Current Research in Psychiatry

Commentary

From complexity to clarity: The Umbrella Collaboration® and the future of tertiary evidence synthesis in psychiatry

Beltran Carrillo, MD, PhD[1,*]

[1]The Umbrella Collaboration®, Madrid, Spain

*Author for correspondence:
Email: bcm@theumbrellacollaboration.org

Received date: July 25, 2025
Accepted date: September 02, 2025

Commentary

The exponential growth in the publication of systematic reviews and meta-analyses (SRs/MAs) over the past two decades has radically transformed the landscape of scientific evidence. In fields like psychiatry, this proliferation is particularly striking: in 2023 alone, more SRs/MAs were indexed in PubMed under psychiatry than randomized controlled trials (RCTs), a reversal of the traditional pattern. For conditions such as depression, anxiety, and schizophrenia, new SRs/MAs emerge daily. This overwhelming and abundant volume, while rich in data, presents a profound challenge: how can clinicians, researchers, and policymakers efficiently interpret and integrate these often divergent findings?

Tertiary evidence synthesis, particularly in the form of umbrella reviews, has emerged as a methodological response to this saturation. Tertiary synthesis refers to the integration and analysis of data derived from SRs/MAs, thus operating one level above primary and secondary studies. Rather than synthesizing original trials or observational studies, tertiary synthesis draws exclusively from previously synthesized evidence. Its goal is to provide a comprehensive, high-level overview that consolidates findings across multiple SRs/MAs addressing similar questions, highlighting consistencies, discrepancies, and knowledge gaps [1–4]. However, traditional umbrella reviews (TURs) are slow, labor-intensive, and frequently outdated by the time they are published. Moreover, they tend to limit their scope to descriptive mapping, rarely performing any statistical aggregation of results. Methodological authorities such as the Joanna Briggs Institute [5] advise against aggregating outcomes across multiple SRs/MAs, while Cochrane overviews allow re-analysis only from primary data and under strict conditions [6]. These approaches, while methodologically cautious, may limit the practical utility of tertiary syntheses in fast-evolving fields.

The Umbrella Collaboration® (TU®) offers a novel model for tertiary synthesis, designed from the ground up to address these, and other, structural limitations. This commentary follows the recent publication of our validation study in JMIR Formative Research, which systematically evaluated the performance of TU® in comparison to traditional umbrella review methods across eight projects in geriatrics [7]. While that study focused on methodological validation, the present work aims to explore broader conceptual and practical implications of TU® when applied to another evidence-intensive discipline: psychiatry. TU® is a semi-automated system that facilitates human-guided synthesis across all phases of the process. Unlike most tools developed to support secondary synthesis, which only assist with specific stages (e.g., screening or data extraction), to our knowledge, TU® is the first tool specifically designed for tertiary evidence synthesis and encompasses the entire workflow. The study referenced above thus represents the first published evaluation of a tool purpose-built for tertiary synthesis. While there are several studies assessing the performance of automation tools in secondary synthesis, no comparable evaluation exists for tertiary synthesis applications. It integrates automation to enhance speed and reproducibility, yet retains human oversight to ensure contextual accuracy, prevent algorithmic bias, and maintain interpretative depth.

This semi-automated design is particularly important given the known limitations of full automation in evidence synthesis. While automation can efficiently handle structured quantitative data, it struggles with qualitative nuance, and its lack of transparency can generate “black-box” concerns. TU® acknowledges these risks and implements safeguards: human reviewers supervise all stages, confirm eligibility, and validate results. The system does not aim to replace expert judgment but to augment it. Semi-automation enables a balance between scalability and rigor, making it feasible to generate robust syntheses without forgoing critical reasoning.

The central role of automation in evidence synthesis is now well established, with recent overviews detailing how AI-enabled tools improve screening, data extraction, and updating while requiring human oversight for reliability [8,9]. While existing tools aim to accelerate specific steps of the systematic review process, such as literature screening or risk-of-bias assessment, few have attempted to orchestrate a comprehensive, continuously updating synthesis. TU® fills this gap by offering a semi-automated system that encompasses the entire workflow of tertiary synthesis, from literature identification to result aggregation, result visualization, and updating. Notably, it performs literature surveillance every 24 hours, well beyond the standard monthly frequency recommended for Living Systematic Reviews by Cochrane [10].

The application of this model to psychiatry is particularly promising. Mental health research is characterized by a high volume of publications, substantial heterogeneity in outcomes, and considerable methodological diversity. Additionally, psychiatric research frequently includes subjective endpoints, population subgroups, and intervention variability, which complicates aggregation and interpretation. The challenge is not just to collect evidence, but to navigate its complexity in a way that is actionable and comprehensible. TU® addresses this by allowing synthesis projects to be built around defined PICO questions and updated with minimal effort, ensuring continued relevance in areas with high publication velocity.

The rationale for tertiary synthesis in psychiatry is further reinforced by the increasing rate of discordance between SRs/MAs answering the same question. Numerous studies have documented significant discrepancies in effect sizes, conclusions, or certainty ratings between reviews covering the same intervention or exposure [11–13]. These divergences arise from varying eligibility criteria, analytical approaches, selective reporting, or even conflicts of interest [14]. While much of the evidence for such discordance comes from other fields of medicine, there is reason to believe these challenges may be even more pronounced in psychiatry due to the inherent complexity of mental health conditions, variability in outcome measures, and subjective interpretations of clinical endpoints. TU® addresses this by treating SRs/MAs as the unit of analysis and calculating a weighted average of the reported effect sizes. This approach does not aim to overwrite individual interpretations, but to offer a centralized, cumulative signal.

Moreover, TU® aligns with international efforts to bridge the gap between research and practice. The WHO [15], along with the European Council [16], has emphasized the importance of knowledge translation in health as a means to facilitate equitable, evidence-based decision-making [17,18]. By providing results in a dynamic, interactive format, TU® enables end-users with varying degrees of statistical literacy to access, interpret, and act upon high-level evidence. This democratization of access is vital in mental health, where decisions often involve multidisciplinary teams and shared decision-making with patients and caregivers.

In the recent validation study conducted in the field of geriatrics [7], TU® was compared directly with TURs across eight projects. The findings demonstrated that TU® successfully replicated 85% (73 of 86) of the outcomes of interest identified by TURs. More importantly, it identified an additional 337 outcomes of interest, reflecting a 4.77-fold increase in the total volume of synthesized evidence. This amplification in detection was not achieved at the cost of quality: full concordance in effect size classification was observed in 50% of cases, and consistent concordance (defined as full agreement plus a one-level deviation) was achieved in 94% of comparisons. The strength of association, measured by Cramér’s V, was moderate (0.339), while Cohen’s kappa indicated fair agreement (K = 0.357). These findings illustrate that TU® not only accelerates and expands synthesis but also maintains methodological rigor in line with accepted standards.

Execution time is another critical dimension where TU® offers transformative potential. While conventional TURs may require 6 to 12 months to complete, TU® projects in the validation study were executed in under ten hours. This extreme reduction in synthesis time creates the possibility of integrating tertiary evidence synthesis into real-time decision-making workflows. In psychiatry, where clinical paradigms evolve rapidly and evidence for new interventions or risk factors emerges daily, this capability may represent a structural advantage.

One point that has drawn particular attention is TU’s use of sentiment analysis to estimate the certainty of evidence. Naturally, this cannot replace the nuanced deliberations of a GRADE assessment [19–21]. But it does offer something else: a standardized, replicable, and scalable approximation. In our validation, these sentiment-based scores showed moderate, statistically significant correlation with GRADE. This tells us that TU may be useful not just for synthesis, but for prioritization, particularly when resources are limited or timeliness is critical.

Likewise, TU’s proprietary effect size metric (R_{TU}) has evolved from a conceptual solution to a functional instrument. In our study, R_{TU} demonstrated high categorical concordance with standardized metrics like Cohen’s *d*. While the correlation at the numerical level was modest, this reflects, in part, the fact that TU works with abstract-level data only. In future versions, we plan to incorporate weighting and contextual modifiers, but even now, R_{TU} seems to perform well as a first-pass indicator of magnitude.

Several technical features distinguish TU® from traditional methods. The platform performs searches, extracts outcomes, estimates effect sizes, assesses direction and statistical significance, and updates certainty assessments using sentiment analysis. It incorporates filters and dynamic tables that enable users to explore results across subgroups or compare related outcomes. All of this is achieved without abandoning human judgment, reviewers remain involved in validation, interpretation, and final inclusion of outcomes.

It is also important to note that TU® does not attempt to replicate every function of a traditional systematic review. Instead, it streamlines those functions that can be standardized while preserving flexibility where expert input is needed. In this way, it reflects the design philosophy of rapid reviews [22–26], where scope is pragmatically defined to yield timely, actionable insights. What differentiates TU® is that it applies these principles at the tertiary level, across a meta-layer of evidence, and does so with an unprecedented level of automation and speed.

The greatest strength of TU® is not its speed, nor even its breadth, but its ability to remain active and keep synthesis projects continuously updated. Once a synthesis project is initiated, TU® continues to monitor the literature, automatically every 24 hours or upon request by the project reviewer, flagging new SRs/MAs that meet the original inclusion criteria. With minimal reviewer input, these findings can be integrated into the live project, maintaining its relevance over time.

This idea, that evidence synthesis can be 'living' without becoming unmanageable, has long been discussed in theory [27–29]. TU® makes it operational. Although this dynamic updating capacity was not explored in our initial validation study, as it requires a different methodological framework and longitudinal design, it represents a central focus for future research. This is planned for future work, as it requires extended reviewer engagement and dedicated monitoring. Nevertheless, the platform already allows for real-time incorporation of new SRs/MAs as they are published, making it one of the few systems capable of functioning as a Living Tertiary Review.

In parallel, we are currently exploring its usability and acceptance among expert reviewers. Early findings suggest that TU® is especially well received by users with experience in tertiary synthesis and by those whose areas of interest produce a high volume of SRs/MAs, such as psychiatry. These users appreciate the ability to interact with evidence in a structured, visual, and continuously evolving environment, something traditional methods cannot offer. Future studies will further characterize patterns of use, barriers to adoption, and potential integration into clinical guideline development.

The justification for tertiary synthesis in psychiatry is no longer merely conceptual; it is a practical necessity. Although the volume of SRs/MAs has grown across all fields of medicine, the pace in psychiatry is particularly intense, with conditions such as depression and anxiety each generating several new SRs/MAs daily. This surge in secondary synthesis has naturally led to an increase in tertiary synthesis publications as a response to the challenges of fragmentation and interpretative inconsistency. Between 2009 and 2020, the number of tertiary syntheses published annually increased eightfold, reaching a frequency of one publication per day by 2020 [30]. This trend underscores both the need and the opportunity for platforms like TU®.

Despite the advances in automation, the human element remains essential. The interpretation of psychiatric evidence often requires nuanced clinical reasoning and contextual awareness. TU® was built with this reality in mind. It does not pretend to automate thought but to assist it. Its design recognizes the importance of supervised automation, the only kind that can be trusted in fields where the consequences of misinterpretation are substantial [9,31].

Recent years have seen a rapid expansion of tertiary evidence in psychiatry across exposures, predictors, and interventions. Umbrella reviews have mapped predictors of treatment response across mental disorders [32], examined psychotherapy and pharmacotherapy efficacy with stricter risk-of-bias filters [33], and synthesized wide-ranging mental-health impacts of environmental exposures, including air pollution and climate change [34]. Other umbrella syntheses in youth populations and during the COVID-19 period further illustrate the field's breadth and heterogeneity [35,36]. Against this backdrop, our commentary adds a platform-oriented perspective: The Umbrella Collaboration® operationalizes a semi-automated, living tertiary-synthesis workflow that can continuously surveil, harmonize, and visualize SRs/MAs across psychiatric topics, addressing fragmentation and timeliness beyond topic-specific umbrella reviews.

As the number and complexity of SRs/MAs continues to grow, the ability to perform reliable, interpretable, and up-to-date tertiary synthesis will become increasingly central to evidence-based psychiatry. TU® is not the final word in this evolution, but it is an operational and scalable starting point. It proposes a model in which semi-automation, methodological transparency, and human oversight are not opposing forces, but mutually reinforcing components. In doing so, it offers not just a tool for synthesis, but a path forward for the science of summarizing science.

References

1. Choi GJ, Kang H. Introduction to umbrella reviews as a useful evidence-based practice. Journal of lipid and atherosclerosis. 2022 Oct 21;12(1):3.
2. Bonczar M, Ostrowski P, D'Antoni AV, Tubbs RS, Iwanaga J, Ghosh SK, et al. How to write an umbrella review? A step-by-step tutorial with tips and tricks. Folia Morphologica. 2023;82(1):1–6.
3. Fusar-Poli P, Radua J. Ten simple rules for conducting umbrella reviews. BMJ Ment Health. 2018 Aug 1;21(3):95–100.
4. Biondi-Zoccai G, editor. Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies. Cham: Springer International Publishing; 2016.
5. Aromataris E, Munn Z. JBI Manual for Evidence Synthesis—JBI Global Wiki [Internet]. [cited 2024 Jun 5]. Available from: https://jbi-global-wiki.refined.site/space/MANUAL.
6. Pollock M, Fernandes RM, Becker LA, Pieper D, Hartling L. Chapter V: overviews of reviews [last updated August 2023]. Cochrane handbook for systematic reviews of interventions version. 2024;6.
7. Carrillo B, Rubinos-Cuadrado M, Parellada-Martin J, Palacios-López A, Carrillo-Rubinos B, Canillas-Del Rey F, et al. Effectiveness of The Umbrella Collaboration Versus Traditional Umbrella Reviews for Evidence Synthesis in Health Care: Protocol for a Validation Study. JMIR research protocols. 2025 Apr 14;14(1):e67248.
8. O'Connor AM, Clark J, Thomas J, Spijker R, Kusa W, Walker VR, et al. Large language models, updates, and evaluation of automation tools for systematic reviews: a summary of significant discussions at the eighth meeting of the International Collaboration for the Automation of Systematic Reviews (ICASR). Systematic reviews. 2024 Nov 27;13(1):290.
9. Ge L, Agrawal R, Singer M, Kannapiran P, De Castro Molina JA, Teow KL, et al. Leveraging artificial intelligence to enhance systematic

reviews in health research: advanced tools and challenges. Systematic reviews. 2024 Oct 25;13(1):269.

10. 201912_LSR_Revised_Guidance.pdf [Internet]. [cited 2024 Aug 15]. Available from: https://community.cochrane.org/sites/default/files/uploads/inline-files/Transform/201912_LSR_Revised_Guidance.pdf.

11. Shrier I, Boivin JF, Platt RW, Steele RJ, Brophy JM, Carnevale F, et al. The interpretation of systematic reviews with meta-analyses: an objective or subjective process?. BMC medical informatics and decision making. 2008 May 21;8(1):19.

12. Valentine JC, Cooper H, Patall EA, Tyson D, Robinson JC. A method for evaluating research syntheses: The quality, conclusions, and consensus of 12 syntheses of the effects of after-school programs. Research Synthesis Methods. 2010 Jan;1(1):20–38.

13. Mueller M, D'Addario M, Egger M, Cevallos M, Dekkers O, Mugglin C, et al. Methods to systematically review and meta-analyse observational studies: a systematic scoping review of recommendations. BMC medical research methodology. 2018 May 21;18(1):44.

14. Ioannidis JP. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. The Milbank Quarterly. 2016 Sep;94(3):485–514.

15. World Health Organization. Bridging-the-know-do-gap.pdf [Internet]. 2005 [cited 2024 Jun 2]. Available from: https://www.measureevaluation.org/resources/training/capacity-building-resources/high-impact-research-training-curricula/bridging-the-know-do-gap.pdf.

16. Evidence-informed Policy Network (EVIPNet) Europe [Internet]. [cited 2024 May 26]. Available from: https://www.who.int/europe/initiatives/evidence-informed-policy-network-(evipnet)-europe.

17. Evidence synthesis for policy A STATEMENT OF PRINCIPLES [Internet]. The Royal Society of Evidence Synthesis; 2018. Available from: https://royalsociety.org/-/media/policy/projects/evidence-synthesis/evidence-synthesis-statement-principles.pdf.

18. Fadlallah R, El-Jardali F, Kuchenmüller T, Moat K, Reinap M, Kheirandish M, et al. Prioritizing policy issues for knowledge translation: a critical interpretive synthesis. Global Health Research and Policy. 2025 Aug 20;10(1):35.

19. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. Bmj. 2008 Apr 24;336(7650):924–6.

20. Murad MH, Verbeek J, Schwingshackl L, Filippini T, Vinceti M, Akl EA, et al. GRADE guidance 38: updated guidance for rating up certainty of evidence due to a dose-response gradient. Journal of Clinical Epidemiology. 2023 Dec 1;164:45–53.

21. Guyatt G, Vandvik PO, Iorio A, Agarwal A, Yao L, Eachempati P, et al. Core GRADE 7: principles for moving from evidence to recommendations and decisions. bmj. 2025 Jun 3;389:e083867.

22. Booth A, Jones-Diette J. Registering the Review. In: Biondi-Zoccai G, Editor. Diagnostic Meta-Analysis: A useful tool for clinical decision-making. Cham: Springer International Publishing; 2018 May 4. pp. 59–75.

23. Klerings I, Robalino S, Booth A, Escobar-Liquitay CM, Sommer I, Gartlehner G, et al. Rapid reviews methods series: guidance on literature search. BMJ Evidence-Based Medicine. 2023 Dec 1;28(6):412–7.

24. Nussbaumer-Streit B, Sommer I, Hamel C, Devane D, Noel-Storr A, Puljak L, et al. Rapid reviews methods series: guidance on team considerations, study selection, data extraction and risk of bias assessment. BMJ Evidence-Based Medicine. 2023 Dec 1;28(6):418–23.

25. Garritty C, Tricco AC, Smith M, Pollock D, Kamel C, King VJ. Rapid reviews methods series: involving patient and public partners, healthcare providers and policymakers as knowledge users. BMJ evidence-based medicine. 2024 Feb 1;29(1):55–61.

26. Gartlehner G, Nussbaumer-Streit B, Devane D, Kahwati L, Viswanathan M, King VJ, et al. Rapid reviews methods series: guidance on assessing the certainty of evidence. BMJ evidence-based medicine. 2024 Feb 1;29(1):50–4.

27. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, et al. Living systematic review: 1. Introduction—the why, what, when, and how. Journal of clinical epidemiology. 2017 Nov 1;91:23–30.

28. Akl EA, Khabsa J, Iannizzi C, Piechotta V, Kahale LA, Barker JM, et al. Extension of the PRISMA 2020 statement for living systematic reviews (PRISMA-LSR): checklist and explanation. bmj. 2024 Nov 19;387:e079183.

29. Hodder RK, Vogel JP, Wolfenden L, Turner T. Living systematic reviews and living guidelines to maintain the currency of public health guidelines. American journal of public health. 2024 Jan;114(1):21–6.

30. Lunny C, Neelakant T, Chen A, Shinger G, Stevens A, Tasnim Set al. Bibliometric study of 'overviews of systematic reviews' of health interventions: evaluation of prevalence, citation and journal impact factor. Research Synthesis Methods. 2022 Jan;13(1):109–20.

31. Department-of-Defense-Responsible-Artificial-Intelligence-Strategy-and-Implementation-Pathway.pdf [Internet]. [cited 2024 May 26]. Available from: https://media.defense.gov/2022/Jun/22/2003022604/-1/-1/0/Department-of-Defense-Responsible-Artificial-Intelligence-Strategy-and-Implementation-Pathway.PDF.

32. Solmi M, Cortese S, Vita G, De Prisco M, Radua J, Dragioti E, et al. An umbrella review of candidate predictors of response, remission, recovery, and relapse across mental disorders. Molecular Psychiatry. 2023 Sep;28(9):3671–87.

33. Leichsenring F, Steinert C, Rabung S, Ioannidis JP. The efficacy of psychotherapies and pharmacotherapies for mental disorders in adults: an umbrella review and meta-analytic evaluation of recent meta-analyses. World psychiatry. 2022 Feb;21(1):133–45.

34. Radua J, De Prisco M, Oliva V, Fico G, Vieta E, Fusar-Poli P. Impact of air pollution and climate change on mental health outcomes: an umbrella review of global evidence. World Psychiatry. 2024 Jun;23(2):244–56.

35. de Pablo GS, Rodriguez V, Besana F, Civardi SC, Arienti V, Garceo LM, et al. Umbrella review: atlas of the meta-analytical evidence of early-onset psychosis. Journal of the American Academy of Child & Adolescent Psychiatry. 2024 Jul 1;63(7):684–97.

36. Al Maqbali M, Alsayed A, Hughes C, Hacker E, Dickens GL. Stress, anxiety, depression and sleep disturbance among healthcare professional during the COVID-19 pandemic: An umbrella review of 72 meta-analyses. PLoS One. 2024 May 9;19(5):e0302597.