JB&JS OPEN ACCESS

A Locally Executed Agentic Artificial Intelligence Framework for Deduplication, Screening, and Structured Data Extraction in Spine Surgery Systematic Reviews

A PRISMA-Transparent Reporting of Artificial Intelligence in Comprehensive Evidence Synthesis–Compliant Evaluation

Sathish Muthu, MS, DNB, MNAMS, PhD, Dhibin Vikash Kolarpatti Ponnusamy, BS, MBBS, Vibhu Krishnan Viswanathan, MS, Sathish Kumar Rajappan Chandra, MD, and Khan Sharun, PhD

Investigation performed at Department of Spine Surgery, Orthopaedic Research Group, Coimbatore, Tamil Nadu, India

Abstract

Systematic reviews and meta-analyses remain time-consuming and labor-intensive. We developed and validated a locally executed agentic artificial intelligence (AI) framework for deduplication, screening, and structured data extraction in systematic reviews, reported following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)-Transparent Reporting of AI in Comprehensive Evidence Synthesis (trAIce) guidelines. A multiagent pipeline of specialized agents for deduplication, title/abstract screening, structured data extraction, and verification was executed entirely locally to ensure data governance and reproducibility. Human-in-the-loop validation compared outputs against dual independent reviewers across 3 spine surgery systematic reviews, assessing accuracy, inter-rater agreement (Cohen's κ), time savings, and clinically critical error rates. Across 6,214 records, deduplication achieved near-perfect agreement with human reviewers (κ = 0.98), title and abstract screening yielded higher concordance than human screening (κ = 0.91) while reducing full-text review volume by 83%, and structured data extraction reached substantial agreement (κ = 0.87). The framework reduced reviewer time by 91.1% (90.8%-91.4%), a mean saving of 19.5 hours per review ($p < 0.001$). Clinically critical discrepancies were rare (<1%) and traceable, with no fabricated or hallucinated data introduced. A locally executed agentic AI framework has the potential to deliver accurate, efficient, and secure automation of systematic review tasks with human oversight, offering a reproducible pathway for trustworthy evidence synthesis under PRISMA-trAIce standards.

Introduction

Systematic reviews and meta-analyses occupy a central role in evidence-based medicine[1]. Unlike single studies, which may be limited by sample size, methodology, or context, systematic reviews integrate diverse data sources to deliver comprehensive and balanced conclusions[2,3]. Meta-analyses further strengthen this process by applying statistical techniques to pool results, quantify effect sizes, and assess heterogeneity[4].

Despite their importance, systematic reviews and meta-analyses are notoriously time-consuming and labor-intensive, resulting in reduced methodological rigor[5,6]. Collectively, these processes can take months or even years to complete, delaying the translation of evidence into practice[7].

The exponential growth of scientific literature has intensified these challenges[8]. The tension between the need for speed and the requirement for rigor has created a pressing methodological bottleneck[9,10].

No funding was utilized in the conduction of the study.

The authors declare no conflict of interest.

Disclosure: The **Disclosure of Potential Conflicts of Interest** forms are provided with the online version of the article (https://links.lww.com/JBJSOA/B309).

Copyright © 2026 The Authors. Published by The Journal of Bone and Joint Surgery, Incorporated. All rights reserved. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

JBJS Open Access • 2026:e26.00102. openaccess.jbjs.org

The era of artificial intelligence (AI) is reshaping this landscape[11]. Advances in machine learning, natural language processing, and large language models have demonstrated the potential to automate discrete tasks within evidence synthesis[12]. AI systems can rapidly identify duplicates, classify abstracts, and extract structured data, offering efficiency gains that were previously unattainable[13]. They are often single-task tools, without integration across the multistage workflow of systematic reviews. Moreover, many operate as opaque cloud-based services, raising concerns about transparency, reproducibility, and data governance[14].

Agentic AI represents a new paradigm[15]. Agentic AI systems are composed of multiple specialized agents, each designed to perform a specific task while coordinating with others through autonomous reasoning, memory retention, and iterative task decomposition[15,16]. This modular architecture mirrors the workflow of systematic reviews, making agentic AI particularly well suited to evidence synthesis. By embedding human-in-the-loop validation, agentic AI frameworks can combine efficiency with oversight.

The objective of this study was to develop and validate a locally executed agentic AI framework capable of supporting systematic reviews and meta-analyses by automating key tasks—deduplication, title and abstract screening, and structured data extraction—while retaining human oversight for full-text screening in a limited area of spine surgery.

Methods

Study Design

This study was conducted as a methods development and validation project to evaluate a locally executed agentic AI framework for systematic reviews and meta-analyses. The study was reported in adherence to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)-Transparent Reporting of AI in Comprehensive Evidence Synthesis guidelines[17].

Development Phase

The agentic AI pipeline was initially developed and refined using 26 previously published systematic reviews authored by the investigators[6,18-42]. Importantly, this development phase was distinct from the validation phase, ensuring that subsequent performance testing was conducted independently.

Validation Phase

Validation was performed using 3 de novo systematic reviews in spine surgery: (1) glucagon-like peptide-1 analogs in spine surgery, (2) giant cell tumors in the cervical spine, and (3) return to work following spinal cord injury. These 3 spine surgery reviews were selected because they were ongoing de novo projects within our group at the time of validation and together spanned 3 distinct review questions—a pharmacological intervention, an oncological condition, and a functional outcome—providing topical heterogeneity within a single clinical domain so that the framework could be tested against varied review logic while keeping the discipline constant. For these reviews, independent human reviewers with no involvement in developing the system established the "*ground truth*." Each record was screened and extracted by 2 blinded reviewers, with discrepancies resolved by consensus.

System Architecture

The framework was implemented as a multiagent pipeline, with each agent responsible for a distinct stage of the review process: a deduplication agent (DOI matching, title similarity, author/year concordance), a screening agent applying protocol-defined criteria to titles and abstracts, an extraction agent capturing structured fields using predefined schemas, and a verification agent that flagged discrepancies for human adjudication (Fig. 1). Agents communicated through a locally executed coordination layer, keeping all processes under researcher control and, unlike cloud-based services, safeguarding data governance and reproducibility.

Data Sources and Preparation

Bibliographic records and full-text PDFs for the 3 validation reviews were retrieved from PubMed, Embase, Web of Science, and the Cochrane Library. Inclusion and exclusion criteria were established a priori, and all records were imported into the pipeline.

Deduplication

The deduplication agent compared records across multiple fields (DOI, title, author, year). In cases of conflict, where duplication could not be confidently determined, both documents were retained to avoid inadvertently removing valid entries. The deduplication agent execution code is given in the Deduplication_agent_link.

Title and Abstract Screening

The screening agent applied inclusion and exclusion criteria to titles and abstracts. A *dual-pass strategy* was used to ensure consistency in the AI decisions for screening, where each screening task was executed twice independently. Only when both outputs matched was the decision finalized and exported into Google Sheets (Google LLC, Workspace v2025.2). We also used a *cross-validation* approach. Specifically, the screening agent's outputs were randomly partitioned into 2 subsets. Separate human reviewers independently adjudicated each subset, and discrepancies were reconciled. The screening agent execution code is given in the Screening_agent_link.

Full-Text Screening

Full-text screening was conducted exclusively by human reviewers to adjudicate nuanced eligibility criteria with clinical judgment. AI agents were not applied at this stage, but the verification agent ensured adjudication logs were complete and traceable.

Structured Data Extraction

The extraction agent applied a predefined schema. Deterministic rules converted reported percentages into counts and

JBJS Open Access • 2026:e26.00102.

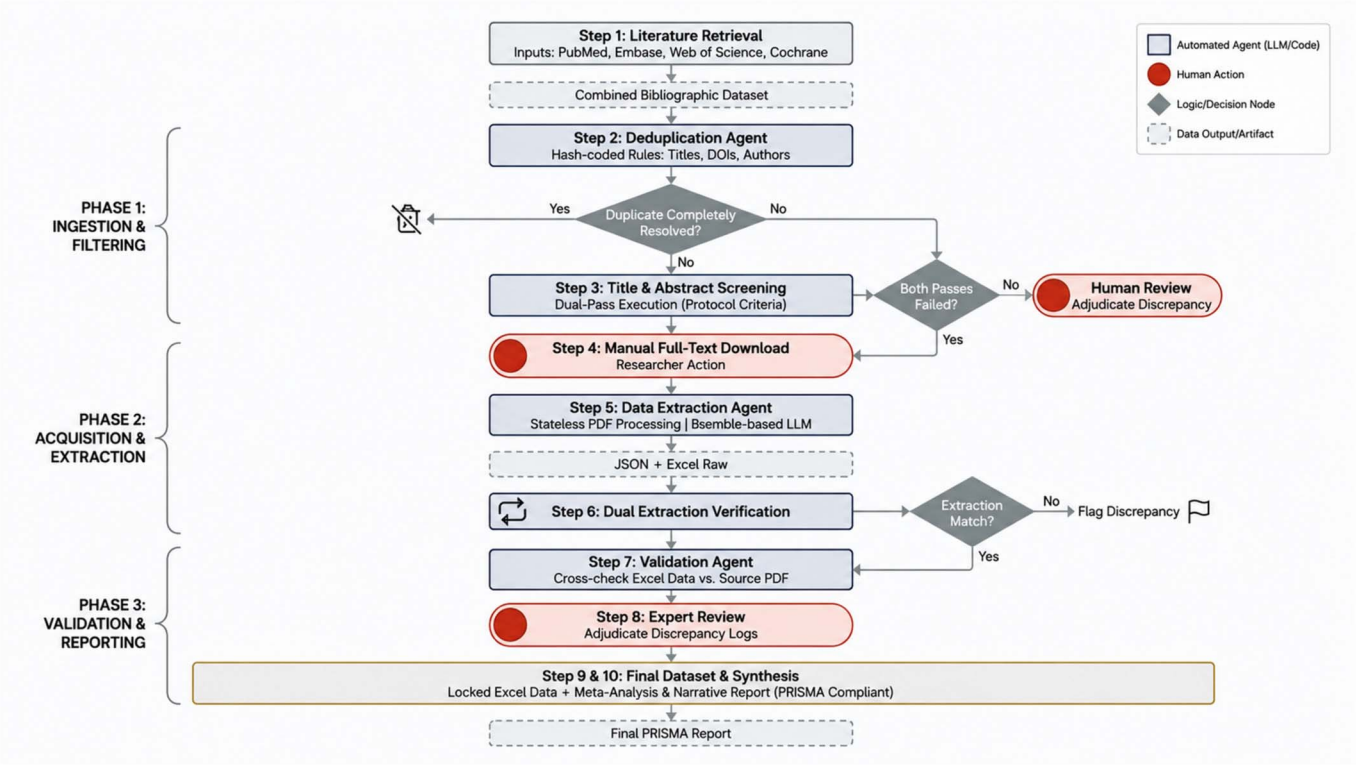


Fig. 1

Automated workflow for systematic review and meta-analysis. Blue boxes indicate LLM-based agents, red circles mark human oversight, gray diamonds represent decision nodes, and dashed boxes show data outputs such as JavaScript Object Notation, Excel, and Preferred Reporting Items for Systematic Reviews and Meta-Analyses reports.

standardized continuous outcomes. Extraction tasks were subjected to cross-agent redundancy: 2 independent extraction agents processed the same data set. Outputs were compared, and discrepancies were flagged for human adjudication. For entries with missing data, justification logs were generated. The data extraction agent execution code is given in the Extraction_agent_link.

Verification and Human Oversight

The verification agent continuously monitored logs, identified discrepancies, and directed them to human reviewers. Dual independent reviewers validated AI outputs for deduplication, screening, and extraction. Audit trails documented the origin of each data element, the agent responsible, and any human intervention. The verification agent execution code is given in the Verification_agent_link.

Output and Validation

All outputs were exported into Google Sheets, the central platform for verification.

Performance Measures

Performance was evaluated across 4 dimensions: accuracy (concordance between AI and human outputs), inter-rater agreement (Cohen's κ), time savings vs. manual workflows, and error classification distinguishing minor discrepancies from clinically critical errors with the potential to alter conclusions.

Statistical Analysis

Analyses were performed in Stata (StataCorp LLC, v17.0). Accuracy was the proportion of concordant decisions between AI agents and human reviewers, and inter-rater agreement was quantified using Cohen's κ, interpreted by Landis and Koch benchmarks (>0.80 = almost perfect). Time efficiency compared AI-assisted with human-only workflows on the same tasks and records. Human reviewer time was captured prospectively through a structured per-session log completed by each reviewer, comprising start and end timestamps, task identifier (deduplication, screening, or extraction), reviewer initials, and brief activity notes; only active minutes were counted, with idle periods, meetings, and breaks excluded. AI time was derived from pipeline execution logs that recorded the wall-clock duration of each agent's run. Time saved per review was computed as the difference between the matched human-only and AI-assisted workflows on the same records and tasks and expressed in hours per review. Paired comparisons used a paired t-test, and discrepancies were classified as minor (non-impactful) or clinically critical. All tests were two-sided, with significance at $p < 0.05$.

Hardware and Software

The pipeline was executed locally using a Gemini agent (Google DeepMind, release 2025.1)[43] with temperature set to 0.2 and max tokens capped at 2,048. Antigravity software (Google,

JBJS Open Access • 2026:e26.00102. openaccess.jbjs.org

v3.4) coordinated agent communication. The configuration included 16 GB of RAM, an integrated graphics processor with 8 GB capacity, and a solid-state drive for storage. The system operated on a stable 5G network connection.

Results

Data Set and Workflow

Across the 3 validation reviews, 6,214 bibliographic records were retrieved from PubMed, Embase, Web of Science, and the Cochrane Library and processed sequentially through deduplication, title and abstract screening, full-text screening, and structured data extraction.

Deduplication Performance

The deduplication agent identified 1,025 duplicate records using DOI matching, title similarity, and author/year concordance. Against human reviewers, accuracy was 98.7% with Cohen's $\kappa = 0.98$, indicating almost perfect agreement comparable with established reference management tools. In cases of uncertainty, duplicates were retained to avoid inadvertent exclusion (Table I).

Title and Abstract Screening

Of 5,189 records subjected to title and abstract screening, reviewers excluded 4,861 records based on eligibility criteria, of which the AI agent correctly excluded 98.5%. The dual-pass strategy yielded $\kappa = 0.91$ (0.89-0.93) vs. human reviewers, reflecting strong concordance, and reduced records requiring full-text review by 92% (Table I). Robustness was enhanced through cross-validation and stratified sampling, with outputs partitioned by publication year and independently adjudicated against human ground truth.

Full-Text Screening

Full-text screening was conducted exclusively by human reviewers. The screening agent forwarded 399 records vs. the 329 reviewers would have forwarded—a deliberate overinclusion of 70 records (21.3%) reflecting a sensitivity-weighted threshold, as false negatives at title/abstract screening are unrecoverable while false positives are inexpensive to remove at full-text adjudication. All 70 additional records were excluded on full-text review, confirming the overinclusion did not compromise the evidence base. Of the 329 records reviewers progressed, 72 met the inclusion criteria.

Structured Data Extraction

The extraction agent processed the 72 included studies, capturing 1,550 structured data fields (sample sizes, binary outcomes, and continuous measures with SDs). Dual-pass extraction yielded $\kappa = 0.87$ (0.86-0.88) with 93% concordance vs. human reviewers, representing substantial agreement. Justification logs were generated for 6.7% of entries with missing or ambiguous data—most often absent SDs, merged table cells, or incomplete subgroup reporting—all directed to human adjudication.

Verification and Human Oversight

The verification agent flagged 104 discrepancies across deduplication, screening, and extraction, all reviewed by dual independent adjudicators with consensus in every case. Audit trails confirmed no fabricated data were introduced, and every extracted element was traceable to its source. Of the 104, 15 were critical (0.9% of total data fields extracted)—for example, swapped cohorts (intervention vs. control mislabeling) or misinterpreted composite outcomes—and were corrected by human reviewers.

Time Efficiency

The framework reduced reviewer time by 91.1% (90.8%-91.4%), a mean saving of 19.9 hours per review ($p < 0.001$), most pronounced in the data extraction phase (Table I).

Error Classification

Minor discrepancies accounted for 86% of flagged outputs (formatting inconsistencies, ambiguous abbreviations, or incomplete metadata), while clinically critical errors (14%; 0.9% of total outputs) typically involved mislabeled cohorts, incorrect denominators, or omitted adverse events—for example, misclassifying "revision surgery" as a primary rather than secondary end point. These were consistently identified by the verification agent and resolved through human adjudication, underscoring the necessity of the human-in-the-loop design.

Discussion

This study demonstrates the feasibility of a locally executed agentic AI framework for systematic reviews and meta-analyses. By integrating specialized agents for deduplication, screening, extraction, and verification into a coordinated pipeline, it achieved high accuracy, strong inter-rater agreement, substantial time savings, and minimal clinically critical errors, while operating entirely on local devices to safeguard data and ensure reproducibility. Validation against independent reviewers not involved in development avoided circularity and ensured unbiased ground truth.

A key contribution is the demonstration that multiagent coordination can replicate and enhance the traditional review workflow: Deduplication was near-perfect, screening achieved strong concordance, and extraction yielded substantial agreement, while the verification agent identified and adjudicated discrepancies at every stage. Agentic AI can thus serve as a reliable partner in evidence synthesis, reducing repetitive manual tasks while preserving clinical integrity.

When considering the broader landscape, it is essential to acknowledge that independent software solutions already exist for each of the discrete tasks performed in this pipeline. *EndNote* (Clarivate Analytics)[44] and *Zotero* (Corporation for Digital Scholarship)[45] are widely used for reference management and deduplication. Screening tasks can be performed using platforms such as *Covidence* (Veritas Health Innovation)[46] or *Rayyan* (Qatar Computing Research Institute)[47]. Structured data extraction is supported by tools such as *DistillerSR* (Evidence Partners)[48]. Verification and

JBJS Open Access • 2026:e26.00102. openaccess.jbjs.org

TABLE I Distribution of Records and Agreement Across 3 Systematic Reviews

Stage of Review	Review 1	Review 2	Review 3	Total
Records retrieved	1,145	2,980	2,089	6,214
Deduplication records removed	Agent: 248	Agent: 436	Agent: 341	Agent: 1,025
	Human: 255	Human: 440	Human: 344	Human: 1,039
	κ = 0.98	κ = 0.97	κ = 0.98	κ = 0.98
Title and abstract screening records removed	Agent: 765	Agent: 2,302	Agent: 1,723	Agent: 4,790
	Human: 780	Human: 2,356	Human: 1,725	Human: 4,861
	κ = 0.89	κ = 0.91	κ = 0.93	κ = 0.91
Full-text screening eligible records	Agent: 132	Agent: 242	Agent: 25	Agent: 399
	Human: 125	Human: 184	Human: 20	Human: 329
	κ = 0.87	κ = 0.86	κ = 0.88	κ = 0.87
Studies included for data extraction	Human: 17	Human: 38	Human: 17	Human: 72
Structured data extraction (concordant fields)	Agent: 284	Agent: 788	Agent: 374	Agent: 1,446
	Human: 306	Human: 836	Human: 408	Human: 1,550
	κ = 0.87	κ = 0.86	κ = 0.88	κ = 0.87
Verification oversight	Agent flagged: 22	Agent flagged: 48	Agent flagged: 34	104
	Human adjudicated: 22	Human adjudicated: 48	Human adjudicated: 34	
Error classification	Minor: 17	Minor: 41	Minor: 31	Minor: 89
	Critical: 5	Critical: 7	Critical: 3	Critical: 15
Time consumed	Agent: 1.3 hours	Agent: 2.8 hours	Agent: 1.7 hours	Agent: 5.8 hours
	Human: 14.2 hours	Human: 32.6 hours	Human: 18.7 hours	Human: 65.5 hours
	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$

adjudication can be managed through collaborative platforms such as *EPPI-Reviewer* (UCL Institute of Education)[49]. While these tools are robust and widely adopted, they are typically subscription-based, requiring institutional or personal licenses.

The reliance on paid software introduces barriers to accessibility, particularly for researchers in resource-limited settings[50]. Subscription costs can be prohibitive, and cloud-based platforms often raise concerns about data privacy and governance[51]. By contrast, the current framework offers a free, locally executable solution that does not compromise personal security. This design choice addresses a critical gap in the current ecosystem, providing an accessible and secure alternative to paid platforms.

These findings carry broader implications: Agentic AI has the potential to standardize evidence synthesis by applying criteria consistently while accelerating the production of timely evidence.

Several limitations must be acknowledged. First, the framework was validated on 3 systematic reviews, all in spine surgery and conducted by overlapping investigators from a single research group; within-group validation cannot establish transferability to other clinical domains, to reviews using different inclusion logic (e.g., diagnostic accuracy, qualitative synthesis, network meta-analysis), or to teams with different reviewer training and adjudication conventions. Independent prospective validation by unaffiliated groups in unrelated specialties—ideally with preregistered protocols and blinded comparators—is required before claims of cross-domain generalizability. Second, while the dual-pass and cross-validation strategies for screening and extraction enhance reliability, they increase computational demands that may limit scalability for extensive data sets; the framework is feasible on midrange consumer devices, though large-scale reviews may benefit from more powerful hardware to optimize processing speed. Third, full-text screening was intentionally retained as a human-only task, reflecting the complexity of adjudicating nuanced eligibility criteria, so the framework does not yet achieve complete automation. Fourth, although the system is free and locally executable, technical expertise is required to implement and maintain the pipeline, which may challenge researchers without computational backgrounds. Finally, reliance on specific versions of proprietary software (Antigravity, Gemini) may limit reproducibility if future updates alter functionality.

Conclusion

This study provides evidence that a locally executed agentic AI framework can substantially reduce the workload of systematic reviews while maintaining transparency, reproducibility, and clinical reliability, achieving high accuracy, strong

inter-rater agreement, significant time savings, and minimal clinically critical errors. Unlike subscription-based, cloud-dependent solutions, it is free to execute locally, ensuring accessibility and safeguarding personal security. Future work should expand validation across diverse domains, refine automation of complex tasks, enhance usability for nontechnical researchers, and continue to keep the human reviewer in the loop as the source of ground truth and as the arbiter of AI output accuracy. Within the boundaries of the spine-surgery validation reported here, the framework offers a working prototype for locally executed agentic evidence synthesis; broader claims about democratizing systematic reviews must await independent multidomain validation.

Data and Code Availability

Source code for the 4 agents is openly available on GitHub under the MIT License: the deduplication agent (Deduplication_agent_link), the screening agent (Screening_agent_link), the data extraction agent (Extraction_agent_link), and the verification agent (Verification_agent_link). Each repository includes orchestration scripts, prompt templates, schema definitions, and the local execution environment specification. De-identified screening and extraction logs supporting the results in Table I are included in the repositories'/data directories. No patient-identifiable information was processed.

Data Availability

Data generated in the study will be made available upon reasonable request to the authors.

Author Contributions

Dr SMuthu and Dr VKV contributed to the conceptualization and design of the research goals and aims. Dr SMuthu developed the methodology and statistical framework. Data validation and ensuring the accuracy of results were undertaken by Dr SMuthu, Dr SKR Chandra and Dr VKV. Dr SMuthu secured the necessary resources for the study, and Dr DVKP curated and organized the study data. Writing the original draft of the manuscript was managed by Dr SMuthu, Dr SK with other authors providing critical revisions and editing. Visualization and creation of figures were executed by Dr SMuthu. Supervision and coordination of the project were led by Dr SMuthu. ■

Sathish Muthu, MS, DNB, MNAMS, PhD[1,2,3]
Dhibin Vikash Kolarpatti Ponnusamy, BS, MBBS[1,3]
Vibhu Krishnan Viswanathan, MS[1,4]
Sathish Kumar Rajappan Chandra, MD[5]
Khan Sharun, PhD[6]

[1]Department of Spine Surgery, Orthopaedic Research Group, Coimbatore, Tamil Nadu, India

[2]Central Research Laboratory, Aarupadai Veedu Medical College and Hospital, Vinayaka Mission's Research Foundation (Deemed to be University), Puducherry, India

[3]Department of Orthopaedics, Government Medical College, Karur, Tamil Nadu, India

[4]Department of Orthopaedic Surgery, University of Alabama, Birmingham, Alabama

[5]Clinical Trial & Research Unit, Interdisciplinary Institute of Indian System of Medicine (IIISM), SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India

[6]Graduate Institute of Medicine, Yuan Ze University, Taoyuan, Taiwan

E-mail address for S. Muthu: drsathishmuthu@gmail.com

References

1. Tenny S, Varacallo MA. Evidence-Based Medicine. Treasure Island, FL: StatPearls Publishing; 2026.
2. Girinivasan C, Sathish M. Analysis of reference practices among practicing orthopaedicians in India. Indian J Orthop. 2021;55(4):869-78.
3. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Based Healthc. 2015;13(3):132-40.
4. Haidich AB. Meta-analysis in medical research. Hippokratia. 2010;14(Suppl 1):29-37.
5. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open. 2017;7(2):e012545.
6. Sathish M, Eswar R. Systematic reviews and meta-analysis in spine surgery—How good are they in methodological quality? A systematic review. Glob Spine J. 2021;11(3):378-99.
7. Simmonds MC, Higginsa JPT, Stewartb LA, Tierneyb JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. Clin Trials. 2005;2(3):209-17.
8. González-Márquez R, Schmidt L, Schmidt BM, Berens P, Kobak D. The landscape of biomedical research. Patterns (N Y). 2024;5(6):100968.
9. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
10. Cumpston M, Li T, Page MJ, Chandler J, Welch VA, Higgins JP, Thomas J. Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions. Cochrane Database Syst Rev. 2019;10(10):ED000142.
11. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthc J. 2021;8(2):e188-e194.
12. Ruan M, Fan J, Liu M, Meng Z, Zhang X, Zhang C. Artificial intelligence for the science of evidence synthesis: how good are AI-powered tools for automatic literature screening? BMC Med Res Methodol. 2025;25(1):199.
13. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med. 2023;29(8):1930-40.
14. Muthu S. The efficiency of machine learning-assisted platform for article screening in systematic reviews in orthopaedics. Int Orthop. 2023;47(2):551-6.
15. Zou J, Topol EJ. The rise of agentic AI teammates in medicine. Lancet. 2025;405(10477):457.
16. Karunanayake N. Next-generation agentic AI for transforming healthcare. Inform Health. 2025;2(2):73-83.
17. Holst D, Moenck K, Koch J, Schmedemann O, Schüppstuhl T. Transparent reporting of AI in systematic literature reviews: development of the PRISMA-trAIce checklist. JMIR AI. 2025;4:e80247.
18. Muthu S, Viswanathan VK, Munisamy S, Kolarpatti Ponnusamy DV, Theiss S, Rajappan Chandra SK, Sharun K, Rajaram Manoharan SR. Role of prophylactic negative pressure wound therapy in reducing surgical-site infections in spine surgery - a systematic review and meta-analysis. Glob Spine J. 2026:21925682261426264.

19. Muthu S, Mavrovounis G, Corluka S, Buser Z, Brodano GB, Wu Y, Meisel HJ, Wang J, Yoon ST, Demetriades AK; AO Spine Knowledge Forum Degenerative. Does the choice of chemoprophylaxis affect the prevention of deep vein thrombosis in lumbar fusion surgery? A systematic review of the literature. Brain and Spine. 2023; 3:102711.

20. Cabrera JP, Camino-Willhuber G, Muthu S, Guiroy A, Valacco M, Pola E. Percutaneous versus open pedicle screw fixation for pyogenic spondylodiscitis of the thoracic and lumbar spine: systematic review and meta-analysis. Clin Spine Surg. 2023;36(1):24-33.

21. Muthu S, Ramakrishnan E, Chellamuthu G. Is endoscopic discectomy the next gold standard in the management of lumbar disc disease? Systematic review and superiority analysis. Glob Spine J. 2021;11(7):1104-20.

22. Muthu S, Ramakrishnan E. Fragility analysis of statistically significant outcomes of randomized control trials in spine surgery: a systematic review. Spine (Phila Pa 1976). 2021;46(3):198-208.

23. Viswanathan VK, Muthu S. Is anterior cervical plating necessary for cage constructs in anterior cervical discectomy and fusion surgery for cervical degenerative disorders? Evidence-based on the systematic overview of meta-analyses. World Neurosurg. 2023;18:100185.

24. Muthu S, Chellamuthu G. How safe is unilateral pedicle screw fixation in lumbar fusion surgery for management of 2-level lumbar degenerative disorders compared with bilateral pedicle screw fixation? Meta-analysis of randomized controlled trials. World Neurosurg. 2020;140:357-68.

25. Rodrigues-Pinto R, Muthu S, Diniz SE, Cabrera JP, Martin CT, Agarwal N, Meisel HJ, Wang JC, Buser Z; AO Spine Knowledge Forum Degenerative. Complications of the use allograft in 1- or 2-level anterior cervical discectomy and fusion: a systematic review. Glob Spine J. 2024;14(2_suppl):70S-77S.

26. Wang D, Wang W, Han D, Muthu S, Cabrera JP, Hamouda W, Ambrosio L, Cheung JPY, Le HV, Vadalà G, Buser Z, Wang JC, Cho S, Yoon ST, Lu S, Chen X, Diwan AD; AO Spine Knowledge Forum Degenerative. Clinical effectiveness of reduction and fusion versus in situ fusion in the management of degenerative lumbar spondylolisthesis: a systematic review and meta-analysis. Eur Spine J. 2024;33(5): 1748-61.

27. Sathish M, Girinivasan C. Is use of topical vancomycin in pediatric spine surgeries a safe option in the prevention of surgical site infections? A meta-analysis and systematic review of the literature. Glob Spine J. 2021;11(5):774-81.

28. Muthu S, Ramakrishnan E, Natarajan KK, Chellamuthu G. Risk-benefit analysis of wound drain usage in spine surgery: a systematic review and meta-analysis with evidence summary. Eur Spine J. 2020;29(9):2111-28.

29. Muthu S, Viswanathan VK, Gangadaran P. Is platelet-rich plasma better than steroids as epidural drug of choice in lumbar disc disease with radiculopathy? Meta-analysis of randomized controlled trials. Exp Biol Med (Maywood). 2025;250: 10390.

30. Muthu S, Jeyaraman M, Gulati A, Arora A. Current evidence on mesenchymal stem cell therapy for traumatic spinal cord injury: systematic review and meta-analysis. Cytotherapy. 2021;23(3):186-97.

31. Muthu S, Viswanathan VK, Annamalai S, Thabrez M. Bilateral erector spinae plane block for postoperative pain relief in lumbar spine surgery: a PRISMA-compliant updated systematic review & meta-analysis. World Neurosurg. 2024;23: 100360.

32. Muthu S, Ćorluka S, Buser Z, Malcolm JG, Luo Z, Gollahalli Shivashankar P, Ambrosio L, Griffoni C, Demetriades AK, Ivandić S, Wu Y, Wang J, Meisel HJ, Yoon TS. Rate of reoperation following decompression-only procedure for lumbar degenerative spondylolisthesis: a systematic review of literature. JB JS Open Access. 2024;9(3):e23.00163.

33. Ambrosio L, Vadalà G, de Rinaldis E, Muthu S, Ćorluka S, Buser Z, Meisel HJ, Yoon ST, Denaro V; AO Spine Knowledge Forum Degenerative. Discectomy versus sequestrectomy in the treatment of lumbar disc herniation: a systematic review and meta-analysis. Spine J. 2025;25(2):211-26.

34. Viswanathan VK, Kalanchiam GP, Chinnappan A, Muthu S. Recovery patterns from C5 palsy after anterior cervical decompression and fusion, posterior cervical decompression and fusion, and laminoplasty for degenerative cervical myelopathy: systematic review and meta-analysis of 748 C5 palsy cases. Asian Spine J. 2025; 19(6):1059-71.

35. Sotiris V, Muthu S, Gary M, Cho S, Kato S, Lewis SJ, Kim HJ, Wang J, Jain A, Yoon ST; AO Spine Knowledge Forum Degenerative. Practice preference and evidence analysis on topical use of tobramycin powder in lumbar spine surgery: a Multi-National AO spine survey with systematic review of the literature. World Neurosurg. 2024;24:100397.

36. Viswanathan VK, Muthu S. Sexual dysfunction in cervical spondylomyelopathy and its prognosis following surgical decompression: a systematic review. Glob Spine J. 2025;15(5):2779-88.

37. Corluka S, Muthu S, Yoon T, Cunha C, Gary M, Vadala G, Barbanto Brodano G, Monetta A, K Demetriades A, Ivandić S, Wu Y, Wang J, Meisel HJ, Buser Z. Decompression-only for lumbar degenerative spondylolisthesis - what are the risk for failure? - a systematic review. Glob Spine J. 2025;15(8):3919-32.

38. Ambrosio L, Schol J, Tamagawa S, Muthu S, Sakai D, Papalia R, Vadalà G, Denaro V. Efficacy and safety of osteobiologics for lumbar spinal fusion: a systematic review and network meta-analysis. J Bone Joint Surg. 2025;107(18):2110-21.

39. Muthu S, Kalanchiam GP, Munisamy S, Viswanathan VK. Incidence of C5 palsy in anterior cervical decompression & fusion, posterior cervical decompression & fusion and laminoplasty for degenerative cervical myelopathy: systematic review and meta-analysis of 21,231 cases. Asian Spine J. 2026;20(1):158-75.

40. Yu A, Tiao J, Cai CW, Huang JJ, Mohamed K, Hoang R, Hong J, Berman D, Lee J, Ambrosio L, Buser Z, Cabrera JP, Chen X, Cini C, Ćorluka S, Demetriades AK, Diwan A, Jain A, Kim JS, Li X, Muthu S, Tavakoli J, Vadalà G, Hsieh PC, Cho SK, Knowledge Forum Degenerative AS. Radiographic assessment of successful lumbar spinal fusion: a systematic review of fusion criteria in randomized trials. Glob Spine J. 2026;16(3):1608-18.

41. Muthu S, Viswanathan VK, Palani B, Theiss S, Rajappan Chandra SK, Sharun K, Rajaram Manoharan SR. Mental health problems in traumatic spinal cord injury patients - an umbrella systematic review. Glob Spine J. 2026;16(4):2028-39.

42. Viswanathan VK, Tanbour H, Muthu S, Sheppard E, Hiatt L, Theiss S, Rajaram Manoharan SR. Intact posterior tension band in spinal fractures in patients with spinal ankylosing disorders - is it a myth or a reality? A systematic review. Glob Spine J. 2026:21925682261416430.

43. Google Gemini. Gemini. Available at: https://gemini.google.com. Accessed January 13, 2026.

44. Hupe M. EndNote X9. J Electron Resour Med Libraries. 2019;16(3–4):117-9.

45. Ahmed KKM, Al Dhubaib BE. Zotero: a bibliographic assistant to researcher. J Pharmacol Pharmacother. 2011;2(4):303-5.

46. Covidence - better systematic review management. Covidence. Available at: https://www.covidence.org/. Accessed January 13, 2026.

47. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—A web and mobile app for systematic reviews. Syst Rev. 2016;5(1):210.

48. DistillerSR. Systematic Review Software | Literature Review Software. DistillerSR. Available at: https://www.distillersr.com/products/distillersr-systematic-review-software. Accessed January 13, 2026.

49. EPPI-Reviewer: systematic review software. Available at: https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2914. Accessed January 13, 2026.

50. Nussbaumer-Streit B, Ellen M, Klerings I, Sfetcu R, Riva N, Mahmić-Kaknjo M, Poulentzas G, Martinez P, Baladia E, Ziganshina LE, Marqués ME, Aguilar L, Kassianos AP, Frampton G, Silva AG, Affengruber L, Spjker R, Thomas J, Berg RC, Kontogiani M, Sousa M, Kontogiorgis C, Gartlehner G; working group 3 in the EVBRES COST Action. Resource use during systematic review production varies widely: a scoping review. J Clin Epidemiol. 2021;139:287-96.

51. Haliassos A, Kasvis D, Karathanos S. The challenges of data privacy and cybersecurity in cloud computing and artificial intelligence (AI) applications for EQA organizations. EJIFCC. 2025;36(4):599-604.