STARD-IONM Checklist Table

	STARD Item	IONM specific commentary
TITLE		
1	Identify the report as a diagnostic accuracy study using at least one accuracy measure. (such as but not limited to sensitivity, specificity, predictive values, or AUC)	When possible, titles should explicitly include terms such as intraoperative and diagnostic accuracy, or refer to predictive performance (e.g., sensitivity, specificity). Including “intraoperative” improves discoverability in systematic reviews. In some cases, especially when describing new monitoring methods, full diagnostic terminology may be difficult to incorporate, but emphasis on predictive metrics is recommended.
2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	The abstract should clearly and concisely state the aim or hypothesis. The methods section must specify study type (e.g., prospective, retrospective, observational), number of patients, and statistical analyses performed. Results should report key findings, including sample size and main outcomes (e.g., odds ratios, confidence intervals, p-values, predictive values). Conclusions should interpret the clinical relevance for IONM practice, indicate whether further studies are needed, and summarize the significance of the findings.
3	Scientific and clinical background, including the intended use and clinical role of the index test	A description of the scientific and clinical background is important, starting with a broad description of the field, of current knowledge and then identification of the knowledge gap. Information about the reference test, and its shortcomings, is important. Information about the rationale for a different or newer test (index test) should be given, and where relevant the knowledge gap. When the reference test is clinical outcome, the authors should clearly state which outcome parameter is used; and ideally explain why there is no intraoperative reference test.
4	Study objectives and hypotheses	Clearly state the primary objective of the study, such as assessing the diagnostic accuracy of an IONM modality (e.g., SSEP, MEP, EEG) to predict perioperative neurological deficits. Hypotheses should reflect expected performance measures like sensitivity, specificity, or predictive values. Pre-specifying primary and secondary hypotheses strengthens the methodological rigor and reduces bias in interpretation. IONM-specific considerations: In studies involving IONM, objectives may include determining whether specific intraoperative changes predict postoperative deficits, or whether the absence of changes reliably indicates neurological preservation. Hypotheses should address the relationship between IONM signal changes (index test) and the occurrence of postoperative outcomes (reference standard), while accounting for the potential influence of intraoperative interventions.
5	Specify whether the study was prospective or retrospective.	This is certainly important for IONM studies. It is particularly needed when interpreting the adequacy of the study design and interpreting the study results. If prospective, then one can determine whether important confounds were avoided by perusing the eligibility criteria and methods; and checking that an adequate sample size calculation was performed. For prospectively collected data, it is important to know whether how the results of the index or reference standard were interpreted, and whether action was taken (i.e. more information needed later in the methods). If the data were collected retrospectively, then subsequent descriptions in the methods and results of attempts to adjust for confounders are important.
6	Eligibility criteria	Clear eligibility criteria—age, sex, diagnosis, comorbidities,—are essential in STARD-IONM to reduce bias, ensure reproducibility, and support valid comparisons and clinical applicability.
7	Describe how potentially eligible participants were identified.	It is important to define eligibility criteria and describe how individuals were selected to ensure the study hypothesis can be appropriately investigated. Authors should specify how and where eligible subjects were identified, and report detailed inclusion and exclusion criteria to allow reproducibility.
8	Specify where and when potentially eligible participants were identified.	Report where and when eligible participants were identified, including details such as patient pool, clinical setting, age range, condition, and team experience. This contextualizes study applicability, minimizes selection bias, and improves reproducibility.
9	Whether participants formed a consecutive, random or convenience series	Random or convenience sampling can introduce selection bias. Consecutive enrollment of subjects, whether in prospective or retrospective studies, helps minimize bias and improves study validity.
Test methods		
10	Describe the index test with technical specifications sufficient for replication.	Describe the index test in sufficient detail to permit replication, including modality, stimulation and recording parameters, equipment, acquisition timing,

		baseline definition, alert criteria, and any patient- or procedure-specific adaptations. Report anesthesia and physiology variables that may affect interpretation, including anesthetic technique, neuromuscular blockade, at baseline and at alerts. Authors should describe the index test methodology with sufficient technical detail to permit replication. This includes specifying the stimulation and recording parameters (e.g., stimulus type, intensity, frequency, recording montage, filter settings), equipment used (manufacturer and model if relevant), and any modifications or special techniques applied during data acquisition. The description should reflect real intraoperative conditions and clarify any patient- or procedure-specific adaptations made to standard protocols. Include anesthesia-relevant factors affecting IONM (technique TIVA/volatile/mixed and NMB level before MEPs, MAP/ at baseline and at alerts).
11	Describe the reference standard(s).	Describe the reference standard in sufficient detail to permit replication, including the neurologic or functional domain assessed, timing of assessment, procedures used, thresholds/categories, assessor qualifications, and whether assessors were blinded to index-test results. When postoperative outcomes are used, specify the time point and whether they were compared with the final IONM signal recorded near the end of surgery.
12a	Define and justify index-test positivity cut-offs/result categories; distinguish prespecified vs exploratory.	Define the index-test positivity threshold(s) or result categories and justify their use, including whether thresholds were pre-specified or data-derived. The interpretation of the index test depends on how positivity is defined. Researchers must specify the exact threshold, cutoff value, or result category used to classify the index test as positive or negative. This includes, for example, amplitude changes, signal loss percentages, or categorical outputs in tests like SSEP, Brainstem auditory evoked potentials, EEG, or MEPs. The rationale for choosing this threshold should be provided, based on physiological reasoning, clinical guidelines, or prior studies. It must also be clearly stated whether the definition was determined before data collection (pre-specified) or derived from the study data itself (exploratory). Pre-specified cut-offs enhance reproducibility and reduce bias, while exploratory cut-offs should be interpreted cautiously and validated in future studies. State how alert criteria account for anesthetic depth, neuromuscular blockade, and physiologic state; specify rules for near-threshold fluctuations.
12b	Define and justify reference standard positivity cut-offs and result categories; distinguish prespecified vs exploratory.	Define the reference-standard positivity threshold(s) or result categories and justify their use, including timing of outcome assessment, diagnostic confirmation methods, and whether categories were pre-specified or exploratory. If assessor judgment was involved, describe steps taken to reduce variability and bias. The reference standard is the benchmark used to judge the accuracy of the index test. Researchers must clearly define what constitutes a positive result, such as the presence of a clinical condition, the timing of evaluation, and use of diagnostic test confirmation (audiogram for hearing, MRI for stroke) or specific clinical outcome (e.g. delirium). The criteria should be explicitly described and supported by clinical rationale or standard definitions (e.g NIH stroke scale, MRC grading scale, ICDSC scale for delirium). Importantly, the report must state whether these definitions were established before the study began (pre-specified) or determined after examining the data (exploratory). When outcome definitions are based on clinical judgment or follow-up, the risk of variability and bias increases—particularly if the assessor is influenced by index test results. Transparent reporting allows readers to assess the reliability and applicability of the study's accuracy estimates.
13a	Specify whether clinical information and reference standard results were available to index-test assessors.	State whether performers/readers of the index test had access to clinical information, operative context, or reference-standard results, and describe how this may have influenced interpretation. In diagnostic accuracy studies, especially those involving tests requiring subjective interpretation—such as imaging, electrophysiological monitoring, or clinician-judged outcomes—it is critical to report whether the interpreters of each test had access to other information that could influence their judgment. Item 13a requires reporting whether clinical details (e.g., symptoms, surgical findings) or reference standard outcomes (e.g., postoperative deficits) were known to the individuals performing or reading the index test. If such information was available, their interpretation might be biased toward the expected result, leading to inflated diagnostic accuracy.
13b	Specify whether index-test results were available to reference-standard assessors.	Item 13b addresses whether those assessing the reference standard (e.g., determining postoperative neurological outcome) had access to the index test results. If so, their judgment might be influenced by knowledge of IONM alerts or lack thereof, artificially increasing agreement between tests. This is

		particularly relevant in IONM studies where surgical teams might anticipate or dismiss subtle deficits based on intraoperative monitoring trends. This reporting item does not prescribe whether blinding is necessary—it simply requires that the presence or absence of blinding be transparently described. This enables readers and reviewers to understand the context of test interpretation and judge the potential risk of bias in accuracy estimates. If full blinding was not feasible, explain what information was available, and why, so the study's generalizability to real-world practice can be assessed.
14	Methods for estimating or comparing measures of diagnostic accuracy	Describe the statistical methods used to estimate or compare diagnostic accuracy, including sensitivity, specificity, predictive values, likelihood ratios, odds ratios, AUC, and confidence intervals. State how variability was assessed, how comparisons between IONM modalities were made, and how ND/PR/RD/ID categories were handled in the analysis. Authors must describe in detail how diagnostic performance was measured. This includes calculating sensitivity, specificity, predictive values (PPV, NPV), likelihood ratios, diagnostic odds ratios, and area under the ROC curve (AUC). Clearly specify the type of each variable analyzed—continuous, ordinal, categorical, or binary—and for discrete variables, report all possible values or ranges. State which statistical methods were used to obtain p-values and confidence intervals for diagnostic measures (e.g., chi-square or Fisher's tests, logistic regression, bootstrap methods for CIs). When comparing multiple IONM modalities, describe comparative tests used (e.g., McNemar's test for paired proportions, DeLong's test for AUC comparisons). Additionally, accuracy analyses should account for reversible IONM changes using extended contingency tables (e.g., 3×2 format as outlined by Holdefer et al.), and regression models (logistic or ordinal) should be named and their use justified when adjusting for covariates. Explicit reporting ensures reproducibility, supports validity, and enables readers to assess susceptibility to bias slightly modified from STARD guidance.
15	How indeterminate index test or reference standard results were handled	In IONM studies, indeterminate results—such as borderline signal changes or technical failures—can significantly impact the assessment of diagnostic accuracy. To ensure transparency and reproducibility, it's essential to: Define and report the frequency and causes of indeterminate results. Specify handling methods, such as exclusion from analysis, categorization as a separate group, or reclassification as false positives/negatives based on clinical outcomes. Justify the chosen approach and discuss its potential impact on diagnostic accuracy estimates. Transparent reporting of these aspects aligns with the STARD 2015 guidelines, which emphasize the importance of addressing indeterminate results to avoid biased estimates of test performance.
16	How missing data on the index test and reference standard were handled	Dealing with missing data is a challenging task. The most common approach is to exclude patients for whom the data of interest is missing. However, this approach is only valid if we have strong reasons to assume that the data is "missing at random" (MAR), meaning that the probability of missing is unrelated to unobserved data and does not introduce selection bias. For instance, if data is consistently missing for patients in the worst clinical condition because their condition prevents them from undergoing certain tests, excluding these patients would systematically omit the worst cases and introduce bias. Examples in IONM are in case of MEP loss, patients may not show up in the follow-up because they have a severe postoperative neurologic condition. Another issue is when the IONM team does not have access to follow-up patient charts, because of limited collaboration with the treating clinic. In such situations, it is highly recommended to consult a statistician to determine the most appropriate method for handling missing data. Alternatives to simple exclusion exist, such as data imputation (e.g. complete with the population mean, assign an existing value, etc.), which can be used to retain information from patients with incomplete data and reduce potential bias. The methods used to deal with missing data should be stated explicitly in the text, or at the very least in the limitations of the study.
17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	If the variance in the entire population is high, it might be a good idea to consider subgroup analyses. Possible explanations for high variances could be Simpson's paradox (when subgroups have different trends than the entire population) or the presence of outliers (beware also of typos!). Outliers can for instance be detected via box plots with whiskers. A stratification should be considered and justified. For example, if patients have preoperative neurologic deficits or motor weaknesses, this might influence the MEPs and the distribution of the outcome

		data. It might therefore be reasonable to stratify the results according to this criterion. It is crucial to distinguish pre-specified analyses from exploratory analyses. Before any data collection, a main hypothesis should be formulated, defining a primary endpoint (or at most two, but ideally not more). Based on this hypothesis, an appropriate statistical test should be selected to address the question raised. This approach enables a power analysis (i.e., determining the required sample size) and ensures that the full statistical power is dedicated to this test, increasing the likelihood of detecting a significant result if one exists.
18	Intended sample size and how it was determined	Sample size calculations are crucial in diagnostic accuracy studies to ensure sufficient precision and align with study objectives and hypotheses. Report when and how the sample size was determined and whether the assumptions made were in line with the study objective
19	Report flow of participants, using a diagram.	A participant flow diagram provides a visual summary of the progression of individuals through the stages of a study. It typically includes the number of participants assessed for eligibility, excluded (with reasons), enrolled, allocated to interventions, followed up, and analyzed. In the context of IONM, such diagrams can illustrate the number of patients undergoing monitoring, those with successful signal acquisition, instances of signal loss, and correlations with postoperative outcomes.
20	Baseline demographic and clinical characteristics of participants	Report baseline demographic and clinical characteristics relevant to IONM interpretation, including age, sex, diagnosis, preoperative neurologic status, and relevant surgical features In intraoperative neurophysiological monitoring (IONM) studies, reporting baseline demographic and clinical characteristics—such as age, sex, and preoperative neurological status—is critical. These factors can influence evoked potential amplitudes, signal reliability, and the likelihood of intraoperative alerts or false positives. For example, older patients or those with preexisting motor deficits may have lower baseline amplitudes or delayed responses, impacting both the index test and outcome classification. Detailed characterization of the surgical population also helps assess whether findings are generalizable to other settings and whether diagnostic accuracy measures (sensitivity, specificity, NPV, PPV) apply equally across patient subgroups. This information supports valid interpretation of IONM performance and helps identify sources of variability. Report anesthesia context relevant to interpretation: technique, depth monitor values at first acquisition, NMB level before MEPs, mean MAP/$EtCO_2$/core temperature during IONM.
21	Distribution of severity of disease in those with the target condition	Report the distribution of severity of the target condition, including degree, duration, and permanence of postoperative neurologic deficit where applicable. In studies using intraoperative neurophysiological monitoring (IONM) to detect or predict postoperative neurological deficits, the target condition can vary in severity—from mild, transient changes to severe, permanent impairments. Reporting this distribution is essential because the sensitivity of IONM may be higher in more severe cases, where signal changes are more pronounced. Stratifying outcomes by degree and duration of deficit allows a more accurate assessment of how IONM performs across the clinical spectrum. Equally important is the characterization of postoperative findings in patients who do not meet the definition of a neurological deficit. These individuals may still present with transient symptoms, subjective complaints (e.g., numbness, paresthesia), or unrelated complications that may account for intraoperative IONM changes. Describing the range and frequency of these alternative outcomes provides crucial context for interpreting false positives and evaluating the test's specificity. Without this information, test performance may be misestimated. Together, these data allow for a realistic, nuanced understanding of IONM accuracy and generalizability in surgical practice.
22	Time interval and any clinical interventions between index test and reference standard	Report the time interval between the index test and the reference standard, and describe any clinical interventions performed in response to IONM changes. Specify the timing of baseline acquisition, the last IONM recording relative to closure, and the timing of anesthetic, physiologic, or surgical interventions after alerts In IONM studies, the index test—typically a neurophysiological alert during surgery—is often compared to a reference standard such as a postoperative neurological exam or longer-term follow-up. It is essential to report the time interval between the index test and the outcome assessment, as changes in the patient's condition may occur due to natural recovery, delayed injury, or interventions performed after the alert. These changes can bias the relationship between the index test and the reference standard, particularly if IONM alerts

		lead to immediate surgical or anesthetic maneuvers that prevent a deficit from manifesting. Therefore, studies should describe not only the timing of follow-up (e.g., immediate post-op, discharge, or 3-month visit), but also any interventions performed in response to IONM alerts, such as stopping dissection, increasing blood pressure, or administering steroids. These actions may influence the outcome and can obscure whether the index test truly predicted the deficit or contributed to its avoidance. Some have proposed using the immediate intraoperative resolution of an IONM alert as an alternative or secondary reference standard, which may provide a closer temporal link and reduce post-surgical bias. Regardless of the model used, transparent reporting of both the follow-up interval and all relevant clinical interventions is critical to interpret the true diagnostic value of IONM. Report time of baseline and time of last IONM recording relative to skin closure, plus timestamps for anesthetic/physiologic interventions after alerts.
23	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	A clear cross-tabulation, typically in a 2×2 table—or a prespecified table should be provided, displaying the number of true positives, false positives, false negatives, and true negatives, based on the index test (e.g., IONM signal changes) and the reference standard (e.g., postoperative neurological outcome). This format allows calculation of sensitivity, specificity, PPV, and NPV, and supports transparent interpretation of the test's diagnostic accuracy. IONM-specific considerations: In IONM studies, the index test may be defined as a significant change in evoked potentials (e.g., >50% amplitude drop), while the reference standard is often a postoperative neurological deficit. Given the clinical implications of real-time intraoperative decisions, providing a complete cross-tabulation enables readers to assess both performance and potential biases in outcome classification. Extended tables may also be used when multiple thresholds or outcomes are assessed. If reversible changes are separated, consider an expanded 3×2 (ND/RD/ID × deficit/no-deficit); state how RD/partial recovery was classified (if available) and show sensitivity analyses for its impact on Se/Sp/PPV/NPV/AUC.
24	Estimates of diagnostic accuracy and their precision	Report estimates of diagnostic accuracy with measures of precision (for example, 95% confidence intervals), and state how RD and PR were handled in the calculations. If sensitivity analyses were performed to show the impact of alternate classifications, report them. Diagnostic accuracy studies should present quantitative estimates—such as sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and odds ratios—along with measures of statistical precision, including 95% confidence intervals. P-values may be included where appropriate, though effect size and precision are more informative. These statistics enable critical appraisal of the reliability and clinical significance of the IONM modality under study. IONM-specific considerations: In IONM research, where sample sizes may be limited or outcomes infrequent, it's essential to report these measures transparently. If statistical estimates are not provided, authors should clearly explain why (e.g., exploratory study, limited sample). Precision measures are particularly important given the variability in intraoperative responses and outcome definitions. State how reversible deterioration/partial recovery was handled in accuracy calculations (separate category vs collapsed to ±) and provide sensitivity analyses showing its impact on Se/Sp/PPV/NPV and AUC. Report 95% CIs and model/test used for all estimates.
25	Any adverse events from performing the index test or the reference standard	Although IONM techniques are generally considered low risk, adverse events may occur due to electrical stimulation, electrode placement, or equipment malfunction. In IONM, index test-related events include tongue lacerations, dental trauma, thermal burns from electrodes, and inadvertent patient movement during stimulation. Reference standard-related events may involve additional imaging, neurological testing, or interventions postoperatively. Accurate reporting of such adverse events is essential to assess the safety profile of IONM and refine monitoring protocols.
26	Study limitations, including sources of potential bias, statistical uncertainty, and generalizability	Definition of Test Criteria Clearly describe the neurophysiological test criteria (e.g., specific warning criteria used, cut-off thresholds for alarms). Specify timing of outcome assessment (e.g., immediate post-op, days after surgery, or long-term follow-up). Statistical Methods For small samples, justify statistical test selection based on assumptions (e.g., chi-square requires ≥5 expected per cell). For very small samples (<12), avoid inferential statistics; instead use descriptive statistics and acknowledge limitations.

		Correction for Multiple Testing Indicate whether multiple testing correction was performed. Specify method used (e.g., Bonferroni, Holm–Bonferroni). If no correction applied, justify and discuss implications. Adjustment for Confounders Report if multivariable regression or other adjustment methods were used to control for confounders. If not adjusted, acknowledge and include in limitations section. Discuss limitations from anesthetic variability (technique shifts, depth drift, NMB fluctuations) and unrecorded physiologic changes as potential accuracy bias.
27	Implications for practice, including the intended use and clinical role of the index test	The clinical relevance of intraoperative neurophysiological monitoring (IONM) depends on differences in baseline characteristics—such as clinical status and neurophysiological data—between study populations and the patients typically encountered in surgical settings. Key considerations should include: (1) the stage of development of IONM, ranging from proof-of-concept validation to accuracy assessment in specific surgical populations, evaluation in the broader target population, and demonstration of its incremental benefit in routine clinical use; and (2) the additional research required to optimize its application for real-world surgical practice. Emphasizing these aspects clarifies IONM's clinical utility, its role in intraoperative decision-making, and its potential to enhance patient safety and surgical outcomes.
28	Registration number and name of registry	Provide trial/registry name and registration number (or state "not registered") and give the registration URL if available.
29	Where the full study protocol can be accessed	Provide a public link or repository identifier for the full protocol (or state unavailability) and summarize deviations from protocol.
30	Sources of funding and other support; role of funders	Disclose funding sources and other support; describe the role of funders/sponsors in design, conduct, analysis, and reporting.