JOURNAL OF THE BEST AVAILABLE EVIDENCE IN MEDICINE

https://doi.org/XXXXXX/XXXX-XXX-XXXXX-X

Rethinking Evidence Synthesis: Why Causally Interpreted Meta-Analysis Should Supplant Conventional Meta-Analysis at the Top of the Evidence Hierarchy in Medicine

Alsanussi T. Elsherif, MBChB, MPH, PhD

Abstract

Meta-analysis has traditionally occupied the highest position in evidence-based medicine because it combines data from multiple studies and appears to provide the most precise summary of treatment effects. Yet precision alone does not guarantee causal relevance. Conventional meta-analysis often synthesizes results across studies that differ in populations, interventions, outcome definitions, adherence patterns, and follow-up structures, while weighting studies primarily by statistical precision rather than relevance to a real-world target population. As a result, pooled estimates may be internally consistent yet poorly aligned with the causal questions.

Causally interpreted meta-analysis (CIMA) offers a more rigorous alternative. Anchored in the potential outcomes framework, it shifts evidence synthesis from associational averaging toward estimation of treatment effects in explicitly defined target populations. This approach reframes heterogeneity as effect modification, incorporates transportability and target trial emulation, and supports robust estimation through methods such as the g-formula, inverse probability weighting, augmented inverse probability weighting, and targeted maximum likelihood estimation. It also provides a principled basis for integrating randomized trials, real-world data, and causal machine learning.

This viewpoint argues that CIMA should be regarded as a higher form of evidence synthesis than conventional meta-analysis because it prioritizes causal interpretability, external validity, and decision relevance. As medicine moves toward precision care, real-world evidence, and automated analytics, evidence hierarchies must evolve accordingly. The future value of meta-analysis will depend not merely on how efficiently it pools evidence, but on how credibly it informs action in the populations where decisions are made.

Invited: 19/05/2026

OPEN ACCESS

Published online: 26/05/2026

Department of Public Health, University of Benghazi, Libya

Corresponding Author:

alsanussi.elsherif@uob.edu.ly

Key Words: Causal meta-analysis; Causal inference; Evidence synthesis; Transportability; Hierarchy of evidence

Viewpoint

Evidence-based medicine has long treated conventional meta-analysis as a pinnacle of evidentiary strength because it pools results across studies and improves statistical precision.[1-3] In practice, however, the central objective of clinical evidence is not simply to generate a narrower confidence interval, but to support valid decisions about what will happen if an intervention is used in a particular population. That is fundamentally a causal question.[1,4] The longstanding elevation of traditional meta-analysis therefore deserves reconsideration. A pooled estimate derived from multiple studies may be statistically elegant, yet still fail to answer the question that matters most: what is the likely causal effect of an intervention in the target population to whom the evidence will be applied?[1,4,5]

The problem is not that conventional meta-analysis lacks technical sophistication. Rather, its dominant framework is primarily associational. It summarizes observed study-level effects, usually by inverse-variance weighting, and often treats between-study heterogeneity as a statistical nuisance to be measured, adjusted for, or down-weighted.[1,2,4] This approach is useful for describing the evidence base, but it does not automatically produce a causally interpretable estimate. Trials differ in eligibility

criteria, co-interventions, adherence, treatment versions, follow-up duration, and outcome ascertainment.[4–6] Even randomized trials are not immune to post-randomization complications, including protocol deviations, informative loss to follow-up, and selective reporting.[4,5] Once observational studies are included, the threats multiply further through confounding, selection mechanisms, and measurement error.[4,7]

These concerns are not merely academic. Clinicians and policymakers rarely want an abstract average effect across a blended collection of study populations. They want to know whether a treatment will benefit patients in a clearly defined real-world context: older adults, multimorbid patients, underrepresented ethnic groups, registry populations, or patients managed in resource-constrained systems.[1,5,8] Traditional meta-analysis often gives disproportionate influence to larger studies with smaller standard errors, even when those studies are less representative of the population of actual interest.[1,4] Precision, in this setting, may come at the expense of relevance.

Causally interpreted meta-analysis addresses this gap by reframing evidence synthesis within the potential outcomes tradition.[1,4,7] Instead of asking only how results may be aggregated, it asks under what assumptions those results can be transported to a specified target population and interpreted causally.[1,5] This is a substantial conceptual advance. It requires investigators to define the estimand explicitly, specify the target population, identify the relevant effect modifiers, and evaluate whether the conditions for valid transportability are plausible.[4,5,8] In this framework, heterogeneity is no longer just unwanted variance; it becomes a source of information about effect modification and population differences[1,4,5]

This framework becomes even more important when considering the choice of estimators. In CIMA, the average treatment effect in the target population is not obtained by simple precision-weighted averaging alone, but by combining trial evidence with models that account for differences between trial and target populations.[1,5,6] Such methods include the g-formula, inverse probability weighting, augmented inverse probability weighting, and targeted maximum likelihood estimation.[1,5,6] These methods differ in assumptions and robustness, but they share a common strength: they explicitly attempt to estimate a causal effect for the population that matters, rather than an abstract pooled association.

Directed acyclic graphs (DAGs) serve as an indispensable tool for making these assumptions visible and analytically tractable within CIMA. A DAG encodes the investigator's substantive knowledge of the causal structure as a set of nodes representing variables and directed edges representing hypothesised causal relationships.[4,7] By formalising the roles of confounders, mediators, colliders, and other covariates, a DAG allows researchers to identify which variables must be adjusted for and which open or close specific causal pathways. DAGs therefore function not merely as illustrative aids but as formal reasoning tools that underpin the credibility of causal effect estimates across synthesised evidence.

The practical deployability of these methods is supported by dedicated software. Wang et al. developed the CausalMetaR package for R, which implements doubly robust and non-parametric efficient estimators to estimate average and subgroup treatment effects in both internal and external target populations using multi-source data.[11]

Table 1: Comparative framework: traditional versus causally interpreted meta-analysis

Feature	Traditional meta-analysis	Causally interpreted meta-analysis
Fundamental goal	To calculate a pooled estimate from multiple studies and improve statistical power	To estimate the causal effect of an intervention in a pre-specified target population
Weighting mechanism	Inverse-variance weighting, prioritizing statistical precision	Causal weighting, prioritizing similarity and relevance to the target population
Target population	Often implicit, ill-defined, or an artificial amalgam of trial participants	Explicitly defined clinical, registry, or real-world population
Handling of heterogeneity	Quantified statistically (e.g., tau^2, I^2)	Reframed as effect modification and used to transport inference
Measurement scale	Frequently assumes pooled relative measures are meaningful summaries	Explicitly addresses non-collapsibility and scale-specific causal interpretation
Inference type	Mainly associational or internal to the body of studies	Counterfactual, transportable, and decision-oriented
Key identifiability assumptions	Largely implicit; design hierarchy assumed to guarantee internal validity; exchangeability across studies rarely stated	Explicit: conditional exchangeability, positivity, consistency, non-interference, and transportability conditions stated and assessed

The package addresses a recognised barrier to adoption: the difficulty of correctly implementing nuisance function estimation, sample splitting, and cross-fitting when using flexible machine learning methods. The authors highlight that conventional meta-analytic methods pool data across sources using precision-based weights, and that the resulting pooled estimate generally does not have a clear causal interpretation because it does not pertain to a well-defined target population.[11] CausalMetaR operationalises the alternative: by explicitly specifying the target population and using transportability-aware estimators, it enables analysts to generate estimates that are

anchored to a real population of interest rather than an abstract amalgam of study participants. The availability of this open-source tool signals that CIMA has crossed from methodological theory into usable analytical practice.

Another major contribution of CIMA is the incorporation of target trial emulation into evidence synthesis.[1,5] Every synthesis should begin with a clear specification of the hypothetical pragmatic trial that would ideally answer the question of interest. This includes eligibility criteria, treatment strategies, assignment mechanism, time zero, outcomes, causal contrast, and handling of censoring.[1,5,7] Once that protocol is defined, trials and other data sources can be judged not merely by availability, but by how well they approximate the target trial. This provides far greater conceptual discipline than the conventional practice of pooling all "eligible" studies and addressing incompatibility later through sensitivity analyses.

Causal interpretation also forces deeper engagement with identifiability assumptions. Conditional exchangeability, positivity, consistency, and non-interference are not optional theoretical add-ons; they are the bridge between observed data and counterfactual claims.[4,7,8] Conventional meta-analysis often operates as though causal meaning were an automatic product of study design hierarchy, especially when randomized trials are involved. CIMA rejects that shortcut. It accepts that causal inference is assumption-dependent, but argues that making assumptions explicit is scientifically preferable to hiding them behind statistical routine.[1,4,7]

One area in which this difference is especially important is the treatment of non-linear effect measures. Relative risks and odds ratios are frequently pooled and interpreted as though they represented straightforward summary effects across populations. Yet such measures are generally non-collapsible, meaning that the pooled relative estimate may not correspond to the causal effect in any real target population.[1,4,5] CIMA brings this issue to the foreground by emphasizing causal aggregation formulas and scale-aware interpretation. This is not a technical luxury; it is central to avoiding misleading conclusions when treatment effects interact with baseline risk.

The case for elevating CIMA becomes even stronger in the era of real-world data, precision medicine, and automated analytics. Modern healthcare increasingly relies on electronic health records, registries, claims databases, digital monitoring systems, and longitudinal observational data.[1,3,5] These data sources create opportunities to define target populations more realistically and to estimate treatment effects that are clinically relevant beyond trial settings. At the same time, they introduce high-dimensional confounding structures and complex forms of treatment heterogeneity that exceed the limits of conventional meta-analytic workflows.[1,5,6] CIMA is better suited to this environment because it can incorporate doubly robust estimators, transportability methods, and causal machine learning to estimate conditional and population-specific effects.[1,5,6]

This naturally leads to the issue of automation, which may become one of the most transformative developments in causal evidence synthesis.[11] Automation in this context should not be understood as replacing methodological judgment, but as enabling consistent, scalable, and transparent implementation of causal workflows. Future causal meta-analytic platforms could automate several essential functions: identification and harmonization of eligibility criteria across studies, extraction of covariates relevant to transportability, detection of overlap and positivity violations, estimator selection, implementation of doubly robust methods, diagnostic assessment, and dynamic updating of causal estimates as new evidence emerges.[1,5,6] Such systems could integrate trial data with real-world target population data and generate estimates tailored to predefined clinical contexts.

Automation also offers an important corrective to one of the main criticisms of CIMA, namely its methodological complexity. At present, the application of transportability weights, TMLE, augmented estimators, cross-fitting, and causal diagnostics requires advanced expertise in both biostatistics and causal inference.[1,5,6] That level of expertise will remain essential for oversight, but user-friendly software can lower the threshold for adoption. In the same way that conventional meta-analysis became widely used after the emergence of packages that automated inverse-variance pooling, forest plots, heterogeneity statistics, and funnel plots, CIMA will expand more rapidly when comparable platforms become available for causal estimands, target trial specification, transportability diagnostics, and automated sensitivity analyses. However, it would be premature to overstate current automated capabilities. Existing tools such as CausalMetaR, while valuable, are research-grade implementations that require expert configuration and are not yet embedded in standard systematic review software environments.[11] Barriers including the absence of harmonized covariate ontologies across trial registries, variability in reporting of eligibility criteria, and the absence of validated pipelines for automated positivity checking represent substantive implementation challenges that the field has not widely resolved.[1,5,6] Progress in this area will depend on coordinated efforts between methodologists, software developers, data curators, and regulatory agencies.

There is also a deeper opportunity here. Automated causal synthesis could help reposition evidence-based medicine from static guideline production toward continuously updated, target-population-specific decision support.[3] As new trials, registries, and observational datasets become available, automated CIMA systems could function as "live causal meta-analyses," updating effect estimates in near real time.[1,3,5] This would be particularly valuable in fast-moving therapeutic fields, rare diseases, long-term safety surveillance, and precision medicine applications where subgroup-specific effects matter greatly. Rather than issuing a single pooled estimate for all settings, automated CIMA could generate a family of estimates tied to explicit target populations and clearly stated assumptions.

Still, automation should not be romanticized. Poorly designed automated systems could simply reproduce the limitations of conventional meta-analysis at greater speed. The quality of automated causal analysis will depend on sound target trial specification, harmonized covariate definitions, rigorous attention to missingness and censoring, and transparent reporting of assumptions and model dependence.[1,5,6] Automation can enhance consistency and scalability, but it cannot rescue invalid design. Human expertise will remain indispensable in defining the question, choosing relevant covariates, assessing plausibility of exchangeability, and interpreting findings in clinical context.

Practical limitations of CIMA should therefore be acknowledged. Access to individual participant data remains challenging; measurement heterogeneity across trials is common; and unmeasured effect modifiers remain a persistent threat.[1,4,5] The requirement for individual partic-

ipant data is particularly consequential: aggregate-level data meta-analyses, which constitute the majority of published syntheses, cannot support the covariate-level modelling that transportability methods require.[5,10] Data-sharing agreements, regulatory restrictions, and proprietary trial databases present formidable practical barriers that limit CIMA's reach in the short term.[1,11] Similarly, automated causal synthesis platforms remain nascent; the CausalMetaR package, while a meaningful advance, currently requires substantial user expertise in causal model specification and is not yet integrated into the systematic review workflows that most applied researchers employ.[11] Scaling these tools to routine clinical practice will require investment in interoperable data infrastructure, training, and validation standards that extend well beyond methodological development.[1,6] Journals and funders should therefore consider whether reporting standards for CIMA applications, analogous to PRISMA for conventional meta-analysis, are now warranted. Yet these limitations do not weaken the argument for causal interpretation. On the contrary, they strengthen it. They show that the path from evidence synthesis to decision-making is more fragile than conventional hierarchies imply. CIMA does not create these problems; it exposes them. In doing so, it provides a more transparent and methodologically mature foundation for evidence appraisal.

If evidence hierarchies are to remain relevant, they must evolve with the science of inference. A hierarchy that places conventional meta-analysis above all other designs without regard to causal interpretability is increasingly difficult to defend.[1–3] It is important to note that this argument does not diminish the internal validity of well-conducted randomized controlled trials, which remain the gold standard for establishing that an intervention can work under controlled conditions.[4,7] The critique is directed not at the individual trial, but at the conventional synthesis framework that collapses across trials with different populations without explicit causal reckoning. CIMA does not replace the RCT; it reframes how RCT evidence is transported, aggregated, and made actionable for decision-relevant populations. In an age of target trial emulation, transportability science, real-world evidence, and causal automation, the strongest evidence is not merely the most pooled, but the most decision-relevant. Causally interpreted meta-analysis should therefore be regarded not as a niche refinement of traditional synthesis, but as its logical successor.

Conclusion

Meta-analysis should no longer be judged solely by its capacity to combine studies with statistical efficiency. Its real value lies in whether it can inform causal decisions for the populations in which care is delivered. Causally interpreted meta-analysis provides that necessary shift. By integrating causal inference, transportability, target trial emulation, and the emerging possibilities of automation, it offers a more rigorous and clinically meaningful model of evidence synthesis. For that reason, it deserves to rank above conventional meta-analysis in the hierarchy of evidence in medicine.[3]

Disclosure Statement

The author declares no conflicts of interest.

Funding Statement

No funding was requested or received for the preparation or writing of this viewpoint.

Ethics Statement

Ethics approval and informed consent were not required for this viewpoint article.

References

1. Berenfeld C, Boughdiri A, Colnet B, van Amsterdam WAC, Bellet A, Khellaf R, et al. Causal meta-analysis: rethinking the foundations of evidence-based medicine. arXiv [Preprint]. 2025. doi:10.48550/arXiv.2505.20168.
2. Calderon Martinez E, Ghattas Hasbun PE, Salolin Vargas VP, García-González OY, Fermin Madera MD, Rueda Capistrán DE, et al. A comprehensive guide to conduct a systematic review and meta-analysis in medical research. Medicine (Baltimore). 2025;104(33):e41868. doi:10.1097/MD.0000000000041868.
3. El Oakley Reida, Elmuttardi Khairya, Altawaty Adel, et al. Redefining the hierarchy of evidence in medicine in the era of the next generation clinical trials and real world evidence: a critical review and the next generation hierarchy. J Best Avail Evid Med. 2025;1(2):25-9. doi:10.63720/v1i2005.
4. Markozannes G, Vourli G, Ntzani E. A survey of methodologies on causal inference methods in meta-analyses of randomized controlled trials. Syst Rev. 2021;10(1):170. doi:10.1186/s13643-021-01726-1.
5. Dahabreh IJ, Petito LC, Robertson SE, Hernán MA, Steingrimsson JA. Toward causally interpretable meta-analysis: transporting inferences from multiple randomized trials to a new target population. Epidemiology. 2020;31(3):334-44. doi:10.1097/EDE.0000000000001177.
6. Alfarone I, Gondan M. Change of direction: current limitations and future directions in psychological meta-analysis. Front Psychol. 2026;17:1717798. doi:10.3389/fpsyg.2026.1717798.
7. Hernán MA, Robins JM. Causal inference: what if. Boca Raton (FL): Chapman & Hall/CRC; 2020. Available from: https://miguelhernan.org/whatifbook
8. Zhu Y, Hubbard RA, Chubak J, Roy J, Mitra N. Core concepts in pharmacoepidemiology: violations of the positivity assumption in the causal analysis of observational data: consequences and statistical approaches. Pharmacoepidemiol Drug Saf. 2021;30(11):1471-85. doi:10.1002/pds.5338.
9. Degtiar I, Rose S. A review of generalizability and transportability. Annu Rev Stat Appl. 2023;10:501-24. doi:10.1146/annurev-statistics-042522-103837.
10. Dahabreh IJ, Robertson SE, Petito LC, Hernán MA, Steingrimsson JA. Efficient and robust methods for causally interpretable meta-analysis: transporting inferences from multiple randomized trials to a target population. Biometrics. 2023;79(2):1057-72. doi:10.1111/biom.13716.
11. Wang G, McGrath S, Lian Y. CausalMetaR: an R package for performing causally interpretable meta-analyses. Res Synth Methods. 2025;16(2):425-40. doi:10.1017/rsm.2025.5.