

# Insights into systematic reviews and meta-analyses: Are we judgmental enough?

## Abstract

Systematic reviews (SRs) and meta-analyses are regarded as the pinnacle of evidence-based medicine, synthesizing fragmented research to guide clinical decision-making. However, the exponential increase in their publication has raised concerns regarding redundancy, methodological flaws, and misleading conclusions. Many SRs lack preregistration, employ inconsistent methodologies, and fail to account for clinical heterogeneity. Additionally, the integration of artificial intelligence (AI) in research poses new challenges, including data fabrication, plagiarism, and AI-induced biases. While AI offers efficiency in processing vast amounts of data, its unregulated use may compromise research integrity. Furthermore, current review processes often overlook critical clinical insights, leading to misinterpretations. To improve SR quality, a rigorous, stepwise approach is necessary, including preregistration, comprehensive literature searches, unbiased data abstraction, and meticulous assessment of confounding factors. AI-generated content must be scrutinized through human oversight to ensure accuracy and reliability. Ultimately, the overreliance on statistical aggregation without critical clinical evaluation may lead to flawed conclusions. Clinicians must actively engage in assessing SRs beyond methodological frameworks, integrating human judgment to enhance the credibility and applicability of findings. Addressing these challenges is crucial for maintaining the integrity of evidence-based medicine.

**Keywords:** systematic review, meta-analysis, scientific credibility

## Introduction

Systematic reviews (SR) and meta-analyses function as key tools for clinicians in their evidence-based decision-making process. Their position has been considered at the top of the evidence-based medicine study hierarchy. They serve as a way of identifying key knowledge gaps, piecing together these fragmented and selectively reported published information for further investigation. SRs aim to assess existing knowledge on a particular subject, whereas a meta-analysis is a statistical method used to combine results from the relevant studies and resynthesize them within a larger sample size [1], [2]. Summarizing data from individual trials increases the statistical power and the precision of the effect estimates [3]. Therefore, clinical decisions could be based on the totality of the best evidence instead of the results of individual studies.

## History of systematic reviews

The evolution of systematic reviews has significantly impacted the lives of preterm newborns suffering from respiratory insufficiency. Collaboration developed in response to Archie Cochrane's request to the medical community to compile "a critical summary, periodically adapted, of

all... pertinent randomized controlled trials." He has initially assessed the impact of steroids on newborns and determined that steroids helped facilitate pulmonary maturation in preterm pregnancies. This has ultimately transformed the future [4]. With courage from his early work, science has shifted into another era to find out if randomized controlled studies and large series really meant scientific value.

## Current status

The credibility of a SR relies heavily on the quality and heterogeneity of the underlying studies, as well as a rigorous assessment of bias and confounding factors guided by a transparent protocol. Unfortunately, many existing SRs and meta-analyses are undermined by redundancy, a focus on trivial research questions, or misleading claims, which ultimately limits their scientific value [5].

## Perspective

Our perspectives regarding this drive include the following. First, the rate of publication of SRs and meta-analyses has dramatically increased in the last decade (2,728%

Ulas Kumbasar<sup>1</sup>

Ahmet Baris Durukan<sup>2,3</sup>

1 Hacettepe University, School of Medicine, Department of Thoracic Surgery, Ankara, Turkey

2 Liv Ankara Hospital, Department of Cardiovascular Surgery, Ankara, Turkey

3 Istanbul İstinye University, School of Medicine, Department of Cardiovascular Surgery, Ankara, Turkey

and 2,635%, respectively; from 1991 to 2014) [1]. This massive increase may reflect the performance of replicating previous studies. Siontis et al. demonstrated in their study that 65% of the overlapping meta-analyses on the same topic did not include any additional outcomes [6]. Second, these replicating/overlapping studies may end up with different conclusions and can be confusing. One can reach variable final conclusions by using different selection criteria of relevant studies, choosing different outcomes of interest, comparing different populations, and using different statistical methods. Even the interpretation of similar results may differ among the researchers/clinicians based on their scientific backgrounds, the ability of critical thinking, and conflicts of interest. A thorough understanding and judgmental analysis of the data is the key factor in the correct interpretation of studies. Third, systematic and transparent conduct and reporting of SRs are still uncommon. Quantitative analysis reveals that solely around 15% to 30% of SRs are prerecorded in registries such as PROSPERO. Although there has been an increase in registrations in recent years, approximately two-thirds of published SRs still bypass this critical step, significantly increasing the risk of outcome switching and redundancy [7], [8]. Lastly, most current SRs involve non-randomized studies, which make the comparison of cohorts unreliable in most instances despite efforts at conducting propensity matching and other methods of adjusting for confounders. Propensity score-based analyses account only for known and observed patient characteristics, and thus, the confounders that are not accounted for remain as a limiting factor [9]. Meta-analyses do not necessarily address this as well. Overall, the quality of evidence and the validity of findings belonging to many SRs and meta-analyses are low.

## Use of artificial intelligence (AI) technologies

The ongoing progress of AI-driven language technology enhances the possibilities for automating the SR process through alternative approaches. Utilizing these AI-driven solutions provides an opportunity to improve the SR process, conserving time and resources while mitigating risks associated with human responses. By evaluating these alternatives, we may optimize workflows and improve the overall efficiency of conducting SR [10]. Nevertheless, this application has certain disadvantages. The use of AI in scientific research has introduced new forms of academic misconduct, mostly data fabrication and plagiarism using AI algorithms [11]. The accessibility of free web-based technology has led several individuals lacking scientific expertise to fabricate data and utilize AI for its analyses. AI technologies themselves do misconduct by copy-pasting, therefore making “enhanced” plagiarism and even fabricating references, so-called “AI hallucination”. The provided references include authors from the relevant field; the topic is significant and supposedly published in subject-related journals. Nonetheless, the

reference in question is nonexistent [12]. These techniques compromise the integrity of the research in question and can readily mislead conclusions. Unfortunately, investigating such misconduct remains a significant challenge for academic institutions, largely due to the technical limitations of current detection tools. Recent studies indicate that available AI-detection software often yields low accuracy rates and high false-positive rates, making it difficult to definitively distinguish between human-written and AI-generated text [13]. As AI technology evolves at a pace that outpaces the development of reliable detection methods, institutions are left without robust tools to enforce ethical standards. Consequently, the emphasis on ethical norms and researcher qualifications risks being overshadowed as the barriers to generating sophisticated, albeit potentially fabricated, content are lowered by these rapidly developing technologies. Emphasizing ethical norms and enhancing qualifications of researchers, which are main duties of communities, are already forgotten since “brilliant-minded new generation science people” have already overcome barriers with light-speed developing AI technologies. Such issues are also observed in SR and meta-analysis conduct since AI technologies are capable of writing papers even in seconds. Therefore, it is crucial to acknowledge that although AI offers invaluable assistance, researchers must rigorously assess the model’s outputs, carefully check information, and verify the accuracy and reliability of the generated content. Moreover, rapid implementation of complete criteria for AI research integrity, including precise protocols for AI utilization in data analysis and publication, is essential to ensure transparency and accountability.

## Pros and cons of AI

Artificial intelligence (AI) offers several important advantages in the conduct of systematic reviews and meta-analyses, particularly in improving efficiency and scalability. AI-driven tools can rapidly screen large volumes of literature, assist in identifying relevant studies across multiple databases, and reduce the time required for title and abstract screening. Natural language processing algorithms can support automated data extraction from published texts and tables, while machine learning models can aid in detecting patterns, inconsistencies, or potential biases within datasets. These capabilities may enhance reproducibility and reduce human error, particularly in repetitive and labor-intensive steps of the review process. Furthermore, AI can facilitate living systematic reviews by continuously monitoring and incorporating newly published evidence, thereby keeping analyses up to date in rapidly evolving fields.

However, these advantages must be balanced against important limitations and risks. AI systems remain dependent on the quality and structure of input data and may propagate existing biases or introduce errors such as hallucinated references, incorrect data extraction, or

**Table 1: Clinician's checklist for appraising systematic reviews**

<b>1. Protocol &amp; transparency</b>
<input type="checkbox"/> Was the review preregistered (e.g., PROSPERO) before the search began?
<input type="checkbox"/> Are the inclusion/exclusion criteria explicitly defined?
<b>2. Search &amp; selection</b>
<input type="checkbox"/> Was the literature search comprehensive (multiple databases + unpublished data)?
<input type="checkbox"/> Did the authors check for "AI Hallucinations" (fake references) in the included studies?
<b>3. Data &amp; analysis</b>
<input type="checkbox"/> Was data extraction performed by at least two independent investigators?
<input type="checkbox"/> Does the analysis distinguish between <i>statistical</i> heterogeneity and <i>clinical</i> diversity?
<b>4. Human oversight</b>
<input type="checkbox"/> Are the conclusions based on clinical significance, not just p-values?
<input type="checkbox"/> Is there a declaration regarding the use of AI tools in drafting or analysis?

misclassification of studies. Overreliance on AI outputs may also contribute to "automation bias," where researchers accept algorithmic results without sufficient critical appraisal. Therefore, the integration of AI into systematic reviews should be viewed as an assistive, rather than autonomous, process. Human oversight remains essential to validate findings, ensure clinical relevance, and interpret results within the appropriate methodological and clinical context. A hybrid model combining AI efficiency with expert clinical judgment is likely to provide the most robust and reliable approach to future evidence synthesis.

## Review process

The reviewers' aspect is an issue as outlined. Reviewing a systematic review or meta-analysis constitutes a specialized area of expertise. Misapprehensions and insufficient technical expertise may result in the early acceptance of manuscripts lacking evident scientific value. A comprehensive understanding for evaluation is needed [14].

## Suggestions

What can be done? Adopting a step-by-step approach to conducting a SR may help to improve the quality of these studies. These include (i) designing a detailed review protocol with explicit inclusion and exclusion criteria by adhering to the PRISMA statements; (ii) preregistration of full study protocol including the statistical analysis plan; (iii) performing a comprehensive systematic literature search of multiple databases as well as the unpublished data to minimize the publication bias, (iv) accurate data abstraction by at least two independent investigators; (v) careful anticipation of clinical, statistical, and methodological heterogeneity; (vi) detailed assessment of the risk of biases and confounding factors in individual studies; (vii) adequate use of statistical methods; and (viii) assessment of the overall quality of evidence (e.g. GRADE) in each study [5], [15]. Moreover, these studies should be designed and conducted by non-conflicted investigators, and the implementation of the results should

be based on transparency and critical thinking. Likewise, AI-generated content must undergo assessment through human supervision by ensuring alignment with set standards, finding and correcting any potential errors or biases, and delivering a thorough and precise interpretation of the content [16]. Table 1 summarizes the clinician's checklist for appraising SRs.

However, the above-mentioned methodological steps do not sufficiently capture the issue. As clinicians, we must go further. Our emphasis on the literature search or the statistical tests makes us think that this can overcome the underlying bias due to clinical issues that are not well spelled out in the papers. In that regard, the general "risk of bias" assessments do not help much either. These tools are more designed to just eliminate some studies rather than to highlight what residual confounding exists and how it might impact our interpretation. The methodologists are doing the best they can in an abstract and general manner. What they are missing is the ability to assess actual issues because they do not have any clinical insight. And the clinicians think that because the methodologists provide the criteria we just need to follow, and do not need to think critically ourselves. Furthermore, meta-analysis inherently lumps all the information together and hides the underlying nuances, and statistical tests for heterogeneity are fundamentally different than clinical heterogeneity. A diagram of stepwise approach to SRs with integrated human oversight has been proposed in Figure 1.

## Conclusion

While systematic reviews and meta-analyses remain the cornerstone of evidence-based medicine, their current value is frequently undermined by redundancy, methodological inconsistencies, and the unregulated use of artificial intelligence. To restore the credibility of these studies, the scientific community must enforce stricter adherence to transparent protocols, including comprehensive preregistration and independent data verification. Furthermore, while AI offers efficiency, it must remain under strict human oversight to prevent the propagation of fabricated or biased data. Ultimately, researchers must

### A Diagram of Stepwise Approach to Systematic Reviews with Integrated Human Oversight

Standard workflow steps, AI-assisted roles, and mandatory human oversight checkpoints across each phase

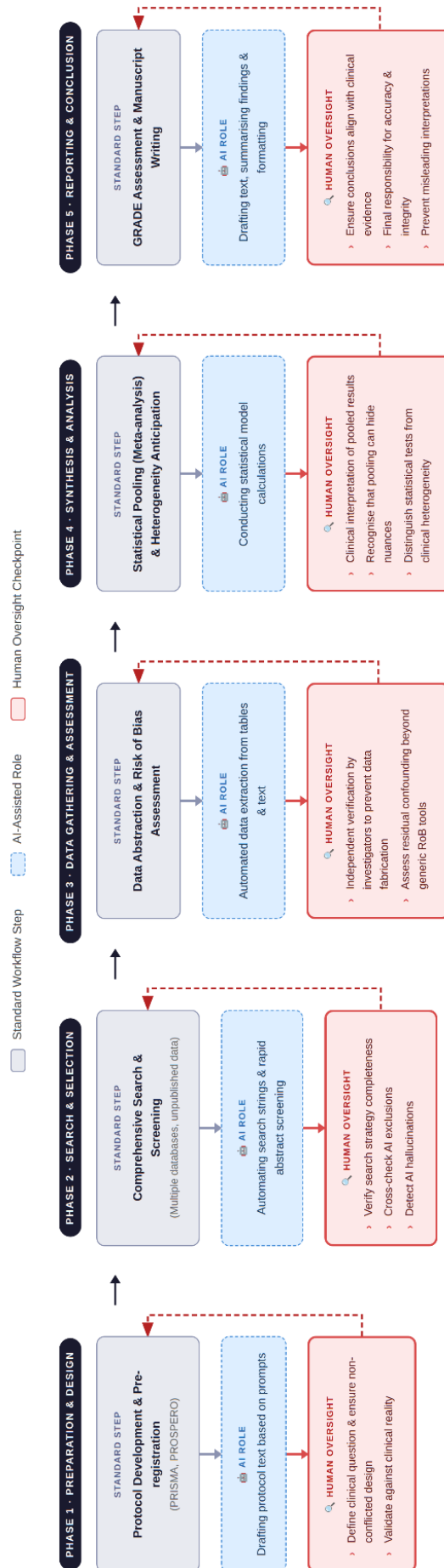


Figure 1: Conceptual flow diagram: stepwise approach to improving SR quality with integrated human oversight in the AI era

recognize that statistical aggregation can obscure underlying nuances; therefore, a rigorous assessment of clinical heterogeneity must accompany statistical testing. By prioritizing expert clinical judgment alongside methodological rigor, we can ensure these studies remain reliable tools for decision-making.

## Notes

### Authors' ORCIDs

- Ulas Kumbasar: 0000-0003-0616-1326
- Ahmet Baris Durukan: 0000-0003-0566-0350

### Author's contributions

- The conception of the work; or the acquisition, analysis, or interpretation of data for the work: UK, ABD
- Drafting the work, reviewing it critically for important intellectual content: UK, ABD
- Final approval of the version to be published: UK, ABD
- Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved: UK, ABD

### Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

### Competing interests

The authors declare that they have no competing interests.

## References

1. Ioannidis JP. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *Milbank Q.* 2016 Sep;94(3):485-514. DOI: 10.1111/1468-0009.12210
2. Møller AM, Myles PS. What makes a good systematic review and meta-analysis? *Br J Anaesth.* 2016 Oct;117(4):428-30. DOI: 10.1093/bja/aew264
3. Ridgeon EE, Bellomo R, Aberlegg SK, Sweeney RM, Varughese RS, Landoni G, Young PJ. Effect sizes in ongoing randomized controlled critical care trials. *Crit Care.* 2017 Jun 5;21(1):132. DOI: 10.1186/s13054-017-1726-x
4. Starr M, Chalmers I, Clarke M, Oxman AD. The origins, evolution, and future of The Cochrane Database of Systematic Reviews. *Int J Technol Assess Health Care.* 2009 Jul;25 Suppl 1:182-95. DOI: 10.1017/S026646230909062X
5. Møller MH, Ioannidis JPA, Darmon M. Are systematic reviews and meta-analyses still useful research? We are not sure. *Intensive Care Med.* 2018 Apr;44(4):518-20. DOI: 10.1007/s00134-017-5039-y
6. Siontis KC, Hernandez-Boussard T, Ioannidis JP. Overlapping meta-analyses on the same topic: survey of published studies. *BMJ.* 2013 Jul 19;347:f4501. DOI: 10.1136/bmj.f4501
7. Rombey T, Doni K, Hoffmann F, Pieper D, Allers K. More systematic reviews were registered in PROSPERO each year, but few records' status was up-to-date. *J Clin Epidemiol.* 2020 Jan;117:60-7. DOI: 10.1016/j.jclinepi.2019.09.026
8. Page MJ, Shamseer L, Tricco AC. Registration of systematic reviews in PROSPERO: 30,000 records and counting. *Syst Rev.* 2018 Feb 20;7(1):32. DOI: 10.1186/s13643-018-0699-4
9. Freemantle N, Marston L, Walters K, Wood J, Reynolds MR, Petersen I. Making inferences on treatment effects from real world data: propensity scores, confounding by indication, and other perils for the unwary in observational research. *BMJ.* 2013 Nov 11;347:f6409. DOI: 10.1136/bmj.f6409
10. Kebede MM, Le Cornet C, Fortner RT. In-depth evaluation of machine learning methods for semi-automating article screening in a systematic review of mechanistic literature. *Res Synth Methods.* 2023 Mar;14(2):156-72. DOI: 10.1002/jrsm.1589
11. Chen Z, Chen C, Yang G, He X, Chi X, Zeng Z, Chen X. Research integrity in the era of artificial intelligence: Challenges and responses. *Medicine (Baltimore).* 2024 Jul 5;103(27):e38811. DOI: 10.1097/MD.00000000000038811
12. Sun Y, Sheng D, Zhou Z, Wu Y. AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content. *Humanit Soc Sci Commun.* 2024;11:1278. DOI: 10.1057/s41599-024-03811-x
13. Weber-Wulff D, Anohina-Naumeca A, Bjelobaba S, Foltýnek T, Guerrero-Dib J, Popoola O, Šigut P, Waddington L. Testing of detection tools for AI-generated text. *Int J Educ Integr.* 2023;19:26. DOI: 10.1007/s40979-023-00146-z
14. Russo MW. How to Review a Meta-analysis. *Gastroenterol Hepatol (N Y).* 2007 Aug;3(8):637-42.
15. Nguyen NH, Singh S. A Primer on Systematic Reviews and Meta-Analyses. *Semin Liver Dis.* 2018 May;38(2):103-11. DOI: 10.1055/s-0038-1655776
16. Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare (Basel).* 2023 Mar 19;11(6):887. DOI: 10.3390/healthcare11060887

### Corresponding author:

Ahmet Baris Durukan  
Liv Ankara Hospital, Department of Cardiovascular Surgery, Kavaklıdere, Bestekar sokak No: 8, 06880, Cankaya/Ankara, Turkey, Phone: +90 532 2273814  
barisdurukan@yahoo.com

### Please cite as

Kumbasar U, Durukan AB. Insights into systematic reviews and meta-analyses: Are we judgmental enough? *GMS Ger Med Sci.* 2026;24:Doc06.  
DOI: 10.3205/000359, URN: urn:nbn:de:0183-0003591

### This article is freely available from

<https://doi.org/10.3205/000359>

Received: 2026-01-12

Published: 2026-06-12

### Copyright

©2026 Kumbasar et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.