

CLINICAL PRACTICE

Machine learning-augmented interventions in perioperative care: a systematic review and meta-analysis

Divya Mehta¹ , Xiomara T. Gonzalez² , Grace Huang³ and Joanna Abraham^{1,4,*} 

¹Department of Anesthesiology, Washington University School of Medicine, St. Louis, MO, USA, ²Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, USA, ³Medical Education, Washington University School of Medicine, St. Louis, MO, USA and ⁴Institute for Informatics, Data Science and Biostatistics (I2DB), Washington University School of Medicine, St. Louis, MO, USA

*Corresponding author. E-mail: joannaa@wustl.edu

Abstract

Background: We lack evidence on the cumulative effectiveness of machine learning (ML)-driven interventions in perioperative settings. Therefore, we conducted a systematic review to appraise the evidence on the impact of ML-driven interventions on perioperative outcomes.

Methods: Ovid MEDLINE, CINAHL, Embase, Scopus, PubMed, and ClinicalTrials.gov were searched to identify randomised controlled trials (RCTs) evaluating the effectiveness of ML-driven interventions in surgical inpatient populations. The review was registered with PROSPERO (CRD42023433163) and conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. Meta-analysis was conducted for outcomes with two or more studies using a random-effects model, and vote counting was conducted for other outcomes.

Results: Among 13 included RCTs, three types of ML-driven interventions were evaluated: Hypotension Prediction Index (HPI) ($n=5$), Nociception Level Index (NoL) ($n=7$), and a scheduling system ($n=1$). Compared with the standard care, HPI led to a significant decrease in absolute hypotension ($n=421$, $P=0.003$, $I^2=75\%$) and relative hypotension ($n=208$, $P<0.0001$, $I^2=0\%$); NoL led to significantly lower mean pain scores in the post-anaesthesia care unit (PACU) ($n=191$, $P=0.004$, $I^2=19\%$). NoL showed no significant impact on intraoperative opioid consumption ($n=339$, $P=0.31$, $I^2=92\%$) or PACU opioid consumption ($n=339$, $P=0.11$, $I^2=0\%$). No significant difference in hospital length of stay ($n=361$, $P=0.81$, $I^2=0\%$) and PACU stay ($n=267$, $P=0.44$, $I^2=0$) was found between HPI and NoL.

Conclusions: HPI decreased the duration of intraoperative hypotension, and NoL decreased postoperative pain scores, but no significant impact on other clinical outcomes was found. We highlight the need to address both methodological and clinical practice gaps to ensure the successful future implementation of ML-driven interventions.

Systematic review protocol: CRD42023433163 (PROSPERO).

Keywords: artificial intelligence; evidence synthesis; predictive modelling; perioperative outcomes; surgery

Editor's key points

- Despite there being numerous machine learning algorithms in perioperative care, their clinical application remains limited. This review identified three machine learning-driven interventions (Hypotension Prediction Index, Nociception Index, and scheduling

system) that improve physiological outcomes but are yet to demonstrate clinical benefits.

- There is statistical and clinical heterogeneity in reporting effectiveness outcomes and limited emphasis on implementation outcomes. Future work should use standardised clinical outcomes to evaluate intervention effectiveness and incorporate clinician feedback for real-world clinical translation.

Received: 3 July 2024; Accepted: 5 August 2024

© 2024 The Author(s). Published by Elsevier Ltd on behalf of British Journal of Anaesthesia. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

For Permissions, please email: permissions@elsevier.com

Over 300 million people undergo surgery annually, with nearly 50 million in the USA. In total, 20% of surgical patients experience major postoperative complications, such as heart attacks, infections, blood clots, and chronic pain.^{1,2} Furthermore, 30-day patient mortality rates following surgeries are between 1% and 5%, and 1-yr rates are between 5% and 10%.³ Although a fraction of these complications cannot be avoided, the majority can be prevented through pre-emptive monitoring and early detection of clinical signs contributing to these risks for complications.^{4,5}

To this end, recent advances in machine learning (ML)-driven models have been leveraged to augment perioperative care delivery by enabling early diagnosis and risk predictions.^{4,6} ML-driven models have been developed to predict surgical case durations^{7,8} and intraoperative and postoperative complications.^{6,9,10}

Several reviews have collated retrospective studies evaluating the performance of ML-driven models supporting perioperative care. A narrative review¹¹ of ML-driven model validation studies within the context of thoracic surgery highlighted that ML algorithms such as support vector machines, convolutional neural networks, and decision trees could potentially enhance the efficiency in diagnosing and classifying pulmonary nodules, enhancing surgical planning and pre-anaesthetic evaluation of these patients. A scoping review of ML-driven models in cardiac surgery anaesthetic care identified that ML-driven models could potentially improve perioperative care in three categories: prediction analysis (e.g. mortality, hospital readmissions, and acute kidney injury), haemodynamic monitoring, and automation of echocardiography.¹² The authors concluded that ML-driven models did not show any benefit in predictive capability over existing clinical scores but demonstrated remarkable performance using dynamic variables, such as haemodynamic monitoring and echo automation. A systematic review of ML-driven models in neurosurgery found that ML-driven models predicted neurosurgical outcomes, such as seizure freedom time, survival, mortality, and symptom improvement, with a median accuracy and an area under the receiver operating characteristic curve (AUROC) of 94.5% and 0.83, respectively.¹³ A similar but broader systematic review of ML-driven models in surgical settings found that these algorithms used for postoperative predictive outcome models and risk stratification were more accurate than validated prognostic scores and traditional statistics.¹⁴ They evaluated standard ML models for predicting perioperative complications, such as mortality, cardiovascular complications, acute kidney injury, surgical complications, and intensive care unit admission, and reported that the best-performing models were random forest and gradient boosting trees, with an area under the curve (AUC) >0.90.

A recent systematic review by Arina and colleagues¹⁵ specifically examined the state of ML tools in predicting complications and prognostication within perioperative medicine. This review encompassed a diverse array of study types, including retrospective analyses, prospective studies, and randomised controlled trials (RCTs). Of 103 included studies, only 13 were prospective, with only one RCT. Although these algorithms have shown promise in predicting postoperative complications, reflecting the significant potential of ML to improve patient outcomes through advanced predictive analytics, the review also highlighted a scarcity of high-quality evidence regarding the effectiveness of ML interventions in the perioperative setting. Despite the increasing number of

original research and reviews on ML-driven models in perioperative care, most reviews have primarily focused on aggregating the evidence on the development and statistical validation of ML-driven models, rather than on real-world effectiveness and implementation studies.

To address this gap, we conducted a systematic review and meta-analysis on evaluation studies of ML-driven interventions in perioperative care to ascertain the impact of ML-driven interventions on effectiveness and implementation outcomes. We aggregate and appraise the empirical evidence to offer insights into the use of ML-driven tools in perioperative care and opportunities for future ML use, implementation, and research directions.

Methods

The review followed the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines¹⁶ and was registered with PROSPERO (CRD42023433163).

Search strategy

A medical librarian (MD) conducted a systematic search of Ovid MEDLINE, CINAHL, Embase, Scopus, PubMed, and [ClinicalTrials.gov](https://www.clinicaltrials.gov) on January 24, 2023, to identify English-language articles on ML and artificial intelligence (AI)-driven interventions used in surgical settings. Combinations of query terms and keywords included the following: (machine learning OR artificial intelligence OR prediction index) AND (surgery OR postoperative OR perioperative) AND (randomized controlled trials OR observational study OR cohort study OR feasibility study OR prospective study OR evaluation study OR implementation study). Manual screening of references under relevant articles supplemented the search. The full search strategy is provided in [Supplement S1](#).

Study screening and selection

Three reviewers (DM, GH, and XG) independently screened article titles and abstracts for eligibility. Eligible articles were then considered for full-text review. Reviewers independently assessed full-text articles for inclusion using the PICOS framework: P (participant population)—adult or paediatric surgical patients; I (intervention)—ML-driven perioperative interventions; C (comparison)—RCTs with clinical trial registration numbers; O (outcome)—perioperative outcomes; and S (study setting)—inpatient settings. Only English-language, original research articles published in peer-reviewed journals were included (see inclusion criteria in [Supplement S2](#)). We excluded retrospective studies, studies on modelling and design of tools, studies including both inpatient and outpatient settings, studies reporting on nonsurgical procedures (e.g. colonoscopy), and studies reporting only qualitative findings ([Supplement S3](#)). Disagreements were discussed and resolved with a fourth reviewer (JA). References from included articles were also screened for eligibility.

Data abstraction and management

One reviewer (DM) extracted and recorded data on the study population, design, setting, intervention details, comparison group, and outcomes. Data discrepancies were reviewed and adjudicated by a second reviewer (JA). The data abstraction form is available in [Supplement S4](#).

Risk of bias assessment

Two reviewers (DM and GH) independently assessed the risk of bias (ROB) of included studies using the Cochrane Collaboration criterion for RCTs.¹⁷ A third reviewer (JA) reviewed ROB scores for any disagreements that were resolved through team discussion. RCTs' ROB across categories was reported using Review Manager 5 (RevMan 5) [Computer program]. Version 5.4.¹⁸

Data coding, synthesis, and analysis

The study characteristics were coded based on country, site type and number, types of participants (e.g. patients and clinicians), patient population, inpatient setting, surgery, study design, ML-driven intervention type, characteristics and functions supporting perioperative phases of care, and outcomes of interest.

Meta-analysis

A meta-analysis across studies was performed to ascertain the cumulative effect of ML-driven interventions on outcomes. Studies that reported similar outcomes (with two or more studies) were included in the meta-analysis. Among these, studies were excluded if they had insufficient reported data for a pooled analysis. However, where possible, missing data such as standard deviation (SD) were estimated through standard error calculations using provided P-values,¹⁹ and mean and SD were calculated through median and interquartile ranges provided using highly reliable calculators.^{20,21} In addition, for studies with missing data, all primary study authors were contacted; however, we did not receive any responses. A random effects model was used, and statistical heterogeneity was assessed using the I^2 test statistic. All analyses were conducted using Review Manager 5.4.¹⁸

Vote counting

Vote counting was conducted for outcomes with a low number of studies (fewer than two studies per outcome) and for subjective or qualitative outcomes (e.g. survey-based subjective assessments). The findings were synthesised based on the direction of the effect and not the statistical significance or the size of the effect. The number of effects showing benefit was compared with the number showing harm. Studies showing benefits were reported as 'improved' if the majority of effects were favourable to the intervention group and 'no difference' if there was no effect.

Results

Study selection

Of the 13 245 articles identified from the search, 14 articles from 13 original RCT studies met the inclusion criteria (Fig. 1). Of the 14 articles, two were from a single RCT.^{22,23}

Study characteristics

Table 1 presents the characteristics of the included studies. All studies were published between 2019 and 2023, with the majority conducted at teaching hospitals in Europe, except for three in the USA^{24–26} and one in Canada²⁷ (Supplement S5).

Three major types of ML-driven interventions were identified among the 13 RCTs: Nociception Level Index (NoL),

Hypotension Prediction Index (HPI), and a scheduling system. NoL and HPI were used intraoperatively, whereas the scheduling system was used preoperatively.²⁴

Population

All studies targeted adult surgical patients. Patient characteristics varied across studies by age, surgery type, and ASA physical status (Table 1).

Interventions

Across the 13 RCTs, seven reported on NoL use,^{26–30,33,36} five on HPI use,^{22,23,25,31,34,35} and one on a scheduling system.²⁴ NoL and HPI were used by anaesthesia teams intraoperatively for pain and blood pressure management, respectively. Schedulers used a scheduling system to streamline surgery schedules. Table 2 presents a summary of the ML-driven interventions.

Comparisons

All 13 RCTs compared ML-driven interventions with standard care. Two studies by Funcke and colleagues^{29,33} included four arms comparing NoL with two other pain monitors and with the anaesthesiology teams' clinical judgements. Wijnberge and colleagues²² conducted a preliminary observational study before using HPI. In this observational study, mean arterial pressure (MAP) goals were maintained in the control arm according to the clinical judgement of the anaesthesiologist. Schneck and colleagues³¹ used a historical cohort's data for comparison with the intervention group to mitigate the potential for Hawthorne bias.

Outcomes

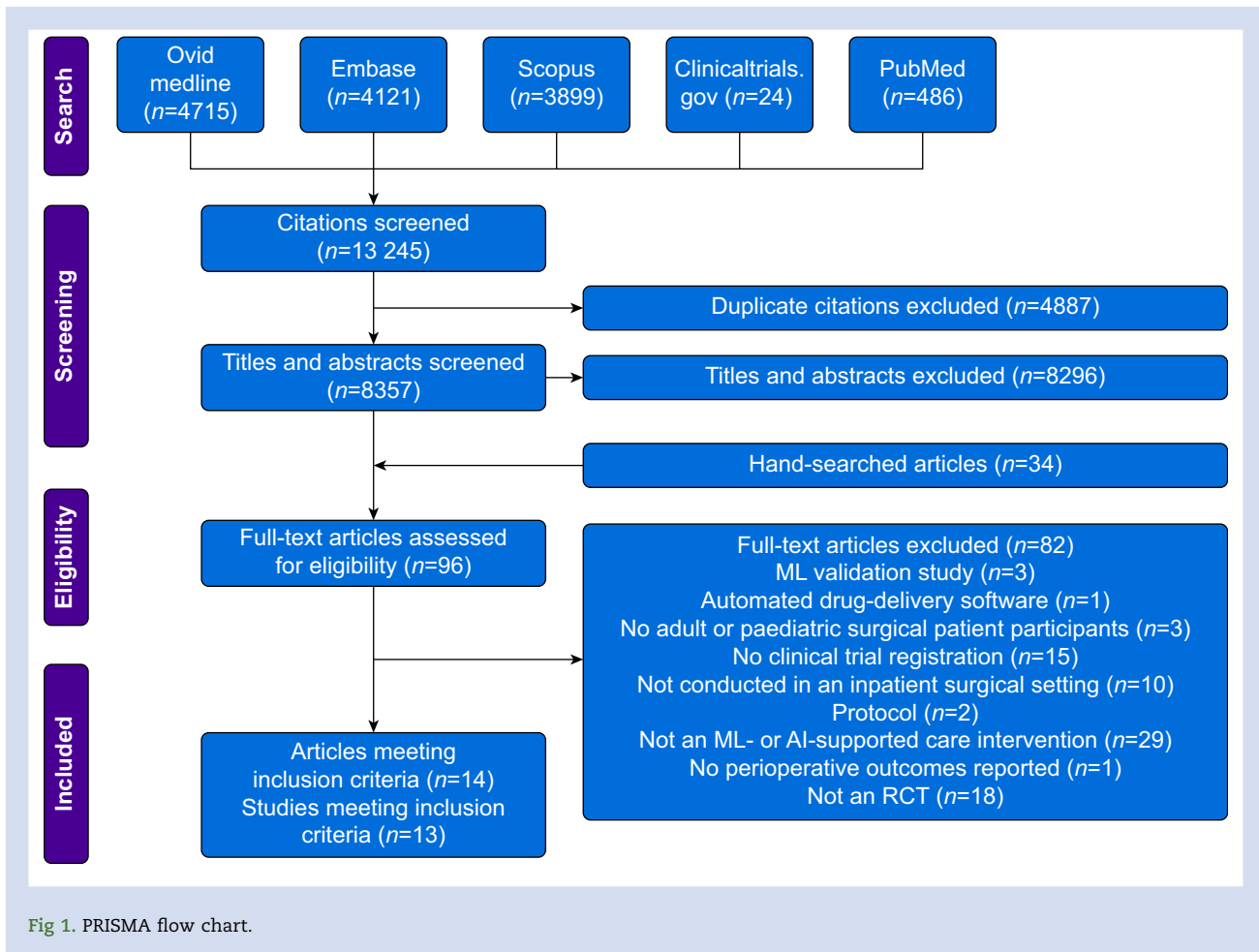
Clinical outcomes from studies pooled for the meta-analysis and vote counting are presented in Table 3. Details of clinical outcomes are presented in Supplement S7.

Impact of machine learning-driven interventions on clinical outcomes compared with standard care: results from meta-analysis

Meta-analysis findings on ML-driven interventions' impact on significant clinical outcomes are presented in Figure 2 (see Supplement S8 for additional outcomes). The study by Schenk and colleagues³² was a sub-study of Wijnberge and colleagues.²² Thus, we merged the intraoperative data reported by Schenk and colleagues³² into our analysis to avoid skewing results. Schenk and colleagues³² assessed postoperative hypotension using HPI, so it could not be pooled into a meta-analysis. The study by Stromblad and colleagues²⁴ was excluded from the pooled meta-analysis, given that the reported outcomes and intervention were tailored explicitly for use by schedulers. In addition, Fuica and colleagues³⁶ and Murabito and colleagues³⁵ reported outcomes that could not be converted into the mean and SD, so they were not pooled for meta-analysis.

Impact of machine learning-driven interventions on hypotension outcomes

Time-weighted average hypotension. Among four studies reporting on time-weighted average (TWA) hypotension, three used the HPI monitor^{22,25,34} and one used NoL.²⁶ No significant



difference was found between studies using ML-driven interventions and standard care ($n=444$, $P=0.25$, $I^2=83\%$) (Supplement S8). Considering only studies using the HPI monitor, there was no significant decrease in TWA hypotension ($n=362$, $P=0.07$, $I^2=83\%$). Effective pain management using NoL can increase the duration of hypotension; therefore, it was excluded from our analysis of ML-driven interventions on blood pressure.

Area under curve mean arterial pressure <65 mm Hg. Four studies reported on AUC MAP <65 mm Hg, of which three used HPI^{22,25,34} and one used the NoL monitor.²⁶ No significant difference was reported in the AUC MAP <65 mm Hg using HPI and the NoL monitor compared with standard care ($n=444$, $P=0.31$, $I^2=85\%$). When considering studies using HPI, there was no significant difference reported in AUC MAP <65 mm Hg between the intervention and standard care groups ($n=372$, $P=0.06$, $I^2=81\%$) (Supplement S8).

Absolute hypotension. Six studies analysed the duration of hypotension, of which four used HPI^{22,25,31,34} and two used NoL.^{26,27} After combining analyses for intervention groups, no statistical significance was found compared with standard care ($n=559$, $P=0.17$, $I^2=85\%$). Sub-group analyses revealed a significant decrease in the duration of hypotension for the HPI group compared with standard care ($n=421$, $P=0.003$,

$I^2=75\%$). In contrast, studies using NoL monitors showed a significant increase in the duration of hypotension, likely attributable to improved pain management ($n=138$, $P=0.02$, $I^2=0\%$).

Relative hypotension. Three studies reported on relative hypotension. Compared with standard care, a significant decrease in relative hypotension for the HPI group was found^{22,31,34} ($n=208$, $P<0.0001$, $I^2=0\%$).

Hypertension

Of seven studies using HPI, two studies reported hypertension. TWA hypertension (AUC MAP >100 mm Hg), absolute hypertension, and relative hypertension were all significantly higher in the HPI groups compared with standard care groups^{22,34} (Fig. 2).

Impact of machine learning-driven interventions on opioid consumption

Intraoperative opioid consumption. No significant difference in intraoperative opioid consumption was found between the NoL monitor and standard care groups across six studies^{26–30,33} ($n=339$, $P=0.31$, $I^2=92\%$) (Supplement S8).

Opioid consumption in the PACU. No significant difference in opioid consumption in the PACU was found between the NoL

Table 1 Characteristics of included studies. HPI, Hypotension Prediction Index; MAP, mean arterial pressure; NoL, Nociception Level Index.

Study author, yr	Population	Surgery	Setting	Intervention	Comparator (standard care)	Conflict of Interest
Meijer and colleagues, 2019 ²⁸	ASA 1–3, Age ≥18–80 yr	Major abdominal, urologic, or gynaecological surgery	Academic, single centre, Netherlands	NoL	Target controlled infusion remifentanyl increased or decreased according to clinical judgement	No
Funcke and colleagues, 2020 ²⁹	Age ≥18 yr, ASA 2 or 3	Radical retropubic prostatectomy	Academic, single centre, Germany	NoL	Clinical judgement, as per the anaesthesia team	No
Maheshwari and colleagues, 2020 ²⁵	Age ≥45 yr, ASA 1–4	Moderate/high-risk noncardiac surgery	Academic, two hospitals of the same university, USA	HPI	MAP>65 mm Hg, as per the anaesthesia team	Yes/Edward Lifesciences
Meijer and colleagues, 2020 ³⁰	Age >17 yr, ASA 1–3	Elective laparoscopic surgery—gynaecological, general, and urological surgery	Academic and referral centre, Netherlands	NoL	Clinical judgement, as per the anaesthesia team	Yes/Medasense
Schneck and colleagues, 2020 ³¹	Age ≥18 yr, ASA 1–3	Elective total hip arthroplasty	Academic, single centre, Germany	HPI	MAP >65 mm Hg, according to the anaesthesia team	Yes/Edward LifeSciences
Wijnberge and colleagues, 2020 ²² ; Schenk and colleagues, 2021 ³²	Age ≥ 18 yr, ASA 3	Open or laparoscopic—gynaecological, gastrointestinal, or other surgeries	Academic, single centre, Netherlands	HPI	MAP>65 mm Hg, as per the anaesthesia team	Yes/Edward LifeSciences
Espitalier and colleagues, 2021 ²⁷	Age 18–75 yr, ASA 1–3	Laparoscopic hysterectomy	Academic, single centre, Canada	NoL	Clinical judgement, according to the anaesthesia team	Yes/Medasense
Funcke and colleagues, 2021 ³³	Males >18 yr, ASA 2 or 3	Radical retropubic prostatectomy	Academic, Germany, single centre	NoL	Clinical judgement for bolusing with remifentanyl and changing the infusion rate	No
Strömbländ and colleagues, 2021 ²⁴	Adults	Gynaecological or colorectal surgery	Academic, USA, single centre, two campuses	Scheduling system	Pre-existing scheduling system within electronic health records supplemented by scheduler and surgeon estimates	No
Tsoumpa and colleagues, 2021 ³⁴	Adults ≥18 yr, ASA 1–3	Elective noncardiac surgery	Academic, single centre, Greece	HPI	MAP >65 mm Hg, as per the anaesthesia team	No
Murabito and colleagues, 2022 ³⁵	Adult ≥18 yr, ASA 1, 2, and 3	Elective major laparoscopic—general, gynaecological, or other surgeries	Academic, single centre, Italy	HPI	Institutional algorithm to maintain MAP >65 mm Hg	Yes/Edward LifeSciences
Ruetzler and colleagues, 2022 ²⁶	Age ≥21–85 yr, ASA 1, 2, or 3	Major open or laparoscopic abdominal noncardiac surgeries	Academic, US, single centre	NoL	Clinical judgement for administering opioid	No disclosure
Fuica and colleagues, 2023 ³⁶	Adult, ASA 1–3	Major laparoscopic abdominal—urologic or gynaecological surgeries	Academic, single centre, Israel	NoL	Clinical judgement, according to the anaesthesia team	No disclosure

Table 2 ML-driven intervention details. HPI, Hypotension Prediction Index; MAP, mean arterial pressure; ML, machine learning; NoL, Nociception Level Index.

ML intervention	Purpose	Description	Commercial monitor	Algorithm	Implementation in included studies
NoL	Objectively measures pain in anaesthetised patients and helps personalise opioid usage during surgery ³⁷	Uses a noninvasive finger probe to acquire multiple parameters, such as photoplethysmography, galvanic skin response, temperature, and accelerometer data, and generate a dimensionless number between 0 and 100. FDA approved	PMD-200 (Medasense, Israel)	Random forest model	The studies maintained NoL between 10 and 25 and administered analgesia for an NoL index greater than 25. The studies varied in type and dosage of opioid used, anaesthesia type (General Anaesthesia vs TIVA), and time to administer analgesia after NoL >25 (see details in Supplement S6)
HPI	Predicts likelihood of hypotensive events (MAP <65 mm Hg) and allows clinicians to pre-emptively prevent it ³⁸	Uses invasive arterial cannula to acquire multiple waveform features (amplitude, slope, and complexity features) to predict hypotensive events. It also provides advanced haemodynamic information, including cardiac output, dynamic arterial elastance, dP/dtmax (systolic slope), and stroke volume—which presumably helps clinicians select optimal treatments. FDA approved	Haemosphere and EVD-1000 (Edwards Lifesciences, USA)	Logistic regression	All studies used a treatment algorithm from additional parameters derived from the monitor. All studies used thresholds >85 for using the treatment algorithm, except one that used HPI >80. The studies varied in their HPI treatment thresholds, treatment algorithm, type of pressors or inotropes (see details in Supplement S6)
Scheduling system	Predicts operating room case duration based on various patient, procedural, surgeon, and operational factors	Uses more than 300 data features involving information of the patient, procedure, surgeon, and operational factors to train an ML model for gynaecology and colorectal surgical services at the Memorial Sloan Kettering Center and help predict operating room case duration	Not available	Random forest	The predictions were generated a day before the surgery and were implemented and published into the live scheduling system by schedulers for the next day

Table 3 Details of outcomes reported. PONV, postoperative nausea and vomiting; TWA, time-weighted average. *Outcomes included in the meta-analysis.

Included studies	Absolute hypertension*	Absolute hypotension*	AUC MAP <65*	AUC MAP >100*	Blood loss*	Colloid s*	Crystalloids*	End of narcotics to extubation time	Hospital readmissions	Incidence of hypotension	Intraoperative opioid consumption*	Length of stay in hospital*	Maximum pain score in PACU*	Mean pain score in PACU*	PACU opioid consumption*	PACU stay duration*	Pain score on arrival in PACU*	PONV	Postoperative complications	Relative Hypertension*	Relative hypotension*	Reversal time	TWA Hypertension*	TWA Hypotension*
Meijer and colleagues, 2019 ²³											✓		✓		✓	✓	✓					✓		
Funcke and colleagues, 2020 ²⁰								✓			✓		✓		✓	✓	✓							
Mareshwari et al 2020 ²⁶		✓	✓		✓	✓	✓		✓			✓							✓					✓
Meijer and colleagues, 2020 ²¹											✓		✓		✓	✓	✓					✓		
Schneck and colleagues, 2020 ²²		✓			✓	✓	✓			✓		✓			✓	✓	✓				✓			
Wijnberge and colleagues, 2020 ²⁵ ; Schenk and colleagues, 2021 ²³	✓	✓	✓	✓		✓	✓		✓	✓										✓	✓		✓	✓
Espitalier and colleagues, 2021 ²³		✓									✓		✓	✓	✓	✓	✓	✓						
Funcke and colleagues, 2021 ²⁴								✓			✓		✓		✓	✓	✓		✓					
Sirombald and colleagues, 2021 ²⁵																								
Tsoumpa and colleagues, 2021 ²⁵	✓	✓	✓	✓	✓	✓	✓			✓		✓							✓	✓		✓	✓	✓
Murabito and colleagues, 2022 ²⁶									✓										✓					
Ruetzler and colleagues, 2022 ²⁷		✓	✓					✓		✓	✓				✓			✓	✓					✓
Fuica and colleagues, 2023 ²⁷										✓			✓									✓		

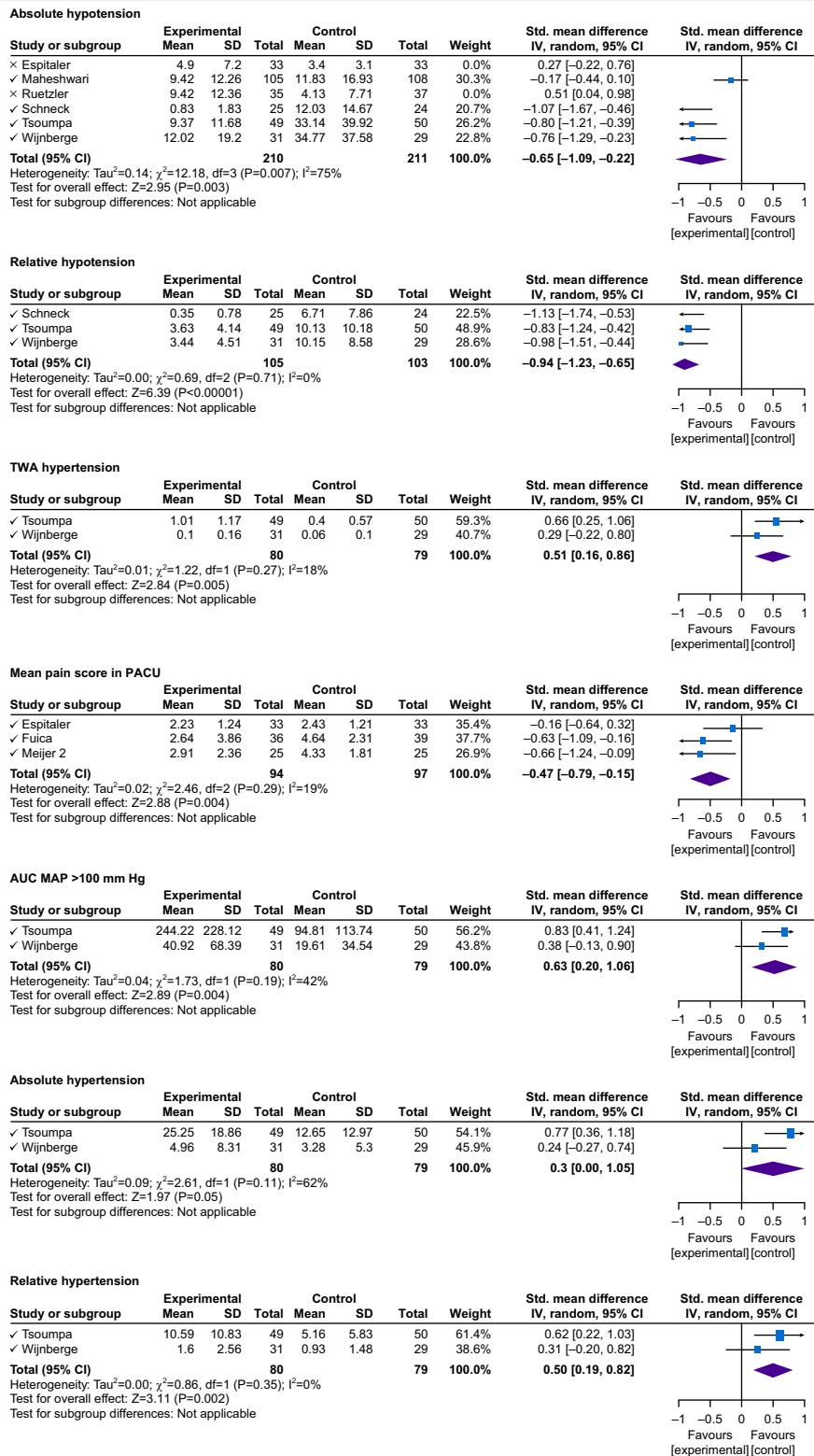


Fig 2. Random Forest plots of significant outcome results. CI, confidence interval; AUC, area under the curve; MAP, mean arterial pressure; TWA, time-weighted average.

monitor and standard care groups across six studies^{26–30,33} (n=339, P=0.11, I²=0%) (Supplement S8).

Impact of machine learning-driven interventions on pain management

Mean pain score in the PACU. The mean pain score in the PACU was significantly lower in the NoL group compared with the standard care group^{27,30,36} (n=191, P=0.004, I²=19%) (Fig. 2).

Maximum pain score in the PACU. No significant difference in the maximum pain score in the PACU was found between the NoL and standard care groups^{28,29,33} (n=151, P=0.31, I²=0%) (Supplement S8).

Pain score upon arrival at the PACU. No significant difference in the pain score upon arrival at the PACU was found between the NoL and standard care groups^{27–29,33} (n=217, P=0.37, I²=0%) (Supplement S8).

Impact of machine learning-driven interventions on the duration of hospital and PACU stay

Length of hospital stay. No significant difference was found in the length of stay in the hospital between the HPI and standard care groups^{25,31,34} (n=361, P=0.81, I²=0%) (Supplement S8).

Length of PACU stay. No significant difference was found in the PACU stay between groups using either NoL or HPI compared with standard care^{27–30,33} (n=267, P=0.44, I²=0) (Supplement S8).

Impact of interventions on clinical outcomes compared with standard care

We included studies evaluating ML-driven interventions' impact on postoperative complications (e.g. mortality, stroke, and acute kidney injury), emergence of anaesthesia, hospital readmissions, postoperative nausea and vomiting (PONV), and incidence of hypotension. These outcomes had varying definitions across the studies, so could not be pooled into a meta-analysis. Therefore, we used the vote counting method to synthesise the direction of effect.

Impact of interventions on postoperative complications

Four studies^{22,25,34,35} investigated the impact of HPI on postoperative complications (e.g. mortality, stroke, and acute kidney injury). None of the studies found significant improvements in the incidence of complications. Each study had distinct definitions of postoperative complications. Maheshwari and colleagues²⁵ observed patients for complications during their hospital stay. Three studies assessed complications during the hospital stay through surveys or phone calls 1 month after surgery.

Impact of interventions on emergence from anaesthesia

Six studies^{26,28–30,33,36} using the NoL monitor investigated the time taken for patients to emerge from anaesthesia. This investigation was based on the theoretical consideration that the NoL monitor might increase opioid consumption, potentially leading to slower wake-ups. However, there was no significant difference between the NOL monitor and standard care groups. The definition of time to emerge from anaesthesia varied across the studies. Two studies by Meijer and

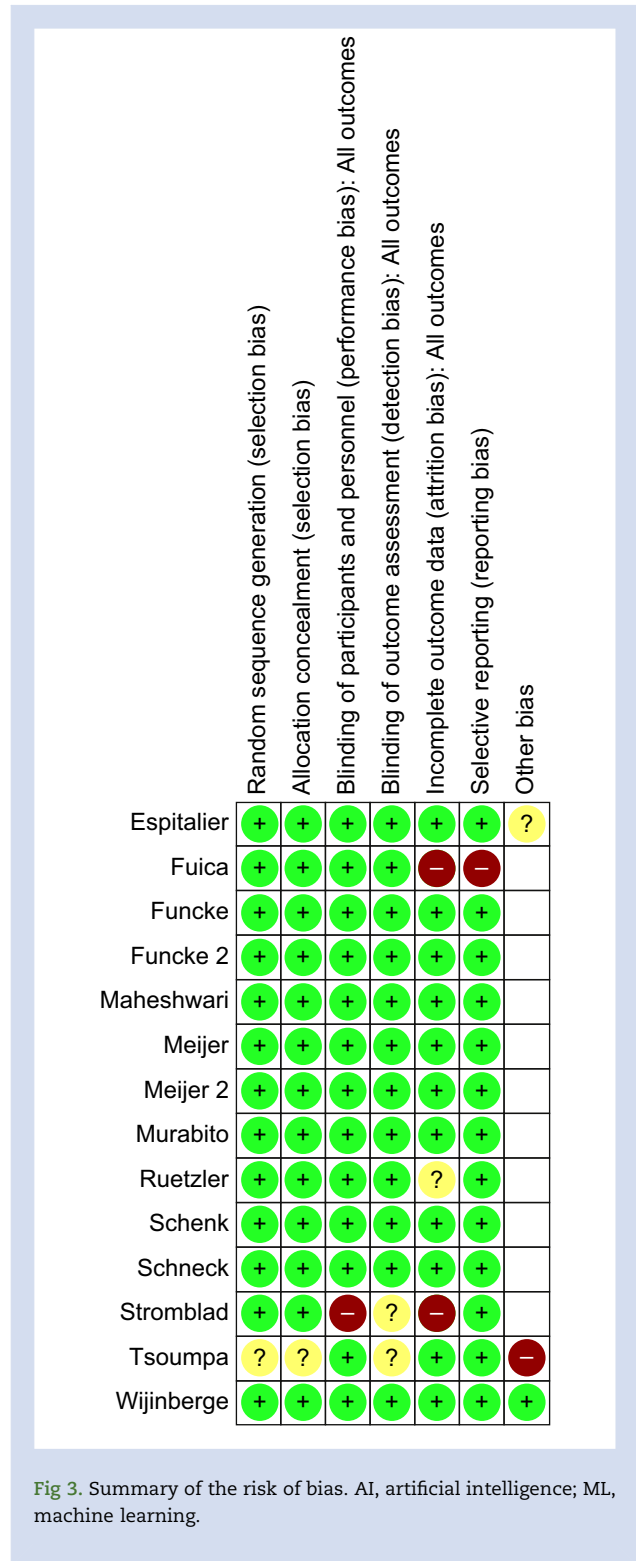


Fig 3. Summary of the risk of bias. AI, artificial intelligence; ML, machine learning.

colleagues^{28,30} and one study by Fuica and colleagues³⁶ reported it as the reversal of relaxant to extubation time. Two studies by Funcke and colleagues^{29,33} defined it as the end of narcotics to extubation time. Ruetzler and colleagues²⁶ defined it as the duration from the last minimum alveola

concentration fraction of ≥ 0.3 to when the patient left the operating room.

Impact of interventions on incidence of hypotension

Five studies^{22,26,31,34,36} investigated the incidence of hypotension using HPI and NoL. The incidence of hypotension was significantly decreased with the use of ML-driven interventions in four studies. One study using a NoL monitor³⁶ did not show any improvement in the incidence of hypotension.

Impact of interventions on postoperative nausea vomiting

Three studies reporting PONV did not show significant improvement with the use of the NoL monitor. Ruetzler and colleagues²⁶ and Funcke and colleagues³³ did not find any significant difference in the number of patients who developed PONV in the PACU. Espitalier and colleagues²⁷ measured counts of PONV events on arrival and 24 h after surgery, finding no significant difference between the intervention and standard care groups.

Impact of interventions on hospital readmissions

Three studies^{22,25,35} reported hospital readmissions. No difference in hospital readmissions was found between the ML-driven intervention and standard care groups.

Risk of bias in included studies

All studies were determined to have an overall low ROB (Fig. 3). Study factors contributing to any ROB were related to incomplete outcome data reporting and blinding of outcome assessors.

Discussion

Thirteen RCT studies investigating the effectiveness of ML-driven interventions in perioperative care settings were synthesised in this systematic review. All studies were published after 2019 and conducted at academic institutions, with the majority being in Europe, except for three in the USA^{24–26} and one in Canada.²⁷ All RCTs were single-centre studies, with three conducted at different sites within a single centre.^{24,25,30} Three types of perioperative interventions were evaluated: NoL, supporting clinicians in pain management decisions; HPI, supporting clinicians in intraoperative blood pressure management; and a surgery scheduling system, supporting preoperative management of operating room resources. Edward Lifesciences funded four^{22,25,31,35} of five studies investigating HPI; similarly, Medasense sponsored two^{27,30} of seven NoL studies, thus highlighting the industry's role in commercialising such ML-driven applications. We found that HPI significantly reduces the incidence and duration of hypotensive events, but not postoperative complications or hospital stay duration, compared with standard care. This result may be attributed to the insufficient statistical power among the included studies to detect significant differences in postoperative outcomes. Most studies showed a decreased incidence of hypotension with HPI, but varying definitions prevented a pooled meta-analysis. In contrast, we found an increase in hypertension duration with HPI use, further highlighting the need for ML algorithm refinement to mitigate adverse events and human-in-the-loop to verify ML-driven clinical decision support outputs.

Our analysis of NoL monitoring for pain management revealed no statistically significant differences in either intraoperative or postoperative opioid consumption between the NoL-guided groups and standard care. Still, the NoL-guided groups had significantly lower mean pain scores in the PACU. These results should be interpreted in the context of the following considerations. First, significant variability in the type of opioids used in the studies (e.g. ultrashort-acting remifentanyl, sufentanil, and fentanyl) may confound postoperative pain scores and opioid consumption. Second, differences in pain assessment methods and timing of administration hinder conclusive findings to determine the clinical benefit of NoL monitoring across studies. Improvements in physiological outcomes using NoL and HPI may be attributed to clinician performance bias, given their awareness of the ML-driven interventions (i.e. the Hawthorne effect). Few included studies^{22,31} mitigated clinician performance bias by comparing with historical controls. However, the learning effect from these interventions should be noted; clinicians using ML-driven interventions might have learned to manage patient pain levels and blood pressure better pre-emptively than clinicians who did not use the interventions. In summary, HPI and NoL have not shown any clinical benefit in improving the length of PACU and hospital stay. Interestingly, few studies^{25,27–31,33,34} reported on the length of hospital and PACU stay, and these studies were not powered sufficiently to draw definitive conclusions about their usefulness in decreasing the length of stay in hospital or PACU. The scheduling system²⁴ used to predict operating room case durations proved more accurate than traditional methods (i.e. estimating durations from electronic health records supplemented with the surgical team's input). The scheduling system also improved operational outcomes, such as patient waiting times and presurgical length of hospital stay. However, predictions using the scheduling system were limited to two surgical services via medical codes in the electronic health records. Any deviation from procedures performed at the site would make case duration estimates obsolete and not generalisable.

Implications for research and practice

Our review highlights six key insights from empirical studies on ML-driven interventions in perioperative care. First, there was limited standardisation in evaluating the effectiveness of ML-driven interventions across studies and across interventions. Although significant progress has been made in ML model development and retrospective validation of ML models across various clinical settings,^{6,9,10,39} the standardisation of evaluation metrics in clinical applications is crucial for comparison and pooling of clinical outcomes. The Standardised Endpoints in Perioperative Medicine (StEP) initiative⁴⁰ identified eight outcomes for measuring the quality, safety, and improvement in perioperative care. These outcomes include surgical site infection at 30 days, stroke within 30 days, death within 30 days of surgery, death within 30 days of cardiovascular surgery, readmission within 30 days, readmission to ICU within 14 days of surgery, and length of hospital stay. Only three studies^{22,25,35} followed up patients for up to 30 days and reported three STEP-COMPAC outcomes, namely mortality, complications, and readmissions. Future ML-driven intervention trials should be designed to measure standardised outcomes (e.g. STEP-COMPAC) to understand their cumulative impact on the quality of care and to enable reproducibility and robustness

in algorithm development, enhancing the reliability and generalisability of research.⁴¹

Second, there was limited emphasis on implementation outcomes across the included studies. Although ML-driven interventions can demonstrate improved outcomes, they would fail to sustain and scale if these interventions were developed in silos, without paying attention to integration and fit within clinical workflows. For example, the use of an elaborate and time-intensive algorithm to treat hypotension with HPI may pose a challenge in the clinical workflow. Hybrid effectiveness implementation trials embedded within the RCT designs and mixed-methods approaches can assess both clinical effectiveness and ML intervention usability, feasibility, and acceptability by clinicians. HPI and NoL have the potential to revolutionise the standard anaesthesia practice, but they still need considerable work to prove that they are better than the current standard of care and can be seamlessly integrated into the clinical workflow. Future studies should obtain feedback from clinicians to gauge perspectives on ML acceptability, feasibility, and appropriateness.⁴² The SALIENT framework by Van der Vegt and colleagues⁴¹ provides some guidance on implementing clinical AI in healthcare settings. It integrates tasks and components, offering checklists for each stage to support AI developers and healthcare leaders in real-time deployment, aiming to optimise perioperative care delivery through rigorous research and clinician engagement.

Third, despite the promise of ML-driven interventions in transforming perioperative care, the included studies have failed to adequately evaluate ML-driven interventions' impact on hard perioperative clinical and patient outcomes that align with the quintuple aim⁴³ of healthcare improvement—enhanced patient experience, improved population health through predictive analytics, reduced costs via operational streamlining, increased health equity with access to healthcare, and enhanced clinician well-being through automation and decision support.

Fourth, to our surprise, none of the studies focused on ML-driven interventions for paediatric surgical populations. Although a recent review highlighted similar perioperative ML-driven interventions for paediatric surgical patients (e.g. adverse event and risk prediction, and depth of anaesthesia), they are still in the development and validation phase. The review further noted that interventions for paediatric surgical patients were comparable to those for adult patients.⁴⁴ However, it should be noted that neonates, toddlers, and older children exhibit distinct physiological responses to surgery and anaesthesia compared with adults, necessitating tailored ML approaches that cannot be directly extrapolated from one age group, therefore making it difficult to generalise these algorithms.

Fifth, the interventions in our included studies were tested mostly among patients with ASA physical status 1–3 undergoing elective noncardiac surgeries, excluding a large proportion of complex, high-risk cardiac or emergency surgeries. The scarcity of empirical research in complex or emergency surgeries could be attributed to the dynamic and fast-paced perioperative environment in complex surgeries, making it challenging to develop, validate, and implement new interventions that are yet to show clinical benefit in routine surgeries. We found a scoping review¹² examining the extent and nature of ML-driven interventions in cardiac surgeries. The review identified 46 articles with a focus of ML-driven interventions in three categories. The majority ($n=41$) of interventions focused on prediction analysis (e.g. readmission,

mortality, and acute kidney injury), three on haemodynamic monitoring, and two on ultrasound guidance. However, most of these applications are still in the development and validation stages and have not yet been tested in clinical practice.

Sixth, we would like to acknowledge that the ROBAs related to the inability to blind personnel to the intervention arm, performance bias, and the learning curve may have introduced variability in the results, making it challenging to draw definitive conclusions about the efficacy of the ML-driven interventions. These factors can lead to an overestimation of the benefits of ML tools, as the observed improvements might be partially attributed to the heightened attention and modified practices of the clinicians rather than the interventions themselves. To mitigate these issues, future research should consider the following approaches. (1) Observational studies: conduct observational studies in addition to RCTs before and after the implementation of ML-driven interventions. Observational studies can provide insights into the real-world application and impact of ML interventions without the artificial constraints of a controlled trial. They can help determine whether changes in practice and outcomes persist over time after the initial implementation of the technology. In addition, comparing with historical cohorts can help reduce the effect of performance bias attributed to the inability to blind the personnel in the intervention arm. (2) Training and education: address the learning curve by providing comprehensive training and ongoing support for clinicians using ML tools. This can help standardise the use of these systems and reduce variability in outcomes attributed to differences in user proficiency. (3) Longitudinal studies: implement longitudinal studies to assess the long-term impact of ML interventions on clinical outcomes. These studies can help understand how the benefits of ML tools evolve as clinicians become more familiar with the technology. (4) Mixed-methods approach: use a mixed-methods approach that combines quantitative and qualitative research. This can provide a more comprehensive understanding of how ML interventions affect clinical practice, including insights into the experiences and perceptions of clinicians using these tools.

The systematic review by Arina and colleagues¹⁵ assessed the reliability, validity, and performance of these ML models using the Prediction model Risk of Bias Assessment Tool (PROBAST). Among the 103 included studies, only a small fraction (13%) had undergone external validation across multiple centres. This review underscores the limited generalisability of the existing ML models and suggests that their application in diverse clinical settings remains uncertain. The review concluded that ML interventions in perioperative medicine are still in their infancy, with significant room for improvement, particularly in terms of model validation and clinical application. In contrast, our study specifically evaluated the usefulness of ML interventions in the perioperative period. We focused on the impact of these interventions on both physiological outcomes, such as pain and blood pressure, and clinical outcomes. Our review found only 13 RCTs where ML interventions were used perioperatively. Importantly, the clinical and long-term benefits of these interventions remain uncertain, as no study has focused on more pragmatic outcomes.

Although the review by Arina and colleagues¹⁵ highlighted the early development stage and lack of external validation in AI applications for anaesthesiology, our study emphasised the limited power of existing RCTs to detect significant improvements in clinical outcomes. Moreover, we observed significant

variability in how clinical outcomes, such as mortality and morbidity, were measured across different studies. This inconsistency further complicates the assessment of ML interventions' effectiveness.

In summary, both reviews underscore the need for more robust, externally validated research to establish the efficacy of ML interventions. However, our study places additional emphasis on the necessity of standardising outcome measures and conducting more powerful RCTs to fully understand the clinical and long-term benefits of ML in perioperative care.

A narrative review by Hashimoto and colleagues⁴⁵ identified 173 articles with six main clinical applications in anaesthesiology: depth of anaesthesia monitoring, adverse event prediction, drug control and administration, pain monitoring, operating room logistics, and imaging techniques in regional anaesthesia. Similarly, another narrative review by Bellini and colleagues⁴⁶ also found the application of ML-driven interventions in categories similar to those of Hashimoto and colleagues.⁴⁵ However, these reviews did not answer questions on the effectiveness or utility of such AI-based technologies in clinical practice but assessed the breadth of AI research that has been conducted in anaesthesiology. In our systematic review, we found that the implementation of the ML-driven interventions only translated to three main categories: adverse event prediction, operating room logistics, and pain monitoring.

We acknowledge our review limitations. First, our search was limited to English-language articles and RCTs. Second, observed data heterogeneity indicated variations in study populations, methodologies, and outcomes. For example, we had to standardise measures to means and sds using the reported information for the meta-analysis. Third, any limitations reflected in the included studies are also limitations of this review. Fourth, pooling primary and secondary outcomes from different studies has contributed to the variability in the results of our meta-analysis. However, the review offers substantial insights into the effectiveness of perioperative ML-driven interventions and provides future directions for perioperative ML research and practice.

Conclusions

Our review found that randomised controlled trials using HPI and NOL were helpful in improving physiological outcomes by decreasing the duration of intraoperative hypotension and mean PACU pain scores, respectively. However, these trials were not powered enough to find any long-term or patient-centred outcomes, such as mortality, morbidity, and readmissions. The scheduling system showed potential in improving operational outcomes, such as patient wait time, but it involved only two surgical services of one hospital and is not generalisable. In addition, there is a lack of clinician feedback on using and implementing ML-driven interventions in the trials. We also had difficulty in pooling outcomes from different studies for meta-analysis as various outcomes had different definitions across various studies.

Advancing the implementation of machine learning-driven interventions in healthcare requires us to address critical challenges: standardising clinical outcomes, refining intervention protocols, and integrating clinician feedback. Establishing clear outcome measures, standardised protocols, and engaging clinicians throughout the process using

mixed-methods studies can enhance intervention effectiveness and adoption. Embracing interdisciplinary collaboration and leveraging implementation science frameworks will be pivotal in navigating real-world complexities and ensuring these innovations benefit diverse patient populations. Ultimately, these efforts will create robust, scalable solutions that align with clinical practice and contribute positively to patient care outcomes in varied perioperative care settings.

Authors' contributions

Had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis: JA

Concept and design: JA

Data acquisition, analysis, or interpretation: XTG, GH, DM

Drafting of the manuscript: DM, JA

Critical review of the manuscript for important intellectual content: JA, DM, XTG, GH

Statistical analysis: JA, DM

Funding: JA

Administrative, technical, or material support: JA

Supervision: JA

Acknowledgements

We would like to thank Michelle Doering (Librarian) from Washington University in Saint Louis for assisting with our search strategy for this review.

Declaration of interest

The authors have declared no conflicts of interest.

Funding

Agency for Healthcare Research and Quality (R01HS029324) to JA.

Data availability statement

The data that support the findings of this study are available from the corresponding author, JA, upon reasonable request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bja.2024.08.007>.

References

1. Story DA, Leslie K, Myles PS, et al. Complications and mortality in older surgical patients in Australia and New Zealand (the REASON study): a multicentre, prospective, observational study. *Anaesthesia* 2010; **65**: 1022–30
2. Visser BC, Keegan H, Martin M, Wren SM. Death after colectomy: it's later than we think. *Arch Surg* 2009; **144**: 1021–7
3. Kertai MD, Palanca BJ, Pal N, et al. Bispectral index monitoring, duration of bispectral index below 45, patient risk factors, and intermediate-term mortality after noncardiac surgery in the B-Unaware Trial. *Anesthesiology* 2011; **114**: 545–56

4. Xue B, Li D, Lu C, et al. Use of machine learning to develop and evaluate models using preoperative and intraoperative data to identify risks of postoperative complications. *JAMA Netw Open* 2021; 4, e212240
5. Abraham J, Bartek B, Meng A, et al. Integrating machine learning predictions for perioperative risk management: towards an empirical design of a flexible-standardized risk assessment tool. *J Biomed Inform* 2023; 137, 104270
6. Fritz BA, Cui Z, Zhang M, et al. Deep-learning model for predicting 30-day postoperative mortality. *Br J Anaesth* 2019; 123: 688–95. <https://doi.org/10.1016/j.bja.2019.07.025>
7. Jiao Y, Sharma A, Ben Abdallah A, Maddox TM, Kannampallil T. Probabilistic forecasting of surgical case duration using machine learning: model development and validation. *J Am Med Inform Assoc* 2020; 27: 1885–93. <https://doi.org/10.1093/jamia/ocaa140>
8. Jiao Y, Xue B, Lu C, Avidan MS, Kannampallil T. Continuous real-time prediction of surgical case duration using a modular artificial neural network. *Br J Anaesth* 2022; 128: 829–37. <https://doi.org/10.1016/j.bja.2021.12.039>
9. Bihorac A, Ozrazgat-Baslanti T, Ebadi A, et al. MySurgeyRisk: development and validation of a machine-learning risk algorithm for major complications and death after surgery. *Ann Surg* 2019; 269: 652–62. <https://doi.org/10.1097/SLA.0000000000002706>
10. Jeong YS, Kim J, Kim D, et al. Prediction of postoperative complications for patients of end stage renal disease. *Sensors (Base)* 2021; 21: 544. <https://doi.org/10.3390/s21020544>
11. Bellini V, Valente M, Del Rio P, Bignami E. Artificial intelligence in thoracic surgery: a narrative review. *J Thorac Dis* 2021; 13: 6963–75. <https://doi.org/10.21037/jtd-21-761>
12. Rellum SR, Schuurmans J, van der Ven WH, et al. Machine learning methods for perioperative anesthetic management in cardiac surgery patients: a scoping review. *J Thorac Dis* 2021; 13: 6976–93. <https://doi.org/10.21037/jtd-21-765>
13. Senders JT, Staples PC, Karhade AV, et al. Machine learning and neurosurgical outcome prediction: a systematic review. *World Neurosurg* 2018; 109: 476–486.e1. <https://doi.org/10.1016/j.wneu.2017.09.149>
14. Bellini V, Valente M, Bertorelli G, et al. Machine learning in perioperative medicine: a systematic review. *J Anesth Analg Crit Care* 2022; 2: 2. <https://doi.org/10.1186/s44158-022-00033-y>
15. Arina P, Kaczorek MR, Hofmaenner DA, et al. Prediction of complications and prognostication in perioperative medicine: a systematic review and PROBAST assessment of machine learning tools. *Anesthesiology* 2024; 140: 85–101. <https://doi.org/10.1097/ALN.0000000000004764>
16. Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ* 2021; 372: n160. <https://doi.org/10.1136/bmj.n160>
17. Higgins JPT, Green S, editors. Assessing risk of bias in included studies. *Cochrane Handbook Syst Rev Interventions* 2008; 1: 187–241. <https://doi.org/10.1136/bmj.d5928>
18. Review Manager (RevMan) [Computer program]. Version 5.4, 2020. The Cochrane Collaboration. Available at revman.cochrane.org. Last accessed on June 10, 2024.
19. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.4 (updated August 2023). Cochrane, 2023. John Wiley & Sons. UK. Available from www.training.cochrane.org/handbook. Last accessed on June 10, 2024.
20. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol* 2014; 14: 135. <https://doi.org/10.1186/1471-2288-14-135>
21. Luo D, Wan X, Liu J, Tong T. Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. *Stat Methods Med Res* 2018; 27: 1785–805. <https://doi.org/10.1177/0962280216669183>
22. Wijnberge M, Geerts BF, Hol L, et al. Effect of a machine learning-derived early warning system for intraoperative hypotension vs standard care on depth and duration of intraoperative hypotension during elective noncardiac surgery: the HYPE randomized clinical trial. *JAMA* 2020; 323: 1052–60. <https://doi.org/10.1001/jama.2020.0592>
23. Wijnberge M, Schenk J, Bulle E, et al. Association of intraoperative hypotension with postoperative morbidity and mortality: systematic review and meta-analysis. *BJs Open* 2021; 5: zraa018. <https://doi.org/10.1093/bjsopen/zraa018>
24. Strömblad CT, Baxter-King RG, Meisami A, et al. Effect of a predictive model on planned surgical duration accuracy, patient wait time, and use of presurgical resources: a randomized clinical trial. *JAMA Surg* 2021; 156: 315–21. <https://doi.org/10.1001/jamasurg.2020.6361>
25. Maheshwari K, Shimada T, Yang D, et al. Hypotension Prediction Index for prevention of hypotension during moderate- to high-risk noncardiac surgery. *Anesthesiology* 2020; 133: 1214–22. <https://doi.org/10.1097/ALN.0000000000003557>
26. Ruetzler K, Montalvo M, Bakal O, et al. Nociception Level Index-guided intraoperative analgesia for improved postoperative recovery: a randomized trial. *Anesth Analg* 2023; 136: 761–71. <https://doi.org/10.1213/ANE.0000000000006351>
27. Espitalier F, Idrissi M, Fortier A, et al. Impact of Nociception Level (NOL) index intraoperative guidance of fentanyl administration on opioid consumption, postoperative pain scores and recovery in patients undergoing gynecological laparoscopic surgery: a randomized controlled trial. *J Clin Anesth* 2021; 75, 110497. <https://doi.org/10.1016/j.jclinane.2021.110497>
28. Meijer FS, Martini CH, Broens S, et al. Nociception-guided versus standard care during remifentanyl–propofol anesthesia: a randomized controlled trial. *Anesthesiology* 2019; 130: 745–55. <https://doi.org/10.1097/ALN.0000000000002634>
29. Funcke S, Pinnschmidt HO, Wessler S, et al. Guiding opioid administration by 3 different analgesia nociception monitoring indices during general anesthesia alters intraoperative sufentanil consumption and stress hormone release: a randomized controlled pilot study. *Anesth Analg* 2020; 130: 1264–73. <https://doi.org/10.1213/ANE.0000000000004388>
30. Meijer F, Honing M, Roor T, et al. Reduced postoperative pain using Nociception Level-guided fentanyl dosing during sevoflurane anaesthesia: a randomised controlled trial. *Br J Anaesth* 2020; 125: 1070–8. <https://doi.org/10.1016/j.bja.2020.07.057>
31. Schneck E, Schulte D, Habig L, et al. Hypotension Prediction Index based protocolized haemodynamic management reduces the incidence and duration of intraoperative hypotension in primary total hip arthroplasty: a single centre feasibility randomised blinded prospective interventional

- trial. *J Clin Monit Comput* 2020; **34**: 1149–58. <https://doi.org/10.1007/s10877-019-00433-6>
32. Schenk J, Wijnberge M, Maaskant JM, et al. Effect of Hypotension Prediction Index-guided intraoperative haemodynamic care on depth and duration of postoperative hypotension: a sub-study of the Hypotension Prediction trial. *Br J Anaesth* 2021; **127**: 681–8. <https://doi.org/10.1016/j.bja.2021.05.033>
 33. Funcke S, Pinnschmidt HO, Brinkmann C, et al. Nociception level-guided opioid administration in radical retropubic prostatectomy: a randomised controlled trial. *Br J Anaesth* 2021; **126**: 516–24. <https://doi.org/10.1016/j.bja.2020.09.051>
 34. Tsoumpa M, Kyttari A, Matiatou S, et al. The use of the Hypotension Prediction Index integrated in an algorithm of goal directed hemodynamic treatment during moderate and high-risk surgery. *J Clin Med* 2021; **10**: 5884. <https://doi.org/10.3390/jcm10245884>
 35. Murabito P, Astuto M, Sanfilippo F, et al. Proactive management of intraoperative hypotension reduces biomarkers of organ injury and oxidative stress during elective non-cardiac surgery: a pilot randomized controlled trial. *J Clin Med* 2022; **11**: 392. <https://doi.org/10.3390/jcm11020392>
 36. Fuica R, Krochek C, Weissbrod R, Greenman D, Freundlich A, Gozal Y. Reduced postoperative pain in patients receiving nociception monitor guided analgesia during elective major abdominal surgery: a randomized, controlled trial. *J Clin Monit Comput* 2023; **37**: 481–91. <https://doi.org/10.1007/s10877-022-00906-1>
 37. Ben-Israel N, Kliger M, Zuckerman G, Katz Y, Edry R. Monitoring the nociception level: a multi-parameter approach. *J Clin Monit Comput* 2013; **27**: 659–68. <https://doi.org/10.1007/s10877-013-9487-9>
 38. Hatib F, Jian Z, Buddi S, et al. Machine-learning algorithm to predict hypotension based on high-fidelity arterial pressure waveform analysis. *Anesthesiology* 2018; **129**: 663–74. <https://doi.org/10.1097/ALN.0000000000002300>
 39. Meyer A, Zverinski D, Pfahringer B, et al. Machine learning for real-time prediction of complications in critical care: a retrospective study. *Lancet Respir Med* 2018; **6**: 905–14. [https://doi.org/10.1016/S2213-2600\(18\)30300-X](https://doi.org/10.1016/S2213-2600(18)30300-X)
 40. Haller G, Bampoe S, Cook T, et al. Systematic review and consensus definitions for the Standardised Endpoints in Perioperative Medicine initiative: clinical indicators. *Br J Anaesth* 2019; **123**: 228–37. <https://doi.org/10.1016/j.bja.2019.04.041>
 41. van der Vegt AH, Scott IA, Dermawan K, Schnetler RJ, Kalke VR, Lane PJ. Implementation frameworks for end-to-end clinical AI: derivation of the SALIENT framework. *J Am Med Inform Assoc* 2023; **30**: 1503–15. <https://doi.org/10.1093/jamia/ocad088>
 42. Plana D, Shung DL, Grimshaw AA, Saraf A, Sung JY, Kann BH. Randomized clinical trials of machine learning interventions in health care: a systematic review. *JAMA Netw Open* 2022; **5**, e2233946. <https://doi.org/10.1001/jamanetworkopen.2022.33946>
 43. Nundy S, Cooper LA, Mate KS. The quintuple aim for health care improvement: a new imperative to advance health equity. *JAMA* 2022; **327**: 521–2. <https://doi.org/10.1001/jama.2021.25181>
 44. Antel R, Sahlas E, Gore G, Ingelmo P. Use of artificial intelligence in paediatric anaesthesia: a systematic review. *BJA Open* 2023; **5**, 100125. <https://doi.org/10.1016/j.bjao.2023.100125>
 45. Hashimoto DA, Witkowski E, Gao L, Meireles O, Rosman G. Artificial intelligence in anesthesiology: current techniques, clinical applications, and limitations. *Anesthesiology* 2020; **132**: 379–94. <https://doi.org/10.1097/ALN.0000000000002960>
 46. Bellini V, Rafano Carna E, Russo M, et al. Artificial intelligence and anesthesia: a narrative review. *Ann Transl Med* 2022; **10**: 528. <https://doi.org/10.21037/atm-21-7031>

Handling Editor: Jonathan Hardman