Biomolecules and Biomedicine
ISSN: 2831-0896 (Print) | ISSN: 2831-090X (Online)
Journal Impact Factor® (2024): 2.2
CiteScore® (2024): 5.2
www.biomolbiomed.com | blog.bjbms.org

REVIEW

Begagić et al: How to review a systematic review?

The role of reviewers in the era of systematic reviews and meta-analysis: A practical guide for researchers

Emir Begagić[1*], Faruk Skenderi[2], Semir Vranić[3]

[1]Department of Neurosurgery, Cantonal Hospital Zenica, Zenica, Bosnia and Herzegovina;

[2]Biomolecules and Biomedicine, Sarajevo, Bosnia and Herzegovina;

[3]College of Medicine, QU Health, Qatar University, Doha, Qatar.

*Correspondence to Emir Begagić: begagicem@gmail.com

DOI: https://doi.org/10.17305/bb.2025.12979

ABSTRACT

A systematic review with meta-analysis (SRMA) represents the pinnacle of evidence, but its validity depends on methodological rigor. This narrative review synthesizes recommendations from major reporting frameworks— Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 (PRISMA-2020), Meta-Analysis of Observational Studies in Epidemiology (MOOSE) and Preferred Reporting Items for Overviews of Reviews (PRIOR)—into a concise checklist for peer reviewers. The checklist addresses common sources of bias that often escape editorial assessment. Initially, it outlines how reviewers should assess the rationale for an SRMA by identifying existing syntheses on the same topic and determining whether the new work provides substantive novelty or a significant update. Best practices are summarized for protocol registration, comprehensive search strategies, study selection and data extraction, risk-of-bias evaluation, and context-appropriate statistical modeling, with a specific focus on heterogeneity, small-study effects, and data transparency. Case examples highlight frequent pitfalls, such as unjustified pooling of heterogeneous designs and selective outcome reporting. Guidance is also provided for formulating balanced, actionable review comments that enhance methodological integrity without extending editorial timelines. This checklist equips editors and reviewers with a structured tool for systematic appraisal across clinical disciplines, ultimately improving the reliability, reproducibility, and clinical utility of future SRMAs.

Keywords: Systematic reviews; Meta-analysis; Guideline adherence; Reproducibility of results; Bias; Risk assessment.

INTRODUCTION

Systematic reviews (SRs) and meta-analyses (SRMAs) are highly regarded in scientific research for synthesizing data from original studies and offering evidence-based recommendations in the medical sciences [1]. They are considered the best available evidence in the hierarchy of evidence-based practice [2]. The data reveals a trend from the first SR documented in the PUBMED/Medline database in 1957 to a substantial total of 38,449 publications by 2022 (Figure 1). The academic community is facing multiple challenges [3], one of which is the rapid proliferation of journals, growing from 10 in the 17th century to over 100,000 by the end of the 20th century. Additionally, the rise of "paper mills" exacerbates the issue. These organizations use artificial intelligence and other tools to mass-produce publications, selling authorship for as little as $200 without any real contribution to the work. Some sites, based in countries like Russia, Iran, and Latvia, claim to have published over 12,000 articles and offer main authorship for €2,000 [4].

Ensuring the quality of SRMAs is crucial given the growing volume of publications. The retraction of 13 papers from the Scottish Medical Journal in April 2024, including ten SRMAs, highlighted significant issues with data extraction integrity [5]. This underscores the critical role of reviewers in detecting misconduct within SRMAs. Additionally, the rise of AI-based chatbots in scientific writing has raised ethical concerns and divided the scientific community [6,7]. Noteworthy instances of academic fraud include a Spanish chemist publishing an article every 37 hours and a Japanese psychiatrist producing 115 articles in a year. There were also reports of 78 journals receiving 300 unethical articles from two Japanese doctors, with half being retracted. The issue is exacerbated by "paper mills" that sell articles and ghostwriting services. A 2022 report estimated that up to 20% of submissions come from paper mills, with analysis showing that ~2.2% of 2.85 million published studies originate from such sources. Over 100 articles were partially written by AI, with a 72% increase in suspected AI-generated content, despite AI's potential for data falsification [8-10]. Thus, reviewers and editorial staff must play an essential role in maintaining academic standards and quality, with a rigorous review process necessary to combat misuse and uphold SRMA integrity.

Ultimately, this review aims to equip reviewers with practical insights and strategies to uphold standards of excellence in academic publishing. By fostering a rigorous and ethical review culture, it seeks to enhance the reliability and impact of SRMAs in shaping evidence-based practices and policies across diverse disciplines.

PRACTICAL RECOMMENDATIONS

Initial evaluation

The initial step in the critical assessment of SRMAs involves establishing the background and justification for conducting the review. Reviewers should begin by determining whether there are existing SRMAs addressing the topic of the manuscript under review. If such reviews exist, it is essential to assess whether there is a valid reason for publishing the current work, whether its findings differ from existing SRMAs, or provide an updated perspective. Additionally, reviewers need to ascertain whether the manuscript adheres to relevant guidelines such as the *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) [11], *Reporting of Meta-Analyses of Observational Studies* (MOOSE) [12], *Reporting for Overviews of Reviews or Umbrella Reviews* (PRIOR) [13], etc. The authors of SRMAs usually follow the PRISMA guidelines, providing a PRISMA checklist with 27 items that they address. This checklist enables reviewers to verify whether the SRMAs adhered to the methodological standards outlined in PRISMA [14]. For reviewers' convenience, Table 1 presents a concise checklist for evaluating SRMAs.

Evaluation of the methodology section

A thorough appraisal of the Methods section is critical, as flaws here inevitably undermine the results. First, reviewers should check whether the SRMA was prospectively registered in an open-access registry (such as PROSPERO, Research Registry, INPLASY, OSF Registries, or protocols.io). Prospective registration is highly recommended to promote transparency, prevent unnecessary duplication, and reduce the risk of bias [15]. A registered protocol (ideally with a registration ID cited) allows the reviewer to compare the planned methods with what was actually done, thus identifying any deviations. Studies have shown that protocol registration and adherence are associated with more reliable outcomes [15,16]. If the authors claim registration, the reviewer should verify the registry entry and ensure consistency between the protocol and the submitted manuscript (for example, are all pre-specified outcomes reported? Were any analyses added post hoc?).

Reviewers must also consider the eligibility criteria (inclusion/exclusion criteria) defined by the authors. A well-structured SRMA uses the PICOS framework – Population, Intervention, Comparator, Outcomes, Study design – to clearly outline what studies were eligible [17]. This framework not only clarifies the scope (e.g., which patient population and interventions are of interest) but also aids in designing the literature search strategy [18]. The

search strategy itself should be described in enough detail to be reproducible. Ideally, the manuscript (or a supplement) will list the specific search queries, the databases used, and the date of the last search [19]. While there is no universal rule on the number of databases, searching at least two robust databases is a minimal requirement, and using multiple databases (e.g., MEDLINE/PubMed, Embase, Web of Science, Scopus, Cochrane Library) is strongly encouraged to capture a broad range of studies [20,21]. In practice, Embase, MEDLINE (PubMed), and Google Scholar or Web of Science are often recommended as a core trio for medical SRMAs. Reviewers should assess whether the authors used an adequate combination of sources and whether the search was likely comprehensive. The Methods should also specify if any language or date restrictions were applied and justify them (unjustified restrictions might omit relevant studies and bias the results).

Transparent reporting of the search and selection process is commonly aided by a PRISMA flow diagram. Reviewers should examine this flow chart to check the number of identified studies, how many were excluded (and for what reasons) at each stage (screening, eligibility), and the final number included. Any discrepancies (like numbers not adding up) should be questioned. Tools like the PRISMA 2020 flow diagram generator are available to authors [22], so the lack of a clear flow diagram in a modern SRMA is a notable oversight.

Another key component is the data extraction and quality assessment procedure. The Methods should list what data were extracted from each study (e.g., participant characteristics, outcomes, follow-up duration, effect measures, etc.) and describe how the risk of bias in the primary studies was assessed. It is standard to use validated tools: for randomized trials, Cochrane's Risk of Bias 2 (RoB 2) is widely used, and for observational studies, tools like ROBINS-I can be employed [23]. If the manuscript is missing a risk-of-bias assessment of included studies, the reviewer should consider this a serious deficiency, as the credibility of a meta-analysis depends on the quality of the evidence synthesized. Moreover, the review should report whether this appraisal was done by at least two independent reviewers with a process for resolving disagreements – a safeguard against errors and bias in study selection and data extraction.

Finally, the methodology evaluation should confirm that the authors have followed their protocol and that all analyses were planned. Outcome measures and statistical methods need to be described clearly. For example, the Methods should state whether a fixed-effect or random-effects meta-analysis model was used, and why that choice was made [24]. A

random-effects model is generally more appropriate when combining studies that are not identical, as it accounts for between-study variability, but it yields wider confidence intervals. Conversely, a fixed-effect model might be justified if the studies are essentially identical in methods and populations (which is rare in practice). Any subgroup analyses or meta-regression planned to explore heterogeneity should be specified in the Methods as well [25]. Reviewers should be cautious if numerous subgroup analyses are presented that were not declared in advance, as this may indicate data dredging. In summary, a methodologically sound SRMA will pre-define its analytical approach; a reviewer's task is to check for adherence to those plans and the appropriateness of the techniques used.

Evaluation of the results section

Once the methodology has been deemed solid, the Results section requires equal scrutiny. At this stage, the reviewer assesses how the data were synthesized and whether the findings are presented clearly and accurately.

It is essential to evaluate whether the authors chose the correct effect measures and statistical models for the meta-analysis [26]. For dichotomous outcomes (like event rates), did they use risk ratios, odds ratios, or risk differences appropriately? For continuous outcomes, are mean differences or standardized mean differences reported with correct units and interpretations? Check if the chosen model (fixed-effect vs. random-effects) was appropriate given the studies' diversity. A random-effects model is generally more conservative when heterogeneity is present, as it assumes the true effect may vary between studies. If the authors conducted a narrative synthesis (e.g., because meta-analysis was not possible), ensure that this narrative is unbiased and that they did not simply count studies ("vote counting") without considering study quality or sample size. All decisions regarding pooling or not certain data should be justified in the text. As a reviewer, consider whether any subset of data might have been inappropriately combined, for instance, pooling results from vastly different study designs (RCTs mixed with observational studies) without a proper rationale. Any such issues should be commented on. It is also worthwhile to see if the authors followed any established guidance for data synthesis (for example, the Cochrane Handbook recommendations for choosing summary measures and models). Deviations from expected practice are not necessarily wrong, but they demand a clear explanation.

One of the most important aspects of a meta-analysis result is the degree of heterogeneity among included studies. The I^2 statistic is typically reported to quantify

heterogeneity, representing the percentage of total variation across studies that is due to real differences rather than chance. As a rule of thumb, I^2 values of 0–25% indicate low heterogeneity, ~50% moderate, and >75% high heterogeneity (although these cutoffs are not absolute). Reviewers should verify that the I^2 is reported and consider its implications [23,27]. Cochran's Q test is another measure (with a p-value for heterogeneity), but it has low power when the number of studies is small and tends to be overly sensitive when there are many studies, so I^2 is usually more informative [28]. If heterogeneity is high, a good SRMA will explore possible reasons rather than ignore it. Look for any analyses of subgroups or meta-regression that attempt to explain variability in results. For example, authors might stratify results by population characteristics, dosage, study quality, or study year. As a reviewer, critically evaluate these subgroup analyses: Were they pre-specified or data-driven? Are there plausible explanations for differences between subgroups? And importantly, did the authors test for interaction (i.e., whether the difference between subgroups is statistically significant)? It is well known that improper subgroup analyses can be misleading – they may yield false positives by chance alone if numerous comparisons are made [29]. Credible subgroup effects generally should be hypothesized a priori, seen consistently across related outcomes, and supported by a significant interaction test rather than just separate p-values for each subgroup. If the manuscript claims a subgroup difference, the reviewer should check these criteria and possibly advise caution in interpretation [29]. Similarly, meta-regression (a technique to relate study-level characteristics to effect size) can be a powerful tool to investigate heterogeneity, but it is prone to false findings when the number of studies is small [30]. Each meta-regression or subgroup analysis should thus be treated as exploratory unless strongly justified [31]. The reviewer should ensure that authors have acknowledged the exploratory nature, if applicable, and have not overstated such findings.

Forest plots are the visual centerpiece of meta-analysis results, and reviewers should examine them closely [32]. Each forest plot should display the effect estimates and confidence intervals for each study and the pooled estimate at the bottom. Pay attention to whether the studies' point estimates largely overlap or not. A quick visual scan can often affirm the I^2 value – if the confidence intervals of most studies overlap with each other and with the pooled estimate, heterogeneity is likely low; if they are widely scattered, heterogeneity is high [33]. Reviewers should check for outliers – studies that deviate substantially from the others. Outliers can have a large influence on the pooled result, especially in a fixed-effect model or if the study has a large weight (often due to a large

sample size). If one or two studies drive the results, the authors should mention this and perhaps conduct a sensitivity analysis excluding them. Ensure that the labels in the forest plot (study names, interventions, etc.) are accurate and that any stratifications (e.g., by subgroup) match what is described in the text. Also, verify that the numeric results in the plot (effect sizes and confidence intervals) match those given in the text or tables. Inconsistencies between the forest plot and written results could indicate an error.

Consistency of direction is another consideration: do all or most studies favor one direction of effect? If a few studies contradict the majority, do the authors discuss why (differences in population or methodology)? The Results section should not only present the numbers but also *translate* them: e.g., "the meta-analysis found a 25% relative risk reduction in outcome X with intervention Y (RR 0.75, 95% CI 0.60–0.95)". A reviewer should confirm that such interpretations are accurate and not exaggerated (e.g., claiming causality from observational data, or clinical importance from a statistically significant but small effect).

Reviewers should check whether the authors assessed the possibility of publication bias, especially if the meta-analysis includes a substantial number of studies (a common rule is to assess it when ≥10 studies are included in the meta-analysis). Methods to do this include funnel plot analysis and statistical tests like Egger's test or Begg's test for funnel plot asymmetry. A funnel plot is a scatter plot of study effect estimates against a measure of their size or precision; in the absence of bias, the plot resembles a symmetric inverted funnel. If smaller studies tend to have more extreme results than larger ones, the plot may be skewed or hollow on one side, suggesting potential publication bias or other small-study effects. Egger's regression test can detect asymmetry by checking if there is a significant intercept when regressing standard normal deviates on precision [34]. As a reviewer, determine if the authors have presented a funnel plot (often in an appendix) or reported Egger's test p-value. If yes, do they interpret it correctly? For instance, a non-significant Egger's test does not prove the absence of bias, especially with few studies; conversely, a significant result suggests bias but could also arise from true heterogeneity or chance [34]. If the authors did not perform any formal assessment of publication bias, consider whether it would have been appropriate to do so. In cases with many studies or suspicion of unpublished negative studies, reviewers could suggest that the authors conduct such an analysis. Some meta-analyses also use the "trim-and-fill" method to estimate the impact of missing studies on the pooled result. If present, the reviewer should see if the trim-and-fill adjusted result differs markedly, which would indicate robustness issues. Overall, ensure the manuscript discusses the possibility of bias in the

results [35]. If a funnel plot is presented, the text should comment on its symmetry or lack thereof, rather than leaving it to readers to infer. It is also worth noting that when only a few studies are included, these tests have little power, and a funnel plot is not very informative [36].

Good SRMAs include sensitivity analyses to test the robustness of the main findings. As a reviewer, check whether the authors performed analyses such as: excluding studies at high risk of bias, using alternative statistical models (e.g., using a fixed-effect model if the main analysis was random-effects, or vice versa), removing outlier studies, or using different effect metrics [37]. For example, if heterogeneity was high, did the authors try a transformation or choose a more conservative model? If one study was much larger than the others, did they analyze the data without it to see if conclusions change? Sensitivity analysis can also involve using a different cut-off for an outcome (like including only studies with a certain follow-up duration). The Results section (or supplement) should describe these tests. Reviewers should pay attention to whether the conclusions hold across these various analyses. If the results are very sensitive – for instance, if removing one study nullifies the effect – then the manuscript should acknowledge this fragility. If no sensitivity analyses were done, a reviewer might suggest at least a basic one, especially if there is a clear dominant study or a mixture of quality in the included studies. The manuscript should also report any secondary analyses, like using an alternative effect measure (risk difference instead of risk ratio, etc., if pertinent) to ensure the effect is consistently demonstrated. These practices boost confidence that the findings are not an artifact of a particular analytical choice [23].

Increasingly, SRs provide an evaluation of the certainty of evidence for each key outcome, commonly using the GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach. GRADE assesses the body of evidence on factors such as risk of bias, inconsistency (heterogeneity), indirectness, imprecision, and publication bias [38]. Based on these, each outcome is rated as high, moderate, low, or very low certainty. As a reviewer, note whether the authors included a Summary of Findings table or at least a narrative GRADE assessment. If they did, check that the justifications for downgrading or upgrading evidence are sound. For example, did they downgrade for high heterogeneity or for most studies being at risk of bias? GRADE guidelines indicate that even if all studies are observational (initially “low” quality), certain factors like a large effect size or dose-response can upgrade the confidence in the effect, whereas limitations in any of the five domains can downgrade it [38]. The reviewer should verify that any GRADE ratings

align with the data presented. If an outcome with wide confidence intervals and some risk of bias is still rated "high certainty," that might be inconsistent with GRADE criteria. Conversely, if the evidence is downgraded, the reasons should be transparent (e.g., "downgraded for imprecision because the total sample size is small and the 95% CI crosses a minimal important difference"). In the absence of a formal GRADE assessment, the reviewer can still judge qualitatively whether the authors' conclusions seem appropriately cautious given the strengths and weaknesses of the evidence. Watch out for language that overstates certainty – for instance, calling evidence "definitive" or "conclusive" when the meta-analysis includes only a few small trials or has significant limitations. Reviewers may need to recommend rephrasing conclusions in line with evidence quality. Ultimately, the results and their interpretation should reflect a balanced consideration of how much confidence we can place in the findings [23]. If the manuscript does not address this, a reviewer might suggest the authors add a statement grading the confidence in estimates or at least discuss the overall evidence quality (possibly using frameworks like GRADE or an alternative if appropriate).

In summary, when reviewing the Results section of an SRMA, one should act almost like a co-pilot, verifying every instrument reading: confirm that the numerical results are sound, the analyses are appropriate and complete, and the interpretations are fair. The Results should be reported with enough clarity and context that the reader (and reviewer) can follow the thread from raw data to pooled analysis to inference, without having to second-guess the integrity at each step. Any red flags, such as unexplained heterogeneity, selective reporting of outcomes, or insufficient examination of bias, should be raised in the review comments. By rigorously evaluating these elements, reviewers help ensure that only reliable and meaningful meta-analytic findings enter the literature.

EMERGING ISSUES IN SYSTEMATIC REVIEWS AND META-ANALYSES

Beyond the standard methodological concerns, reviewers of SRMAs today must also be vigilant about emerging issues that threaten the credibility of published research. The proliferation era of SRs has unfortunately been accompanied by various forms of scientific misconduct and dubious practices, which can compromise evidence synthesis. This section discusses several pressing issues – from paper mill activities to AI-generated content – and the role of reviewers and journals in addressing them.

The rise of paper mills and their impact

Paper mills are unethical, for-profit entities that produce and sell fabricated or low-quality manuscripts to researchers who need publications. These operations often churn out SRs and SRMAs on demand, since such articles can be produced relatively quickly by recycling content and do not require new data collection [39]. The impact of paper mills on the scientific literature has been alarming. Committee on Publication Ethics (COPE) defines paper mills as "profit-oriented, unofficial and potentially illegal organizations that produce and sell fraudulent manuscripts that resemble genuine research" [40]. They frequently manipulate the publishing process by fabricating data, plagiarizing text, and even providing fake peer reviews to journals [39]. For SRMAs, a paper mill might generate a review by stringing together generic text, using automatic translation, or even employing AI to paraphrase existing reviews, all while offering guaranteed authorship to paying clients who had no role in the work.

Recent investigations have revealed the extent of this problem. A 2022 COPE and STM report estimated that between 2% and 46% of manuscripts submitted to certain journals in the years 2019–2021 could be traced to paper mill activity [39]. In 2022, a major publisher (Wiley) discovered that some of its journals had been compromised by a network of paper mill submissions, especially through *guest editors* of special issues. This led to an unprecedented mass retraction – 511 papers retracted in one announcement – with an ongoing review targeting an additional ~1,200 suspect papers [41]. Many of these retracted papers were literature reviews or meta-analyses that passed superficial checks but were essentially illegitimate. Such mass retractions underscore that the paper mill problem is not hypothetical or rarefied; it is affecting the scientific record on a large scale. In another instance, the *Scottish Medical Journal* retractions in 2024 (noted earlier) were largely attributed to data integrity issues that may hint at paper mill or at least unscrupulous practices [5].

For systematic reviews specifically, one concern is contamination of evidence bases by fraudulent primary studies. A meta-analysis is only as good as the studies it includes – if paper mill products (which can include fake clinical studies) slip into the pool of evidence, the meta-analysis may be distorted. Reviewers should be on the lookout for suspicious patterns: for example, an SRMA that includes many studies from the same region or author cluster with unlikely high positive results could indicate that some of those primary studies were bogus. A recent cross-sectional study in *JAMA Network Open* examined life science SRs for citations of retracted paper mill articles [42]. It found that out of 200,000 SRMAs,

299 had inadvertently incorporated at least one retracted paper-mill-derived article into their analysis (a contamination rate of 0.15%) [42]. While this is a relatively small fraction, the concerning finding was that the rate increased over time, and some reviews included multiple fraudulent papers. Moreover, about one-third of citations to these retracted articles occurred even after the articles had been retracted [42]. This highlights a gap in current peer review and editorial oversight – these contaminated reviews remained uncorrected and continue to propagate false data. As a reviewer, one cannot realistically validate each included study in an SRMA, but one can raise questions if, say, a large portion of included studies come from obscure journals of questionable repute or if certain data look too consistent or "too good to be true." Reviewers should not hesitate to use tools at their disposal: a quick check in the Retraction Watch database or a plagiarism screening of suspicious text could unveil problems. Even a simple Google search of a few study titles can sometimes reveal if an included study has been flagged or retracted elsewhere. Ultimately, while detecting a well-crafted fraudulent paper is difficult, reviewers should maintain a healthy skepticism and be aware that SRs themselves can be vehicles for scientific fraud if the input data or the writing process is corrupted.

Scientific misconduct and detection strategies

Scientific misconduct in SRMAs can take many forms beyond the overt paper mill scenario [43]. Plagiarism is, unfortunately, common in low-quality reviews – authors might copy large parts of the background or methodology from previous publications [44]. Reviewers can often catch this by noticing shifts in writing style or content that seem out of place. Journals usually run plagiarism-detection software, but reviewers can augment this by spot-checking suspicious sections. If the writing suddenly switches voice or includes details not relevant to the current review's results, it might be a sign that the text was lifted from elsewhere. If a reviewer suspects plagiarism or self-plagiarism (authors recycling their past work without citation), this should be communicated privately to the editor for further investigation.

Another area of concern is data manipulation or falsification [45]. While SRMAs do not generate new raw data, authors could manipulate the extracted numbers or analyses. For instance, they might cherry-pick outcomes or time points that yield favorable results while ignoring others. They could miscalculate effect sizes or p-values to exaggerate significance. Reviewers should recalculate key numbers when possible: for example, verify event counts from the included studies if provided, or see if the forest plot visually matches the reported

summary. If something does not add up (e.g., the text claims a significant effect, but the confidence interval crosses 1.0 in the figure), it could be a deliberate misrepresentation or an error – either way, it needs addressing. Some reviewers with statistical expertise even re-run meta-analyses when data are provided, to ensure the results are reproducible. While not every reviewer can do this, a close reading can catch many anomalies.

Ghost authorship and author misconduct are additional subtle issues. Ghost authorship refers to significant contributions from individuals not listed as authors (or conversely, listed authors who did not contribute) [46]. In the context of SRMAs, this often ties back to paper mills or professional writing services, where the people doing the work are not the ones on the author list. Reviewers might infer this if, for example, the manuscript quality is high, but the cover letter or prior publications of the authors are of much lower quality, or if the author names are known from past suspicious submissions. While it is hard for a reviewer to know for sure, any inconsistency in author qualifications and content mastery could be flagged to the editor. Journals are increasingly requiring author contribution statements and even disclaimers about the use of professional writers or AI (more on AI below), which helps shine light on who prepared and drafted the manuscript.

To detect misconduct, reviewers and editors have developed various strategies and tools. COPE has published guidance on recognizing peer review manipulation patterns (such as unusual reviewer email domains suggesting fake reviewer identities) [47]. While that pertains more to editors, reviewers should be conscious of the environment, e.g., if you receive a review request and notice the manuscript has signs of having been handled in a sketchy way (like an unusual barrage of similar papers in the journal issue), it might warrant extra caution. Some journals enlist statistical reviewers to specifically check for inconsistencies or implausible data patterns (for example, identical means and standard deviations across independent studies might indicate data fabrication). Reviewers with content expertise might notice when multiple included studies have overlapping text or figures, which could mean one or more are copied or even invented. In such cases, raising a query like “Study X and Y have strikingly similar results or phrasing – are they perhaps duplicate publications or from the same data source?” can prompt editors to investigate further.

It is also worth mentioning the emergence of new tools that can help identify red flags. For example, image forensics software can detect duplicated images in published papers (more relevant to lab studies than SRMAs). There are automated plagiarism scanners and even programs that can detect statistical implausibility (like the *GRIM* test for checking

consistency of reported means and sample sizes). While an SRMA reviewer might not run these systematically, being aware of their existence is useful. In some cases, performing a simple check like ensuring all citations are real and relevant is important – paper mill products sometimes include irrelevant or fake references to look legitimate. Bhattacharyya et al. (2023) famously found a high rate of fabricated or inaccurate references in ChatGPT-generated medical content. Likewise, a reviewer might find references in a suspect SRMA that do not support the claim in the text, suggesting the writers inserted references without reading them (a common paper mill tactic). Thus, cross-verifying a few critical references can be revealing.

In summary, scientific misconduct in SRMAs is a growing concern. Reviewers are the gatekeepers who can often spot the subtle clues of such misconduct. By being thorough with cross-checking data, verifying the originality of text, and trusting their scientific intuition, reviewers can catch many issues before publication. It is always better to voice a concern (even if it turns out to be a false alarm) than to let a potentially fraudulent article slip through. Journals have mechanisms to confidentially investigate concerns, and reviewers should use those channels when needed, rather than directly confronting authors.

Abuse and misuse of AI tools in scientific writing

The advent of advanced AI language models like ChatGPT has introduced both opportunities and pitfalls in scientific writing and publishing. On one hand, AI tools can aid in literature search, summarizing findings, or even editing text for clarity. On the other hand, there is growing evidence of its misuse, where AI is used to generate content that authors then present as their writing, or to fabricate elements of papers (like references or even data) [48]. Reviewers now must consider the possibility that a manuscript (especially the narrative portions) may have been partially or wholly written by an AI.

The ethical dilemma centers on transparency and accuracy. Most journals and editorial guidelines, including the International Committee of Medical Journal Editors/ICMJE/ recommendations, have now stated that AI tools cannot be listed as authors and that any use of AI in manuscript preparation should be disclosed in the acknowledgments or cover letter [49]. This is because AI cannot take responsibility for the content, and it certainly cannot attest to not having introduced plagiarism or errors. Reviewers, therefore, should check if the journal requires such disclosure and whether the authors have provided one. An absence of disclosure does not mean AI was not used – many authors might not admit it – but a disclosure is a helpful signal. If a reviewer notices phrasing in the text that

feels overly generic, repetitive, or stylistically inconsistent with the rest of the article, it could be AI-generated. Some common tells include overly fluent but factually shallow sentences or a strange detachment in tone in certain sections. The presence of fabricated facts or references is a major warning sign. As mentioned above, one study found that ChatGPT frequently produced nonexistent references that at a glance looked real. A reviewer who encounters a reference that seems odd (e.g., a journal or year that does not make sense for the topic) can quickly try to look up the reference. If it does not exist, that strongly suggests it was auto-generated. In one case, reviewers caught an article that had references that were entirely AI-invented; the submission was rejected for fraud.

Another misuse of AI is to generate plagiarized composites – an AI could take paragraphs from various sources and stitch them together after minor paraphrasing. This can sometimes evade plagiarism checkers that look for exact matches. However, the content might still ring a bell to an expert who has read similar reviews. If a reviewer suspects this, they can attempt a targeted search of a unique phrase in Google; if it appears in another paper, that is evidence of patchwriting via AI. Furthermore, AI might be used to polish language for non-native writers, which is not inherently unethical, but it blurs the line if entire sections are produced by AI. Journals generally permit language editing (by a human or AI) but expect the intellectual content to be the authors' own. The concern is when AI contributes ideas or text that the authors do not fully understand or verify.

One particularly worrisome scenario is using AI to generate fake data or analyses that are then included in a review. For example, an author could ask ChatGPT to fabricate a meta-analysis of studies X, Y, and Z and produce a summary. If a reviewer sees results quoted from studies that they know do not match the actual study findings, that could indicate the authors relied on an AI summary that was incorrect or even hallucinated. Majovský et al. (2023) demonstrated that GPT-3 could generate a *fully fabricated scientific article* on neurosurgery that looked quite convincing [50]. While expert readers found errors on closer inspection (especially in the references and some factual inaccuracies), a cursory review might have missed these. This illustrates that AI can produce "too good to be true" manuscripts – well-structured and formatted, but with subtle nonsensical or erroneous content. Reviewers should thus approach a slick piece of writing with healthy skepticism and focus on substance: do the data and arguments check out?

From an ethical standpoint, the misuse of AI erodes the trust in scientific communication. Journals have responded by formulating policies (e.g., Nature and Elsevier journals require disclosure and ban AI from being an author; *Science* went further to temporarily ban any text generated by ChatGPT). Tools to detect AI-generated text (like GPTZero, Originality.ai, etc.) exist, but they are not foolproof and can be evaded or yield false positives. One study analyzing conference abstracts found that in 2023, abstracts were significantly more likely (roughly two-fold increase) to contain AI-generated content compared to 2021, evidencing the rapid uptake of these tools [51]. As AI becomes more embedded, the onus is on reviewers and editors to ensure transparency. If a reviewer suspects undisclosed AI use, they might ask the editor if the journal has run an AI-detection tool on the submission as part of screening. Some publishers do this for all submissions now, flagging those above a certain threshold of "AI probability." A reviewer can also simply ask, in their comments to the editor, whether the prose appears AI-generated and suggest the authors clarify their writing process.

It is also worth noting potential benefits and acceptable uses of AI in SRMAs, as the goal is not to ban technology but to manage it ethically. AI can assist in screening literature (machine learning tools to sift through thousands of citations for relevant studies) [6], or even in drafting simpler sections of a manuscript (like a first pass at a plain language summary). If disclosed and verified by human authors, this can speed up the review process without compromising integrity. However, the line is crossed when AI is used to do the authors' thinking for them, for example, writing the discussion or interpreting results. Reviewers should encourage honesty about such contributions. A possible comment could be: "If any AI-assisted technology was used in preparing the manuscript (for writing or data analysis), please provide a disclosure of how it was used, following journal policy." This signals to authors and editors that the reviewer is attentive to this issue.

In conclusion, AI tools like ChatGPT present a double-edged sword in scientific publishing. Reviewers must adapt by learning to recognize AI's fingerprints and by pushing for transparency. The misuse of AI, whether to generate fraudulent papers or simply to do sloppy, unchecked writing, ultimately undermines the scholarly record. By remaining vigilant and advocating for clear disclosure and responsible use of AI, peer reviewers uphold the integrity of the publication process in this new era.

Ethical responsibilities and journal policies

The challenges of paper mills, misconduct, and AI misuse all point to a broader theme: the ethical responsibilities of reviewers and the policies that journals must enforce. Peer reviewers are not just evaluating content; they are also guarding the gate of scientific quality. With the surge of questionable practices, reviewers should feel empowered to act on ethical concerns, not only methodological ones. COPE's *Ethical Guidelines for Peer Reviewers* emphasize that reviewers play a pivotal role in maintaining the integrity of the scholarly record [52]. This includes confidentiality, objectivity, and vigilance against ethical lapses.

Reviewers should be aware of and aligned with the journal's policies on these matters. Many journals now require conflict of interest disclosures from reviewers (to avoid peers having undeclared ties to authors or topics), and some have policies on how to handle a manuscript that appears to violate ethical norms (e.g., suspected undisclosed duplicate publication or ethical issues in included studies). If a reviewer suspects something like data fabrication, the appropriate action is to inform the editor confidentially, providing any evidence or reasons for concern. The reviewer should not directly accuse the authors in their report, as this could lead to legal issues; rather, they should flag it to the editor to handle via the journal's procedures. Most reputable journals will take such flags seriously and may initiate an investigation or ask authors for raw data or clarification.

Journal policies have also been evolving to combat paper mills and AI abuse. For example, some journals have implemented screening checks: verifying authors' identities (to prevent fake author accounts), requiring ORCID IDs for all authors, and using software to scan for image or text duplication across submissions. As part of the peer review process, an editor might have already done some checks by the time the manuscript reaches the reviewer. Sometimes, the editor will alert reviewers: "Please be advised we have had issues with manuscripts in this topic area and to look carefully for [specific anomaly]." Reviewers should take such notes from editors seriously, as they often come from patterns noticed at the editorial level.

Another developing area is how journals handle AI tool disclosures. A reviewer might see in the manuscript a statement like, "We used ChatGPT to improve the English of this manuscript." According to ICMJE and other guidelines, this should be acceptable if properly disclosed, but the reviewer may still consider whether the use of AI could have introduced

errors. It is within a reviewer's remit to say, "Please ensure that all content generated with assistance from [AI tool] has been thoroughly validated by the authors for accuracy and originality." Journals count on reviewers to help enforce these standards pragmatically. The GAMER (Guidelines for Artificial Intelligence in Medical Research) checklist further emphasizes the need for transparency, ensuring that AI tools are appropriately disclosed, their contributions are clear, and the content is validated for both accuracy and originality, in line with ethical research practices [53].

From an ethical standpoint, reviewers should also introspect on their own biases and limitations. With contentious issues like suspected misconduct, one must balance skepticism with fairness. If a reviewer has only a hunch but no clear evidence, they might request additional information or data from the authors via the editor, rather than outright condemnation. The process should be one of clarification and ensuring trustworthiness. Ethical reviewing also means not misusing one's position – for instance, not delaying a review to benefit one's own work, or not stealing ideas from an unpublished manuscript. Given the intensity of problems like paper mills, it can be easy to become overly suspicious; reviewers should strive to remain objective and evidence-focused in their evaluations.

Lastly, the proliferation of low-quality SRMAs has led some journals to implement stricter triage criteria. For instance, journals might desk-reject SRs that are not registered or that do not strictly adhere to PRISMA. This is a positive development, and reviewers can reinforce it by mentioning in their reviews if a submission falls short of such standards. For example: "This review was not registered, and the authors give no compelling reason for its necessity, considering existing reviews on the topic. The journal's policy might be to decline such submissions." This backs the editor in making tough decisions and signals to authors the expectations of the community.

CONCLUSION

The role of reviewers in the current proliferation era of SRs and SRMAs is more crucial than ever. With tens of thousands of SRMAs being published each year, the scientific community relies on diligent peer review to sift the valid, high-quality evidence syntheses from those that are redundant, flawed, or even fraudulent. A conscientious reviewer approaches an SRMA with a blend of methodological rigor and healthy skepticism, verifying that the review asks a meaningful question, that it was conducted according to best practices, and that its results and conclusions are reliable. This involves checking the fundamentals

(clear rationale, protocol registration, comprehensive search, proper analysis, and transparent reporting of results) as well as delving into the details (assessing heterogeneity, bias, and evidence quality).

At the same time, reviewers must serve as the last line of defense against emerging threats to research integrity. Whether it is the subtle influence of paper mills, instances of plagiarism or data manipulation, or the creeping use of AI to generate content, the reviewer's vigilance can prevent these issues from polluting the scientific literature. By staying informed about issues like publication bias, ghostwriting, and AI ethics – and by leaning on established guidelines and one's own informed judgment – reviewers can detect warning signs and take appropriate action. It is a responsibility that extends beyond simply improving a manuscript; it is about safeguarding the credibility of evidence-based science.

In conclusion, high-quality SRMAs are indispensable for informing clinical practice and policy. Ensuring their quality is a collective effort, but peer reviewers play an outsized role in this endeavor. This practical guide has highlighted strategies and considerations for reviewing SRMAs effectively. By meticulously evaluating methodology and results and by remaining alert to misconduct and ethical issues, reviewers can uphold the standards of excellence in academic publishing. In doing so, they help ensure that SRMAs fulfill their promise: to reliably summarize evidence for the betterment of healthcare and scientific understanding. The task is challenging and often underappreciated, but by embracing this role, reviewers become key contributors to the integrity and utility of the scientific literature.

Ultimately, fostering a rigorous and ethical review culture will enhance the reliability and impact of SRMAs, allowing them to truly inform and shape evidence-based practice across disciplines, even amid an era of information overload and evolving challenges.

Conflicts of interest: The authors declare no conflicts of interest.

Funding: Authors received no specific funding for this work.

Submitted: 20 July, 2025

Accepted: 20 July, 2025

Published online: 20 July, 2025

REFERENCES

1. Ahn E, Kang H. Introduction to systematic review and meta-analysis. Korean J Anesthesiol. 2018 Apr;71(2):103-112.
2. Sargeant JM, Brennan ML, O'Connor AM. Levels of Evidence, Quality Assessment, and Risk of Bias: Evaluating the Internal Validity of Primary Research. Front Vet Sci. 2022;9:960957.
3. Zerem E, Vranić S, Hanjalić K, et al. Scientometrics and academia. Biomol Biomed. 2024 Mar 11;24(2):207-209.
4. Vasconez-Gonzalez J, Izquierdo-Condoy JS, Naranjo-Lara P, et al. Integrity at stake: confronting "publish or perish" in the developing world and emerging economies [Perspective]. Frontiers in Medicine. 2024 2024-July-17;11.
5. Retraction notice. Scottish Medical Journal. 2024:00369330241240826.
6. Blaizot A, Veettil SK, Saidoung P, et al. Using artificial intelligence methods for systematic review in health sciences: A systematic review. Res Synth Methods. 2022 May;13(3):353-362.
7. Bečulić H, Begagić E, Skomorac R, et al. ChatGPT's contributions to the evolution of neurosurgical practice and education: a systematic review of benefits, concerns and limitations. Med Glas (Zenica). 2024 Feb 1;21(1).
8. Chakraborty C, Pal S, Bhattacharya M, et al. Overview of Chatbots with special emphasis on artificial intelligence-enabled ChatGPT in medical science. Front Artif Intell. 2023;6:1237704.
9. Metze K, Morandin-Reis RC, Lorand-Metze I, et al. The Amount of Errors in ChatGPT's Responses is Indirectly Correlated with the Number of Publications Related to the Topic Under Investigation. Ann Biomed Eng. 2023 Jul;51(7):1360-1361.
10. Bhattacharyya M, Miller VM, Bhattacharyya D, et al. High Rates of Fabricated and Inaccurate References in ChatGPT-Generated Medical Content. Cureus. 2023 May;15(5):e39238.
11. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021 Mar 29;372:n71.
12. Brooke BS, Schwartz TA, Pawlik TM. MOOSE Reporting Guidelines for Meta-analyses of Observational Studies. JAMA Surg. 2021 Aug 1;156(8):787-788.

13. Pollock M, Fernandes RM, Pieper D, et al. Preferred Reporting Items for Overviews of Reviews (PRIOR): a protocol for development of a reporting guideline for overviews of reviews of healthcare interventions. Syst Rev. 2019 Dec 23;8(1):335.
14. Sohrabi C, Franchi T, Mathew G, et al. PRISMA 2020 statement: What's new and the importance of reporting guidelines. Int J Surg. 2021 Apr;88:105918.
15. Pieper D, Rombey T. Where to prospectively register a systematic review. Syst Rev. 2022 Jan 8;11(1):8.
16. Pieper D, Rombey T. Where to prospectively register a systematic review. Systematic Reviews. 2022 2022/01/08;11(1):8.
17. Frandsen TF, Bruun Nielsen MF, Lindhardt CL, et al. Using the full PICO model as a search tool for systematic reviews resulted in lower recall for some PICO elements. J Clin Epidemiol. 2020 Nov;127:69-75.
18. Eriksen MB, Frandsen TF. The impact of patient, intervention, comparison, outcome (PICO) as a search strategy tool on literature search quality: a systematic review. J Med Libr Assoc. 2018 Oct;106(4):420-431.
19. Alharbi A, Stevenson M. Refining Boolean queries to identify relevant studies for systematic review updates. J Am Med Inform Assoc. 2020 Nov 1;27(11):1658-1666.
20. Vassar M, Yerokhin V, Sinnett PM, et al. Database selection in systematic reviews: an insight through clinical neurology. Health Info Libr J. 2017 Jun;34(2):156-164.
21. Bramer WM, Rethlefsen ML, Kleijnen J, et al. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. Syst Rev. 2017 Dec 6;6(1):245.
22. Haddaway NR, Page MJ, Pritchard CC, et al. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. Campbell Syst Rev. 2022 Jun;18(2):e1230.
23. Cheungpasitporn W, Wathanavasin W, Thongprayoon C, et al. Critical appraisal of systematic reviews and meta-analyses: a step-by-step guide for nephrologists. Ren Fail. 2025 Dec;47(1):2476736.
24. Stogiannis D, Siannis F, Androulakis E. Heterogeneity in meta-analysis: a comprehensive overview. Int J Biostat. 2024 May 1;20(1):169-199.
25. Choi GJ, Kang H. Heterogeneity in meta-analyses: an unavoidable challenge worth exploring. Korean J Anesthesiol. 2025 Feb 14.

26. Higgins JPT, Li T, Deeks JJ. Choosing effect measures and computing estimates of effect. Cochrane Handbook for Systematic Reviews of Interventions2019. p. 143-176.
27. Rücker G, Schwarzer G, Carpenter JR, et al. Undue reliance on I2 in assessing heterogeneity may mislead. BMC Medical Research Methodology. 2008 2008/11/27;8(1):79.
28. Cheungpasitporn W, Wathanavasin W, Thongprayoon C, et al. Critical appraisal of systematic reviews and meta-analyses: a step-by-step guide for nephrologists. Renal Failure. 2025 2025/12/31;47(1):2476736.
29. Farrokhyar F, Skorzewski P, Phillips MR, et al. When to believe a subgroup analysis: revisiting the 11 criteria. Eye. 2022 2022/11/01;36(11):2075-2077.
30. Mathur MB, VanderWeele TJ. Meta-regression methods to characterize evidence strength using meaningful-effect percentages conditional on study characteristics. Res Synth Methods. 2021 Nov;12(6):731-749.
31. Morton SC, Adams JL, Suttorp MJ, et al. AHRQ Technical Reviews. Meta-regression Approaches: What, Why, When, and How? Rockville (MD): Agency for Healthcare Research and Quality (US); 2004.
32. Verhagen AP, Ferreira ML. Forest plots. J Physiother. 2014 Sep;60(3):170-3.
33. Sarkar S, Baidya DK. Meta-analysis - interpretation of forest plots: A wood for the trees. Indian J Anaesth. 2025 Jan;69(1):147-152.
34. Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. Bmj. 1997 Sep 13;315(7109):629-34.
35. Page MJ, Sterne JAC, Higgins JPT, et al. Investigating and dealing with publication bias and other reporting biases in meta-analyses of health research: A review. Res Synth Methods. 2021 Mar;12(2):248-259.
36. Godavitarne C, Robertson A, Ricketts DM, et al. Understanding and interpreting funnel plots for the clinician. Br J Hosp Med (Lond). 2018 Oct 2;79(10):578-583.
37. Mathur MB, VanderWeele TJ. Robust Metrics and Sensitivity Analyses for Meta-analyses of Heterogeneous Effects. Epidemiology. 2020 May;31(3):356-358.
38. Balshem H, Helfand M, Schünemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011 Apr;64(4):401-6.
39. Byrne JA, Abalkina A, Akinduro-Aje O, et al. A call for research to address the threat of paper mills. PLoS Biol. 2024 Nov;22(11):e3002931.

40. Stone JAM. Artificial Intelligence-Generated Research in the Literature: Is It Real or Is It Fraud? Med Acupunct. 2023 Jun 1;35(3):103-104.
41. Addressing paper mills and a way forward for journal security.
42. Tang G, Cai H. Citation Contamination by Paper Mill Articles in Systematic Reviews of the Life Sciences. JAMA Netw Open. 2025 Jun 2;8(6):e2515160.
43. Phogat R, Manjunath BC, Sabbarwal B, et al. Misconduct in Biomedical Research: A Meta-Analysis and Systematic Review. J Int Soc Prev Community Dent. 2023 May-Jun;13(3):185-193.
44. Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PLoS One. 2009 May 29;4(5):e5738.
45. Resnik DB. Data fabrication and falsification and empiricist philosophy of science. Sci Eng Ethics. 2014 Jun;20(2):423-31.
46. Yadav S, Rawal G. Ghostwriters in the scientific world. Pan Afr Med J. 2018;30:217.
47. How to spot potential manipulation of the peer review process. Committee on Publication Ethics; 2017.
48. Bernard N, Sagawa Y, Jr., Bier N, et al. Using artificial intelligence for systematic review: the example of elicit. BMC Med Res Methodol. 2025 Mar 18;25(1):75.
49. Teixeira da Silva JA. The ICMJE recommendations: challenges in fortifying publishing integrity. Ir J Med Sci. 2020 Nov;189(4):1179-1181.
50. Májovský M, Černý M, Kasal M, et al. Artificial Intelligence Can Generate Fraudulent but Authentic-Looking Scientific Medical Articles: Pandora's Box Has Been Opened. J Med Internet Res. 2023 2023/5/31;25:e46924.
51. Howard FM, Li A, Riffon MF, et al. Characterizing the Increase in Artificial Intelligence Content Detection in Oncology Scientific Abstracts From 2021 to 2023. JCO Clin Cancer Inform. 2024 May;8:e2400077.
52. Min SK. Ethics and Responsibilities of Peer Reviewers to the Authors, Readers, and Editors. Vasc Specialist Int. 2021 Mar 31;37(1):1-3.
53. Luo X, Tham YC, Giuffrè M, et al. Reporting guideline for the use of Generative Artificial intelligence tools in MEdical Research: the GAMER Statement. BMJ Evidence-Based Medicine. 2025:bmjebm-2025-113825.

TABLES AND FIGURES WITH LEGENDS

Table 1. Reviewer Checklist for Systematic Reviews and Meta-Analyses (SRMAs)

Section	Item No.	Checklist Item	Response (Yes/No/NA)	Comments
A. Initial Evaluation	A1	Is there a clear justification for conducting this SRMA (e.g., update/new evidence vs. redundancy)?		
	A2	Has the manuscript attached and addressed a PRISMA/MOOSE/PRIOR checklist?		
B. Methodology	B1	Is the review protocol prospectively registered (e.g., PROSPERO) with registration number provided?		
	B2	Are inclusion/exclusion criteria clearly defined using the PICOS framework?		
	B3	Is the search strategy (databases, dates, keywords, Boolean operators) fully detailed and reproducible?		
	B4	Is a PRISMA flow diagram included, with numbers and reasons for exclusions at each step?		
	B5	Were data extracted by at least two independent reviewers, with a reconciliation process described?		
	B6	Has risk of bias in included studies been assessed using an		

		appropriate tool (e.g., RoB 2, ROBINS-I)?
	B7	Is the planned analytical approach (fixed- vs. random-effects, subgroup/meta-regression) pre-specified in Methods?
C. Results	C1	Are effect measures appropriate for the data (e.g., RR/OR for dichotomous, MD/SMD for continuous outcomes)?
	C2	Is heterogeneity quantified (I^2 and/or Q-test) and interpreted?
	C3	Have subgroup analyses or meta-regression been conducted or justified to explore heterogeneity?
	C4	Do forest plots accurately reflect study data and pooled estimates, with outliers identified?
	C5	Has publication bias been assessed (e.g., funnel plot, Egger's test), and are limitations of these tests discussed?
	C6	Were sensitivity analyses performed (e.g., excluding high-risk or dominant studies, alternative models)?
	C7	Is the overall quality/certainty of evidence rated (e.g., GRADE), with justification for downgrading/upgrading?
D. Emerging Issues & Ethics	D1	Are all author contributions transparently declared (no ghost or guest authorship)?

	D2	Is any use of AI tools or professional writing assistance disclosed?
	D3	Are included studies checked for possible paper-mill origin or suspicious patterns (e.g., retractions, abnormal positive rates)?
	D4	Is plagiarism or patchwriting screened for and addressed?
	D5	Are fabricated or "hallucinated" references verified against original sources?
	D6	Are any conflicts of interest declared by authors or reviewers?
	D7	Does the manuscript comply with the journal's specific policies (e.g., registration requirement, ORCID for all authors, AI disclosure)?

Note for reviewers:

1. Mark each item as Yes, No, or N/A (Not available).
2. For any No or N/A, provide a brief comment or question for the authors.

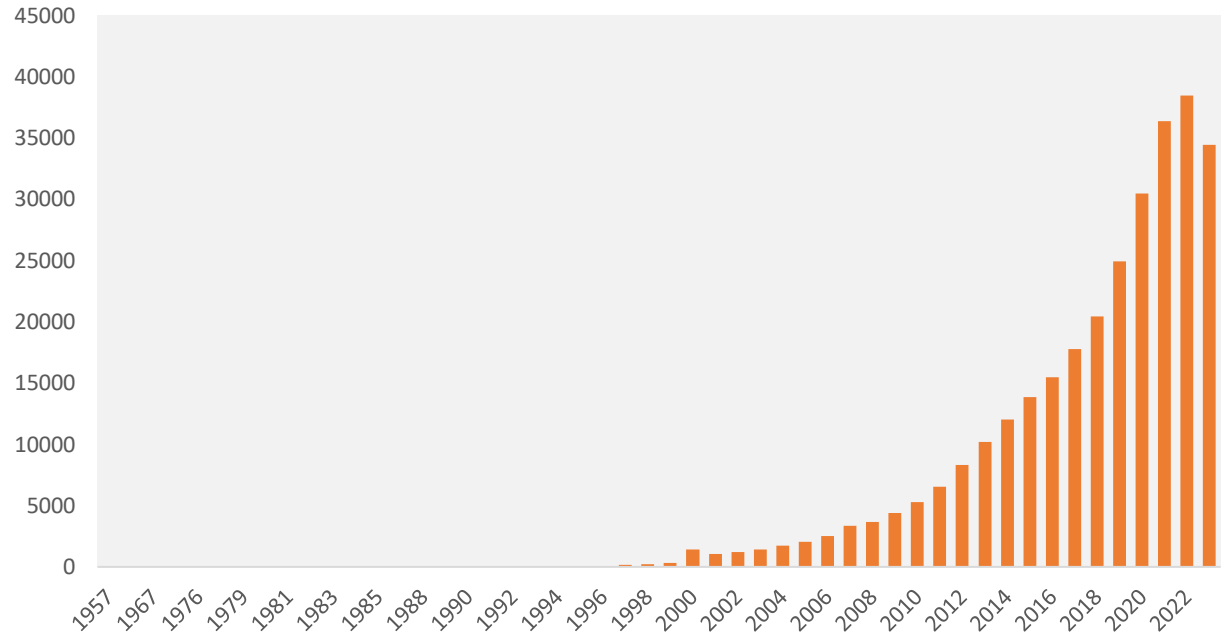

Figure 1. Number of Published Systematic Reviews and Meta-Analyses. Data were retrieved from PubMed/MEDLINE using the "systematic review" and "meta-analysis" filters on December 14, 2024.