WOŹNIAK, Łukasz, LOMPART, Albert, WABISZCZEWICZ, Michał, KOSAREWICZ, Albert and KRYSIAK, Patrycja. Enhancing the Integrity of Evidence-Based Medicine: A Literature Review on Reporting Quality in Systematic Reviews and Randomised Controlled Trials. Quality in Sport. 2025;44:62873. eISSN 2450-3118.
https://doi.org/10.12775/QS.2025.44.62873
https://apcz.umk.pl/QS/article/view/62873

The journal has been awarded 20 points in the parametric evaluation by the Ministry of Higher Education and Science of Poland. This is according to the Annex to the announcement of the Minister of Higher Education and Science dated 05.01.2024, No. 32553. The journal has a Unique Identifier: 201398. Scientific disciplines assigned: Economics and Finance (Field of Social Sciences); Management and Quality Sciences (Field of Social Sciences).
Punkty Ministerialne z 2019 - aktualny rok 20 punktów. Załącznik do komunikatu Ministra Szkolnictwa Wyższego i Nauki z dnia 05.01.2024 l.p. 32553. Posiada Unikatowy Identyfikator Czasopisma: 201398. Przypisane dyscypliny naukowe: Ekonomia i finanse (Dziedzina nauk społecznych); Nauki o zarządzaniu i jakości (Dziedzina nauk społecznych).
© The Authors 2025.
This article is published with open access under the License Open Journal Systems of Nicolaus Copernicus University in Torun, Poland. Open Access: This article is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited. This is an open access article licensed under the terms of the Creative Commons Attribution Non-commercial Share Alike License (http://creativecommons.org/licenses/by-nc-sa/4.0/), which permits unrestricted, non-commercial use, distribution, and reproduction in any medium, provided the work is properly cited.
The authors declare that there is no conflict of interest regarding the publication of this paper.
Received: 24.06.2025. Revised: 21.07.2025. Accepted: 06.08.2025. Published: 09.08.2025.

Enhancing the Integrity of Evidence-Based Medicine: A Literature Review on Reporting Quality in Systematic Reviews and Randomised Controlled Trials

Łukasz Woźniak, https://orcid.org/0009-0009-6452-3066
nototwojmail@wp.pl
West Pomeranian Center for the Treatment of Severe Burns and Plastic Surgery,
ul. Niechorska 27 Gryfice, Poland

Albert Lompart, https://orcid.org/0009-0006-7591-4765
albert.lompart@gmail.com
Saint Wojciech Hospital (Copernicus Hospital) al. Jana Pawła II 50, 80-462 Gdańsk, Poland

Michał Wabiszczewicz, https://orcid.org/0009-0006-0339-5628
michalwabi@gmail.com
Medunit Primary Care Clinic, Marii Skłodowskiej-Curie 5, 80-210 Gdansk, Poland

Albert Kosarewicz, https://orcid.org/0009-0004-9108-1754
kosarewicz.albert@wp.pl
University Clinical Centre of Gdańsk Medical University, Debinki 7, 80-952 Gdansk, Poland

Patrycja Krysiak, https://orcid.org/0009-0006-5777-3751
krysiak.patrycja00@gmail.com
Medical University of Łódź, al. Tadeusza Kościuszki 4, 90-419 Łódź, Poland

Corresponding author
Łukasz Woźniak, nototwojmail@wp.pl

Abstract

This literature review synthesises evidence on reporting quality and integrity in evidence-based medicine (EBM). It critically examines challenges undermining the trustworthiness of systematic reviews (SRs) and randomised controlled trials (RCTs) and explores solutions and frameworks to address these issues. Based on 22 selected academic articles, including reviews, guidelines, studies, and opinions, the analysis highlights themes like reporting standards (PRISMA, CONSORT, SPIRIT), appraisal tools (GRADE, AMSTAR), technology (AI, telehealth, electronic data), and expert collaboration. Findings show a gap between reporting guidelines and practice, causing a "trustworthiness crisis". Despite guideline adoption, poor reporting, data issues, and bias persist. Key solutions include involving information science experts to refine searches, using AI and automation to boost efficiency and spot integrity issues, adopting new tools like Evidence Gap Maps (EGMs), and improving trial design with Decentralised Clinical Trials (DCTs). Strengthening EBM's integrity needs more than guideline compliance: it demands a cultural shift towards transparency, combined with methodological rigour, technology, and interdisciplinary collaboration. Only this integrated approach ensures that clinical and policy decisions rest on the most robust and credible evidence possible.

Keywords: Evidence-Based Medicine, Systematic Reviews, Randomised Controlled Trials, Reporting Guidelines, PRISMA, CONSORT, Research Integrity, Methodological Quality, Evidence Synthesis, Publication Bias.

Introduction

Evidence-based medicine (EBM) is a cross-sectional fundamental paradigm of modern health care, which is understood as the thinking, careful, and obvious use of current best evidence in making decisions regarding patient care (Butt et al., 2024). EBM needs good quality science to direct such an endeavour to attain the ultimate patient outcomes and produce regular quality care. The most obviously identified part of the EBM model is the evidence pyramid. This system is currently rated as the most thorough methodological approach and can yield the least biased results. Randomised controlled trials (RCTs) are the most solid output of such hierarchy in measuring the efficacy of the clinical intervention. In RCT, the design is more

rigorous, where subjects are randomly assigned to receive intervention or control, which hugely limits the selection and confounding bias that might result in biased results (Clyne et al., 2020; Butt et al., 2024). Even better than one RCT, it merges many quality trials. Systematic reviews (SRs) and meta-analyses (MAs) systematically identify, appraise, and summarise all available indexed research data produced on an intervention or topic, offering more rigorous evidence. Meta-analyses pool data across studies, allowing for greater statistical power and more accurate estimates of an intervention's effect, making them the sound foundation on which clinical practice guidelines (CPGs) are developed and disseminated (Mancin et al., 2022; Platz, 2021).

However, evidence-based medicine is built on a highly flimsy base: trust. Such evidence will only be as valuable as the quality of the primary research, the clarity with which it is reported, and the integrity of its synthesis. This basis has been undermined recently; others have called this a "crisis of trustworthiness." These have been documented in a rich and growing literature about fundamental and ongoing deficiencies in clinical research design, conduct, and reporting. Such limitations give rise to avoidable research waste, whereby resources are wasted in studies that do not yield either reproducible or helpful information, and the scientific principle of reproducibility of knowledge is severely hampered (Clyne et al., 2020). Even more concerning, the crisis is exacerbated by cases of outright research misconduct, including data fabrication; the explosion of so-called zombie trials that may have never been conducted; and an alarming increase in the number of retracted publications that have eroded trust in the medical literature (Núñez et al., 2023).

An international scientific community has developed a vast family of reporting guidelines and methodological frameworks to help maximise transparency of reporting, completeness of research reporting, and quality of research reporting to help address this expanding problem. These include the Consolidated Standards of Reporting Trials (CONSORT) of RCTs, Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) of trial protocols, and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) of systematic reviews and meta-analyses, and are among the most used. These give checklists and the flow diagrams of a minimum set of objects that a research report must report to make the reader know whether the researcher has done enough to understand what they were likely to accomplish and what was likely to be done, what was done, what was found and how to judge the validity of the study critically. These tools enable users to systematically assess the level of evidence quality and methodological rigour of a review, thus

supporting the translatability of research findings into recommendations in a credible way (Platz, 2021; Shaheen et al., 2023).

Although these essential tools are widely endorsed and easily accessible, their usefulness has been hampered by a significant hurdle: poor uptake (Clyne et al., 2020). To do so, the following literature review will draw upon the evidence of 22 recent academic publications to build a comprehensive picture of reporting quality and integrity in SRs and RCTs. The review will do so by interrogating the persistent problems that have undermined the evidence base and by interrogating the imaginative and, in some cases, revolutionary solutions being proposed to recreate the trust. In this space are some of the topics we are exploring the potential of technology to reshape. For example, we are using artificial intelligence (AI) for evidence synthesis and decentralised clinical trials (DCTs) to enhance the efficiency and diversity of trials. It will also address the importance of interdisciplinary collaboration, particularly the collaboration with specialists in information science, and the development of new methodological solutions that are necessary to guarantee that the techniques we test, evaluate, and refine improve the utility, transparency, and reliability of the evidence syntheses that underpin modern medicine.

The Pillars of Evidence-Based Medicine

The Role of Randomised Controlled Trials (RCTs)

Randomised Trials (RCTs) are considered the highest type of study design to ascertain clinical practice and policy interventions about efficacy (Nunez-Nunez et al., 2025). Their uniqueness lies in the fact that participants in these studies are randomised to either receive or not receive an experimental intervention, thereby reducing the possibility of selection bias while controlling confounding variables to a much greater extent and hence allowing for much stronger statements of causality concerning the effects of an intervention (Clyne et al., 2020). Once properly carried out and given due importance, RCT results ought to be published and reported to be useful in administrative decision-making; however, these processes are mostly unfairly obstructed. Poor randomisation methods, lack of allocation concealment, and imperfect masking can invalidate the results by introducing serious bias (Shaheen et al., 2023). Besides that, publishing a considerable proportion of the trials is not done, especially in cases with negative or null results: this creates publication bias and tends to inflate treatment effects in our literature (Clyne et al., 2020).

Synthesising Evidence: The Power of Systematic Reviews (SRs) and Meta-Analyses (MAs)

Given the ocean of published research, any practitioner will find it almost impossible to keep up with all relevant evidence. The SRs were provided as a solution to this cluttered information landscape. An SR is a research approach that entails collecting and considering all empirical evidence that meets pre-specified eligibility criteria for answering a research question (Mancin et al., 2023). SRs, with systematic and explicit methods, identify, select, and critically appraise relevant research and extract and analyse data from the included studies or trials (Shaheen et al., 2023). Analysis can be undertaken to provide a more precise estimate of the effect of the treatment.

It is a primitive overview of EBM. CPGs rely on evidence provided by SRs and allow for filling any gaps in the literature, offering a detailed picture of a particular subject, which negates the necessity of conducting a redundant study (Platz, 2021). A strict approach to conducting SRs demands a pre-specified protocol, exhaustive search, two-reviewer selection, information harvesting, and systematic risk-of-bias evaluation of the selected studies (Randles & Finnegan, 2023).

The "Trustworthiness Crisis" and Research Integrity

The worth of RCTs, as well as SRs, is the combination of reliability. Among the trends observed in the literature that is worrying is a lack of confidence in published studies (Butt, et al., 2024). The increased number of retractions compounds such a crisis, statements on the authenticity of RCTs, and the presence of notorious zombie trials- experiments that can be fabricated or falsified or have such severe methodological issues that it is not valid (Núñez-Núñez et al., 2023). These problematic studies can slip through the peer-review system and become incorporated into SRs, thereby polluting the evidence base and potentially leading to harmful clinical recommendations (Butt, Nunez-Nunez, et al., 2025).

Research integrity, defined as adherence to ethical and professional standards in the conduct of research, is therefore a critical concept (Núñez-Núñez et al., 2023). It does not equal methodological quality; a study can be methodologically poor because of good-faith errors. However, the absence of integrity implies that it either willfully or negligently disregards the set standards, such as in the fabrication or falsification of data. That is stated in the literature, and it has become evident to the systematic reviewers that they could no longer implicitly trust the editorial and peer-review systems of journals to guarantee integrity. Instead, a

proactive approach to assessing the integrity of included trials is becoming a necessary step in the SR process (Butt et al., 2024).

Quality Deficit: Reporting and Methodological Guidelines

The research community has developed guidelines and appraisal tools to combat the scourges of ineffective reporting and integrity issues. The tools are intended to promote transparency and methodological rigour and allow end-users to assess evidence quality and trustworthiness.

PRISMA, CONSORT, and SPIRIT: Transparent Reporting Frameworks.

The primary focus of modern research is transparency and reproducibility, as these notions are the cornerstones of a respectable body of scientific proof. The most essential tools to achieve this are reporting guidelines, since they provide a structural framework to authors. The guidelines are also not intended as an instrument to evaluate the quality of the research. As such, they are lists of the lowest common denominator of items that must be included in a research report. Their fundamental purpose is to ensure that authors provide sufficient and clear information, thereby empowering readers to fully understand the methodology, critically appraise the study's validity, and reproduce the work if necessary. By standardising the information presented, these frameworks enhance the clarity, completeness, and overall value of scientific publications across various disciplines (Boaye Belle & Zhao, 2023).

·CONSORT: Helping to increase the transparency of randomised controlled trials

The CONSORT (Consolidated Standards of Reporting Trials) statement is a crucial evidence-based framework specifically developed to improve the reporting of randomised controlled trials (RCTs). Initial CONSORT statements were published in 1996 and updated every few years; they provide a long 25-item checklist and a mandatory flow diagram. The checklist directs authors to present vital information regarding the design, conduct, analysis, and interpretation of the trial, including the methods used to generate randomisation, allocation concealment of participants, the process of blinding, and a precise description of primary and secondary outcomes. This is provided with the flow diagram, which demonstrates how the participant will move through the phases of trials: enrollment, intervention assignment, follow-up, and analysis. This flowchart shows the subjects in each stage and considers those excluded or lost to follow-up. The strong adoption rate of CONSORT in research institutions and medical journals is correlated with substantial increases in the quality of reporting, including completeness and transparency, which is essential in determining a trial's internal or external validity (Clyne et al., 2020).

- **SPIRIT: Offering Transparency even at the Protocol Level**

Where the CONSORT guideline concentrates on the final publication of a clinical trial, the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) guideline deals with the other part of the trial protocol, which is also crucial. A protocol is an elaborate study recipe that contains details regarding all the planned aspects of the design and survey undertaken before the recruitment of the initial subject. Such protocols have been made to be complete and transparent through the development of a SPIRIT statement. It provides a 33-point checklist of the minimum contents to be included in a clinical trial protocol, including the trial's objectives, administrative information, how participants should be recruited, how the data will be collected, and the plan for how the data will be analyzed statistically. Such an open and thorough protocol is important because it pre-determines the procedures of the trial and the results. This practice is vital for reducing the risk of selective reporting bias (where researchers only report favourable outcomes) and preventing undeclared post-hoc changes to the study design, which can undermine the credibility of the results (Platz, 2021). SPIRIT is to be utilised along with CONSORT since it helps to enhance the quality of protocols and, hence, assists in covering the integrity of the whole circle of a clinical trial.

· **Synthesizing Reporting of Evidence: PRISMA**

Shifting focus from primary studies to evidence synthesis, the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement is the leading framework for reporting systematic reviews and meta-analyses. They represent the synthesis of the findings of several primary studies on a given subject and occupy the highest level on the evidence pyramid. The PRISMA statement, most recently updated in 2020, consists of a detailed 27-item checklist and a four-phase flow diagram designed to help authors report their work with exceptional clarity and completeness (Page et al., 2021). The checklist addresses all areas of the review, including the title and abstract, methods, results, and discussion, whereby the authors are expected to elaborate their search strategy, the criteria they will use to select studies, the process of extracting data, and how they plan to address the risk of bias in studies that will make it to the review. The four-phase flow diagram (Identification, Screening, Eligibility, and Included) visually documents the entire review process, showing how the authors systematically filtered many initial records down to the final set of studies included in the synthesis. The ultimate intention behind using PRISMA is to allow readers to rigorously assess the strengths and weaknesses of a review and, consequently, determine the trustworthiness of its findings and conclusions (Nawijn et al., 2019).

Systems of Methodological Appraisal: GRADE and AMSTAR

Beyond ensuring that research is transparently reported, a crucial next step in EBM is critically appraising the evidence. This changes the focus of assessment from the quality of reporting to methodological quality and assurance of the evidence. Reporting guidelines, including PRISMA and CONSORT, provide information about what was done. Conversely, appraisal guidelines, including AMSTAR and GRADE, give us information regarding the quality with which the results have been completed and to what extent we can believe them. Thus, there are numerous tools.

AMSTAR (A Measurement Tool to Assess Systematic Reviews)

AMSTAR (A Measurement Tool to Assess Systematic Reviews) is a tool to measure the methodological quality of validated systematic reviews (SRs) (Shaheen et al., 2023). Its primary function is to provide a structured and standardised way to evaluate the process by which a review is conducted. It does not assess the results of the review or the quality of the primary studies within it, but rather the quality of the review's methodology. AMSTAR 2, the updated version, is a 16-item checklist that has expanded its scope to apply not only to SRs of randomised controlled trials (RCTs) but also to those including non-randomised studies of interventions, a critical update given the increasing use of real-world evidence.

The tool helps users differentiate between high- and low-quality reviews by focusing on critical domains of review conduct. These domains include:

· A Priori Design: Did the review authors register a protocol before commencing the review? This is critical for preventing post-hoc changes to methods and selective reporting of outcomes (Muka et al., 2023).

· The comprehensiveness of the Search: Was the literature search strategy adequate and comprehensive, including searches of multiple databases and efforts to identify grey literature to mitigate publication bias?

· Study Selection and Data Extraction: Were these key steps performed in duplicate by two independent reviewers to minimise error and bias in the selection and extraction process?

· Risk of Bias Assessment: Did the review authors assess the risk of bias in the included primary studies using an appropriate tool (e.g., the Cochrane Risk of Bias tool for RCTs)? This is fundamental to understanding the validity of the primary evidence base (Shaheen et al., 2023).

· Appropriateness of Meta-Analysis: If a meta-analysis was performed, did the authors use appropriate statistical methods and adequately investigate the potential impact of heterogeneity?

By systematically addressing these and other items, AMSTAR provides a transparent assessment of the methodological strengths and weaknesses of an SR, helping users gauge its trustworthiness.

GRADE (Grading of Recommendations, Assessment, Development and Evaluations)

While AMSTAR assesses the quality of the SR's methodology, the GRADE approach is used for the distinct but related purpose of rating the certainty (or quality) of the evidence and the subsequent strength of recommendations derived from that evidence (Platz, 2021). GRADE has become the international standard for developing clinical practice guidelines.

A key feature of GRADE is that it is "outcome-centric." This means that evidence is not given a single quality rating for an entire review; rather, the certainty of evidence is rated for each specific outcome across a body of studies (Jafari & Sahebkar, 2025). For example, in a review of a new drug, the certainty of evidence for its effect on mortality might be "high," while the certainty for its impact on quality of life might be "low."

The GRADE framework starts by categorising evidence from RCTs as high-quality and evidence from observational studies as low-quality. This initial rating is then modified based on five key domains that can decrease certainty and three that can increase it. The five downgrading factors are:

1. Risk of Bias: Serious limitations in the design or conduct of the primary studies.
2. Inconsistency: Unexplained heterogeneity or variability in the results across studies.
3. Indirectness: Differences between the research question and the available evidence regarding population, interventions, comparators, or outcomes (PICO).
4. Imprecision: Results have wide confidence intervals, indicating statistical uncertainty.
5. Publication Bias: There is a high likelihood that studies with negative or null findings have not been published.

Based on an assessment of these factors, the evidence for each outcome is rated as high, moderate, low, or very low. This transparent rating system is essential for translating the evidence from SRs into actionable clinical practice guidelines, allowing guideline panels and clinicians to understand the confidence level they can place in the estimated effects of an intervention (Platz, 2021).

AMSTAR and GRADE serve distinct, complementary roles in the EBM ecosystem. AMSTAR evaluates the methodological quality of the systematic review process, while GRADE evaluates the certainty of the synthesised body of evidence. A review can be of high methodological quality (a high AMSTAR score) but conclude that the certainty of the evidence for a particular outcome is very low (a low-grade rating). The combined use of these

tools is therefore critical for a comprehensive appraisal, ensuring that decisions in healthcare are based on evidence that is systematically synthesised and rigorously and transparently evaluated for its trustworthiness.

The Enduring Gap: Guideline Endorsement vs. Practical Adherence

While journals and funding bodies widely endorse these guidelines, their implementation in practice remains a significant challenge. Several articles in this review highlight a persistent gap between standards and reality. For instance, Nawijn et al. (2019) found that in a sample of SRs in emergency medicine, while journals requiring PRISMA endorsement published reviews with a statistically significantly higher quality of reporting, no single review fulfilled all PRISMA criteria. Similarly, Clyne et al. (2020), in an observational study of RCTs in Irish Health Research, found that despite widespread endorsement of reporting standards, reporting remains suboptimal in key domains. For example, only 32% of RCTs published since the 2005 mandate for trial registration provided a registration number, and only 16% made specific reference to the CONSORT statement.

This lack of adherence is not trivial; it directly impacts the usability and trustworthiness of the evidence. Butt et al. (2025) conducted a systematic review of recommendation documents for clinical trial integrity. They found that 78% were of poor quality according to appraisal checklists, highlighting a systemic failure to produce robust guidance. The body of knowledge clearly shows that though guidelines are the needed initial step, the mere endorsement is not enough to bridge the gap in reporting quality.

Unrelenting Difficulties in Carrying out Quality Literature Reviews

While robust reporting guidelines like PRISMA provide an essential blueprint for structuring a systematic review (SR), they do not eliminate the profound practical and methodological hurdles inherent in the process. The process of SR is much more than a box-ticking exercise; it is a high-fidelity scientific initiative fraught with difficulties that have only been exacerbated by the contemporary information environment. These challenges run the full review cycle, from the Herculean effort of locating all the pertinent literature to the essential critique of primary data flaws to the intricate manoeuvring of statistical and methodological heterogeneity. How well the review team survives these nagging problems will determine the validity and reliability of the conclusions made in the SR and risk the introduction of bias and uncertainty at every stage.

Finding the Way through the Deluge: The Search and Scope Problem

The first and most obvious obstacle the contemporary systematic reviewer could encounter is the geometric increase in published research. This "deluge of information" presents a monumental obstacle to one of the core tenets of a systematic review: the comprehensive identification of all relevant studies (Stevens & Laynor, 2024). To do that, reviewers should design and implement a carefully crafted search strategy that is sensitive enough to identify all the relevant articles and specific enough to identify no more than an unmanageable number of irrelevant articles. It is a very technical activity and needs great expertise. The strategy should be appropriately transferred to various electronic databases, including PubMed, Embase, and the Cochrane Library, where each database has its specific search syntax, indexing system, and controlled vocabulary. For example, a search string built using Medical Subject Headings (MeSH terms) for PubMed must be re-engineered using Emtree terms for Embase, a nuance that, if overlooked, can lead to the failure to identify critical studies and the introduction of significant bias into the review (Shaheen et al., 2023).

Compounding this issue is the "grey literature," a vast and nebulous body of evidence outside traditional academic publishing channels. These are conference proceedings, dissertations, government and non-governmental organisation reports, and clinical trial registries. This literature is infamously complex and expensive to search because it is not indexed in standard databases. Nevertheless, it is vital to the integrity of an SR to include it. Studies that show null or negative results are statistically less acceptable to peer-reviewed journals and, maybe, grey literature. Their omission creates publication bias, distorting the evidence available by focusing on positive findings (Sablerolles et al., 2022). Therefore, a truly comprehensive search must venture into these less-structured domains. This task demands considerable time and resources but is essential for creating an unbiased and complete picture of the evidence (Stevens & Laynor, 2024).

Data Quality and Integrity: The "Garbage In, Garbage Out" Problem

The systematic review is essentially a secondary analysis, and its quality is inseparably associated with the quality of the primary studies that it contains. One of the most serious and widespread problems is the low quality of the source data on which such reviews are constructed. This issue is particularly acute in research that relies on real-world data, such as information extracted from electronic health records (EHRs). As highlighted in a review by Lighterness et al. (2024), while EHRs offer a massive reservoir of clinical data for research, they are fundamentally administrative tools, not research databases. As a result, there are

usually serious problems of incompleteness, inconsistency, and inaccuracy of the data they carry. For example, a patient's record may have missing data points for key variables, different clinicians may use non-standardised terms for the same diagnosis, or data entry errors may lead to factually incorrect information.

This presents a "garbage in, garbage out" dilemma for systematic reviewers. An SR that adheres to the PRISMA guidelines may still reach utterly misleading conclusions if it summarises the data of primary studies that, in turn, are founded on incorrect EHR data. This problem is further complicated because the standardised frameworks to measure the quality of this source data and enhance it are in dire shortage. Lighterness et al. (2024) found significant heterogeneity in how primary studies attempted to address data quality, with many failing to tackle the root causes of the poor data. This implies that reviewers are commonly left with insufficient information to assess the validity of the data included in the studies they incorporate, and they have to either include questionable information or omit a significant part of available real-world evidence, neither of which is an ideal situation.

The Heterogeneity Conundrum and the Omneurous Specter of Bias

Moreover, even when high-quality studies are successfully located, the reviewers face the complicated task of combining their results. One is heterogeneity, defined as the variation or diversity of the studies in a review. This variation may take several forms. Clinical heterogeneity occurs because of variations in the populations of patients, the details of the intervention, or the outcomes of interest. Methodological heterogeneity results from disparities in the study design, conduct, or risk of bias. These two forms often lead to statistical heterogeneity, where the observed treatment effects vary between studies more than expected by chance alone (Shaheen et al., 2023). The investigation and evaluation of heterogeneity is a serious process in any review. Simply pooling highly dissimilar studies—often described as mixing "apples and oranges"—in a single meta-analysis can produce a statistically precise but clinically meaningless result. Reviewers must use statistical tools (like the I^2 statistic) and analytical methods (like subgroup analysis) to investigate the sources of heterogeneity and decide whether pooling is appropriate.

Publication bias is also one of the greatest threats to the validity of any evidence synthesis. As previously discussed, this is the phenomenon where the direction or statistical significance of a study's results influences its likelihood of being published. Those studies that demonstrate positive or exciting new effects are more likely to be written up by authors, accepted by journals, and consequently found by systematic reviewers. This systemic issue can lead to a

literature base that paints an overly optimistic picture of an intervention's benefits. As Nawijn et al. (2019) emphasise, this bias undermines the trustworthiness of a review's findings. Although reviewers use statistical techniques such as funnel plots to identify the possibility of publication bias and make great efforts to locate unpublished studies, this is a widespread and hard-to-eliminate problem that can skew the findings of an otherwise fine systematic review.

Ingenious Solutions and the Way Ahead

The literature indicates a convergent complex of solutions involving human expertise and technological and methodological development to address these ongoing issues.

Critical Role of Human Skills: Information Science

The sophistication of current literature searching supports the requirement of specific expertise. Stevens and Laynor (2024) compellingly argue for the essential collaboration between research teams and library and information science experts. Such specialists can create strict and thorough search strategies, deal with the complexities of various databases and their controlled vocabularies, and successfully seek grey literature. Their role goes beyond administration and is essential to achieving methodological rigour and completeness of a systematic review. Studies have shown that reviews with librarian co-authors report higher-quality search strategies (Stevens & Laynor, 2024). Such cooperation is effective and inexpensive in improving the quality of evidence synthesis.

In Evidence Synthesis: Technological Innovations

Technology can provide strong instruments to meet the EBM lifecycle's scale, efficiency, and integrity challenges.

· **Artificial Intelligence (AI) and Automation**: AI and machine learning are poised to revolutionise evidence synthesis. These technologies can assist in multiple stages of the review process, including screening titles and abstracts, extracting data, and assessing the risk of bias (Stevens & Laynor, 2024; Boaye Belle & Zhao, 2023). Han et al. (2024), in a scoping review of RCTs evaluating AI in clinical practice, note the rapid expansion of this field. However, they also caution that the evidence base for AI's effectiveness and optimal implementation is still developing. In the case of SRs, AI-based technologies have the potential to considerably decrease the burden of manual screening, leaving reviewers with higher-order tasks of interpretation and synthesis. AI can also be used for integrity assessment, for example, by detecting anomalous data patterns that may suggest fabrication or falsification

(Butt, Fawzy, et al., 2024). However, the critical need to validate these AI tools remains a key theme (Han et al., 2024).

· **Decentralised Clinical Trials (DCTs) and Telehealth:** Technology is also transforming primary research. Goodson et al. (2025) describe an academic-industry model for implementing DCTs, using technology like telehealth, electronic consent (e-consent), and remote data capture to conduct trials outside traditional brick-and-mortar research sites. This model has the potential to bring research closer to various and geographically scattered populations, which may finally solve the under-representation problem in clinical trials. It has also solved the digital divide dilemma that the researchers were experiencing, as they now have the technological framework to carry out contemporary and efficient trials. Additional examples of how digital platforms can bridge the gap between clinical research and community-level health needs are electronic community resource referral systems reviewed by Drewry et al. (2023).

New Synthesis and Visualisation Methods

The scientific research field is distinguished by the vertical accumulation of data and by the lack of precedent in terms of the necessity to develop methods of efficient synthesis of this evidence and its provision to individuals who can use it, most clinicians, policymakers, researchers, and patients (Hamilton et al., 2023). The standard systematic review (SR) is a traditional evidence synthesis method upon which many rely, but it has several limitations; it is often not timely, comprehensive, or accessible. A single SR can take over a year to arrive at and typically addresses a focused clinical question, leaving decision-makers to piece a larger image together by handing out numerous, frequently conflicting reports. The evidence synthesis techniques are evolving to address these problems and meet the demands of modern evidence-based practice. These innovations are not intended to replace more traditional reviews but to complement them with nimbler, comprehensive, and easy-to-use tools that can help accelerate the translation of research into actionable intelligence. The emerging practices under development relate to the visualisation of complex landscapes of evidence, integration of different forms of evidence, and subjecting pressing global health issues to stringent synthesis procedures.

· **Visualising the Landscape: The Power of Evidence and Gap Maps (EGMs)**

The Evidence and Gap Map is relatively new in evidence synthesis and is one of the strong players. Compared to a traditional SR's narrow brief, it aims to provide a broader, panoramic overview of the body of evidence in a given area of interest. The map is a graphical grid that

formally catalogues existing evidence according to interventions plotted along relevant outcomes. The visual framework enables the stakeholders to view it and know where the evidence is strong, developing, and missing. The symbols or colours placed in these map cells typically indicate the amount and nature of the available evidence, like completed SRs, ongoing SRs, or primary studies, and typically some measure of the quality or certainty of that evidence. An example of such methodology is found in the study of Zhang et al. (2025), who suggest applying an EGM protocol to formally assess and visualise the vast amount of evidence on the different nutrient interventions in treating depression. EGM is a priceless asset in a complicated area where the possible interventions are countless (e.g., Omega-3 fatty acids, Vitamin D, B vitamins, zinc) and outcomes (e.g., symptom severity, remission rates, quality of life). It enables researchers to strategically locate the literature gaps, focusing on future research and preventing unnecessary research. An executive summary in an easy-to-read format, EGM can also supply policymakers and funding organisations with an easy-to-read executive summary that could be used directly to influence policy and research funding in the areas where it is most needed and potentially most beneficial. The EGM becomes a potent strategic instrument by turning a bewildering ocean of stand-alone studies into a navigational map (Zhang et al., 2025).

· **Filling the Gap: The Systematic Integration of Guidelines and Reviews**

One of the most critical problems of evidence-based medicine is the gap between current research knowledge and the suggestions included in clinical practice guidelines (CPGs). CPGs are meant to be among the primary sources of practical recommendations for practising clinicians. However, they may rapidly go out of date with more superior SRs published. Conversely, new SRs may not directly affect clinical practice unless they are included in a formal guideline. Systematic innovation of methodologies to integrate these two valuable sources of evidence to deal with this disjunction exists. One forward-looking method is suggested by Mancini et al. (2023), who propose to synthesise both CPGs and SRs on a specific topic. Such integrative viewpoints consider that CPGs and SRs are not mutually exclusive but complementary clinical evidence foundations. By comparing the recommendations of available CPGs with the latest findings of SRs, this method can be used to determine areas of concurrence, in which practice is current with the evidence, and areas of disagreement, in which guidelines may be outdated with the science. Such combined synthesis develops a more holistic and current view of the clinical evidence, virtually surpassing the drawbacks of utilising fragmented sources (Dermody et al., 2020). For a busy clinician, this provides a single, powerful resource that summarises the evidence's state and

contextualises it within current practice recommendations, facilitating more informed and timely clinical decision-making (Mancin et al., 2023).

· **Precision Syntheses of Big-Impact Health Issues**

In addition to the growth in broad-scoping instruments, such as EGMs, there is a trend in the evidence synthesis field towards smaller, nimbler, and more methodologically robust reviews, which address narrow, high-impact health issues. This kind of development means there is a need to provide evidence, both timely and confident, on new clinical and population health-related issues, where it is impossible to delay decisions. These reviews tended to employ complex assessment tools and novel synthesis procedures to enhance the quality and dependability of their findings. An example is that Jafari and Sahebkar (2025) describe a GRADE-rated SR protocol dealing with ginger's impact on cardiovascular health. Here, the specific application of the GRADE (Grading of Recommendations, Assessment, Development, and Evaluations) framework is essential since this tool offers a straightforward and methodical approach to grading the certainty of evidence. It is by no means just a reportage of what the studies found out; it is information to the reader as to the extent of confidence they may attach to the findings, and this is critical information to offer in prescribing treatment. In a different context, addressing the monumental impact of a global pandemic, Muka et al. (2023) conducted an umbrella review of existing SRs to synthesise the vast and rapidly accumulating evidence of the effects of the COVID-19 pandemic on cancer care. An umbrella review is a highly effective way of bringing together information when many SRs are already available, and a high-level summary can be achieved without re-analysing all of the primary studies. By synthesising the findings of multiple reviews, this approach can be used to get a robust and detailed image of interruption diagnosis, treatment, and patient outcomes. These instances reflect a strategic change to using the most advanced synthesis techniques on pressing questions so that decision-makers can access high-quality, pertinent, and focused evidence when required the most (Jafari & Sahebkar, 2025; Muka et al., 2023).

Conclusion

The fidelity of evidence-based medicine is not fixed in history but must instead be seen as "a series of quality improvements." Positioned within a dynamic tension, the present literature review comprised a meta-analysis of 22 articles demonstrating the double nature of the forces. This conflict only underlines the manifest, persistent "crisis of trustworthiness": indicative reporting, questionable data integrity, and the paradox between ideals ensconced in reporting guidelines and their application in the field. Further implications of this ongoing crisis are

research waste, discouragement of public backing, and the probability of making wrong clinical decisions and lawmaking.

Conversely, the discipline is answering back with a strong surge of creativity. The way ahead does not lie in one answer but in a multi-faceted, synergistic approach. It necessitates a rejuvenated adherence to the principal tenets of methodological rigour, as advocated by initiatives such as PRISMA, CONSORT, and GRADE. It requires a more profound inclusion of human expertise, especially of information science professionals who are needed to manage the complicated information environment. Most importantly, it entails the mindful and verified utilisation of technology. AI and automation promise to make evidence synthesis more efficient and robust, and telehealth and decentralised trial platforms can make primary research more inclusive and accessible.

Author`s contribution

Conceptualization; Łukasz Woźniak, Albert Kosarewicz,
methodology; Albert Lompart, Michał Wabiszczewicz
software, ; Albert Kosarewicz, Albert Lompart, Patrycja Krysiak
check; Albert Lompart
formal analysis; Łukasz Woźniak
investigation; Łukasz Woźniak, Albert Kosarewicz,
resources; Patrycja Krysiak, Albert Kosarewicz,
data curation; Albert Lompart
writing - review and editing; Michał Wabiszczewicz, Patrycja Krysiak
supervision; Michał Wabiszczewicz
project administration; Patrycja Krysiak, Michał Wabiszczewicz, Łukasz Woźniak

All authors have read and agreed with the published version of the manuscript.

Financing statement:

This research received no external funding.

Institutional Review Board Statement:

Not applicable.

Informed Consent Statement:

Not applicable.

Data Availability Statement:

Not applicable.

Conflict of interest:

The authors deny any conflict of interest.

Declaration of the use of generative AI and AI-assisted technologies in the writing process.

In preparing this work, the authors used ChatGPT for the purpose of improving language and readability. After using this tool, the authors have reviewed and edited the content as needed and accept full responsibility for the substantive content of the publication

References

1. Belle. B, A., & Zhao, Y. (2023). Evidence-based decision-making: Using systematic cases to check the compliance of reviews with reporting guidelines such as PRISMA 2020. Expert Systems with Applications, 217, 119569. https://doi.org/10.1016/j.eswa.2023.119569
2. Butt, F. A., Fawzy, M., Wattar, A., Bueno-Cavanillas, A., Khan, K. S., & Khalaf, Y. (2024). The randomised clinical trial trustworthiness crisis. Middle East Fertility Society Journal, 29(1). https://doi.org/10.1186/s43043-023-00161-7
3. Butt, F. A., Nunez-Nunez, M., Juhász, B., Bueno-Cavanillas, A., & Khan, K. S. (2025). The quality and reporting of recommendation documents to enhance the integrity of clinical trials: A systematic review and critical appraisal. Medicina de Familia. SEMERGEN, 51(2), 102333. https://doi.org/10.1016/j.semerg.2024.102333
4. Clyne, B., Boland, F., Murphy, N., Murphy, E., Moriarty, F., Barry, A., Wallace, E., Devine, T., Smith, S. M., Devane, D., Murphy, A., & Fahey, T. (2020). Quality, scope, and reporting standards of randomised controlled trials in Irish Health Research: an observational study. Trials, 21(1). https://doi.org/10.1186/s13063-020-04396-x

5. Dermody, G., Whitehead, L., Wilson, G., & Glass, C. (2020). The role of virtual reality in improving health outcomes for community-dwelling older adults: systematic review. Journal of Medical Internet Research, 22(6), e17331–e17331. https://doi.org/10.2196/17331
6. Drewry, M. B., Yanguela, J., Khanna, A., O'Brien, S., Phillips, E., Bevel, M. S., McKinley, M. W., Corbie, G., & Dave, G. (2023). A systematic review of electronic community resource referral systems. American Journal of Preventive Medicine, 65(6), 1142–1152. https://doi.org/10.1016/j.amepre.2023.06.001
7. Goodson, N., Wiltrout, C., Warren, J., Reis, S., & Massart, M. (2025). The research visit of the future: An academic-industry model for telehealth, electronic clinical outcomes assessments, and Decentralised Clinical Trials (DCTs). Trials, 26(1). https://doi.org/10.1186/s13063-025-08886-8
8. Hamilton, D. G., Hong, K., Fraser, H., Rowhani-Farid, A., Fidler, F., & Page, M. J. (2023). Prevalence and predictors of data and code sharing in the medical and health sciences: Systematic review with meta-analysis of individual participant data. BMJ, e075767–e075767. https://doi.org/10.1136/bmj-2023-075767
9. Han, R., Acosta, J. N., Shakeri, Z., Ioannidis, J. P. A., Topol, E. J., & Rajpurkar, P. (2024). Randomised controlled trials evaluating artificial intelligence in clinical practice: a scoping review. The Lancet Digital Health, 6(5), e367–e373. https://doi.org/10.1016/s2589-7500(24)00047-5
10. Jafari, A., & Sahebkar, A. (2025). Can ginger improve cardiovascular health indices? A protocol for a GRADE-assessed systematic review and planned dose-response meta-analysis of randomised controlled trials. Systematic Reviews, 14(1). https://doi.org/10.1186/s13643-025-02867-3
11. Lighterness, A., Adcock, M., Scanlon, L. A., & Price, G. (2024). Data Quality–Driven Improvement in Health Care: Systematic Literature Review. Journal of Medical Internet Research, 26, e57615. https://doi.org/10.2196/57615
12. Mancin. S., Sguanci, M., Andreoli, D., Soekeland, F., Anastasi, G., Michela Piredda, & Grazia, M. (2023). Systematic review of clinical practice guidelines and systematic reviews: A method for conducting comprehensive analysis. MethodsX, 12, 102532–102532. https://doi.org/10.1016/j.mex.2023.102532
13. Muka, T., Li, J. J., Farahani, S. J., & Ioannidis, J. P. (2023). An umbrella review of systematic reviews on the impact of the COVID-19 pandemic on cancer prevention

and management, as well as patient needs. ELife, 12. https://doi.org/10.7554/elife.85679

14. Nawijn. F., Ham, W., Houwert, R. M., Rolf, Hietbrink, F., & Diederik P. J. Smeeing. (2019). Quality of reporting of systematic reviews and meta-analyses in emergency medicine based on the PRISMA statement. BMC Emergency Medicine, 19(1). https://doi.org/10.1186/s12873-019-0233-6
15. Núñez-Núñez. M., Cano-Ibáñez, N., Zamora, J., Bueno-Cavanillas, A., & Khan, K. S. (2023). Assessing the Integrity of Clinical Trials Included in Evidence Syntheses. International Journal of Environmental Research and Public Health, 20(12), 6138–6138. https://doi.org/10.3390/ijerph20126138
16. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Asbjørn Hróbjartsson, L., M. M., L., T., L., E. W., M.-W., E., M., S., & McGuinness, L. A. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ, n71–n71. https://doi.org/10.1136/bmj.n71
17. Platz, T. (2021). Methods for the Development of Healthcare Practice Recommendations Using Systematic Reviews and Meta-analyses. Frontiers in Neurology, 12. https://doi.org/10.3389/fneur.2021.699968
18. Randles, R., & Finnegan, A. (2023). Guidelines for writing a systematic review. Nurse Education Today, 125, 105803–105803. https://doi.org/10.1016/j.nedt.2023.105803
19. Sablerolles, R. S. G., Rietdijk, W. J. R., Goorhuis, A., Postma, D. F., Visser, L. G., Geers, D., Schmitz, K. S., Garcia Garrido, H. M., Koopmans, M. P. G., Dalm, V. A. S. H., Kootstra, N. A., Huckriede, A. L. W., Lafeber, M., van Baarle, D., GeurtsvanKessel, C. H., de Vries, R. D., & van der Kuy, P. H. M. (2022). Immunogenicity and Reactogenicity of Vaccine Boosters after Ad26.COV2.S Priming. New England Journal of Medicine, 386(10), 951–963. https://doi.org/10.1056/nejmoa2116747
20. Shaheen, N., Shaheen, A., Ramadan, A., Mahmoud Tarek Hefnawy, Ramadan, A., Ibrahim, I. A., Hassanein, M. E., Ashour, M. E., & Flouty, O. (2023). Appraising systematic reviews: A comprehensive guide to ensuring validity and reliability. Frontiers in Research Metrics and Analytics, 8. https://doi.org/10.3389/frma.2023.1268045
21. Stevens, E. R., & Laynor, G. (2024). Enhancing the quality and efficiency of regulatory science literature reviews through innovation and collaboration with library

and information science experts. Frontiers in Medicine, 11. https://doi.org/10.3389/fmed.2024.1434427

22. Zhang, D., Li, C., Tian, X., Miao, G., Shi, W., Lu, H., Qu, Y., Cai, M., Tang, Y., Shen, H., & Gao, B. (2025). Evidence map of nutrient interventions for major depression: A systematic review protocol. Systematic Reviews, 14(1), 1–12. https://doi.org/10.1186/s13643-025-02868-2