately evaluated the risk of bias or correctly implemented statistical methods.

In addition to being a useful tool not only for readers of SRs but also for the teams that conduct them, AMSTAR 2 is, in essence, an educational tool[13]. It is important to incorporate it into the academic training of health sciences students so that the critical appraisal of evidence syntheses becomes a routine practice in professional life. However, AMSTAR 2 is not without limitations: its application requires expert judgment and a solid understanding of SR methodology, which may lead to variability among assessors and limit its use among less experienced users. Furthermore, it does not generate an overall quantitative score, but rather a domain-based classification, which may complicate comparisons between SRs. Finally, some items are open to interpretation or difficult to apply to certain types of reviews, which may reduce their consistency and practical usefulness.

Another tool developed to guide the critical appraisal of SRs is the checklist from the *Critical Appraisal Skills Programme (CASP)*. This tool helps determine the reliability of the results and the relevance of an SR for applying its findings in clinical practice, thereby facilitating decision-making through a set of structured questions with simple responses. The CASP checklist for SRs is divided into four sections, each with a specific analytical objective (Table 2). The first section aims to determine whether the SR has a valid design, with a clearly defined question. For example, an SR that evaluated the effects of different formulations of enteral nutrition in patients with acute pancreatitis[15] clearly defined the components of the PICO question (Population, Intervention, Comparator, Outcomes), including in the population all adult patients with acute pancreatitis; in the intervention, any type of enteral nutrition with a specific formulation; in the comparator, different types of enteral nutrition, placebo, or no intervention; and in the outcomes, all-cause mortality, among others. In contrast, a detailed description of the research question was not found in another SR on the same topic[16].

The second section examines the search for studies, ideally across more than one relevant database (e.g., MEDLINE/PubMed, Embase), and the proper reporting of this process. It also considers the use of unjustified language restrictions and additional search strategies,

Table 2. Sections of the Critical Appraisal Skills Programme (CASP) tool for systematic reviews

	Key question	Guiding questionsa
Section A	Is the basic design of the study valid for a systematic review?	– Did the systematic review address a clearly focused research question? - Did the authors search for appropriate study designs to answer the research question?
Section B	Is the systematic review methodologically sound?	– Is the search strategy comprehensive and clearly reported? – Did the authors define appropriate selection, inclusion, and exclusion criteria?
Section C	What are the results?	– Were the results of the systematic review reported comprehensively and interpreted appropriately? – Was the total number of participants sufficient to detect an effect?
Section D	Will the results help in the local context?	– Can the results of the systematic review be applied to your local population/context? – What resources will be required to implement the intervention?

such as manual screening of reference lists of included studies or consultation with experts. For example, an SR on the global overview of dengue[17] restricted eligibility to studies published in English. This may lead to the omission of important studies published in other languages and, consequently, to an incomplete synthesis of the available evidence, particularly given that relevant research on endemic diseases such as dengue is often also published in Spanish or Portuguese.

This section also focuses on evaluating the internal validity of the included studies, that is, whether their design and conduct were appropriate to minimize bias[18]. While many systematic reviews omit this assessment, Cochrane reviews routinely include it through risk-of-bias assessment tools. It also examines whether the statistical methods used in the meta-analysis were appropriate, properly applied, and adequately reported.

The third section of the CASP checklist for SRs analyzes how results are presented and interpreted, and evaluates whether the effect estimates from the analyses are consistent with their narrative interpretation. It also considers the precision of these effect estimates, assessing whether the total number of participants across the included studies was sufficient to detect an effect and whether the authors reported confidence intervals and p-values to determine statistical significance. For example, a Cochrane review on the use of a monoclonal antibody for the prevention of respiratory syncytial virus[19] raised concerns about the effect of the intervention due to the limited number of participants in the included studies. However, another review on the same topic[20] did not include an analysis of these aspects when interpreting its findings.

The final section focuses specifically on the applicability of the results in the local context. For example, a Cochrane review[21] on inactivated oral cholera vaccines showed that two vaccines (Dukoral® and Shanchol®) reduced cholera cases and were safe and well tolerated. Although the results of the review are valid and methodologically sound, the applicability of these vaccines in other settings with limited resources and challenges in maintaining the cold chain may be reduced due to logistical, economic, and cultural barriers.

It is essential that readers of an SR assess whether the intervention under analysis provides an acceptable balance between benefits and harms for patients, whether the resources required for its implementation are available, and whether its use is reasonable in the given context. This analysis is crucial both in clinical practice and in evidence-based decision-making.

The main limitation of the CASP checklist for SRs lies in its reliance on qualitative, open-ended questions without specific scoring criteria or weighting of items, which may lead to subjective interpretations and variability among assessors. In addition, it provides an overall assessment of the credibility and applicability of the review but does not allow for precise identification of the specific methodological aspects that affect its quality, thereby limiting its usefulness for comparing SRs.

CONCLUSION

The synthesis of evidence through SRs is fundamental for health-related decision-making and for the development and implementation of health policies, as it provides reliable and objective information. However, the critical appraisal of these SRs is essential to ensure the validity, quality, and applicability of their results. Tools such as AMSTAR 2 and CASP are valuable for this purpose and are highly useful for both professionals and students in medical and health-related fields. They also constitute a fundamental component in health technology assessment and in the development of clinical practice guidelines.

There are common elements across the tools discussed. It is important that health professionals acquire critical appraisal skills and use these tools effectively to ensure that scientific evidence is translated into high-quality patient care, into sound and reliable arguments for medical decision-making, and ultimately into robust public policies supported by scientific and methodological evidence.

Author contributions: Conceptualization, Investigation, Writing – original draft, Writing – review and editing, Visualization, Project administration (EQ, JPR, LIG). Supervision (LIG).

Conflicts of interest: The authors declare no conflicts of interest related to the content of this work.

Funding: The authors declare that this study did not receive funding from any external source.

REFERENCES

1. Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions. Hoboken, NJ: John Wiley & Sons; 2011. 553 p.
2. Walker E, Hernandez AV, Kattan MW. Meta-analysis: its strengths and limitations. Cleve Clin J Med. 2008;75(6):431-439. https://doi.org/10.3949/ccjm.75.6.431.
3. Mendis K. What is evidence-based medicine? Ceylon Med J. 2021;66(3):154-156. https://doi.org/10.4038/cmj.v66i3.9495.
4. Río-de-la-Loza-Zamora JG, López-Ortiz G. Barreras para el desarrollo de investigación en medicina familiar en Iberoamérica: revisión sistemática. Rev Mex Med Fam. 2022;30;9(2):49-58. https://doi.org/10.24875/RMF.21000125.
5. Sgarbossa N, Ibáñez Cobaisse M, González Cianciulli G, et al. Revisiones sistemáticas: conceptos clave para profesionales de la salud. Medwave. 2022;22(9):2622. http://dx.doi.org/10.5867/medwave.2022.09.2622.
6. Critical Appraisal Skills Programme. How to use the CASP checklists [Internet]. Oxford: CASP; 2025 [consulta 2025 may 20]. Disponible en: https://casp-uk.net/how-to-use-checklist/.
7. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. https://doi.org/10.1136/bmj.n71.
8. Whiting P, Savović J, Higgins JP, et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol. 2016;69:225-234. https://doi.org/10.1016/j.jclinepi.2015.06.005.
9. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007;7:10. https://doi.org/10.1186/1471-2288-7-10.
10. Ciapponi A. AMSTAR-2: herramienta de evaluación crítica de revisiones sistemáticas de estudios de intervenciones de salud. Evid Actual Pract Ambul . 2018;21(1):4-13. https://doi.org/10.51987/evidencia.v21i1.6834.

11. von Luckner A, Riederer F. Magnesium in migraine prophylaxis-is there an evidence-based rationale? A systematic review. Headache. 2018;58(2):199-209. https://doi.org/10.1111/head.13217.
12. Talandashti MK, Shahinfar H, Delgarm P, et al. Effects of selected dietary supplements on migraine prophylaxis: a systematic review and dose-response meta-analysis of randomized controlled trials. Neurol Sci. 2025;46(2):651-670. https://doi.org/10.1007/s10072-024-07794-0.
13. Oxman AD, Cook DJ, Guyatt GH. Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. JAMA. 1994;272(17):1367-1371. https://doi.org/10.1001/jama.272.17.1367.
14. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017;358:j4008. https://doi.org/10.1136/bmj.j4008.
15. Poropat G, Giljaca V, Hauser G, et al. Enteral nutrition formulations for acute pancreatitis. Cochrane Database Syst Rev. 2015;2015(3):CD010605. https://doi.org/10.1002/14651858.CD010605.pub2.
16. Petrov MS, Loveday BP, Pylypchuk RD, et al. Systematic review and meta-analysis of enteral nutrition formulations in acute pancreatitis. Br J Surg. 2009;96(11):1243-1252. https://doi.org/10.1002/bjs.6862.
17. Khan MB, Yang ZS, Lin CY, et al. Dengue overview: an updated systemic review. J Infect Public Health. 2023;16(10):1625-1642. https://doi.org/10.1016/j.jiph.2023.08.001.
18. Gérvas J, Pérez Andrés C. Evaluación de las intervenciones en salud: la búsqueda del equilibrio entre la validez interna de los resultados y la validez externa de las conclusiones. Rev Esp Salud Pública. 2008;82(6):577-579. https://doi.org/10.1590/s1135-57272008000600001.
19. Garegnani L, Styrmisdóttir L, Roson Rodriguez P, et al. Palivizumab for preventing severe respiratory syncytial virus (RSV) infection in children. Cochrane Database Syst Rev. 2021;11(11):CD013757. https://doi.org/10.1002/14651858.CD013757.pub2. Update in: Cochrane Database Syst Rev. 2025;7:CD013757. https://doi.org/10.1002/14651858.CD013757.pub3.
20. Gonzales T, Bergamasco A, Cristarella T, et al. Effectiveness and safety of palivizumab for the prevention of serious lower respiratory tract infection caused by respiratory syncytial virus: a systematic review. Am J Perinatol. 2024 ;41(S 01):e1107-e1115. https://doi.org/10.1055/a-1990-2633.
21. Saif-Ur-Rahman KM, Mamun R, Hasan M, et al. Oral killed cholera vaccines for preventing cholera. Cochrane Database Syst Rev. 2024;1(1):CD014573. https://doi.org/10.1002/14651858.CD014573.