

BMJ Open Intraoperative evoked potential monitoring for detecting cerebral injury during adult aneurysm clipping surgery: a systematic review and meta-analysis of diagnostic test accuracy

Fang Zhu,^{1,2} Jason Chui,¹ Ian Herrick,¹ Janet Martin^{1,2,3}

To cite: Zhu F, Chui J, Herrick I, *et al.* Intraoperative evoked potential monitoring for detecting cerebral injury during adult aneurysm clipping surgery: a systematic review and meta-analysis of diagnostic test accuracy. *BMJ Open* 2019;**9**:e022810. doi:10.1136/bmjopen-2018-022810

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-022810>).

FZ and JC contributed equally. IH and JM contributed equally.

Received 14 March 2018
Revised 25 November 2018
Accepted 18 December 2018



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Jason Chui;
jason.chui@lhsc.on.ca

ABSTRACT

Objectives We aim to evaluate the diagnostic test accuracy (DTA) of intraoperative evoked potential (EP) monitoring to detect cerebral injury during clipping of cerebral aneurysms.

Design Systematic review.

Data sources Major electronic databases including MEDLINE, EMBASE, LILACS.

Eligibility criteria We included studies that reported the DTA of intraoperative EP monitoring during intracranial aneurysm clipping procedures in adult patients.

Data extraction and synthesis After quality assessment, we performed a meta-analysis using the bivariate random effects model, and calculated the possible range of DTA point estimates using a new best-case/worst-case scenario approach to quantify the impact of rescue intervention on DTA.

Results A total of 35 studies involving 4011 patients were included. The quality of the primary studies was modest and the heterogeneity across studies was high. The pooled sensitivity and specificity for predicting postoperative neurological deficits for the somatosensory evoked potential (SSEP) monitoring was 59% (95% CI: 39% to 76%; I^2 : 76%) and 86% (95% CI: 77% to 92%; I^2 : 94%), for motor evoked potential (MEP) monitoring was 81% (95% CI: 58% to 93%; I^2 : 54%) and 90% (95% CI: 86% to 93%; I^2 : 81%), and for combined SSEP and MEP monitoring was 92% (95% CI: 62% to 100%) and 88% (95% CI: 83% to 93%). The best-case/worst-case range for the pooled point estimates for sensitivity and specificity for SSEP was 50%–63% and 81%–100%, and for MEP was 59%–74% and 93%–100%, and for combined SSEP and MEP was 89%–94% and 83%–100%.

Conclusions Due to the modest quality and high heterogeneity of the existing primary studies, it is not possible to confidently support or refute the diagnostic value of EP monitoring in cerebral aneurysm clipping surgery. However, combined SSEP and MEP appears to provide the best DTA for predicting postoperative stroke. Contrary to popular assertion, the modest sensitivity of SSEP monitoring is not explained by the use of rescue intervention.

PROSPERO registration number CRD42015016884.

Strengths and limitations of this study

- This systematic review provides the most comprehensive evaluation of the quality and limitations of the primary studies investigating the diagnostic test accuracy of all evoked potential monitoring modalities used during clipping of intracranial aneurysm surgery.
- We reported our results according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses diagnostic test accuracy checklist (2018).
- This study is the first to use a best-case/worst-case (conservative/liberal) approach to the analysis of diagnostic test accuracy to quantify the magnitude of change in pooled sensitivity and specificity that were potentially attributable to the use of rescue intervention.
- The results of this study were limited by the modest quality, inherited methodological limitations and high heterogeneity of the existing primary studies.

INTRODUCTION

Intraoperative cerebral injury can result from various surgical and/or anaesthetic causes during surgical clipping of an intracranial aneurysm.¹ Despite advances in surgical and anaesthetic technique, it is estimated that a new postoperative stroke occurs up to 11% of patients undergoing aneurysm clipping.^{1,2} Interest in mitigating the risk of intraoperative cerebral injury associated with aneurysm clipping has led to the deployment of intraoperative evoked potential (EP) monitoring during these procedures. Somatosensory evoked potential monitoring (SSEP) was first used in the mid-1980s³ in the intracranial anterior circulation aneurysm surgery. SSEP is performed by applying an electrical stimulation to a specific mixed nerve or dermatome to generate sensory stimulus and recording the responses along the ascending neural pathway and the sensory cortex. In

the context of aneurysm clipping surgery, a significant increase in latency and reduction in amplitude signifies conduction delay, hence cerebral injury of somatosensory cortex or its ascending pathway. Brain-stem auditory evoked potential monitoring (BAEP) was later used specific for the posterior cerebral circulation aneurysm surgery.^{4,5} BAEP was performed by applying an auditory stimulus in the ear and recording the far-field potentials of neural generators (ie, nucleus) along the auditory pathway. As the auditory pathway (eg, superior olivary complex and lateral lemniscus) are located in brainstem, an abnormal BAEP response signifies injury to brainstem. More recently, the motor evoked potential (MEP) monitoring was introduced for improving the detection of subcortical injury.^{6–8} It is performed by applying an electrical stimulation on the motor cortex and recording the distal muscle responses in the upper and lower limbs. An abnormal MEP response signifies cerebral injury to motor cortex and its descending pathway.

Current evidence pertaining to the accuracy of modalities used for EP monitoring during cerebral aneurysm clipping procedures is conflicting. Although there is a substantial body of literature related to EP monitoring during these procedures, there is a lack of prospective, randomised studies evaluating EP monitoring and its impact on clinically relevant outcome measures. All observational studies addressing the diagnostic test accuracy (DTA) of EP monitoring have fundamental methodological flaws that confound the evaluation of diagnostic performance because the intraoperative identification of EP changes is typically accompanied by a rescue intervention.⁹ While the clinical rationale for rescue intervention is compelling (as the name suggests), the practice leads to misclassification bias in the study results as it is frequently unclear how outcome is altered as a result. None of the primary studies^{4,5,8,10–41} has adequately accounted for this misclassification bias in reporting the diagnostic accuracy of EP monitoring with some investigators choosing to classify effective rescue interventions (eg, associated with improvement or reversal of EP changes) as an averted cerebral injury and other investigators choosing to ignore the effects of rescue interventions. As the magnitude of the impact of this misclassification bias remains unquantified, discrepancies in DTA measurements that arise with the effects of rescue interventions continue to represent a central controversy regarding the role of EP monitoring during these procedures. In addition, there is a lack of published systematic reviews that adequately assess the quality and limitations of the included primary studies to evaluate the DTA and clinical relevance of different EP monitoring modalities in the context of aneurysm clipping procedures.

In view of these methodological concerns, we undertook a rigorous systematic review and meta-analysis to transparently evaluate the shortcomings in the existing literature and to quantify the effect that the use of rescue intervention is likely to have on DTA with the intent of informing future research priorities in this area.

Objectives

The primary objective of this systematic review was to provide an unbiased and complete assessment of the diagnostic value of EP monitoring in aneurysm clipping surgery, based on available evidence from clinical studies, while also relating clinically important outcomes to diagnostic performance. For the following reasons, this systematic review and meta-analysis extends beyond pre-existing reviews^{42,43} on this topic: i) it assesses the quality and limitations of the primary studies, ii) it quantifies the effect of intraoperative rescue interventions on the DTA and iii) it compares the DTA of all EP monitoring modalities for detecting intraoperative cerebral injury that results in an adverse neurological outcome and/or new radiological change that is consistent with stroke following intracranial aneurysm clipping surgery. The secondary objective of this review was to explore potential effect modifiers (eg, subgroups of patients, procedures or combinations of neuromonitoring modalities), which can maximise the DTA and may help to establish evidence-informed standards of appropriateness for clinical application of EP monitoring.

METHODS

Protocol and registration

This systematic review was registered in PROSPERO (International prospective register of systematic review) prior to conducting the review (CRD42015016884).

Eligibility criteria

Types of studies

Prospective and retrospective clinical studies were included if they reported DTA of at least one EP monitoring modality and the number of cerebral injuries (either clinical or radiological) observed. We excluded case-control studies, registry data and data derived from an unpublished hospital database.

Participants

We included studies involving adult patients (>18 years of age) who underwent intracranial aneurysm clipping for ruptured or unruptured cerebral aneurysms under general anaesthesia irrespective of the pathology, location or size of the aneurysm. We excluded studies investigating patients undergoing endovascular coiling procedures or concomitant extracranial-intracranial bypass procedures or procedures involving cardiopulmonary bypass.

Index tests

Intraoperative EP monitoring was defined as the use of SSEP, MEP, BAEP or any combination of these modalities. Other modalities such as electroencephalography, cranial nerve monitoring were not included. The choice of stimulation technique (sites, montages and cut-off values) was not restricted.

Target conditions

The target condition of this review was defined as intraoperative cerebral injury. Intraoperative cerebral injury may

be caused by multiple mechanisms, surgical or anaesthetic-related, ischaemic or non-ischaemic, which can result in different outcomes postoperatively.

Reference standards

The reference standard was the identification of new postoperative clinical neurological deficits or radiological changes indicating clinical or radiological cerebral injury (or stroke), respectively. The diagnosis of cerebral injury (or stroke) is usually established at the initial postoperative physical examination and confirmed with radiological imaging (eg, CT or MRI). If multiple outcome measures at different time points were reported in the primary studies, we used the outcome reported at the earliest time point of measurement as the reference standard in this review. The effect of different timing of outcome assessment on DTA was further examined by meta-regression to explore whether effect sizes differed over time.

Definition of events

An intraoperative rescue intervention is frequently employed when an EP abnormality is detected and prior to postoperative assessment for confirmation of neurological injury. As a consequence, there is a significant risk of outcome misclassification bias in all EP studies that evaluate DTA because neurological outcome may be altered as a result of the rescue intervention. Most studies do not address this potential major source of bias that arises when outcome is assigned (positive or negative) irrespective of whether an intraoperative therapeutic intervention was applied.

To quantify the risk of outcome misclassification bias, this systematic review quantified the treatment effect of rescue intervention in response to EP signal changes on DTA. We calculated the best-case scenario (a more liberal approach that assumed all intraoperative interventions had full beneficial treatment effects)^{44 45} and the worst-case scenario (a more conservative approach that assumed no beneficial treatment effect was produced by intraoperative rescue interventions, ie, by ignoring the potential impact of rescue interventions)) to generate an upper and lower range of plausible estimates of the potential impact of the rescue interventions. Using these two approaches, a range of pooled sensitivity and specificity was calculated that represented the upper and lower range of the pooled point DTA. In this innovative approach, we used data from the subset of studies reporting data regarding the reversibility of detected EP changes to evaluate the best-case and worst-case scenario assuming the impact of rescue intervention on DTA. Details of these two approaches are described in table 1. In essence, the two main differences of these two approaches were:

- ▶ In the best-case scenario, a significant intraoperative EP signal change that is reported to be reversed by an intraoperative rescue intervention and not followed by a new postoperative neurological deficit or radiological change was defined as a true positive. Because all reports of reversal of the EP signal change(s) were assumed to reflect mitigation of injury as a consequence of the intervention.
- ▶ In the worst-case scenario, a false-negative was defined as a reported new postoperative neurological deficit

Table 1 Definitions of positive and negative events in the best-case/worst-case scenario approaches for the analysis of diagnostic test accuracy

		Index test (significant EP changes)		
		Yes		No
		Reversible EP changes	Irreversible EP changes	
Conservative approach (worst-case)	Reference test*	Yes	TP	FN
		No	FP	TN
Liberal approach (best-case)		Yes	FN	TP
		No	TP	FP

*Either new-onset postoperative neurological deficits or radiological changes. The definition(s) for liberal approach calculation are as follows: True-positives were defined as either: (a) A significant intraoperative EP signal change that is irreversible in spite of all rescue intervention(s) employed, and followed by a new-onset postoperative neurological deficit or radiological change. (b) A significant intraoperative EP signal change that is reported to be reversed by an intraoperative rescue intervention and not followed by a new postoperative neurological deficit or radiological change. All reports of reversal of the EP signal change(s) were assumed to reflect mitigation of injury as a consequence of the intervention.

False-positives were defined as no new postoperative neurological deficit or radiological change despite irreversible intraoperative EP signal changes indicating potential neurological injury that were not reversed to baseline by surgical or anaesthetic rescue intervention(s).

True-negatives were defined as no significant intraoperative EP signal changes accompanied by no new postoperative neurological deficits or radiological changes.

False-negatives were defined as a reported new postoperative neurological deficit or radiological change associated with: (a) No significant intraoperative EP signal change. (b) Significant EP signal changes that were reversed to baseline following intraoperative surgical or anaesthetic rescue intervention.

EP, evoked potential; FN, false negative; FP, false positive; TP, true positive, TN, true negative.

or radiological change associated with a significant EP signal changes that were reversed to baseline following intraoperative surgical or anaesthetic rescue intervention.

Information sources and search

A systematic search was performed from 1 January 1960 to 5 January 2016 (last updated on 27 June 2018) in seven major electronic database including MEDLINE, EMBASE, LILACS, IndMed and a variety of other sources including hand searching, snowballing of references lists, conference proceedings and other grey literature databases. The complete list search terms and strategies are summarised in online supplementary appendix 1.

Study selection

Two authors (FZ, JC) independently scanned the titles and abstracts to identify potentially relevant studies. Full-text versions of all relevant articles were retrieved. Two authors (FZ, JC) independently selected the articles that met the predefined inclusion criteria using a standardised study inclusion form (pilot tested by two independent assessors on a sample of three papers). All disagreements were sorted by consensus of two authors (FZ, JC).

Data collection process

Data from each included study were independently extracted and entered by two authors (FZ, JC). All discrepancies were resolved by consensus. Details of the study population, diagnostic test and outcomes were extracted using a standardised electronic data extraction form. The data extraction form was tested and refined by two independent assessors on a sample of three papers. Translations of four articles (one in French, one in Japanese and two in Chinese) were required for data extraction. The translation was performed by two independent translators, who are fluent in the corresponding language and has experiences in performing data extraction for other systematic reviews.

Risk of bias and applicability

Study quality was assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 tool.⁴⁶ Signalling questions were tailored to this review, and were pretested and refined by two independent assessors on a sample of three papers. Two authors (FZ, JC) assessed and graded the quality of the primary studies independently. All disagreements were sorted by consensus of two authors (FZ, JC). All data were tabulated and displayed graphically in the methodological quality summary and study quality graph.

Synthesis of results

A 2×2 cross-table of postoperative neurological deficits versus EP changes was constructed for each study to calculate the corresponding sensitivity and specificity for each study. A meta-analysis using a bivariate random effects model was used to summarise the pooled effect estimates for each EP modality (SSEP, MEP, BAEP) and summary receiver operating characteristic (SROC) curves

were plotted.^{47 48} Area under the ROC curve (AUC) >0.9 was considered to reflect high diagnostic performance.⁴⁸ Positive likelihood ratio (LR+) and negative LR (LR-), diagnostic OR (DOR), positive predictive value and negative predictive value were calculated, and the LR scattergram was plotted. LR+ >10 and LR- <0.1 were considered to be convincing confirmation and exclusion test results, respectively. LR+ >5 and LR- <0.2 were considered to reflect moderate diagnostic test accuracy.^{45 46} A DOR >100 (corresponding to pairing of LR+ of 10 and LR- of 0.1) and >25 (corresponding to pairing of LR+ of 5 and LR- of 0.2) were considered evidence of high and adequate diagnostic test performance, respectively.⁴⁹ The Fagan nomogram was used to illustrate the pretest probability to post-test probability changes in the LR of each EP modality. We assessed heterogeneity using the I² statistic and Galbraith plots. We plotted the standardised log-transformed diagnostic OR against the reciprocal of its SE to look for potential outlying studies.

Sensitivity analysis was performed by sequential removal of outlying studies and/or the most influential studies. Funnel plots were used to explore evidence for potential publication bias and small study effects. For each EP monitoring modality, meta-regression was performed to explore the following potential effect modifiers on DTA parameters: timing of outcome assessment, length of follow-up, ruptured or unruptured aneurysm status, type of anaesthesia, the use of neuroprotection strategy, year of publication, sample size and study design.

Analysis was repeated to calculate the upper and lower range of the DTA in each EP modality (to reflect the potential effect of intraoperative rescue intervention). Ranges for DTA for each modality were calculated to quantify the effect of rescue intervention. All statistical analyses were executed using user written command MIDAS in Stata IC (V.13.1) or RevMan V.5.3 as appropriate.

Patient and public involvement

Patients and public were not involved in this study.

Results

We identified 3095 titles and abstracts from our systematic search in the electronic database, grey literature and hand searching. After exclusion of duplicate works and after screening of titles and abstracts, 156 full-text articles were retrieved for further evaluation. A total of 35 studies^{4 5 8 10–41} reported in 36 publications involving 4011 patients met criteria for inclusion in the meta-analysis (online supplementary appendix 2).

Study and patient characteristics

All 35^{4 5 8 10–41} included studies were observational, with a sample size ranging from 15 to 685, and were performed in 11 countries between 1987 and 2017 (table 2). There were no randomised controlled studies. Most studies (24 studies)^{4 5 8 10 12 14 16–20 24 26 28 32–41} used clinical examination as the outcome reference standard. Five studies used radiological imaging as the reference standard. (CT=4 studies,

Table 2 Study characteristics

Study ID	Study design	Consecutive series of patients	Study period	No. of patients		Country	Follow-up time
				studied (n)	analysed (n)		
Kashkoush <i>et al</i> ¹⁰ 2017	Retro	Yes	2006–2013	267	177	USA	Postoperative within 3 days
Starmann <i>et al</i> ^{11*} 2017	Retro	Yes	January 2010–April 2014	123	133	USA	Postoperative till discharge
Kim <i>et al</i> ¹² 2016	Retro	Yes	January 2013–December 2014	726	685	Korea	Postoperative till discharge
Takebayashi <i>et al</i> ¹⁵ 2014	Pro	NR	January 2006–December 2011	50	50	Japan	Postoperative and at discharge
Yue <i>et al</i> ¹³ 2014	Pro	Yes	January 2009–August 2011	44 (MEP+SSEP)	43	China	Postoperative, at discharge, 3 months and 1 year after discharge (up to 30.3–33.5 months)
Sasaki <i>et al</i> ¹⁴ 2014	Pro	Yes	September 2009–April 2013	270 (177 MEP, 75 SSEP, 33 VEP)	177	Japan	Postoperative
Kang <i>et al</i> ¹⁶ 2013	Unclear	NR	NR	47	45 (SSEP), 37 (MEP)	China	Postoperative day 7
Wicks <i>et al</i> ¹⁸ 2012	Retro	Yes	January 1991–December 2009	663	663	USA	At discharge, and within 2 years after discharge
Maruta <i>et al</i> ¹⁷ 2012	Pro	Yes	NR	22	18	Japan	Postoperative, at discharge
Ni <i>et al</i> ¹⁹ 2012	Pro	Yes	December 2009–February 2011	23	23	China	Postoperative till discharge
Motoyama <i>et al</i> ²² 2011	Retro	NR	April 2007–November 2009	48	48	Japan	Postoperative
Krayenbuhl <i>et al</i> ²⁰ 2011	Pro	NR	December 2005–March 2007	36	36	Switzerland	Postoperative 2 days
Lin <i>et al</i> ²¹ 2011	Retro	Yes	January–August 2010	45	45	China	Postoperative, postoperative 1 week, at discharge
Yeon <i>et al</i> ²³ 2010	Pro	Yes	January 2007–December 2008	164	98	Korea	Postoperative, postoperative 1 day, and at discharge
Irie <i>et al</i> ²⁴ 2010	Retro	No	September 2003–March 2009	111	111	Japan	Postoperative, and on further symptoms
Szelényi <i>et al</i> ²⁵ 2007	Retro and Pro	No	NR	318	108	USA and Germany	Postoperative, at discharge and 6 months after discharge
Penchet <i>et al</i> ^{26*} 2007	Retro	NR	1995–2002	122	122	France	Postoperative within 48 hours, and at 1 month
Szelényi <i>et al</i> ²⁷ 2006	Retro and Pro	Yes	NR	119	113	Germany	Postoperative 1 day, and 1 week
Schick <i>et al</i> ²⁹ 2005	Others	NR	NR	76	76	Germany	Postoperative, 1 week and 2 weeks
Horiuchi <i>et al</i> ²⁸ 2005	Pro	No	NR	53	53	Japan	Postoperative
Neuloh and Schramm ³⁰ 2004	Pro	Yes	NR	95	95 (SSEP); 81 (MEP)	China	Postoperative, at discharge, at postoperative 3 months
Suzuki <i>et al</i> ⁸ 2003	Pro	Yes	January 1997–February 2002	108	108	Japan	Postoperative

Continued

Table 2 Continued

Study ID	Study design	Consecutive series of patients	Study period	No. of patients		Country	Follow-up time
				studied (n)	analysed (n)		
Min <i>et al</i> ³¹ 2001	Pro	NR	NR	87	60	China	Postoperative, and >3 months after surgery
Holland ³² 1998	Pro	NR	1996	45	45	USA	Postoperative
Sako <i>et al</i> ³³ 1998	Unclear	No	NR	15	15	Japan	Postoperative, at 3 months
Sako <i>et al</i> ³⁴ 1995	Pro	Yes	September 1992–October 1993	25	25	China	Postoperative
Manninen <i>et al</i> ^{4*} 1994	Others	No	1985–1991	67	67	Japan	Postoperative with 24 hours, at discharge
Mizoi and Yoshimoto ³⁵ 1993	Pro	Yes	July 1986–June 1989	97	97	Japan	Postoperative
Lazorthes <i>et al</i> ³⁶ 1992	Unclear	NR	March 1991–September 1991	30	30	China	Postoperative within 48 hours, at 6 months after discharge
Matsuda <i>et al</i> ³⁷ 1991	Pro	NR	NR	49	49	Korea	Postoperative 1–2 days
Manninen <i>et al</i> ⁶ 1990	Pro	NR	NR	157	97	Japan	Immediately postoperative, at discharge
Schramm <i>et al</i> ³⁸ 1990	Pro	NR	NR	102	94	USA and Germany	Early, later and late postoperative
Buchthal <i>et al</i> ³⁹ 1988†	Pro	NR	NR	25	25	France	NR
Momma <i>et al</i> ⁴¹ 1987	Pro	NR	NR	40	40	Germany	Postoperative, and postoperative 6 months
Kidooka <i>et al</i> ⁴⁰ * 1987	Pro	NR	NR	28	31	Japan	Postoperative and >2 weeks

*Repeated procedures on the same patient in some studies. The number of procedure was reported.

†Duplicate publication found.

EP, evoked potential; MEP, motor evoked potential; NR, not reported; Pro, prospective cohort; Retro, retrospective cohort; SSEP, somatosensory evoked potential.

MR=1 study).^{21 22 25 27 31} Six studies used both clinical and radiological methods to define outcome.^{11 13 15 23 29 30}

Most study patients were between 48 and 66 years of age and represented a mixed population of ruptured and unruptured cerebral aneurysms (Hunt and Hess grades 0–5). Of the 35 included studies, 18 studies reported only on patients with anterior circulation aneurysms, and 14 studies reported on patients with both anterior and posterior circulation aneurysms (online supplementary table 1).

Evoked potential monitoring and anaesthetic characteristics

Most studies (16 studies) investigated SSEP monitoring,^{5 18 20 26 29 31–41} 9 studies investigated MEP monitoring,^{8 12 14 15 17 22–24 27} and 9 studies investigated SSEP and MEP alone and combined.^{10 11 13 16 19 21 25 28 30} Only one study investigated the combined use of SSEP and BAEP⁴ (online supplementary table 2). In the primary studies, the use of upper limb and/or lower limb SSEP and MEP monitoring was mainly determined by the location of the aneurysm. Diagnostic criteria for SSEP generally included both amplitude reduction and conduction delay. In studies that used MEP monitoring, both transcranial (tc) and direct cortical (dc) MEPs were investigated, and most studies (13 of 17 studies) used amplitude reduction as the primary diagnostic criterion. All studies reported high rates of successful recording; 69%–100% for SSEP and 60%–100% for MEP. None of the studies specified the use of certified neurophysiologists for EP monitoring.

Most recent studies used a total intravenous anaesthetic technique, particularly when investigating MEP monitoring, and avoided the use of nitrous oxide (online supplementary table 2). In comparison, earlier trials typically combined the use of inhaled anaesthetics and nitrous oxide. Despite the well-known anaesthetic-induced suppression of EP monitoring signals, five studies did not report the anaesthetic regimen used.

Summary of events

Despite the use of EP monitoring and timely intervention, 8.3% (302 of 3621 patients) and 11.2% (105 of 941 patients) were reported to have new postoperative neurological deficits (clinical stroke) and radiological features of stroke, respectively. Neuroprotection strategies, including pharmacological burst suppression and passive hypothermia, were employed routinely in some studies (table 3). In cases with intraoperative EP signal changes, most studies reported that surgical and/or anaesthetic interventions were applied with the intention to mitigate possible cerebral injury. Reported manoeuvres included release of temporary clips, repositioning of permanent clips or brain retractor(s) and augmentation of blood pressure to improve cerebral perfusion pressure.

Risk of bias and applicability

The methodological quality of the studies was of concern (online supplementary figure 1–2). Almost half of the primary studies had significant deficiencies in reported

information for all quality domains. The most common domains with reporting deficiencies related to the recruitment process, inclusion and exclusion criteria as well as execution of reference standards (ie, evaluation of neurological outcome). The timing of application of reference standards was variable (from immediately after surgery to 1 year postoperatively) and was also generally poorly reported across studies.

Diagnostic test accuracy of SSEP

Twenty-two studies^{4 5 10 11 16 18 20 26 28–41} reported the DTA of SSEP monitoring; 17 studies^{4 5 10 11 16 18 28 30 31 33–39 41} (1765 patients) assessed the DTA for predicting new postoperative neurological deficits (clinical stroke) and 6 studies^{11 18 20 26 28 29} (474 patients) assessed the DTA for predicting postoperative radiological evidence of stroke (radiological stroke). The pooled sensitivity and specificity of SSEP alone for predicting postoperative stroke was 59% (95% CI: 39% to 76%; I²: 76%) and 86% (95% CI: 77% to 92%; I²: 94%), respectively (figure 1). The LR+ was 4.17 (95% CI: 2.26 to 7.68) and LR– was 0.48 (95% CI: 0.30 to 0.78) (table 4). The AUC in the SROC curve was 0.83 (95% CI: 0.79 to 0.86) (online supplementary figure 3). The pooled sensitivity for predicting radiological stroke was 55% (95% CI: 0.39% to 0.7%; I²: 0%) and the pooled specificity was 89% (95% CI: 84% to 93%; I²: 58%). However, this was based on only six studies reporting on radiological stroke (table 4).

Sensitivity analysis and heterogeneity of SSEP

Analysing the sensitivity results for SSEP, the DTA is only slightly changed after sequential removal of the two most influential studies^{34 35}; the pooled sensitivity and specificity slightly decreased to 58% (95% CI: 43% to 72%) and 85% (95% CI: 78% to 90%) (online supplementary table 3) and the LR+ dropped slightly to 3.93 (95% CI: 2.93 to 5.27) and the LR– increased to 0.49 (95% CI: 0.36 to 0.66). The DOR decreased slightly from 8.66 to 8.01. There was no publication bias or small study effect (p=0.82) identified. In the meta-regression, the use of inhaled anaesthetics was identified to significantly affect the pooled DTA of SSEP in predicting postoperative neurological deficits (p<0.001) compared with combined inhaled and intravenous anaesthetics. Of note, publication year, sample size, location of aneurysm, mean age, ruptured or unruptured aneurysm status, the choice of SSEP diagnostic criteria, the use of nitrous oxide and timing of outcome assessment did not change the DTA of SSEP.

Diagnostic test accuracy of MEP

Fourteen studies investigated the DTA of MEP monitoring; for predicting postoperative neurological injury in 14 studies^{8 11 12 14–17 22 24 25 27 28 30} (1762 patients), and for predicting postoperative radiological stroke in 9 studies^{8 11 13 15 16 23 24 27 28} (740 patients). The pooled sensitivity of MEP was 81% (95% CI: 58% to 93%; I²: 54%) (higher than SSEP) and the pooled specificity was 90%

Table 3 Summary of events

Study ID	No. of patients requiring temporary clipping, n (%)	Mean duration of temporary clipping (min)	Neuroprotection strategy	Intraoperative manoeuvres to treat cerebral ischaemia	No. of cases with clinical neurological deficit, n (%)	No. of radiographic strokes, n (%)	Total no. of cases with EP changes, n (%)*	No. of cases with irreversible changes, n (%)*	No. of cases with reversible changes, n (%)*	No. of cases with partial changes, n (%)*
Kashkoush 2017 <i>et al</i> ¹⁰	177	11.7	NR	NR	10 (6)	NR	29 (16)	NR	NR	NR
Staarmann 2017 <i>et al</i> ^{11,†}	123	NR	PBS (propofol)	NR	3 (2)	2 (2)	15 (11)	3 (20)	12 (80)	0 (0)
Kim 2016 <i>et al</i> ¹²	NR	NR	NR	Hyperventilation	36 (5)	NR	43 (6.3)	13 (30)	30 (70)	Included in irreversible cases.
Takebayashi 2014 <i>et al</i> ¹⁵	NR	NR	NR	NR	10 (20)	17 (34)	19 (38)	4 (21)	15 (79)	NR
Yue 2014 <i>et al</i> ¹³	NR	NR	NR	ROSC, RR	8 (20)	7 (16)	15 (17)	1 (7)	14 (93)	NR
Sasaki 2014 <i>et al</i> ¹⁴	NR	NR	NR	ROSC	6 (3)	NR	22 (8)	2 (9)	20 (91)	NR
Kang 2013 <i>et al</i> ¹⁶	NR	NR	20% mannitol	NR	SSEP: 3 (8) MEP: 5 (11)	NR	SSEP: 12 (27); MEP: 8 (19)	NR	NR	NR
Wicks 2012 <i>et al</i> ¹⁸	NR	NR	PBS (propofol), PH	ROSC	2 (11)	NR	45 (7)	18 (40)	27 (60)	0 (0)
Maruta 2012 <i>et al</i> ¹⁷	NR	NR	NR	NR	6 (26)	NR	5 (23)	NR	NR	NR
Ni 2012 <i>et al</i> ¹⁹	NR	12.4±4.7	PBS	ROSC	1 (2)	NR	6 (26)	1 (17)	5 (83)	NR
Motoyama 2011 <i>et al</i> ²²	15 (31)	4.6	NR	NR	NR	5 (10)	5 (10)	1 (20)	4 (80)	NR
Krayenbuhl 2011 <i>et al</i> ²⁰	NR	5.8	NR	ROSC	3 (7)	NR	5 (14)	NR	NR	NR
Lin 2011 <i>et al</i> ²¹	16 (36)	5.3±3.3	NR	Others	NR	8 (8)	12 (27)	0 (0)	10 (80)	2 (20)
Yeon 2010 <i>et al</i> ²³	98 (60)	NR	PBS (thiopentone)	NR	3 (3)	7 (7)	12 (12)	NR	NR	NR
Irie 2010 <i>et al</i> ²⁴	NR	NR	NR	NR	dc-MEP: 4 (4) tc-MEP: 5 (5)	NR	6 (6)	1 (17)	5 (83)	NR
Szelényi 2007 <i>et al</i> ²⁵	NR	NR	NR	NR	NR	15 (11)	14 (12)	9 (64)	5 (36)	NR
Penchet 2007 <i>et al</i> ^{26,†}	118 (97)	3m25s	NR	ROSC	17 (15)	NR	34 (26)	2 (6)	25 (74)	7 (21)
Szelényi 2006 <i>et al</i> ²⁷	71 (60)	5–11.8	NR	ROSC	NR	28 (35)	14 (12)	4 (27)	10 (73)	NR
Schick 2005 <i>et al</i> ²⁹	37 (49)	8.1	PBS (methohexital), PH	ROSC, RR, ABP	SSEP: 4 (8) MEP: 4 (7)	SSEP: 2 (4) MEP: 5 (9)	19 (25)	2 (11)	6 (32)	11 (58)
Horiuchi 2005 <i>et al</i> ²⁸	NR	NR	NR	NR	MEP: 11 (12) SSEP: 12 (13)	NR	10 (19)	1 (10)	9 (90)	0 (0)
Neuloh and Schramm 2004 ³⁰	NR	NR	NR	Others	5 (5)	2 (2)	SSEP: 7 MEP: 21	NR	NR	NR
Suzuki 2003 <i>et al</i> ⁸	NR	NR	NR	NR	9 (15)	NR	20 (19)	1 (5)	19 (95)	0 (0)
Min 2001 <i>et al</i> ³¹	NR	10.2±5.7	PBS (thiopentone), BPA	ABP	NR	7 (16)	14 (16)	NR	NR	NR
Holland 1998 ³²	27 (60)	NR	PBS (propofol)	NR	3 (20)	NR	8 (18)	2 (25)	6 (75)	0 (0)
Sako 1998 <i>et al</i> ³³	15 (100)	17.2	20% Mannitol	NR	2 (8)	NR	9 (60)	0 (0)	9 (100)	NR
Sako 1995 <i>et al</i> ³⁴	NR	NR	20% Mannitol	NR	19 (24)	NR	12 (48)	2 (17)	10 (83)	NR
Manninen 1994 <i>et al</i> ^{4,†}	NR	NR	NR	ROSC, RR	8 (8)	NR	SSEP: 14 (20) BAEP: 10 (14)	SSEP: 3 (21) BAEP: 2 (20)	SSEP: 11 (79) BAEP: 8 (80)	NR
Mizoi and Yoshimoto ³⁵ 1993	97 (100)	17.9	NR	NR	10 (33)	NR	42 (43)	3 (7)	39 (93)	NR

Continued

Table 3 Continued

Study ID	No. of patients requiring temporary clipping, n (%)	Mean duration of temporary clipping (min)	Neuroprotection strategy	Intraoperative manoeuvres to treat cerebral ischaemia	No. of cases with clinical neurological deficit, n (%)	No. of radiographic strokes, n (%)	Total no. of cases with EP changes, n (%)	No. of cases with irreversible changes, n (%)	No. of cases with reversible changes, n (%)	No. of cases with partial changes, n (%)
Lazorthes 1992 <i>et al</i> ³⁶	18 (60)	2–39	NR	NR	12 (22)	NR	20 (67)	9 (45)	11 (55)	NR
Matsuda 1991 <i>et al</i> ³⁷	NR	NR	NR	NR	23 (24)	NR	15 (27)	NR	NR	NR
Manninen 1990 ⁵	97 (62)	NR	NR	NR	20 (35)	NR	23 (15)	8 (35)	15 (65)	0 (0)
Schramm 1990 <i>et al</i> ³⁸ †	22 (22)	NR	BPA	Others	4 (16)	NR	16 (18)	1 (12)	15 (88)	0 (0)
Buchthal 1988 ³⁹	25 (100)	6.3–52	PH	NR	12 (30)	NR	9 (36)	NR	NR	NR
Momma 1987 ⁴¹	40 (100)	1–69	NR	NR	12 (43)	NR	21 (53)	21 (100)	0 (0)	0 (0)
Kidooka 1987 <i>et al</i> ⁴⁰ ‡	15 (54)	NR	PH	NR	10 (6)	NR	13 (46)	NR	NR	NR

*The total numbers of cases with EP changes are based on the numbers of cases (not events) reported in the primary studies. The total numbers include all intraoperative EP monitoring modalities used in the primary study (if data are available); however, this may vary across different studies.

†Repeated procedures on the same patient in some studies. The number of procedure was reported.

‡One patient with unreliable SSEP was excluded in the analysis.

BAEP, brain-stem auditory evoked potential monitoring; BPA, blood pressure augmentation; dc-MEP, direct cortical-motor evoked potential; EP, evoked potential; MEP, motor evoked potential; NR, not reported; PBS, pharmacological burst suppression; PH, passive hypothermia; RR, retractor repositioning; ROSC, removal of surgical clip; SSEP, somatosensory evoked potential; tc-MEP, transcranial-motor evoked potential.

(95% CI: 86% to 93%; I²: 81%) (figure 2). The DOR of MEP was 38.1 (95% CI: 13.04 to 111.4); the LR+ was 8.19 (95% CI: 5.78 to 11.6) and LR- was 0.21 (95% CI: 0.09 to 0.52); the AUC in the SROC curve was 0.93 (95% CI: 0.90 to 0.95) (online supplementary figure 4), indicating moderate diagnostic accuracy and high overall diagnostic performance. The false negative rate of MEP (2.2%) was lower than the false negative rate of SSEP (6.3%). On subgroup analysis, dc-MEP (98%; 95% CI: 10% to 100%; I²: 63%) achieved higher pooled sensitivity than tc-MEP (58%; 95% CI: 44% to 71%; I²: 0%). Pool specificities were similar for tc-MEP and dc-MEP, which were 92% (95% CI: 87% to 95%; I²: 84%) and 87% (95% CI: 82% to 92%; I²: 9%), respectively. In contrast to SSEP, MEP was found to have a lower pooled sensitivity of 63% (95% CI: 51% to 74%; I²=0%) for predicting radiological evidence of neurological injury; however, there were only nine studies reporting radiological outcomes. The pooled specificity was similar at 90% (95% CI: 81% to 95%; I²=82%).

Sensitivity analysis and heterogeneity of MEP

Sensitivity analysis did not identify a significant change after removing the outlying study (online supplementary table 3). There was no publication bias. In the univariate meta-regression, year of publication, sample size, mean age, location of aneurysm, anaesthetic regimens, use of nitrous oxide, the time of outcome assessment, ruptured or unruptured aneurysm status were not found to have an interaction with the DTA of MEP monitoring. The use of muscle relaxant and location of aneurysm were identified to significantly affect the pooled DTA of MEP in predicting postoperative neurological deficits (p<0.01).

Diagnostic test accuracy of combined SSEP and MEP

Only three studies of 201 patients^{11 19 21} assessed the DTA of combined SSEP and MEP to diagnose postoperative neurological deficits. Based on this limited literature reporting the combined use of SSEP and MEP monitoring, the pooled sensitivity was 92% (95% CI: 62% to 100%), the specificity was 88% (95% CI: 83% to 93%) and the pooled DOR was 83.5 (table 4). However, this results was only based on 3 studies with 201 patients.

Diagnostic accuracy of BAEP and combined SSEP and BAEP

A meta-analysis was not conducted on BAEP data as only on one study⁴ reported the DTA of BAEP alone, or combined SSEP and BAEP during surgery for clipping posterior circulation aneurysms. This single study⁴ reported sensitivity and specificity of BAEP alone was 42% (95% CI: 20% to 67%) and 89% (95% CI: 78% to 95%), respectively, whereas the combined SSEP and BAEP were 84% (95% CI: 60% to 97%) and 79% (95% CI: 66% to 88%), respectively (table 4).

Impact of rescue intervention on DTA

The preceding calculations of DTA used a worst-case scenario (conservative) approach that ignored any potential beneficial effect of rescue interventions. As not all studies reported data regarding the reversibility

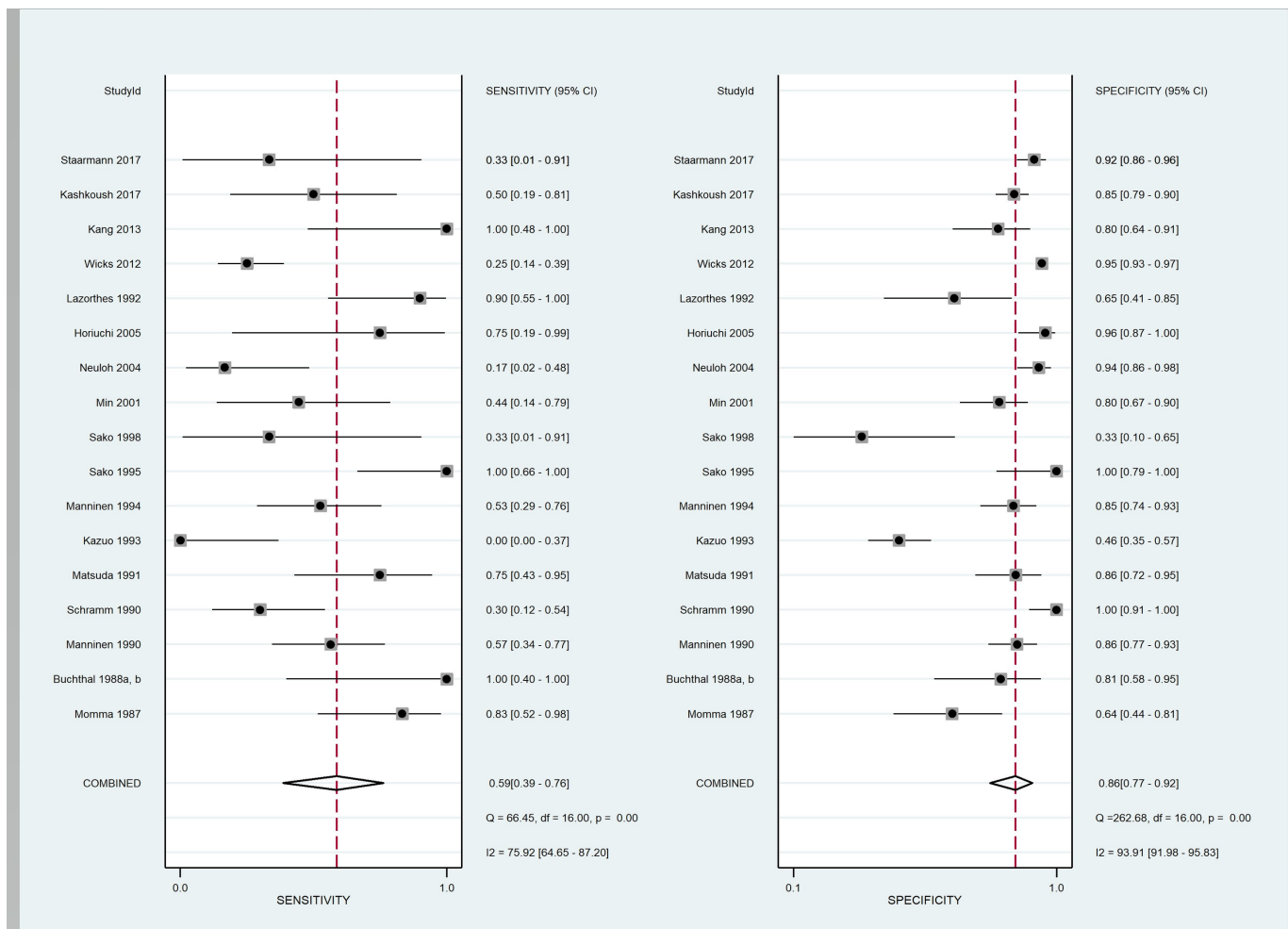


Figure 1 Forest plot of somatosensory evoked potential for predicting postoperative neurological deficit.

of detected EP changes, analysis of the impact of rescue intervention on DTA was conducted on a subset of studies that reported reversible signal changes (10 studies (1262 patients) for SSEP; 5 studies (1152 patients) for MEP) (table 5).

Using this methodology, the pooled sensitivity and specificity of SSEP using a best-case scenario (liberal) approach that included the potential impact of rescue intervention were 63% and 100%, respectively. The pooled sensitivity and specificity of SSEP using a worst-case scenario (conservative) approach that excluded the potential impact of rescue intervention were 50% and 81%, respectively. Thus, the range of pooled sensitivity and specificity for SSEP for predicting postoperative neurological deficits was 50%–63% and 81%–100%, respectively (ie, worst-case and best-case scenario of DTA values that reflect the inclusion or exclusion of the potential impact of rescue intervention on outcome). These results indicate that the low sensitivity observed with SSEP is not attributed to the effect of rescue intervention on misclassification because the pooled sensitivity of SSEP was, at best, only 63%.

Similarly, the pooled sensitivity and specificity of MEP using a best-case (liberal) approach were 74% and 100%,

respectively. The pooled sensitivity and specificity of MEP based on a worst-case scenario (conservative) approach (with exclusion of the potential impact of rescue intervention) were 59% and 93%, respectively. The range of pooled sensitivity and specificity for MEP monitoring was 59%–74% and 93%–100%, respectively, for predicting postoperative neurological deficits.

The range of pooled sensitivity and specificity for combined SSEP and MEP monitoring was 89%–94% and 83%–100%, respectively, for predicting postoperative neurological deficits.

DISCUSSION

Principal findings

This meta-analysis provides the most comprehensive systematic review of the current literature related to the use of EP monitoring during intracranial aneurysm surgery and adds several new insights to the current EP monitoring literature. First, we identified that the quality of the primary studies was modest and marked heterogeneity was encountered. Due to these uncertainties in the existing evidence base, it is not possible to confidently support or refute the diagnostic value of EP monitoring

Table 4 Summary of results

Test	n	TP	FP	FN	TN	Sensitivity	Specificity	LR+	LR-	PPV	NPV
Postoperative neurological deficits (clinical stroke)											
SSEP	17	104	193	111	1368	0.59 (0.39 to 0.76)	0.86 (0.77 to 0.92)	4.17 (2.26 to 7.68)	0.48 (0.3 to 0.78)	0.35	0.92
MEP	14	75	131	38	1518	0.81 (0.58 to 0.93)	0.9 (0.86 to 0.93)	8.19 (5.78 to 11.6)	0.21 (0.09 to 0.52)	0.36	0.98
tc-MEP	9	39	84	32	1222	0.58 (0.44 to 0.71)	0.92 (0.87 to 0.95)	6.81 (4.37 to 10.6)	0.46 (0.34 to 0.63)	0.32	0.97
dc-MEP	5	22	45	2	294	0.98 (0.1 to 1)	0.87 (0.82 to 0.9)	7.3 (5.4 to 9.87)	0.02 (0 to 8.78)	0.33	0.99
BAEP	1	8	7	11	54	0.42 (0.20 to 0.67)	0.89 (0.78 to 0.95)	3.67	0.65	0.53	0.83
SSEP+MEP	3	11	22	1	167	0.92 (0.62 to 1.00)	0.88 (0.83 to 0.93)	7.88 (5.13 to 12.08)	0.09 (0.01 to 0.62)	0.33	0.99
SSEP+BAEP	1	16	13	3	48	0.84 (0.60 to 0.97)	0.79 (0.66 to 0.88)	3.95	0.2	0.55	0.94
Postoperative radiological features of stroke											
SSEP	6	34	51	25	383	0.55 (0.39 to 0.7)	0.89 (0.84 to 0.93)	5.1 (3.33 to 7.8)	0.51 (0.36 to 0.72)	0.4	0.94
MEP	9	43	71	25	601	0.63 (0.51 to 0.74)	0.9 (0.81 to 0.95)	6.23 (3.23 to 12.05)	0.41 (0.3 to 0.56)	0.38	0.96

BAEP, brainstem auditory evoked potential; dc-MEP, direct cortical-motor evoked potential; FP, false positive; FN, false negative; LR+, positive likelihood ratio; LR-, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; SSEP, somatosensory evoked potential; tc-MEP, transcortical-motor evoked potential; TP, true positive; TN, true negative.

in anterior circulation aneurysm clipping surgery. As such, caution is advocated when deploying EP monitoring mainly to predict intraoperative conditions associated with neurological injury.

Second, this meta-analysis is the first to address the longstanding controversy surrounding the potential impact of rescue interventions on outcome using an best-case/worst-case analysis of DTA that adjusted the classification of intraoperative events based on the response to intervention. This approach quantified the magnitude of change in pooled sensitivity and specificity potentially attributable to the rescue intervention. For SSEP monitoring, the possible ranges of pooled sensitivity and specificity were 50%–63% and 81%–100%, respectively. For MEP monitoring, the possible ranges of pooled sensitivity and specificity for MEP monitoring were 59%–74% and 93%–100%. These results indicate that the modest sensitivity observed with SSEP or MEP alone monitoring are not attributed to the effect of rescue intervention on misclassification.

Third, based on current evidence our meta-analysis indicates that combined SSEP and MEP monitoring may have superior DTA compared with other EP modalities used alone for predicting neurological injury during craniotomy for cerebral aneurysm clipping. The DOR was 83.5, indicating moderate-to-high diagnostic performance. This finding supports the general opinion that multimodality EP monitoring improves diagnostic accuracy compared with use of single EP modality monitoring.^{6–8} It must be emphasised however, that this result is based on a small number of studies. The DTA for each EP monitoring modality was higher for predicting postoperative neurological deficits than for predicting postoperative radiological stroke, although the number of studies using a radiological end point was small.

The results of this meta-analysis are focused on the diagnostic accuracy; however, it must be recognised that the choice of EP modality is not exclusively a function of test performance. For example, although our results suggest that MEP monitoring alone has superior DTA compared with SSEP monitoring alone, other relevant clinical factors such as the intermittent nature of MEP monitoring, challenges associated with obtaining satisfactory MEP signals, stimulated movement during intracranial microvascular surgery, restrictions on anaesthetic choices (particularly the use of inhaled anaesthetics and muscle relaxants) and the potential relevance of other information that can be obtained from SSEP monitoring (eg, brachial plexus or ulnar nerve injury due to positioning) remain relevant factors in the clinical choice of monitoring modality.

Furthermore, specificity is relatively high for all modalities and the absence of an EP signal change is associated with a high probability that the patient will not awaken with a new neurological deficit. Many operative teams find value in the reassurance afforded by the absence of EP signal change. However, sensitivity is modest at best for most of these modalities, particularly SSEP monitoring alone. Contrary to conventional opinion, our

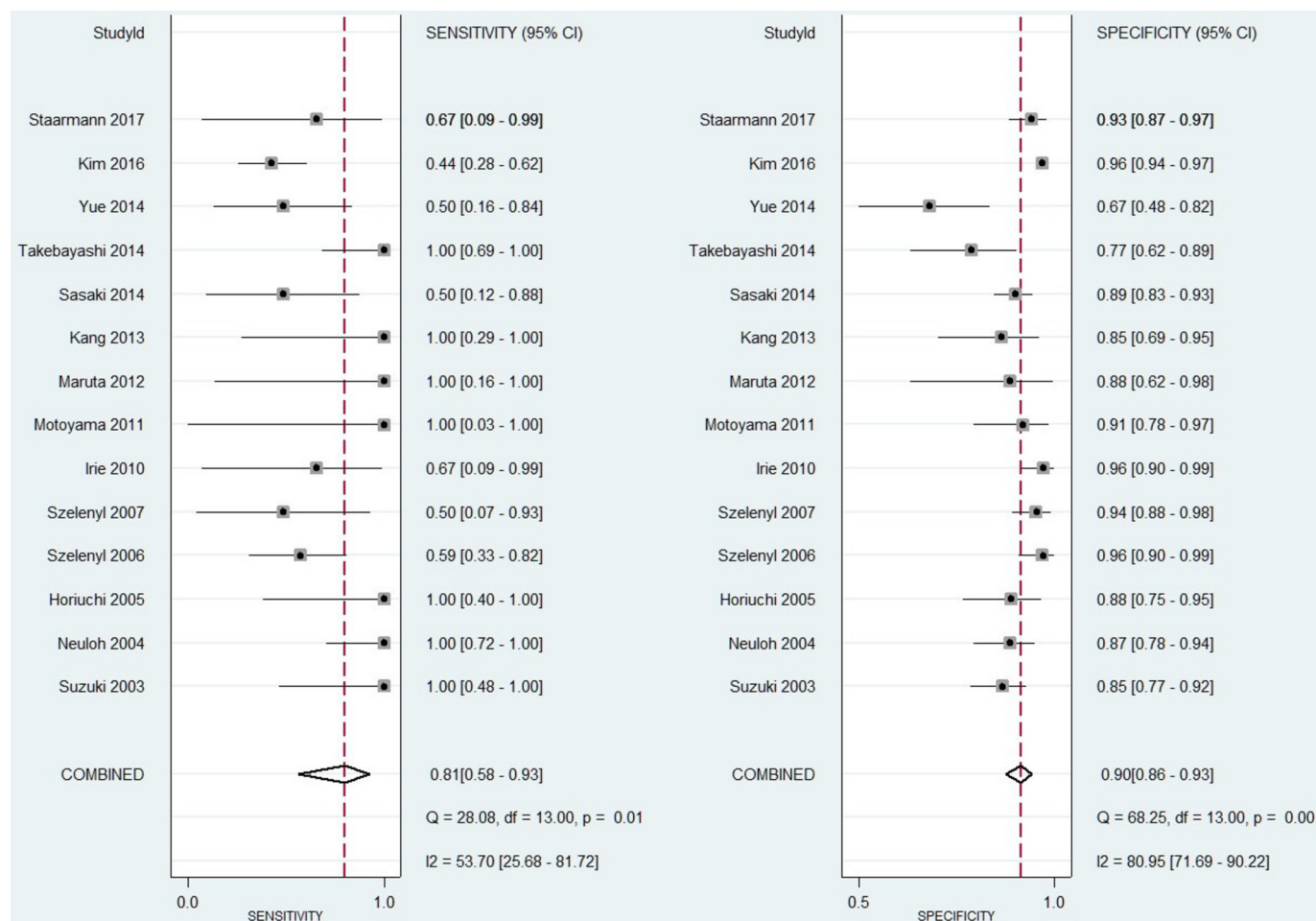


Figure 2 Forest plot of motor evoked potential for predicting postoperative neurological deficit.

results suggest that this lower sensitivity is not attributed to the effect of rescue intervention on misclassification. As a consequence, this modest sensitivity may be relevant when EP monitoring is used during aneurysm surgery to identify cerebral injury and/or guide intervention particularly when the triggering event is not readily identified.

A recently published systematic review⁴² examined SSEP monitoring used alone during aneurysm surgery. This review, based on only 13 studies, reported similar pooled sensitivity (56.8%) and specificity (84.5%) and a DOR of 7.8. The authors concluded that patients with neurological deficits following aneurysm surgery are seven times more likely to have developed an

intraoperative change in SSEP, and that SSEP is highly specific for predicting impending stroke. The low sensitivity observed was attributed to rescue intervention. Notwithstanding these conclusions, it needs to be noted that a DOR of 7.8 reflects inadequate test performance since DOR >100 and >25 are reported to reflect high and adequate diagnostic test performance, respectively.⁴⁹ Although the authors speculate, like many other EP monitoring studies⁵⁰ that the low sensitivity reported with SSEP monitoring reflects successful rescue intervention, our results suggest that rescue intervention does not substantively alter the pooled sensitivity for SSEP monitoring. Furthermore, the previous review⁴² did not

Table 5 Summary of estimated ranges of DTA in predicting postoperative neurological deficit using a best-case scenario and worst-case scenario approach

Evoked potential monitoring	No. of study	No. of patients	Best-case scenario DTA estimates (% (95% CI))		Worst-case scenario DTA estimates (% (95% CI))	
			Sn	Sp	Sn	Sp
SSEP	10	1262	63 (49 to 75)	100 (91 to 100)	50 (30 to 69)	81 (66 to 90)
MEP	5	1152	74 (55 to 87)	100 (90 to 100)	59 (41 to 76)	93 (89 to 95)
SSEP and MEP	3	201	94 (80 to 99)	100 (98 to 100)	89 (52 to 100)	83 (71 to 92)

DTA, diagnostic test accuracy; MEP, motor evoked potential; Sn, sensitivity; Sp, specificity; SSEP, somatosensory evoked potential.

account for the modest quality of the primary studies or heterogeneity across the studies.

The other recently published systematic review⁴³ examined the DTA of SSEP or MEP monitoring used alone during aneurysm surgery. This review, examining only two EP modalities, reported that MEP monitoring has a higher DTA than SSEP monitoring alone in predicting postoperative neurological deficits (consistent with our findings). However, the review was based on a small subset of studies (eight for SSEP and five for both tc-MEP and dc-MEP) and did not account for the quality of primary studies or heterogeneity across the studies.

Limitations

An important weakness of this systematic review is the potential bias due to the modest quality and incomplete reporting of the primary studies. The most significant source of bias is the execution of the reference standard to identify patients with clinical and radiological strokes. There was no consistent method used to assess outcome across the primary studies. In addition, most of the included studies reported incomplete results and/or inadequate methodological detail (online supplementary figure 1-2). However, the DTA estimates for all EP modalities in this analysis were not significantly changed when selected indicators of quality were accounted for in sensitivity analyses.

Another important weakness is the paucity of direct comparisons of EP monitoring modalities within the individual studies, necessitating indirect comparisons across studies within the meta-analysis through bivariate analysis. While this is common for DTA meta-analyses, and reflects the inherent paucity of direct comparisons in the field of diagnostics, it still requires readers to apply cautious interpretation of comparative DTA estimates from meta-analysis since indirect comparisons across studies are likely to be further confounded by differences in a combination of identifiable and unidentifiable factors (such as differences in baseline patient characteristics and co-interventions across studies, etc). Nevertheless, despite the limitations, it is crucial to emphasise that DTA estimates derived from rigorous systematic review and meta-analyses are considered the best possible level of evidence and are considered superior to relying on incomplete assessment of selected studies without aggregation through meta-analysis.

High heterogeneities were encountered during the analysis of DTA of SSEP and MEP. Unfortunately, we were unable to fully account for the high heterogeneity. The DTA estimates did not significantly change in sensitivity analyses attempting to explore these potential moderators. Exploration of funnel plots did not suggest a small study effect or detectable evidence of publication bias. The sources of high heterogeneity are very likely multifactorial and resulted from both clinical (eg, different evoked potential monitoring set-up, diagnostic criteria, different co-interventions or thresholds for rescue interventions, different

risk factors across patient groups) and methodological heterogeneity (eg, time of outcome assessment, reference standards, loss to follow-up) between the primary studies.

Implications

Our systematic review suggests that, based on the current literature, the available data are insufficient to either definitively support or refute the diagnostic value of EP monitoring in anterior circulation aneurysm clipping surgery. Despite three decades of using and investigating EP monitoring in this clinical context, this inadequacy remains a reflection of the modest quality of primary studies, non-homogeneity among studies and underlying methodological flaws in study design. As a consequence of this gap in the literature base, fundamental questions regarding the value of EP monitoring for the detection of intraoperative cerebral injury during these procedures remain inadequately addressed; the development of evidence-based guidelines for the deployment of these modalities in the context of cerebral aneurysm surgery remains elusive and decision support in relation to the comparative efficacy of rescue interventions, particularly in circumstances when a triggering event is not immediately apparent, remains undefined.

Appropriately designed prospective studies will be required to evaluate the impact of EP-guided intervention before evidence-based guidance can be provided. In the context of the current literature base, controversy surrounds ethical concerns regarding the randomisation of access to EP monitoring or failure to intervene if an EP change is detected.⁹ Methodological options to address these concerns have been advanced including the use of pragmatic active-controlled head-to-head study designs or clustered randomisation⁵¹ in centres where EP monitoring is not a routine practice or a large-scale observational study with propensity matching⁹ to compare EP monitored versus non-monitored patients with the benefit of mitigating ethical concerns but at the expense of introducing potential biases from unknown/neglected confounders. A third option that has not yet been adequately addressed would be to conduct a large-scale pragmatic factorial trial where all patients are randomised to active monitoring: SSEP or MEP or SSEP+MEP or BAEP. This would allow for superiority among modalities to be explored. If no differences were found after testing in an adequately powered pragmatic factorial trial, then there may be room for discussion about whether the modalities are equally efficacious or equally non-efficacious, before subsequent randomised trials are conducted to evaluate remaining equipoise.

CONCLUSION

This extensive systematic review and meta-analysis explored the current literature pertaining to the use of EP monitoring to predict the risk of stroke during intracranial aneurysm surgery and introduced a new methodological

approach to evaluate the impact of rescue intervention on DTA. Despite an extensive literature base, modest quality and high heterogeneity across studies is reported to hinder evidence-based decision making and present opportunities for further research in this area. Based on the limited evidence available, combined SSEP and MEP monitoring appears to provide the best DTA for predicting postoperative stroke compared with SSEP or MEP alone. Using a new best-case/worst-case scenario approach, we found the modest sensitivity of SSEP monitoring alone is not explained by misclassification bias introduced by the use of rescue intervention during these procedures. We suggest that future studies might consider using this new methodology to address a prominent methodological controversy that continues to hinder the interpretation of DTA for intraoperative EP monitoring.

Author affiliations

¹Department of Anesthesia and Perioperative Medicine, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada

²Centre for Medical Evidence Decision Integrity and Clinical Impact (MEDICI), Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada

³Department of Epidemiology and Biostatistics, Western University, London, Ontario, Canada

Acknowledgements The authors would like to thank Brie McConnell, Medical Librarian, for her assistance in preparing the search strategy for this systematic review. The authors would also like to acknowledge Maurico Giraldo and Rizq Alamri, neuroanaesthesia fellows, for their help in validating the study inclusion form and data collection form.

Contributors All authors have contributed significantly to the design, conduct, analysis and reporting of this systematic review. FZ and JC contributed to the data collection and screening of publications. All authors gave final approval for submission.

Funding The Evidence-Based Perioperative Clinical Outcomes Research (EPICOR) centre, Western University, Canada and The Centre for Medical Evidence, Decision Integrity and Clinical Impact (MEDICI), Western University, provided statistical and technical support for this systematic review.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- Wiebers DO, Whisnant JP, Huston J, *et al*. Unruptured intracranial aneurysms: natural history, clinical outcome, and risks of surgical and endovascular treatment. *Lancet* 2003;362:103–10.
- Thompson BG, Brown RD, Amin-Hanjani S, *et al*. Guidelines for the management of patients with unruptured intracranial aneurysms. *Stroke* 2015;46:2368–400.
- Symon L, Wang AD, Costa e Silva IE, *et al*. Perioperative use of somatosensory evoked responses in aneurysm surgery. *J Neurosurg* 1984;60:269–75.
- Manninen PH, Patterson S, Lam AM, *et al*. Evoked potential monitoring during posterior fossa aneurysm surgery: a comparison of two modalities. *Can J Anaesth* 1994;41:92–7.
- Manninen PH, Lam AM, Nantau WE. Monitoring of somatosensory evoked potentials during temporary arterial occlusion in cerebral aneurysm surgery. *J Neurosurg Anesthesiol* 1990;2:97–104.
- Guo L, Gelb AW. The use of motor evoked potential monitoring during cerebral aneurysm surgery to predict pure motor deficits due to subcortical ischemia. *Clin Neurophysiol* 2011;122:648–55.
- Quiñones-Hinojosa A, Alam M, Lyon R, *et al*. Transcranial motor evoked potentials during basilar artery aneurysm surgery: technique application for 30 consecutive patients. *Neurosurgery* 2004;54:916–24.
- Suzuki K, Kodama N, Sasaki T, *et al*. Intraoperative monitoring of blood flow insufficiency in the anterior choroidal artery during aneurysm surgery. *J Neurosurg* 2003;98:507–14.
- Howick J, Cohen BA, McCulloch P, *et al*. Foundations for evidence-based intraoperative neurophysiological monitoring. *Clin Neurophysiol* 2016;127:81–90.
- Kashkoush AI, Jankowitz BT, Gardner P, *et al*. Somatosensory evoked potentials during temporary arterial occlusion for intracranial aneurysm surgery: predictive value for perioperative stroke. *World Neurosurg* 2017;104:442–51.
- Staarman B, O'Neal K, Magner M, *et al*. Sensitivity and specificity of intraoperative neuromonitoring for identifying safety and duration of temporary aneurysm clipping based on vascular territory, a multimodal strategy. *World Neurosurg* 2017;100:522–30.
- Kim SH, Jin SJ, Karm MH, *et al*. Comparison of false-negative/positive results of intraoperative evoked potential monitoring between no and partial neuromuscular blockade in patients receiving propofol/remifentanyl-based anesthesia during cerebral aneurysm clipping surgery: A retrospective analysis of 685 patients. *Medicine* 2016;95:e4725.
- Yue Q, Zhu W, Gu Y, *et al*. Motor evoked potential monitoring during surgery of middle cerebral artery aneurysms: a cohort study. *World Neurosurg* 2014;82:1091–9.
- Sasaki T, Kon H, Saito A, *et al*. Intraoperative supporting tools in aneurysm surgery: roles of electrophysiological monitoring. *Japanese Journal of Neurosurgery* 2014;23:716–20.
- Takebayashi S, Kamiyama H, Takizawa K, *et al*. The significance of intraoperative monitoring of muscle motor evoked potentials during unruptured large and giant cerebral aneurysm surgery. *Neurol Med Chir* 2014;54:180–8.
- Kang D, Yao P, Wu Z, *et al*. Ischemia changes and tolerance ratio of evoked potential monitoring in intracranial aneurysm surgery. *Clin Neurol Neurosurg* 2013;115:552–6.
- Maruta Y, Fujii M, Imoto H, *et al*. Intra-operative monitoring of lower extremity motor-evoked potentials by direct cortical stimulation. *Clin Neurophysiol* 2012;123:1248–54.
- Wicks RT, Pradilla G, Raza SM, *et al*. Impact of changes in intraoperative somatosensory evoked potentials on stroke rates after clipping of intracranial aneurysms. *Neurosurgery* 2012;70:1114–24.
- Ni W, Chen L, Xu G, *et al*. Application of multiple intraoperative monitoring techniques in microsurgery for anterior communicating aneurysms. *Chin J Contemp Neurol Neurosurg* 2012;12:11–15.
- Krayenbühl N, Sarthein J, Qinas M, *et al*. MRI-validation of SEP monitoring for ischemic events during microsurgical clipping of intracranial aneurysms. *Clin Neurophysiol* 2011;122:1878–82.
- Lin J, Zhao J, Zhao Y, *et al*. Multiple intraoperative monitoring-assisted micro-neurosurgical treatment for anterior circulation cerebral aneurysm. *J Int Med Res* 2011;39:891–903.
- Motoyama Y, Kawaguchi M, Yamada S, *et al*. Evaluation of combined use of transcranial and direct cortical motor evoked potential monitoring during unruptured aneurysm surgery. *Neurol Med Chir* 2011;51:15–22.
- Yeon JY, Seo DW, Hong SC, *et al*. Transcranial motor evoked potential monitoring during the surgical clipping of unruptured intracranial aneurysms. *J Neurol Sci* 2010;293:29–34.
- Irie T, Yoshitani K, Ohnishi Y, *et al*. The efficacy of motor-evoked potentials on cerebral aneurysm surgery and new-onset postoperative motor deficits. *J Neurosurg Anesthesiol* 2010;22:247–51.
- Szelényi A, Langer D, Beck J, *et al*. Transcranial and direct cortical stimulation for motor evoked potential monitoring in intracerebral aneurysm surgery. *Neurophysiol Clin* 2007;37:391–8.
- Penchet G, Arné P, Cuny E, *et al*. Use of intraoperative monitoring of somatosensory evoked potentials to prevent ischaemic stroke after surgical exclusion of middle cerebral artery aneurysms. *Acta Neurochir* 2007;149:357–64.
- Szelényi A, Langer D, Kothbauer K, *et al*. Monitoring of muscle motor evoked potentials during cerebral aneurysm surgery: intraoperative changes and postoperative outcome. *J Neurosurg* 2006;105:675–81.
- Horiuchi K, Suzuki K, Sasaki T, *et al*. Intraoperative monitoring of blood flow insufficiency during surgery of middle cerebral artery aneurysms. *J Neurosurg* 2005;103:275–83.

29. Schick U, Döhnert J, Meyer JJ, *et al.* Effects of temporary clips on somatosensory evoked potentials in aneurysm surgery. *Neurocrit Care* 2005;2:141–9.
30. Neuloh G, Schramm J. Monitoring of motor evoked potentials compared with somatosensory evoked potentials and microvascular Doppler ultrasonography in cerebral aneurysm surgery. *J Neurosurg* 2004;100:389–99.
31. Min KT, Kim JH, Shin YS, *et al.* The monitoring of somatosensory evoked potentials and neurologic complications in aneurysm surgery. *Yonsei Med J* 2001;42:227–32.
32. Holland NR. Subcortical strokes from intracranial aneurysm surgery: implications for intraoperative neuromonitoring. *J Clin Neurophysiol* 1998;15:439–46.
33. Sako K, Nakai H, Kawata Y, *et al.* Temporary arterial occlusion during anterior communicating or anterior cerebral artery aneurysm operation under tibial nerve somatosensory evoked potential monitoring. *Surg Neurol* 1998;49:316–23.
34. Sako K, Nakai H, Takizawa K, *et al.* [Aneurysm surgery using temporary occlusion under SEP monitoring]. *No Shinkei Geka* 1995;23:35–41.
35. Mizoi K, Yoshimoto T. Permissible temporary occlusion time in aneurysm surgery as evaluated by evoked potential monitoring. *Neurosurgery* 1993;33:434–40.
36. Lazorthes Y, Sonilhac F, Lagarrigue J, *et al.* [Monitoring of somatosensory evoked potentials during surgery for aneurysms of the sylvian artery]. *Neurochirurgie* 1992;38:333–46.
37. Matsuda M, Kidooka M, Nakazawa T, *et al.* Intraoperative monitoring of somatosensory evoked potentials in patients with cerebral aneurysm—correlation between central conduction time and postoperative neurological status. *Neurol Med Chir* 1991;31:13–17.
38. Schramm J, Koht A, Schmidt G, *et al.* Surgical and electrophysiological observations during clipping of 134 aneurysms with evoked potential monitoring. *Neurosurgery* 1990;26:61–70.
39. Buchthal A, Belopavlovic M, Mooij JJ. Evoked potential monitoring and temporary clipping in cerebral aneurysm surgery. *Acta Neurochir* 1988;93:28–36.
40. Kidooka M, Nakasu Y, Watanabe K, *et al.* Monitoring of somatosensory-evoked potentials during aneurysm surgery. *Surg Neurol* 1987;27:69–76.
41. Momma F, Wang AD, Symon L. Effects of temporary arterial occlusion on somatosensory evoked responses in aneurysm surgery. *Surg Neurol* 1987;27:343–52.
42. Thirumala PD, Udesch R, Muralidharan A, *et al.* Diagnostic value of somatosensory-evoked potential monitoring during cerebral aneurysm clipping: a systematic review. *World Neurosurg* 2016;89:672–80.
43. Thomas B, Guo D. The diagnostic accuracy of evoked potential monitoring techniques during intracranial aneurysm surgery for predicting postoperative ischemic damage: a systematic review and meta-analysis. *World Neurosurg* 2017;103:829–40.
44. Hilibrand AS, Schwartz DM, Sethuraman V, *et al.* Comparison of transcranial electric motor and somatosensory evoked potential monitoring during cervical spine surgery. *J Bone Joint Surg Am* 2004;86-A:1248–53.
45. Schwartz DM, Auerbach JD, Dormans JP, *et al.* Neurophysiological detection of impending spinal cord injury during scoliosis surgery. *J Bone Joint Surg Am* 2007;89:2440–9.
46. Whiting PF, Rutjes AW, Westwood ME, *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
47. Macaskill P, Gatsonis C, Deeks J. *Cochrane handbook for systematic reviews of diagnostic test accuracy: Analysing and Presenting Results*, 2010.
48. Rosman AS, Korsten MA. Application of summary receiver operating characteristics (sROC) analysis to diagnostic clinical testing. *Adv Med Sci* 2007;52:76–82.
49. Deeks J. Systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001;323:157–62.
50. Nwachuku EL, Balzer JR, Yabes JG, *et al.* Diagnostic value of somatosensory evoked potential changes during carotid endarterectomy. *JAMA Neurol* 2015;72:73–2.
51. Choudhry NK, Avorn J, Glynn RJ, *et al.* Full coverage for preventive medications after myocardial infarction. *N Engl J Med* 2011;365:2088–97.