Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Diagnostic Accuracy of Intraoperative Neurophysiological Monitoring in Spine Surgery: An Umbrella Review and Study-Level Reanalysis

Sathish Muthu

drsathishmuthu@gmail.com

Orthopaedic Research Group, Coimbatore, India https://orcid.org/0000-0002-7143-4354

Dhibin Vikash Kolarpatti Ponnusamy

Karur Government Medical College https://orcid.org/0000-0003-1071-0084

H. Francis Farhadi

University of Kentucky

Mark Klion

SBH Health System (St Barnabas Hospital), Bronx, NY

Vibhu Krishnan Viswanathan

SBH Health System (St Barnabas Hospital), Bronx, NY https://orcid.org/0000-0002-3804-1698

Systematic Review

Keywords: Intraoperative neurophysiological monitoring, Motor evoked potentials, Somatosensory evoked potentials, Diagnostic accuracy, Spine surgery, Umbrella review

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11668775/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Preprint. This manuscript has not been peer reviewed. It is under consideration at a journal.

Title Page

Diagnostic Accuracy of Intraoperative Neurophysiological Monitoring in Spine Surgery: An Umbrella Review and Study-Level Reanalysis

Running title: IONM accuracy in spine surgery

Article type: Review Article

Authors

Sathish Muthu[1,2*]
Dhibin Vikash Kolarpatti Ponnusamy[3]
H. Francis Farhadi[4]
Mark Klion[5]
Vibhu Krishnan Viswanathan[5]

Affiliations
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, Karur Government Medical College, Karur, Tamil Nadu, India.
4. Department of Neurosurgery, University of Kentucky, Lexington, Kentucky, USA.
5. SBH Health System (St Barnabas Hospital), Bronx, New York, USA.

Corresponding author
Dr Sathish Muthu, MS, PhD

Department of Spine Surgery, Orthopaedic Research Group, Coimbatore, Tamil Nadu, India.
E-mail: drsathishmuthu@gmail.com | Telephone: +91 9600856806
ORCID: 0000-0002-7143-4354

ORCID identifiers
Sathish Muthu — 0000-0002-7143-4354
Dhibin Vikash Kolarpatti Ponnusamy — 0000-0003-1071-0084
H. Francis Farhadi — 0000-0002-4668-8512
Mark Klion — NA
Vibhu Krishnan Viswanathan — 0000-0002-3804-1698

Author e-mail addresses
Sathish Muthu — drsathishmuthu@gmail.com
Dhibin Vikash Kolarpatti Ponnusamy — dhibinvikash@outlook.com
H. Francis Farhadi — Francis.Farhadi@uky.edu
Mark Klion — mklion@sbhny.org
Vibhu Krishnan Viswanathan — drvibu007@gmail.com

Counts

Abstract: 242 words. Main text (Introduction to Conclusion): 3924 words. Tables: 3.

Figures: 4. References: 56. Electronic supplementary material: 1

Acknowledgments

None

Author Contributions

Sathish Muthu: conceptualization, methodology, software, formal analysis, investigation, data curation, visualization, project administration, writing – original draft, writing – review and editing. Dhibin Vikash Kolarpatti Ponnusamy: data curation (data extraction), validation, writing – review and editing. H. Francis Farhadi: data curation (data extraction), validation, writing – review and editing. Mark Klion: data curation (data extraction), validation, writing – review and editing. Vibhu Krishnan Viswanathan: data curation (data extraction), validation, writing – review and editing. All authors read and approved the final version of the manuscript.

Statements and Declarations

Ethical considerations. Not applicable. This study analysed data from published systematic reviews and their supplementary material; no human participants were involved and ethics approval was not required.

Consent to participate. Not applicable.

Consent for publication. Not applicable.

Declaration of conflicting interest. The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding statement. The author(s) received no financial support for the research, authorship, and/or publication of this article.

Data availability. All extracted data, the study-level 2×2 tables, the reconstruction checks and the analysis outputs are provided in the Electronic Supplementary Material.

Use of artificial intelligence. A large language model was used to validate the human reviewer decisions and support record deduplication, screening, eligibility assessment and data extraction within a published framework for auditable AI-assisted evidence synthesis. All automated determinations were logged and independently adjudicated by the authors, who accept full responsibility for the accuracy and integrity of the work.

Prior presentation / preprint. None

Diagnostic Accuracy of Intraoperative Neurophysiological Monitoring in Spine Surgery: An Umbrella Review and Study-Level Reanalysis

Running title: *IONM Accuracy in Spine Surgery*

Abstract

Purpose: Overlapping systematic reviews report conflicting accuracy estimates for intraoperative neurophysiological monitoring (IONM). In this umbrella review with study-level re-analysis, we summarized modality accuracy for new postoperative deficit by procedure and tested whether published estimates are reproducible.

Methods: PubMed, Scopus and Web of Science were searched to 12 October 2026 for systematic reviews reporting the accuracy of somatosensory evoked potentials (SSEP), motor evoked potentials (MEP), D-wave, free-run or triggered electromyography (EMG), or multimodal IONM in spine surgery. Quality was rated with AMSTAR 2 and overlap with the corrected covered area. One index meta-analysis was selected per modality-procedure cell; where study-level data were available, primary studies were deduplicated and re-pooled with a bivariate model. GRADE was applied.

Results: Thirty-seven reviews (31 critically low, 6 low confidence) informed 30 cells; 26 were re-pooled. Specificity was high (88.0%–98.8%) except in intramedullary tumour surgery (80.2%–83.7%). Sensitivity varied: MEP 57.0%–84.0% (mixed spine: 84.0% [80.2–87.1]), multimodal 63.1%–89.1%, SSEP 31.9%–75.6%, free-run EMG 46.8%, intramedullary D-wave 26.6%, and triggered EMG for screw malposition 31.3%–54.9%. The largest review pooled non-independent entries; counting one entry per cohort lowered SSEP sensitivity from 71.4% to 60.7% and MEP from 90.2% to 84.0%. Certainty was moderate in 3 cells, low in 12 and very low in 15.

Conclusion: IONM alerts were specific, but a stable tracing did not reliably exclude a deficit, particularly with SSEP alone or triggered EMG. MEP-based multimodal monitoring showed the highest sensitivity. With mostly low or very low certainty, these estimates should inform, not replace, clinical judgement.

Keywords: Intraoperative neurophysiological monitoring; Motor evoked potentials; Somatosensory evoked potentials; Diagnostic accuracy; Spine surgery; Umbrella review

Introduction

Intraoperative neurophysiological monitoring (IONM) aims to identify neurological compromise, while corrective action remains possible. Somatosensory evoked potentials (SSEP), motor evoked potentials (MEP), D-wave recordings and electromyography assess different neural pathways across procedures with different mechanisms of injury [1, 2].

Existing systematic reviews provide a fragmented evidence base. Broad syntheses combine diverse spinal procedures, whereas focused reviews examine individual modalities, pathologies or alarm criteria [2, 3]. High specificity does not establish that an unchanged recording reliably excludes injury. Similarly, accuracy for imaging-confirmed screw malposition differs from prediction of postoperative neurological deficit [4, 5].

The growing number of reviews creates a problem: apparently consistent findings may come from repeated analysis of the same primary studies. Overlap can exaggerate independent confirmation, while differences in study selection, alarm thresholds, outcome timing and statistical methods can produce conflicting estimates [2, 6]. A previous umbrella review considered the broader safety and efficacy of IONM [7], but procedure-specific accuracy also requires examination of the underlying study data and of whether published estimates can be reproduced.

We therefore conducted an umbrella review of IONM diagnostic accuracy in spine surgery, organized by modality and procedure. We quantified overlap, assessed review quality and evidence certainty, and reconstructed and deduplicated study-level data for re-analysis wherever available.

Methods

Protocol, Reporting and Registration: This umbrella review forms part of a pre-planned programme examining intraoperative neurophysiological monitoring (IONM) in spine surgery through four complementary overviews: diagnostic accuracy, patient outcomes, false alerts and an evidence map. All four reviews share a common search. Reporting followed the Preferred Reporting Items for Overviews of Reviews (PRIOR) statement [8] and the relevant PRISMA diagnostic test accuracy extension (PRISMA-DTA)[9]. The review was conducted within the JBI methodological framework for umbrella reviews [10], with the protocol registered retrospectively on OSF Registries on 27 October 2026 (https://doi.org/10.17605/OSF.IO/WNZM5).

Search Strategy: PubMed, Scopus and Web of Science Core Collection were searched from inception to 12 October 2026, combining concept blocks for IONM, spine surgery, and systematic reviews or meta-analyses, without language or date restrictions (Supplementary Table S1). Reference lists of included reviews and of a previous umbrella review [7] were also examined.

Eligibility Criteria: Eligible systematic reviews searched at least one bibliographic database and reported the accuracy of an IONM modality during spine surgery in patients of any age. Meta-analysis was not required. Acceptable data included pooled or study-level sensitivity, specificity, likelihood ratios, diagnostic odds ratios, area under the receiver operating characteristic curve, or reconstructable 2×2 tables.

The principal target condition was a new postoperative neurological deficit; for triggered EMG (tEMG), imaging-confirmed pedicle screw malposition was analysed separately. We excluded narrative reviews, guidelines without a systematic search, conference abstracts without a full report, non-spinal procedures, reviews without separately extractable IONM accuracy data, and non-English full texts.

Study Selection and Data Extraction: Records were deduplicated using DOI, PubMed identifier and normalized title. Titles and abstracts were screened, followed by full-text assessment. Every full text underwent a second independent eligibility assessment, with disagreements adjudicated. Screening and data extraction were supported by a large language model operating within a previously described framework for the transparent and auditable application of artificial intelligence to evidence synthesis, in which every automated decision is logged with its supporting rationale and independently verified by human reviewers [11]. All automated determinations in the present review were adjudicated by the reviewers before being accepted. The efficiency and reliability of machine-learning-assisted screening in orthopaedic evidence synthesis has been characterized previously [12].

Extracted items included review characteristics, populations, procedures, modalities, alarm criteria, reference standards, study and participant counts, risk-of-bias and statistical methods, and every accuracy estimate, each linked to its modality, procedure, target condition, alarm criterion and outcome timing. Where values differed within a review, the best-supported estimate was retained and the decision documented (Supplementary Table S5).

Methodological Quality: Review quality was assessed with AMSTAR 2 [13], with items 2, 4, 7, 9, 11, 13 and 15 treated as critical (Supplementary Table S6). Primary-study risk-of-bias assessments, mostly QUADAS-2 [14], were taken from the source reviews rather than repeated.

Overlap of Primary Studies: Primary studies were matched across reviews by first-author surname, year and title. Overlap was quantified with the corrected covered area (CCA) across the review set and within each modality–procedure cell, and interpreted as slight (0%–5%), moderate (>5%–10%), high (>10%–15%) or very high (>15%)[6] (Supplementary Table S7).

Data Synthesis: Estimates were organized into modality–procedure cells: SSEP, MEP, D-wave, multimodal monitoring, free-run EMG and tEMG, across mixed spine, cervical, thoracic, lumbar, deformity, intramedullary and extramedullary tumour, and pedicle screw procedures. The primary comparison was any significant signal change against a new deficit on immediate postoperative examination; threshold-specific, reversibility, complete-loss and longer-term estimates were tabulated separately.

Because overlapping reviews re-use the same primary studies, their pooled estimates were not combined. An index meta-analysis was selected for each cell by a pre-specified hierarchy: most primary studies, then fewest AMSTAR 2 critical flaws, then most recent [15]. Where available, 2×2 tables were recovered from review tables, text, supplements or forest plots and combined across reviews, retaining the index review's data for duplicated studies.

Where several non-independent entries represented one cohort, one entry was retained [16, 17], prioritizing any significant signal change or a ≥50% amplitude reduction against immediate postoperative deficit; disjoint subgroups were kept and exact duplicates removed (Supplementary Table S5).

Re-pooling used the bivariate random-effects model of Reitsma et al. [18], fitted by restricted maximum likelihood on the logit scale. A continuity correction of 0.5 was applied to all four cells of a study containing a zero cell. Studies without target-condition events were excluded from joint pooling because sensitivity could not be estimated. Summary receiver operating characteristic parameters followed the hierarchical formulation of Harbord et al. [19]. Univariate I^2 values described heterogeneity in sensitivity and specificity [20]. The implementation was checked against two published bivariate analyses, reproducing their reported estimates [21, 22].

A pre-specified sensitivity analysis excluded an implausible entry in which true positives exceeded the reported number of patients with deficits. A post hoc analysis examined index-review selection by prioritizing methodological quality before study count ***(Supplementary Table S8 and Figure S8)***. This assessed the stability of published index estimates rather than constituting a complete re-extraction under alternative source priorities.

Summary sensitivity, specificity and likelihood ratios are reported with 95% confidence intervals. Re-pooled estimates were used when at least three studies with target-condition events were available; otherwise the index estimate was retained. Illustrative counts per 1,000 patients assumed a deficit prevalence of 2%, informed by large cohorts in the included reviews [21, 23]; accuracy may vary with prevalence [24]. Analyses used Python (NumPy, SciPy).

Certainty of Evidence: Certainty was rated per cell with GRADE for test accuracy [25, 26], starting at high. All cells were downgraded one level for risk of bias (predominantly retrospective evidence and unblinded postoperative assessment). Further downgrades were applied for inconsistency (sensitivity or specificity I^2 above 75%, or not reported); imprecision (one level when the sensitivity confidence interval exceeded 25 percentage points, two when it exceeded 50 points or fewer than 30 events were available); indirectness (imaging-defined malposition informing neurological outcomes); and publication bias (funnel-plot asymmetry reported by the index review) (Supplementary Table S9).

Results

Study Selection and Review Characteristics: Of 1,680 records, 1,080 remained after removing 600 duplicates; 82 reports were sought and 79 assessed in full, of which 13 were excluded (Supplementary Table S2). The shared programme included 66 reviews, of which 37 reported diagnostic accuracy and form this analysis (Figure 1).

The reviews were published between 2009 and 2026; 22 (59%) appeared from 2021 onward (Table 1). Nine addressed intradural tumours [27–35], seven cervical surgery [3, 23, 36–40], six deformity [22, 41–45], six pedicle screw placement [3, 5, 46–49], four lumbar or lateral procedures [40, 50–52], one thoracic surgery [53], and four mixed spine surgery [1, 2, 21, 54]. Each review included 5–163 primary studies. Thirteen reported prospective registration, and 26 used QUADAS-2. Nineteen supplied an index estimate for at least one modality–procedure cell, seven contributed additional study-level data, and 11 provided only ranges, narrative summaries, criterion-specific or long-term estimates.

Methodological Quality and Evidence Overlap: AMSTAR 2 confidence was critically low in 31 reviews and low in six; none achieved moderate or high confidence (Table 1; Supplementary Table S6)[2, 30, 31, 34, 36, 55]. The most frequent critical weaknesses were missing lists of excluded studies with reasons (36 reviews), absence of a registered protocol (25), inadequate consideration of risk of bias when interpreting findings (20), and insufficient investigation of publication bias (seven).

The 37 reviews contained 856 study inclusions representing 442 unique primary studies. Overall overlap was slight (CCA 2.6%), but substantially greater within individual modality–procedure cells. Among 14 cells informed by multiple reviews, median CCA was 25.9% (range 0%–66.7%), with very high overlap (>15%) in 11 ***(Figure 4; Supplementary Table S7)***. The two most frequently included primary studies each appeared in eight reviews.

Reproduction and Re-analysis of Published Estimates: Study-level data were recovered from 20 reviews, yielding 641 2×2 tables and enabling 26 of 30 cells to be re-pooled from 590 deduplicated study entries (Supplementary Table S10)[2, 3, 5, 21–23, 27, 28, 30, 35, 36, 40–42, 44–46, 50, 51, 53, 55]. Of 36 published pooled estimates examined, 15 were reproduced within 0.5 percentage points and another 10 within three points (Supplementary Table S4). Eleven differed: seven involved different models or entry sets—including Stata midas, DerSimonian–Laird, univariate analyses, crude pooled proportions or combined stimulation thresholds—and four arose in one review containing non-independent cohort entries.

That review, the largest synthesis (163 studies), contained 118 MEP entries, including up to 10 alternative alarm thresholds from one cohort, separate spinal-segment and tract rows, D-wave rows and two exact duplicates [2]. With one entry per cohort (75 MEP, 52 SSEP, 16 free-run EMG and 70 multimodal entries), sensitivity was lower than published for SSEP (60.7% versus 71.4%) and MEP (84.0% versus 90.2%), with smaller differences for multimodal monitoring (82.4% versus 83.5%) and free-run EMG (46.8% versus 48.3%); specificity differed by less than two percentage points (Supplementary Figure S7).

Sensitivity to Index-Review Selection: Prioritizing methodological quality over study count changed the index review in seven of 30 cells (Supplementary Table S8 and Figure S8); re-pooled estimates were unaffected. Published index sensitivities differed by more than 10 percentage points in four cells with few studies: intramedullary SSEP (48.4% versus 80.8%) and MEP (72.5% versus 83.8%), cervical SSEP (46.0% versus 32.9%) and extramedullary SSEP (48.8% versus 64.3%)[23, 27, 28, 30, 34–36]; lumbar SSEP differed less (44% versus 36.4%)[50, 55]. The alternative intramedullary reviews lacked study-level data, so these discrepancies could not be resolved.

Diagnostic Accuracy by Modality

Table 2 and ***Figures 2 and 3*** summarize all 30 cells. Study-level results for the 26 re-pooled cells appear in ***Supplementary Figures S1–S6***; criterion-specific, long-term and non-pooled estimates are presented in ***Supplementary Table S3***.

Somatosensory evoked potentials: SSEP showed high specificity outside intramedullary tumour surgery (93.0%–97.4%), but sensitivity varied from 31.9% to 75.6% (Supplementary Figure S1). In cervical surgery, 31 studies with 315 deficits yielded sensitivity of 42.0% (95% CI 32.3–52.3) and specificity of 97.4% (95% CI 95.7–98.4), with LR+ 16.1 and LR− 0.60 [23, 36]. Sensitivity was 31.9% (95% CI 18.7–48.8) in lumbar surgery [50, 55], approximately 48% in both intramedullary and extramedullary tumours [27, 28, 30], 65.5% in thoracic surgery [53], and 75.6% (95% CI 59.8–86.6) in adolescent and paediatric deformity [41, 42]. Two intramedullary reviews without available study-level data reported higher pooled sensitivities [32, 34]. In mixed spine surgery, re-analysis yielded sensitivity of 60.7% (95% CI 48.9–71.3) and specificity of 95.9%[2].

Motor evoked potentials: MEP sensitivity point estimates exceeded those for SSEP in every comparable procedure group, although some confidence intervals overlapped and comparisons involved different study sets (Supplementary Figure S2). In mixed spine surgery, combining studies from two reviews produced 86 studies with 1,609 deficits: sensitivity was 84.0% (95% CI 80.2–87.1), specificity 94.9% (95% CI 93.1–96.2), LR+ 16.3 and LR− 0.17 [2, 21]. Sensitivity was 79.7% in thoracic surgery [53], approximately 71% in both intradural tumour groups [27, 28, 30, 35], 65.0% in lumbar surgery [51], 62.9% in ossification of the posterior longitudinal ligament [40], and 57.0% (95% CI 46.1–67.3) in cervical surgery [3, 36]. Specificity ranged from 89.8% to 95.5% outside intramedullary tumour surgery, where it was 83.7% (95% CI 77.9–88.2)[35].

Two smaller intramedullary reviews reported higher MEP sensitivities (Supplementary Table S8)[32, 34]. The sole idiopathic scoliosis meta-analysis reported sensitivity of 91% (95% CI 34–100), but its study-level data could not be reconciled with its tables [44]. In extramedullary tumours, a ≥50% amplitude

decline yielded sensitivity of 94.1% and specificity of 98.4%[31]. Intramedullary estimates varied with outcome timing and alarm definitions, including reversibility and complete versus incomplete signal loss [29, 35].

D-wave: In intramedullary tumour surgery, seven studies yielded sensitivity of 26.6% (95% CI 15.1–42.6) and specificity of 94.0% (95% CI 86.9–97.4) for a new immediate postoperative deficit (Supplementary Figure S3). For long-term motor outcomes, irreversible D-wave loss had sensitivity of 64.2% and specificity of 95.0%[35]. Identical estimates reported by two reviews from the same group were counted once [27, 35]. Extramedullary evidence remained limited to six studies with 20 deficits [28, 30].

Multimodal monitoring: Multimodal IONM generally combined SSEP and MEP, with or without EMG (Supplementary Figure S4). Sensitivity was 89.1% (95% CI 82.6–93.4) in all-age deformity surgery (36 studies, 589 deficits; LR− 0.11)[45], 86.0% in adolescent idiopathic scoliosis [22], 82.4% (95% CI 77.3–86.5) in mixed spine surgery [2], 82.0% in thoracic surgery [53], 83.9% in intramedullary tumours [27, 34], 77.1% in cervical surgery [36], and 63.1% in extramedullary tumours [28, 30]. Specificity ranged from 92.3% to 97.0%, except in intramedullary tumours (82.9%; 95% CI 66.2–92.3).

Excluding an implausible deformity-study entry changed sensitivity and specificity by less than one percentage point, to 88.2% and 95.0%, respectively. Five lateral lumbar interbody fusion studies yielded sensitivity of 99.3%, but with a wide confidence interval (95% CI 13.7–100)[52]. In anterior cervical surgery, additional modalities increased sensitivity at the expense of specificity [37]. An extramedullary tumour review assessing deficits at six months reported sensitivity of 77.9% and specificity of 91.1%[33].

Electromyography: Across 16 cohorts, free-run EMG yielded sensitivity of 46.8% (95% CI 32.1–62.0), specificity of 91.2% (95% CI 81.5–96.0), and LR− 0.59 (Supplementary Figure S5)[2]. For imaging-defined pedicle screw malposition, re-pooled tEMG sensitivity was 54.9% (95% CI 29.2–78.2) in lumbar and 31.3% (95% CI 9.7–65.9) in thoracic instrumentation, with specificities of 96.1% and 95.8%, respectively (Supplementary Figure S6)[5, 46]. A review combining stimulation thresholds across spinal levels reported sensitivity of 78% and specificity of 94%[48]. Separately, tEMG predicted postoperative neurological deficit after lumbar fusion with sensitivity of 49% and specificity of 88%[4].

Additional Narrative and Criterion-Specific Evidence: Eleven reviews contributed only ranges, narrative findings, threshold-specific or longer-term results (Supplementary Table S3). SSEP sensitivity ranged from 22% to 100% in degenerative cervical surgery [38], and a mixed-spine review reported mean SSEP sensitivity of 67.7% and MEP sensitivities of 87%–100% across thresholds [54]. An earlier review reported multimodal sensitivity above 90% in most studies [1], whereas a cervical guideline review

judged evoked-potential deterioration sensitive but not specific [39]. tEMG sensitivity ranged from 42% to 100% in thoracic deformity [47], thresholds of at least 5 mA were proposed for minimally invasive lumbar screws [49], and three reviews examined alarm thresholds [29, 31, 43].

Clinical Translation and Certainty of Evidence: At an assumed neurological-deficit prevalence of 2%, mixed-spine MEP monitoring would identify approximately 17 of 20 deficits per 1,000 patients, with 50 alerts unaccompanied by postoperative deficit. Cervical SSEP would identify eight of 20 deficits, with 26 such alerts; multimodal monitoring in deformity surgery would identify 18 of 20, with 47 such alerts ***(Table 2)***.

GRADE certainty was moderate in three cells (SSEP in intramedullary and extramedullary tumours, MEP in intramedullary tumours), low in 12 and very low in 15. All cells were downgraded for risk of bias; further downgrades mainly reflected inconsistency in specificity (I^2 79%–98% in most re-pooled cells), imprecision where deficits were few, and indirectness for imaging-defined malposition.

Discussion

The central message: match monitoring to the neurological risk:IONM should be selected according to the structures at risk, the procedure and the outcome being predicted. Across 37 reviews, specificity was generally high but sensitivity varied considerably. MEP-containing strategies detected more deficits than SSEP alone, but neither an unchanged recording nor additional modalities guaranteed neurological preservation.

Deformity and cord-risk surgery: include motor pathway monitoring. The clearest application is surgery in which spinal cord injury is a major concern. Multimodal monitoring detected approximately nine in ten deficits in all-age deformity surgery, supporting an MEP-containing strategy when correction places motor pathways at risk. This interpretation is consistent with AO Spine/Praxis guidance recommending IONM for high-risk procedures, including complex deformity, severe cord compression with myelopathy and intramedullary tumor surgery [45, 56].

For cervical and thoracic procedures, the findings similarly favor including MEP when motor injury is the principal concern. Cervical SSEP sensitivity was only 42%, indicating that preserved sensory responses provide incomplete reassurance about postoperative neurological function. However, cervical MEP sensitivity was also imperfect. Monitoring should therefore complement anatomical assessment and surgical judgement, rather than determine whether a maneuver is safe [3, 23]. Multimodal monitoring offers complementary information, but these indirect comparisons do not establish its superiority over MEP alone in every procedure.

Intramedullary tumours: distinguish immediate warning from long-term prognosis. Intramedullary surgery requires a different interpretation because immediate postoperative deterioration and persistent motor impairment are not interchangeable outcomes. The low sensitivity of D-wave monitoring for immediate deficit should not be interpreted as evidence that it lacks clinical value. The included evidence indicates that myogenic MEP and D-wave changes provide different information, with D-wave findings particularly relevant to longer-term motor outcomes [35].

Where technically feasible, this supports using D-wave alongside myogenic MEP within a multimodal strategy, rather than substituting one for the other. Alarm severity, reversibility and outcome timing should guide interpretation [29, 35]. The relatively lower specificity in intramedullary surgery also warrants caution: an alert without subsequent deficit may reflect transient dysfunction or successful intervention, rather than an unnecessary warning. Conversely, preserved D-wave responses do not exclude immediate postoperative weakness.

Lumbar surgery and instrumentation: recognize the limits of reassurance. For lumbar procedures, the results do not establish a universal indication for monitoring. SSEP sensitivity was low, and MEP estimates remained uncertain; the apparently excellent sensitivity reported for lateral lumbar interbody fusion had a confidence interval too wide to support dependable reassurance [50–52]. Monitoring decisions should therefore consider the neurological structures exposed, procedural complexity and whether an alert can prompt a meaningful response.

Triggered EMG serves a more specific purpose during instrumentation. Its relatively high specificity supports investigating an abnormal response, but its limited sensitivity means that a negative test cannot confirm correct screw placement [5, 46]. It should complement anatomical assessment and imaging rather than replace them. Moreover, accuracy for radiographic malposition must remain separate from accuracy for postoperative neurological deficit [4]. Free-run EMG likewise cannot serve as a comprehensive safeguard against neurological injury.

What this review adds to existing evidence:Beyond a procedure-specific summary, this review tested the reliability of published estimates. High within-cell overlap showed that many reviews re-analysed the same evidence, and re-analysis lowered SSEP and MEP sensitivity in the largest review, underlining the need for transparent cohort selection [2].

Equally, high specificity should not be confused with a high probability of injury after every alert. That probability also depends on baseline risk. An unchanged recording is reassuring only to the extent permitted by the modality’s sensitivity and the neurological injury being considered.

Limitations: Most evidence was of low or very low certainty, and no source review achieved moderate or high AMSTAR 2 confidence. Modality comparisons were indirect, across different populations, alarm thresholds and outcome definitions. Data came mainly from the reviews, sometimes reconstructed from figures, without checking every primary publication, and author-based matching may have missed overlapping cohorts.

Exclusion of studies without deficits and continuity corrections may have influenced pooled estimates. GRADE judgments relied on operational thresholds and inherited risk-of-bias assessments. Non-English reports were excluded; and Embase and the Cochrane Library were not searched. Finally, postoperative outcomes are influenced by responses to alerts, complicating conventional false-positive classification. Neither diagnostic accuracy nor this review establishes the causal benefit of monitoring.

Recommendations and future directions: For spine surgeons, the most defensible approach is to prioritize MEP-containing monitoring when the cord is at substantial risk, consider D-wave as complementary information during intramedullary tumour surgery, and use EMG as a targeted adjunct. Stable SSEP or negative tEMG findings should not override anatomical concern. Monitoring should be embedded within an agreed multidisciplinary response protocol [56].

Future studies should compare modalities within the same patients, standardize alarm criteria, distinguish immediate from persistent deficits, document rescue interventions and share cohort-level data, so that the monitoring strategy and response that improve outcomes can be identified for each surgical setting.

Conclusion

IONM accuracy varies substantially by modality and surgical setting. MEP-containing strategies generally offer greater sensitivity than SSEP alone, but stable recordings do not guarantee neurological preservation. Overlapping reviews and predominantly low-certainty evidence limit confidence in precise estimates and comparisons between modalities. The central clinical message is that IONM provides procedure-specific information about neurological risk, rather than a universal assurance of safety.

Figure legends

Fig. 1 PRISMA 2020 flow diagram of review selection. The search was shared by four companion overviews; the 37 reviews reporting diagnostic accuracy form the present analysis.

Fig. 2 Sensitivity and specificity by modality and procedure. Filled markers show the index meta-analysis for each cell (chosen by the most primary studies, then the fewest AMSTAR 2 critical flaws, then the most recent); open markers show the deduplicated bivariate re-pooled estimate where study-level data were available. Horizontal lines are 95% confidence intervals. The index review, its number of studies (k) and the GRADE certainty are shown on the right.

Fig. 3 Summary operating points of the index meta-analyses in receiver operating characteristic space, by modality. Points closer to the upper-left corner indicate better discrimination.

Fig. 4 Pairwise overlap of primary studies between the 37 reviews, expressed as the corrected covered area (CCA). Values of 15% or more (very high overlap) are printed.

Table 1. Characteristics of the 37 Systematic Reviews Reporting the Diagnostic Accuracy of IONM in Spine Surgery.

Table 2. Summary of Findings: Diagnostic Accuracy of IONM Modalities for New Postoperative Neurological Deficit, by Procedure.

Table 3. Practical interpretation of IONM evidence by surgical setting

References

1. Fehlings MG, Brodke DS, Norvell DC, Dettori JR (2010) The evidence for intraoperative neurophysiological monitoring in spine surgery: does it make a difference? Spine 35:S37-46. https://doi.org/10.1097/BRS.0b013e3181d8338e

2. Alvi MA, Kwon BK, Hejrati N, et al (2024) Accuracy of Intraoperative Neuromonitoring in the Diagnosis of Intraoperative Neurological Decline in the Setting of Spinal Surgery-A Systematic Review and Meta-Analysis. Glob Spine J 14:105S-149S. https://doi.org/10.1177/21925682231196514

3. Reddy RP, Singh-Varma A, Chang R, et al (2024) Transcranial Motor Evoked Potentials as a Predictive Modality for Postoperative Deficit in Cervical Spine Decompression Surgery - A Systematic Review and Meta-Analysis. Glob Spine J 14:1609–1628. https://doi.org/10.1177/21925682231219224

4. Reddy RP, Chang R, Coutinho DV, et al (2022) Triggered Electromyography is a Useful Intraoperative Adjunct to Predict Postoperative Neurological Deficit Following Lumbar Pedicle Screw Instrumentation. Glob Spine J 12:1003–1011. https://doi.org/10.1177/21925682211018472

5. Lee CH, Kim HW, Kim HR, et al (2015) Can triggered electromyography thresholds assure accurate pedicle screw placements? A systematic review and meta-analysis of diagnostic test accuracy. Clin Neurophysiol Off J Int Fed Clin Neurophysiol 126:2019–2025. https://doi.org/10.1016/j.clinph.2014.11.026

6. Pieper D, Antoine SL, Mathes T, et al (2014) Systematic review finds overlapping reviews were not mentioned in every other overview. J Clin Epidemiol 67:368–375. https://doi.org/10.1016/j.jclinepi.2013.11.007

7. Rezapour A, Derakhshani N, Pouyan SN, et al (2025) Safety and Efficacy of Intraoperative Neuromonitoring: An Umbrella Review. Health Sci Rep 8:e71370. https://doi.org/10.1002/hsr2.71370

8. Gates M, Gates A, Pieper D, et al (2022) Reporting guideline for overviews of reviews of healthcare interventions: development of the PRIOR statement. BMJ 378:e070849. https://doi.org/10.1136/bmj-2022-070849

9. McInnes MDF, Moher D, Thombs BD, et al (2018) Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA 319:388–396. https://doi.org/10.1001/jama.2017.19163

10. Aromataris E, Fernandez R, Godfrey CM, et al (2015) Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Based Healthc 13:132–140. https://doi.org/10.1097/XEB.0000000000000055

11. Muthu S, Kolarpatti Ponnusamy DV, Viswanathan VK, et al (2026) 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. JB JS Open Access 11:e26.00102. https://doi.org/10.2106/JBJS.OA.26.00102

12. Muthu S (2023) The efficiency of machine learning-assisted platform for article screening in systematic reviews in orthopaedics. Int Orthop 47:551–556. https://doi.org/10.1007/s00264-022-05672-y

13. Shea BJ, Reeves BC, Wells G, et al (2017) AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 358:j4008. https://doi.org/10.1136/bmj.j4008

14. Whiting PF, Rutjes AW, Westwood ME, et al (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 155:529–536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

15. Pollock M, Fernandes RM, Newton AS, et al (2019) A decision tool to help researchers make decisions about including systematic reviews in overviews of reviews of healthcare interventions. Syst Rev 8:29. https://doi.org/10.1186/s13643-018-0768-8

16. Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y (2023) Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy, 1st ed. Wiley

17. Cochrane Handbook for Systematic Reviews of Interventions | Cochrane. https://www.cochrane.org/authors/handbooks-and-manuals/handbook. Accessed 16 Oct 2026

18. Reitsma JB, Glas AS, Rutjes AW, et al (2005) Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 58:982–990. https://doi.org/10.1016/j.jclinepi.2005.02.022

19. Harbord RM, Deeks JJ, Egger M, et al (2007) A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics 8:239–251. https://doi.org/10.1093/biostatistics/kxl004

20. Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. BMJ 327:557–560. https://doi.org/10.1136/bmj.327.7414.557

21. Thirumala PD, Huang J, Brahme IS, et al (2017) Alarm criteria for motor evoked potentials. Neurol India 65:708–715. https://doi.org/10.4103/neuroindia.NI_1195_16

22. Thirumala PD, Huang J, Thiagarajan K, et al (2016) Diagnostic Accuracy of Combined Multimodality Somatosensory Evoked Potential and Transcranial Motor Evoked Potential Intraoperative Monitoring in Patients With Idiopathic Scoliosis. Spine 41:E1177–E1184. https://doi.org/10.1097/BRS.0000000000001678

23. Reddy RP, Chang R, Rosario BP, et al (2021) What is the predictive value of intraoperative somatosensory evoked potential monitoring for postoperative neurological deficit in cervical spine

surgery?-a meta-analysis. Spine J Off J North Am Spine Soc 21:555–570. https://doi.org/10.1016/j.spinee.2021.01.010

24. Leeflang MM, Bossuyt PM, Irwig L (2009) Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. J Clin Epidemiol 62:5–12. https://doi.org/10.1016/j.jclinepi.2008.04.007

25. Schünemann HJ, Mustafa RA, Brozek J, et al (2020) GRADE guidelines: 21 part 1. Study design, risk of bias, and indirectness in rating the certainty across a body of evidence for test accuracy. J Clin Epidemiol 122:129–141. https://doi.org/10.1016/j.jclinepi.2019.12.020

26. Schünemann HJ, Mustafa RA, Brozek J, et al (2020) GRADE guidelines: 21 part 2. Test accuracy: inconsistency, imprecision, publication bias, and other domains for rating the certainty of evidence and presenting it in evidence profiles and summary of findings tables. J Clin Epidemiol 122:142–152. https://doi.org/10.1016/j.jclinepi.2019.12.021

27. Addanki RND, Golish S, Thomas S, et al (2025) Evaluating the Diagnostic Benefit of Adding Somatosensory Evoked Potentials to Transcranial Motor-Evoked Potentials Compared to Either Modality Alone in Intramedullary Spinal Cord Tumor Resection: A Systematic Review and Meta-Analysis. World Neurosurg 202:124360. https://doi.org/10.1016/j.wneu.2025.124360

28. Addanki RND, Lee BB, Anetakis KM, et al (2025) Precision in Practice: A Systematic Review and Meta-Analysis of Intraoperative Neurophysiological Monitoring for Optimizing Outcomes in Extramedullary Spinal Cord Tumor Resection. J Pers Med 15:513. https://doi.org/10.3390/jpm15110513

29. Addanki RND, Thomas S, Balasundaram S, et al (2025) Systematic review of alarm criteria and reversibility in motor-evoked potentials during intramedullary spinal cord tumor resection. Clin Neurophysiol Off J Int Fed Clin Neurophysiol 178:2110966. https://doi.org/10.1016/j.clinph.2025.2110966

30. Antkowiak L, Antkowiak M, Kasperczuk A, et al (2025) Accuracy of intraoperative neurophysiological monitoring in predicting postoperative neurological decline in intradural

extramedullary spinal tumor surgery: a systematic review and meta-analysis. Neurosurg Rev 49:5. https://doi.org/10.1007/s10143-025-03926-y

31. Antkowiak Ł, Antkowiak M, Kasperczuk A, et al (2026) Comparative diagnostic performance of motor-evoked potential alarm thresholds for predicting postoperative neurological decline in intradural extramedullary spinal tumor surgery: a systematic review and meta-analysis. Neurosurg Rev 49:564. https://doi.org/10.1007/s10143-026-04490-9

32. Azad TD, Pendharkar AV, Nguyen V, et al (2018) Diagnostic Utility of Intraoperative Neurophysiological Monitoring for Intramedullary Spinal Cord Tumors: Systematic Review and Meta-Analysis. Clin Spine Surg 31:112–119. https://doi.org/10.1097/BSD.0000000000000558

33. Ishida W, Casaos J, Chandra A, et al (2019) Diagnostic and therapeutic values of intraoperative electrophysiological neuromonitoring during resection of intradural extramedullary spinal tumors: a single-center retrospective cohort and meta-analysis. J Neurosurg Spine 30:839–849. https://doi.org/10.3171/2018.11.SPINE181095

34. Rijs K, Klimek M, Scheltens-de Boer M, et al (2019) Intraoperative Neuromonitoring in Patients with Intramedullary Spinal Cord Tumor: A Systematic Review, Meta-Analysis, and Case Series. World Neurosurg 125:498-510.e2. https://doi.org/10.1016/j.wneu.2019.01.007

35. Thomas S, Addanki RND, Murugan CK, et al (2026) A Systematic Review and Meta-Analysis to Compare the Diagnostic Accuracy of Direct-Waves and Myogenic Motor-Evoked Potential Neuromonitoring During Intramedullary Spinal Cord Tumor Resection. Oper Neurosurg 30:513–524. https://doi.org/10.1227/ons.0000000000001705

36. Aghaei MM, Babaie A, Bagherzadeh S, et al (2026) Intraoperative neuromonitoring in degenerative cervical myelopathy surgery: a systematic review and meta-analysis of diagnostic accuracy and postoperative neurological outcomes. Eur Spine J Off Publ Eur Spine Soc Eur Spinal Deform Soc Eur Sect Cerv Spine Res Soc 35:5236–5251. https://doi.org/10.1007/s00586-026-10187-w

37. Ajiboye RM, Zoller SD, Sharma A, et al (2017) Intraoperative Neuromonitoring for Anterior Cervical Spine Surgery: What Is the Evidence? Spine 42:385–393. https://doi.org/10.1097/BRS.0000000000001767

38. Di Martino A, Papalia R, Caldaria A, et al (2019) Should evoked potential monitoring be used in degenerative cervical spine surgery? A systematic review. J Orthop Traumatol Off J Ital Soc Orthop Traumatol 20:19. https://doi.org/10.1186/s10195-019-0524-4

39. Resnick DK, Anderson PA, Kaiser MG, et al (2009) Electrophysiological monitoring during surgery for cervical degenerative myelopathy and radiculopathy. J Neurosurg Spine 11:245–252. https://doi.org/10.3171/2009.2.SPINE08730

40. Velur Prasuna L, Vibha V, Marvin G (2026) Diagnostic test accuracy meta-analysis of transcranial motor evoked potential monitoring for the prediction of postoperative neurological deficits in ossification of the posterior longitudinal ligament surgery. Clin Neurophysiol Pract 11:426–433. https://doi.org/10.1016/j.cnp.2026.05.008

41. Thirumala PD, Cheng HL, Loke YK, et al (2016) Diagnostic accuracy of somatosensory evoked potential monitoring during scoliosis fusion. J Clin Neurosci Off J Neurosurg Soc Australas 30:8–14. https://doi.org/10.1016/j.jocn.2016.01.017

42. Gorijala VK, Reddy RP, Anetakis KM, et al (2024) Diagnostic utility of different types of somatosensory evoked potential changes in pediatric idiopathic scoliosis correction surgery. Eur Spine J Off Publ Eur Spine Soc Eur Spinal Deform Soc Eur Sect Cerv Spine Res Soc 33:1644–1656. https://doi.org/10.1007/s00586-023-08063-y

43. Angelliaume A, Alhada TL, Parent HF, et al (2023) Intraoperative neurophysiological monitoring in scoliosis surgery: literature review of the last 10 years. Eur Spine J Off Publ Eur Spine Soc Eur Spinal Deform Soc Eur Sect Cerv Spine Res Soc 32:3072–3076. https://doi.org/10.1007/s00586-023-07837-8

44. Thirumala PD, Crammond DJ, Loke YK, et al (2017) Diagnostic accuracy of motor evoked potentials to detect neurological deficit during idiopathic scoliosis correction: a systematic review. J Neurosurg Spine 26:374–383. https://doi.org/10.3171/2015.7.SPINE15466

45. O'Reilly BT, Dunker AM, Domagala DM, et al (2026) Diagnostic Accuracy of Intraoperative Neurophysiologic Monitoring in Spinal Deformity Surgery: Systematic Review and Meta-Analysis. Glob Spine J 21925682261492292. https://doi.org/10.1177/21925682261492292

46. Fonseca P, Goethel M, Vilas-Boas JP, et al (2022) A systematic review with meta-analysis of the diagnostic test accuracy of pedicle screw electrical stimulation. Eur Spine J Off Publ Eur Spine Soc Eur Spinal Deform Soc Eur Sect Cerv Spine Res Soc 31:1599–1610. https://doi.org/10.1007/s00586-022-07218-7

47. Oner A, Ely CG, Hermsmeyer JT, Norvell DC (2012) Effectiveness of EMG use in pedicle screw placement for thoracic spinal deformities. Evid-Based Spine-Care J 3:35–43. https://doi.org/10.1055/s-0031-1298599

48. Mikula AL, Williams SK, Anderson PA (2016) The use of intraoperative triggered electromyography to detect misplaced pedicle screws: a systematic review and meta-analysis. J Neurosurg Spine 24:624–638. https://doi.org/10.3171/2015.6.SPINE141323

49. Wu Y, Vázquez-Barquero A (2017) Stimulus-Evoked Electromyographic Monitoring During Minimally Invasive Transpedicular Implantation of Screws in Lumbosacral Spine: Threshold Value, Methodology and Clinical Effectiveness. World Neurosurg 98:146–151. https://doi.org/10.1016/j.wneu.2016.10.122

50. Chang R, Reddy RP, Coutinho DV, et al (2021) Diagnostic Accuracy of SSEP Changes During Lumbar Spine Surgery for Predicting Postoperative Neurological Deficit: A Systematic Review and Meta-Analysis. Spine 46:E1343–E1352. https://doi.org/10.1097/BRS.0000000000004099

51. Reddy RP, Gorijala VK, Kaithi VR, et al (2023) Utility of transcranial motor-evoked potential changes in predicting postoperative deficit in lumbar decompression and fusion surgery: a systematic review and meta-analysis. Eur Spine J Off Publ Eur Spine Soc Eur Spinal Deform Soc Eur Sect Cerv Spine Res Soc 32:3321–3332. https://doi.org/10.1007/s00586-023-07879-y

52. Velur P, Velur V, Gong M (2026) Utility of intraoperative neuromonitoring changes in predicting postoperative neurologic deficit in lateral lumbar interbody fusion: A systematic review and meta-

analysis. J Clin Neurosci Off J Neurosurg Soc Australas 151:112039. https://doi.org/10.1016/j.jocn.2026.112039

53. Addanki RND, Ezhuvathra P, Al-Qudah AM, et al (2025) Diagnostic accuracy of intraoperative neuromonitoring during non-tumor thoracic spine surgeries: A systematic review and meta-analysis. J Clin Neurosci Off J Neurosurg Soc Australas 141:111567. https://doi.org/10.1016/j.jocn.2025.111567

54. Basu S, Gohil K (2023) Intraoperative Neuromonitoring in Spine Surgery: Does it Reduce Neural Complications? Current Evidence. Indian Spine J 6:15–26. https://doi.org/10.4103/isj.isj_64_22

55. Velur PL, Velur V, Gong M (2026) Diagnostic accuracy of somatosensory evoked potentials in lumbar spine surgery: A systematic review and meta-analysis. Clin Neurophysiol Pract 11:579–588. https://doi.org/10.1016/j.cnp.2026.07.005

56. Fehlings MG, Alvi MA, Evaniew N, et al (2024) A Clinical Practice Guideline for Prevention, Diagnosis and Management of Intraoperative Spinal Cord Injury: Recommendations for Use of Intraoperative Neuromonitoring and for the Use of Preoperative and Intraoperative Protocols for Patients Undergoing Spine Surgery. Glob Spine J 14:212S-222S. https://doi.org/10.1177/21925682231202343

Table 1. Characteristics of the 37 Systematic Reviews Reporting the Diagnostic Accuracy of IONM in Spine Surgery.

	Procedure focus	Modalities	Primary studies, n	Search end	Registered	Risk-of-bias tool	AMSTAR 2 confidence	Role in sy
	Intramedullary tumour (alarm criteria)	MEP; D-wave	25	12/2024	Yes	QUADAS-2	Critically low	Narrative/c
	Thoracic (non-tumour)	SSEP; MEP; multimodal	19	06/2025	Yes	QUADAS-2	Critically low	Index
	Intramedullary tumour	SSEP; MEP; D-wave; multimodal	29	03/2025	Yes	QUADAS-2	Critically low	Index
	Extramedullary tumour	SSEP; MEP; multimodal	20	04/2025	Yes	QUADAS-2	Critically low	Index
	Degenerative cervical myelopathy	SSEP; MEP; multimodal	44	08/2025	Yes	QUADAS-2	Low	Index
	Anterior cervical	SSEP; MEP; EMG; unimodal vs multimodal	10	NR	No	MINORS	Critically low	Narrative/c
	Mixed spine	SSEP; MEP; EMG; multimodal	163	09/2022	Yes	QUADAS-2	Low	Index
	Scoliosis	SSEP; MEP	43	01/2022	No	Level of evidence	Critically low	Narrative/c
	Extramedullary tumour	MEP; SSEP; D-wave; multimodal	17	02/2025	Yes	QUADAS-2	Low	Index
	Extramedullary tumour (MEP thresholds)	MEP; MEP within multimodal	16	11/2025	Yes	QUADAS-2	Low	Narrative/c
	Intramedullary tumour	MEP; SSEP	21	06/2016	No	Level of evidence	Critically low	Alternate
	Mixed spine	SSEP; MEP	50	2022	No	Level of evidence	Critically low	Narrative/c
	Lumbar	SSEP	11	NR	No	QUADAS-2	Critically low	Index
	Degenerative cervical	SSEP; MEP; MIOM	8	07/2017	No	Level of evidence	Critically low	Narrative/c
	Mixed spine	SSEP; MEP; multimodal IOM	32	12/2008	No	Level of evidence	Critically low	Narrative/c
	Pedicle screws	tEMG	12	05/2021	No	QUADAS-2	Critically low	Alternate
	Paediatric idiopathic scoliosis	SSEP	16	05/2022	Yes	QUADAS-2	Critically low	Index
	Extramedullary tumour	multimodal	5	09/2018	No	Level of evidence	Critically low	Narrative/c
	Pedicle screws	tEMG	11	04/2014	No	QUADAS-2	Critically low	Index
	Pedicle screws	tEMG	26	06/2014	No	Downs and Black chec	Critically low	Index
	Spinal deformity (all ages)	MEP; SSEP; multimodal; DNEP	43	05/2026	No	QUADAS-2	Critically low	Index
	Pedicle screws (thoracic deformity)	tEMG	8	07/2011	No	Level of evidence	Critically low	Narrative/c
	Cervical	SSEP	23	03/2020	No	QUADAS-2	Critically low	Index

	Procedure focus	Modalities	Primary studies, n	Search end	Registered	Risk-of-bias tool	AMSTAR 2 confidence	Role in sy
	Lumbar pedicle screws	tEMG	13	02/2020	No	QUADAS-2	Critically low	Index
	Lumbar	MEP	8	2022	No	QUADAS-2	Critically low	Index
	Cervical	MEP	19	NR	No	QUADAS-2	Critically low	Alternate
	Degenerative cervical	SSEP; MEP; EMG; SCEP	27	2007	No	Level of evidence	Critically low	Narrative/c
	Intramedullary tumour	SSEP; MEP; multimodal	31	02/2018	Yes	QUADAS-2	Low	Alternate
	Idiopathic scoliosis	SSEP	15	01/2014	No	QUADAS-2	Critically low	Alternate
	Idiopathic scoliosis	multimodal	7	01/2015	No	QUADAS-2	Critically low	Index
	Idiopathic scoliosis	MEP	12	10/2014	No	QUADAS-2	Critically low	Index
	Mixed spine	MEP	25	01/2014	No	QUADAS-2	Critically low	Alternate
	Intramedullary tumour	MEP; D-wave	26	2024	Yes	QUADAS-2	Critically low	Index
	Cervical/thoracic OPLL	MEP	5	01/2026	No	QUADAS-2	Critically low	Index
	Lumbar	SSEP	5	01/2026	Yes	QUADAS-2	Low	Alternate
	Lateral lumbar interbody fusion	Free-run/triggered EMG; MEP; saphenous SSEP; multimodal	7	01/2026	Yes	QUADAS-2	Critically low	Index
	MIS lumbar pedicle screws	tEMG	8	12/2015	No	None	Critically low	Narrative/c

Abbreviations: AMSTAR 2, A MeaSurement Tool to Assess systematic Reviews 2; EMG, electromyography; MEP, motor evoked potentials; NR, not reported; OPLL, ossification of the posterior longitudinal ligament; QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies 2; SSEP, somatosensory evoked potentials; tEMG, triggered EMG. Role: Index, the review selected for at least one modality-by-procedure cell; Alternate, contributed study-level data to a cell; Narrative/criterion only, reported ranges, narrative synthesis, threshold-specific or long-term estimates only.

Table 2. Summary of Findings: Diagnostic Accuracy of IONM Modalities for New Postoperative Neurological Deficit, by Procedure.

	Index review (k)	Re-pooled studies (deficits)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	LR+	LR−	Per 1,000: TP / FN / FP	Certainty (reason
d potentials								
	Reddy 2021[23] (23)	31 (315)	42.0 (32.3–52.3)	97.4 (95.7–98.4)	16.1	0.60	8 / 12 / 26	Low (a, b)
tric)	Gorijala 2024[42] (16)	16 (78)	75.6 (59.8–86.6)	97.3 (94.6–98.7)	27.7	0.25	15 / 5 / 26	Very low (a, b, c)
-	Addanki 2025d[28] (6)	12 (93)	48.2 (38.4–58.1)	95.6 (93.2–97.2)	10.9	0.54	10 / 10 / 43	Moderate (a)
	Addanki 2025c[27] (9)	10 (140)	48.4 (37.6–59.4)	80.2 (70.9–87.0)	2.4	0.64	10 / 10 / 194	Moderate (a)
	Chang 2021[50] (11)	14 (120)	31.9 (18.7–48.8)	95.8 (91.5–98.0)	7.5	0.71	6 / 14 / 41	Very low (a, b, c)
	Alvi 2024[2] (52)	52 (692)	60.7 (48.9–71.3)	95.9 (93.8–97.3)	14.8	0.41	12 / 8 / 40	Low (a, b)
	Addanki 2025b[53] (8)	10 (194)	65.5 (45.7–81.2)	93.0 (87.5–96.1)	9.2	0.37	13 / 7 / 69	Very low (a, b, c)
als								
	Aghaei 2026[36] (32)	37 (381)	57.0 (46.1–67.3)	94.3 (92.0–96.0)	10.0	0.46	11 / 9 / 56	Low (a, b)
	Velur 2026a[40] (5)	5 (68)	62.9 (27.3–88.4)	89.8 (74.0–96.4)	5.9	0.41	13 / 7 / 100	Very low (a, b, c²)
tric)	Thirumala 2017a[44] (12)	Index only	91 (34–100)	96 (92–98)	22.8	0.09	18 / 2 / 39	Very low (a, b, c²)
-	Addanki 2025d[28] (13)	27 (282)	70.8 (60.9–79.1)	95.5 (92.7–97.3)	15.6	0.31	14 / 6 / 44	Low (a, b)
	Thomas 2026[35] (20)	27 (443)	70.9 (59.3–80.2)	83.7 (77.9–88.2)	4.3	0.35	14 / 6 / 160	Moderate (a)
	Reddy 2023[51] (8)	6 (36)	65.0 (44.2–81.4)	93.8 (77.6–98.5)	10.4	0.38	13 / 7 / 60	Very low (a, b, c)
	Alvi 2024[2] (68)	86 (1609)	84.0 (80.2–87.1)	94.9 (93.1–96.2)	16.3	0.17	17 / 3 / 50	Low (a, b)
	Addanki 2025b[53] (12)	12 (226)	79.7 (70.0–86.9)	92.1 (87.2–95.2)	10.0	0.22	16 / 4 / 77	Low (a, b)
-	Antkowiak 2025[30] (2)	6 (20)	53.2 (30.7–74.4)	98.5 (96.0–99.4)	34.0	0.48	11 / 9 / 15	Very low (a, c², e)
	Thomas 2026[35] (8)	7 (86)	26.6 (15.1–42.6)	94.0 (86.9–97.4)	4.4	0.78	5 / 15 / 59	Low (a, c)
	Aghaei 2026[36] (17)	14 (120)	77.1 (63.9–86.5)	92.3 (84.5–96.4)	10.1	0.25	15 / 5 / 75	Low (a, b)
tric)	Thirumala 2016b[22] (7)	6 (19)	86.0 (64.0–95.5)	97.0 (96.2–97.7)	28.4	0.14	17 / 3 / 29	Very low (a, c²)
	O'Reilly 2026[45] (36)	36 (589)	89.1 (82.6–93.4)	95.2 (92.6–96.9)	18.4	0.11	18 / 2 / 47	Low (a, b)
-	Antkowiak 2025[30] (10)	14 (112)	63.1 (48.7–75.5)	97.0 (94.8–98.2)	20.7	0.38	13 / 7 / 30	Very low (a, c, e)
	Addanki 2025c[27] (16)	14 (185)	83.9 (76.8–89.2)	82.9 (66.2–92.3)	4.9	0.20	17 / 3 / 167	Low (a, b)
	Velur 2026c[52] (5)	Index only	99.3 (13.7–100)	98.8 (80.2–99.9)	82.7	0.01	20 / 0 / 12	Very low (a, b, c²)
	Alvi 2024[2] (69)	69 (871)	82.4 (77.3–86.5)	93.3 (90.0–95.5)	12.2	0.19	16 / 4 / 66	Low (a, b)
	Addanki 2025b[53] (6)	9 (123)	82.0 (66.7–91.2)	94.7 (88.4–97.6)	15.3	0.19	16 / 4 / 52	Low (a, b)
	Alvi 2024[2] (16)	16 (162)	46.8 (32.1–62.0)	91.2 (81.5–96.0)	5.2	0.59	9 / 11 / 87	Very low (a, b, c)
rence standard: screw malposition on imaging)								
	Mikula 2016[48] (18)	Index only	78 (73–82)	94 (94–95)	13.0	0.23	16 / 4 / 59	Very low (a, b, d)
)	Reddy 2022[4] (13)	Index only	49 (36–63)	88 (80–93)	4.1	0.58	10 / 10 / 118	Very low (a, c, d)
	Lee 2015[5] (7)	9 (256)	54.9 (29.2–78.2)	96.1 (92.3–98.1)	13.9	0.47	11 / 9 / 38	Very low (a, b, c,
	Lee 2015[5] (7)	6 (189)	31.3 (9.7–65.9)	95.8 (92.1–97.8)	7.4	0.72	6 / 14 / 41	Very low (a, b, c²,

Estimates are the deduplicated bivariate re-pool where study-level data were available ("Re-pooled studies"), otherwise the index meta-analysis. Index review chosen by the most primary studies, then the fewest AMSTAR 2 critical flaws, then the most recent. Per-1,000 figures assume a deficit prevalence of 2% (20 patients with and 980 without a new deficit). GRADE downgrading: a, risk of bias; b, inconsistency; c, imprecision (c², two levels); d, indirectness; e, publication bias. Abbreviations: CI, confidence interval; FN, false negatives; FP, false positives; LR+/LR−, positive/negative likelihood ratio; TP, true positives; other abbreviations as in Table 1. Long-term outcomes, alarm thresholds and reversibility are reported in Supplementary Table S3.

Table 3. Practical interpretation of IONM evidence by surgical setting

Surgical setting	Monitoring approach supported by the evidence	Clinical value
Complex spinal deformity	Prioritize an **MEP-containing multimodal strategy**, commonly MEP plus SSEP[39,54]	Provides complementary motor and sensory information during correction; multimodal monitoring showed relatively high sensitivity.
Cervical or thoracic surgery with substantial cord risk	Include **MEP when motor pathways are threatened**, with SSEP providing complementary information.[16,32,48,54]	Assesses motor pathway compromise that SSEP alone may miss.
Intramedullary tumour resection	Consider **myogenic MEP with D-wave, where feasible**, within a multimodal strategy.[22,28]	MEP provides information about evolving motor dysfunction; D-wave adds information relevant to longer-term motor outcome.
Extramedullary tumour resection	Consider **MEP-based monitoring**, adding other modalities according to the structures at risk.[21,23]	Provides functional information during manipulation of the cord and adjacent neural structures.
Lumbar decompression and fusion	Individualize monitoring according to neurological risk, surgical exposure and the ability to respond to an alert.[45,46]	May provide additional information during procedures that threaten motor pathways or nerve roots.
Lateral lumbar interbody fusion	Tailor monitoring to the neural structures exposed; avoid selecting a strategy solely from the pooled sensitivity estimate.[47]	May assist assessment of neural compromise during the approach and instrumentation.
Pedicle screw instrumentation	Use **triggered EMG as an adjunct** to anatomical assessment and imaging.[40,41,53]	An abnormal response can prompt further assessment of screw position and potential neural risk.
Principles applying across settings: Agree on alarm criteria and a multidisciplinary response protocol before surgery. Interpret changes alongside technical, anaesthetic, physiological and surgical factors. An alert without a subsequent deficit may represent successful rescue rather than a false alarm. An unchanged recording should not override anatomical concern.[50,54]		
Abbreviations: IONM, intraoperative neurophysiological monitoring; MEP, motor evoked potentials; SSEP, somatosensory evoked potentials; EMG, electromyography.		

Figures

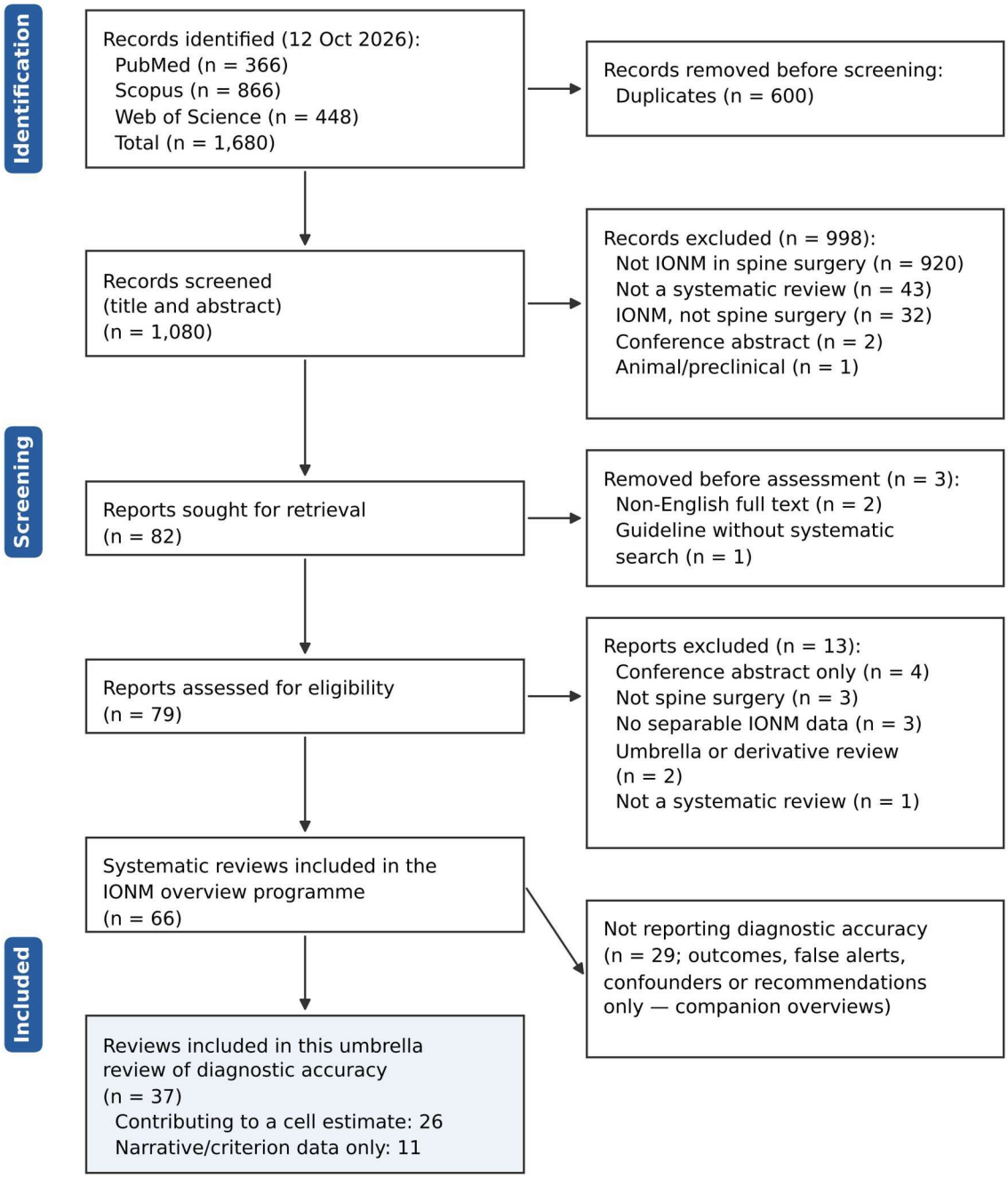

Identification
Records identified (12 Oct 2026):
PubMed (n = 366)
Scopus (n = 866)
Web of Science (n = 448)
Total (n = 1,680)
Records removed before screening:
Duplicates (n = 600)
Screening
Records screened
(title and abstract)
(n = 1,080)
Records excluded (n = 998):
Not IONM in spine surgery (n = 920)
Not a systematic review (n = 43)
IONM, not spine surgery (n = 32)
Conference abstract (n = 2)
Animal/preclinical (n = 1)
Reports sought for retrieval
(n = 82)
Removed before assessment (n = 3):
Non-English full text (n = 2)
Guideline without systematic
search (n = 1)
Reports assessed for eligibility
(n = 79)
Reports excluded (n = 13):
Conference abstract only (n = 4)
Not spine surgery (n = 3)
No separable IONM data (n = 3)
Umbrella or derivative review
(n = 2)
Not a systematic review (n = 1)
Included
Systematic reviews included in the
IONM overview programme
(n = 66)
Not reporting diagnostic accuracy
(n = 29; outcomes, false alerts,
confounders or recommendations
only — companion overviews)
Reviews included in this umbrella
review of diagnostic accuracy
(n = 37)
Contributing to a cell estimate: 26
Narrative/criterion data only: 11

Fig. 1 PRISMA 2020 flow diagram of review selection. The search was shared by four companion overviews; the 37 reviews reporting diagnostic accuracy form the present analysis.

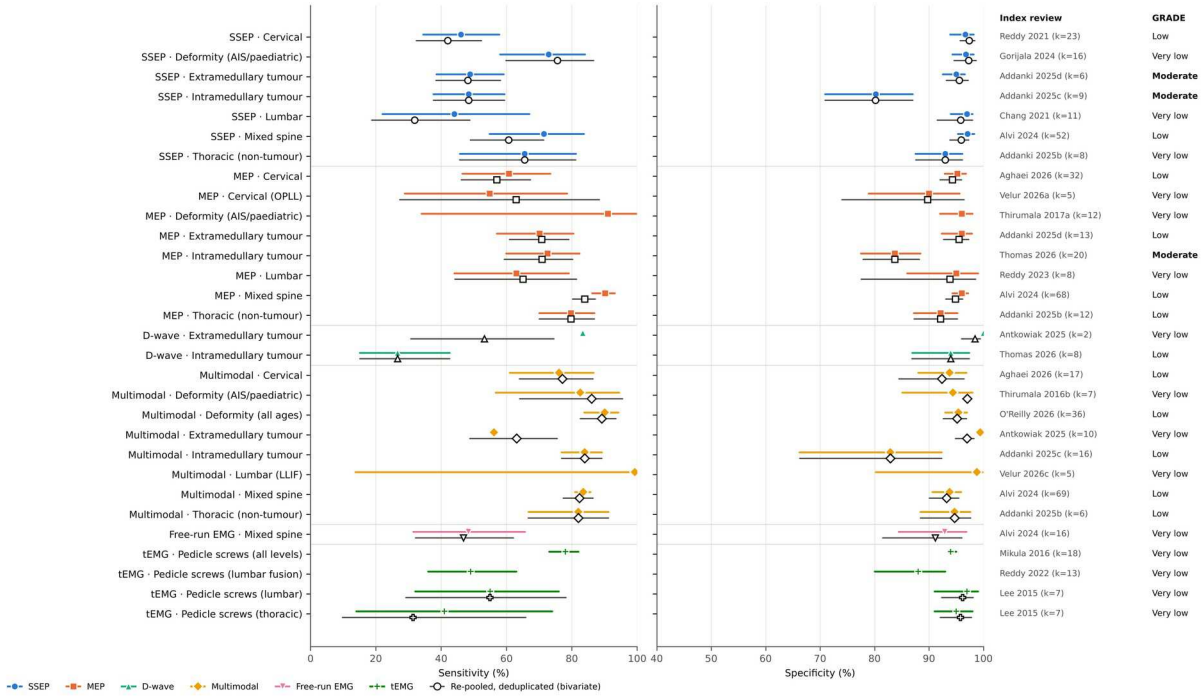


Fig. 2 Sensitivity and specificity by modality and procedure. Filled markers show the index meta-analysis for each cell (chosen by the most primary studies, then the fewest AMSTAR 2 critical flaws, then the most recent); open markers show the deduplicated bivariate re-pooled estimate where study-level data were available. Horizontal lines are 95% confidence intervals. The index review, its number of studies (k) and the GRADE certainty are shown on the right.

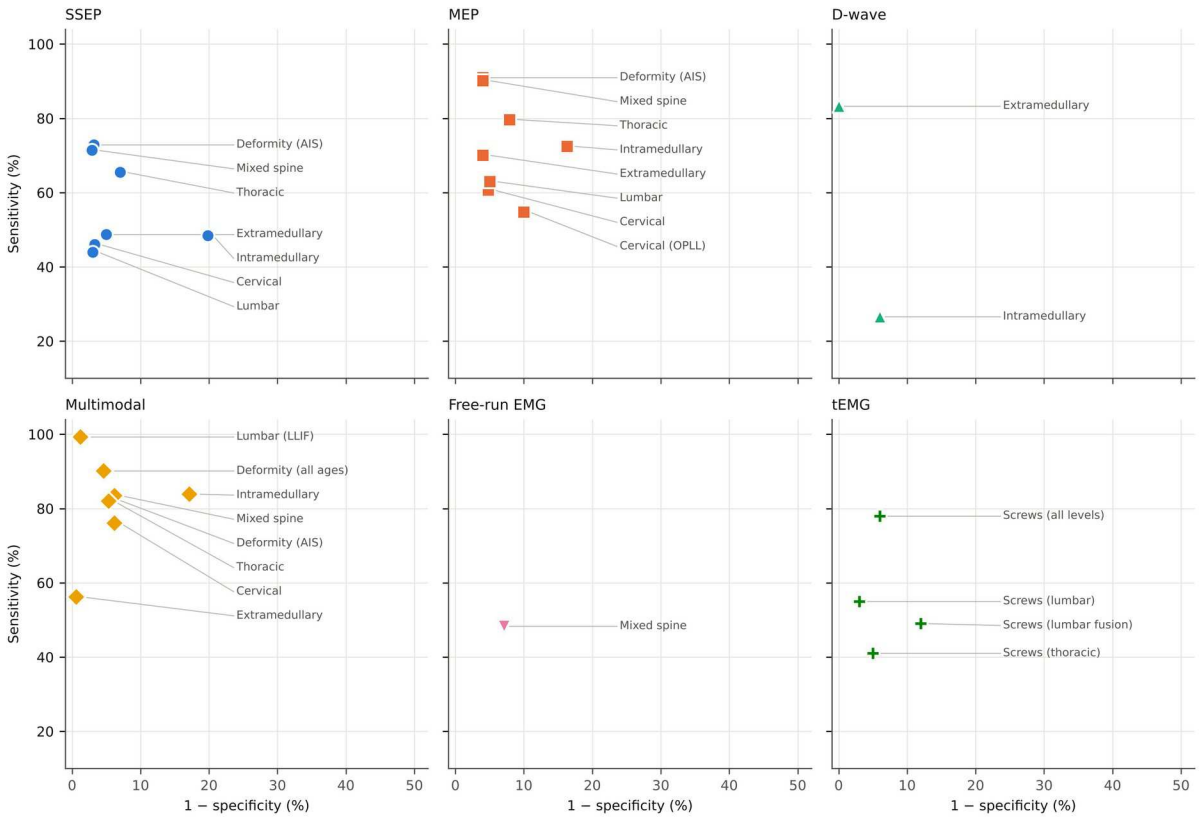


Fig. 3 Summary operating points of the index meta-analyses in receiver operating characteristic space, by modality. Points closer to the upper-left corner indicate better discrimination.

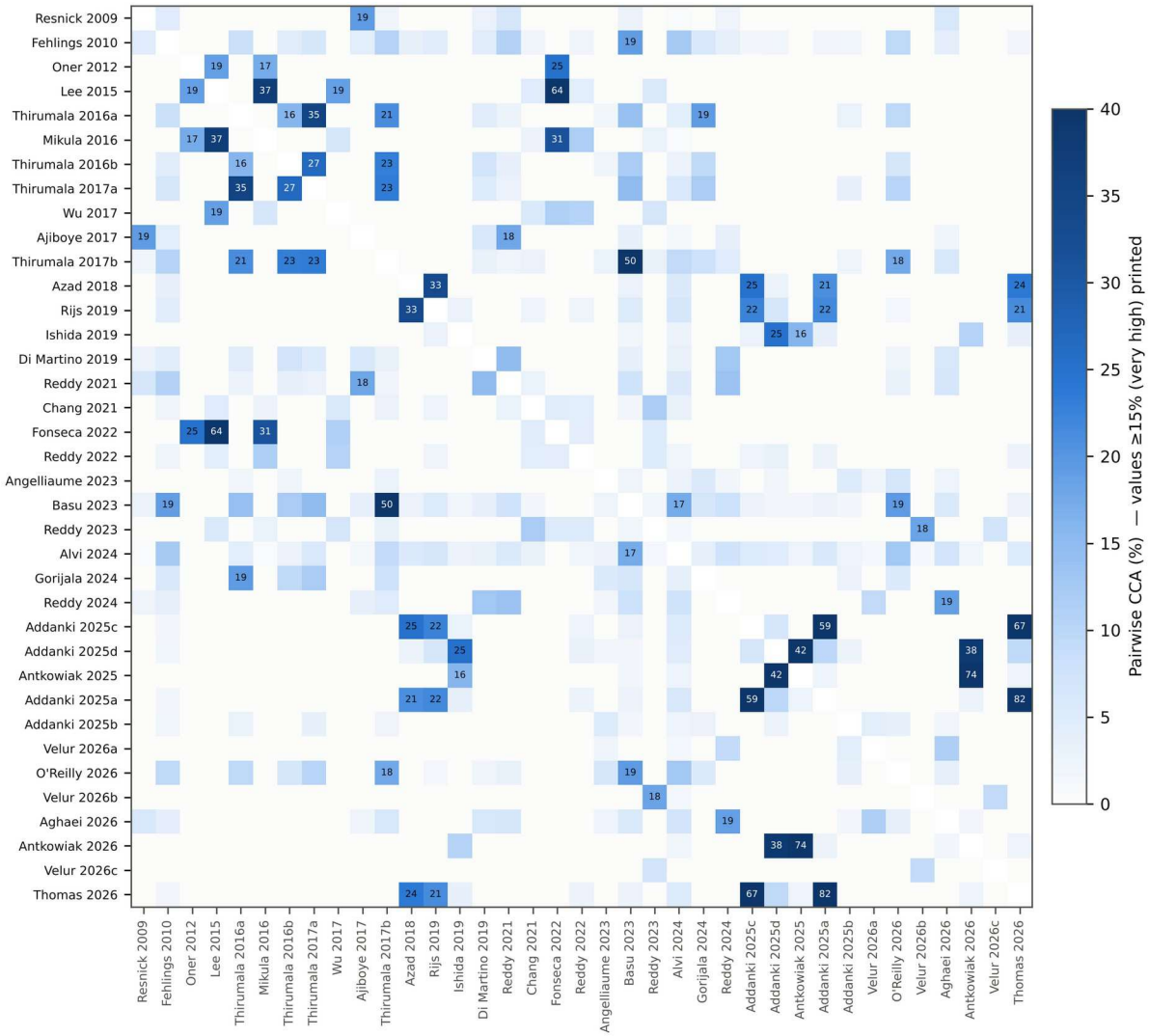


Fig. 4 Pairwise overlap of primary studies between the 37 reviews, expressed as the corrected covered area (CCA). Values of 15% or more (very high overlap) are printed.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESJDiagnosticAccuracySupplementary.docx