

ORIGINAL ARTICLE

Training Level and Bispectral Index Below 40 After Propofol Induction by Anesthesiology Residents in Indonesia: A Prospective Observational Study

Rizki Suhadayanti¹, Isngadi¹, Buyung Hartiyo Laksono¹, Ristiawan Muji Laksono¹¹ Department of Anesthesiology and Intensive Therapy, Faculty of Medicine, Universitas Brawijaya/Dr. Saiful Anwar Regional General Hospital, Malang, Indonesia

ARTICLE INFO

Article history

Received: August 22, 2025
Revised: September 20, 2025
Accepted: October 18, 2025
Available online: October 27, 2025
Published: October 29, 2025

Keywords

Anesthesia, General
Bispectral Index
Clinical Competence
Internship and Residency
Propofol

DOI

<https://doi.org/10.37275/jacr.v6i2.982>

Corresponding author

Rizki Suhadayanti
Suhadayanti.rizki@gmail.com

ABSTRACT

Background. Excessively deep hypnosis after induction is detectable with the bispectral index (BIS), but its relation to residency training level is incompletely quantified.

Objective. To compare induction depth across seven semesters of anesthesiology residency.

Methods. This prospective observational study enrolled 21 residents (three per semester, I–VII), each managing five ASA I–II adults (n=105) in Malang, Indonesia, during April–December 2024. A standardized fentanyl–midazolam–propofol protocol was used under attending supervision. The primary outcome was BIS 2 min after intubation. Group-level analyses included sensitivity analyses for clustering and unmeasured joint distributions.

Results. In post hoc analyses, mean BIS rose by 2.47 units per semester across semesters I–V (95% CI 1.00–3.94; p=0.001), with no significant subsequent change. BIS <40 occurred after 46.7% of junior-led and 11.1% of senior-led inductions (risk difference 35.6 percentage points, 95% CI 15.0–53.9; Holm p=0.003). The difference persisted under the reported clustering and confounding sensitivity analyses. Juniors used more propofol (2.4 vs 1.9 mg/kg; p<0.001). Hypotension did not differ significantly (30.0% vs 17.8%; p=0.265); its association with BIS <40 was not identifiable.

Conclusion. Junior residents more often induced BIS <40 and used larger propofol doses. Post-induction BIS is a candidate metric for formative feedback; these observational data do not establish patient harm.

Access licence: Open access. This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Licence (CC BY-NC-SA 4.0), which permits non-commercial use, sharing, adaptation and reproduction in any medium, provided the original author and source are credited and derivative works are licensed under identical terms.

Copyright: © 2025 The Author(s). Published by HM Publisher in collaboration with CMHC Research Center. e-ISSN 2745-9497.

1. Introduction

Excessive hypnotic depth is one of the most frequent and least visible deviations in everyday general anesthesia. In a departmental audit of 7,031 anesthetics monitored with the bispectral index (BIS), a quarter of cases spent at least 15 min with BIS below 35 while the delivered volatile dose exceeded the age-adjusted minimum alveolar concentration.¹ Whether such depth harms patients has been tested in large trials with mixed results. Targeting a BIS of 50 rather than 35 reduced postoperative delirium from 28% to 19% in 655 at-risk patients (odds ratio [OR] 0.58),² whereas electroencephalography (EEG)-guided titration did not reduce delirium in the ENGAGES and ENGAGES-Canada trials despite halving the duration of EEG suppression,^{3,4} and lighter versus deeper BIS-targeted anesthesia did not significantly change 1-year mortality in the 6,644-patient BALANCED trial.⁵ Meta-analyses place the benefit of processed-EEG guidance on delirium at an OR of 0.78 to a risk ratio of 0.81.^{6,7} The EEG signatures most consistently linked with delirium — burst suppression and loss of frontal alpha power — are the same patterns that a low BIS is designed to detect,^{8,9} and they occur inadvertently and frequently during propofol-based anesthesia.^{10,11}

Induction is the phase in which depth is most easily overshoot. The hypnotic effect of a propofol bolus depends on the

dose, on the speed of injection and on patient factors: requirement for loss of consciousness falls by 15–39% with advancing age,¹² and faster injection increases the induction dose, the incidence of burst suppression and the lowest BIS reached.¹³ The clinical endpoint commonly used to stop injecting, loss of the eyelash reflex, occurs at BIS values of about 61–65,¹⁴ so the window between an adequate and an excessive dose is narrow, and intubation itself arouses the EEG within the following minutes.¹⁵ Propofol induction has also been associated with post-induction hypotension,^{16–18} whose incidence ranges from 31% to more than 80% depending on its definition and is associated with age, baseline pressure and body mass index.^{19–23}

Much of the variation in anesthetic practice is attributable to the clinician rather than to the patient. Risk-adjusted intraoperative hypotension can be measured at the level of the individual clinician, with reliable differences between clinicians and between fellows and attending staff in one multicenter evaluation, although another found little clinician-level variation;^{24,25} 13–26% of the variance in neuromuscular blocker use lies between providers;²⁶ 44% of the variation in peripheral nerve block use is attributable to the anesthetist;²⁷ and vasopressor selection and inotrope use also vary between clinicians.^{28,29} If clinician-level variation is this large among specialists, variation by stage of training is likely to be larger still, yet it has rarely been measured with a physiological endpoint.

Residency programs increasingly describe progress in terms of milestones and entrustable professional activities (EPAs), but the tools that support these judgements are mostly observational ratings. Workplace-based EPA ratings in anesthesiology need at least 12 observations to reach acceptable reliability, and only two of seven EPAs discriminated between residents in a recent evaluation;^{30,31} longitudinal milestone data from five programs showed variation in residents' starting levels and rates of improvement across both residents and programs.³² Objective learning curves exist mainly for procedural skills — flexible bronchoscopic intubation after about 15 procedures, hyperangulated videolaryngoscopy after a median of 12 attempts, caudal block after 9–13 attempts and transesophageal echocardiography after about 20 studies — and cumulative-sum (CUSUM) methods have limited accuracy for high-stakes decisions.^{33–37} Structured teaching and EPA-based feedback can accelerate these curves.^{38,39} Titration of a hypnotic, by contrast, is a continuous pharmacological skill without a pass–fail endpoint, and the degree to which supervision shapes it is unknown: higher attending staffing ratios worsen surgical outcomes, novices value direct attending supervision, and postanesthesia care training requires supervised patient assessment and safe transfer of care.^{40–42}

Processed EEG offers a candidate metric. A 35-min training session improved clinicians' recognition of EEG depth patterns from 60% to 78%,⁴³ an EEG-guided curriculum improved residents' knowledge,⁴⁴ and institutional education programs increased intraoperative use of processed EEG to 61% of general anesthetics.⁴⁵ A decision-support display reduced the time spent at BIS <40 and the incidence of severe hypotension,⁴⁶ Bayesian advisory tools for propofol have been proposed as teaching aids for novices,⁴⁷ BIS-guided closed-loop propofol delivery has been associated with lower propofol consumption than manual control,⁴⁸ and just-in-time coaching of inexperienced clinicians improves first-attempt success while lowering cognitive load.⁴⁹ Education alone, however, did not reduce hypotension where software guidance did.⁵⁰

These questions matter particularly in Indonesia and Southeast Asia. Physician anesthesia provider density is below five per 100,000 population in 76 countries and around one per 100,000 in Thai provinces, so residents deliver a large share of anesthetics.^{51,52} Only 13.7% of 531 anesthesiologists surveyed in Singapore, Thailand and Korea monitored depth in every patient,⁵³ although EEG-guided care has reduced anesthetic exposure and burst suppression in a regional trial.⁵⁴ Indonesian studies have shown a training-level gradient in residents' clinical reasoning and a benefit of simulation for teaching anesthetic pharmacokinetics, while the international ANTS framework provides a basis for assessing nontechnical skills; these approaches do not directly quantify post-induction anesthetic depth.^{55–57} We therefore examined, in a teaching hospital in East Java, how BIS 2 min after intubation, propofol dose, time to loss of consciousness and post-induction hypotension differed across seven semesters of residency. We hypothesized that junior residents would induce deeper hypnosis than senior residents and that the difference would be robust to the nesting of patients within residents and to patient characteristics.

2. Methods

2.1. Study design and reporting

This was a prospective, single-center observational study comparing residents at different stages of training, each observed during the same standardized induction. It is reported according to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement.

2.2. Setting

The study was conducted in the central operating theaters of Dr. Saiful Anwar Regional General Hospital, Malang, East Java, Indonesia, a provincial tertiary referral center and the principal teaching hospital of the Faculty of Medicine, Universitas Brawijaya, between April and December 2024. The anesthesiology residency is organized in semesters; semesters I–II form the foundational phase and semesters V–VII the advanced, pre-specialist phase. BIS monitoring is used routinely in the department, so residents and supervisors were familiar with the device before the study.

2.3. Participants and eligibility

Residents in semesters I to VII were eligible if they had completed the department's foundational training and had passed the institutional Direct Observation of Procedural Skills assessment for tracheal intubation; residents who had not completed the mandatory induction-protocol training were excluded. To obtain equal representation of each training level, the first three residents in each semester who had an eligible patient during the study period were enrolled (21 residents). Each resident then managed five consecutive eligible patients (15 per semester; 105 in total). Patients aged 18–65 years with American Society of Anesthesiologists (ASA) physical status I or II scheduled for elective surgery under general anesthesia with tracheal intubation were eligible. Patients were excluded for known or suspected central nervous system disease (epilepsy, stroke, dementia), current psychoactive medication, a pre-operative Glasgow Coma Scale score below 15, allergy to any study drug, contraindications to frontal sensor placement (facial trauma, skin infection) or unstable hemodynamics.

2.4. Anesthetic protocol

On arrival in the operating room, five-lead electrocardiography, noninvasive blood pressure cycling every 3 min, pulse oximetry and, after intubation, capnography were applied. A four-electrode BIS Quatro sensor was placed on the forehead after skin preparation with isopropyl alcohol, impedance was confirmed to be below 5 k Ω , and the signal from a stand-alone BIS monitor (Covidien) was allowed to stabilize for at least 3 min before the baseline value was recorded. No sedative premedication was given. The resident administered fentanyl 2 μ g/kg and midazolam 0.03 mg/kg intravenously and, 1 min later, titrated propofol to loss of the eyelash reflex, which defined loss of consciousness; the total propofol dose was recorded. Atracurium 0.5 mg/kg was then given, the trachea was intubated, and placement was confirmed by auscultation and sustained end-tidal carbon dioxide. Anesthesia was maintained with sevoflurane in 50% oxygen in air, and the resident adjusted the end-tidal concentration to maintain hemodynamic stability and a BIS of 40–60.

2.5. Supervision

Every case was supervised in the room by one of ten board-certified anesthesiologists (mean 12.5 years of specialist

experience). Supervisors were not matched to resident seniority; the attending on call for the operating room supervised the case. All supervisors were briefed on the protocol and asked to observe without verbal or physical intervention unless a predefined safety threshold was about to be crossed: mean arterial pressure (MAP) below 55 or above 120 mmHg, heart rate below 40 or above 120 beats/min, or oxygen saturation below 90%, sustained for more than 1 min.

2.6. Outcomes and operational definitions

A research assistant who was not part of the clinical team recorded all data prospectively. BIS was recorded at three time points: T1, immediately before induction; T2, exactly 2 min after successful intubation; and T3, 5 min after skin incision. The primary outcome was BIS at T2, analyzed as a continuous value and in three categories: deep (BIS <40), target (40–60) and light (>60). Secondary outcomes were the weight-normalized propofol induction dose (mg/kg), the time from the start of propofol injection to loss of consciousness (s), the proportion of patients with BIS <40 at T2, and post-induction hypotension, defined as a fall in MAP of more than 20% from baseline (the T1 value) at any time within 10 min of induction. MAP and heart rate were also recorded at T1–T3.

2.7. Sample size

The sample was fixed by design (three residents per semester, five patients per resident) rather than by a formal power calculation. Its precision is therefore reported explicitly: the width of the 95% confidence interval (CI) for each key proportion, the number of patients that would be required for half-widths of 10 and 5 percentage points, the smallest difference in hypotension detectable with 80% power, and the number of residents needed for a future cluster-aware study.

2.8. Ethics

The study protocol was approved by the CMHC Research Center Ethics Committee, Indonesia (Approval No. 2024/8854) in 2024, before enrollment began. Written informed consent was obtained from every participating resident and patient before any study procedure, covering the observational nature of the study, monitoring procedures, confidentiality and the right to withdraw. The study was conducted in accordance with the Declaration of Helsinki.

2.9. Data available for analysis

The analyses in this article were performed on the study's group-level data: semester-level means and standard deviations (SDs) of BIS at T1–T3, patient age and body mass index; semester-level counts of patient sex and ASA status; resident age and sex by semester; training-level counts of BIS categories at T2; and junior-versus-senior summaries of propofol dose, time to loss of consciousness, hypotension and BIS <40. Patient-level records linking BIS category, dose, hypotension and patient characteristics within individuals, resident and supervisor identifiers, and semester-level values of the secondary outcomes were not available. MAP and heart rate at T1–T3 were therefore not analyzable. For the intermediate group, the number of patients with BIS <40 was available only as approximately 25% of 30 (7 or 8 patients); every analysis involving that group was run under both admissible counts. There were no missing values for any variable analyzed. Because semester-level values of the secondary outcomes were unavailable, secondary outcomes could be compared only between

the junior and senior groups, an extreme-groups contrast that tends to overstate standardized effects relative to the full training range.

2.10. Statistical analysis

Continuous variables are presented as mean $\pm$ SD with 95% CIs, and categorical variables as counts and percentages with Wilson score 95% CIs. Comparisons across the seven semesters used one-way analysis of variance (ANOVA) and Welch's ANOVA computed from the semester means, SDs and sizes, which are sufficient statistics for these tests under the assumption of approximately normal within-semester distributions; effect size is reported as η^2 with a 95% CI obtained by inverting the noncentral F distribution, and as ω^2 . A linear contrast with semester scores 1–7 tested for a trend and yielded the slope in BIS units per semester; the residual between-semester variation tested deviation from linearity, and the patient-level Pearson correlation between semester and BIS was computed exactly from the same quantities. Pairwise differences between semesters were tested with the Games–Howell procedure, which allows unequal variances and controls the family-wise error over all 21 comparisons, with Tukey–Kramer results shown for comparison.

Training levels were defined a priori as junior (semesters I–II; 6 residents, 30 patients), intermediate (III–IV; 6 residents, 30 patients) and senior (V–VII; 9 residents, 45 patients); pooled means and SDs for each level were derived exactly from the semester statistics. Junior and senior residents were compared with Welch's *t* test, mean differences with 95% CIs and Hedges *g* for continuous outcomes, and with Fisher's exact test, risk differences (RD) with Newcombe hybrid-score CIs, risk ratios (RR) and conditional maximum-likelihood ORs with exact CIs for binary outcomes; Cohen *h* was used to place binary outcomes on the same standardized scale as *g*. The distribution of BIS categories across the three levels was tested with the Fisher–Freeman–Halton exact test by complete enumeration of all tables with the observed margins, Cramér's *V*, the linear-by-linear association test and the Cochran–Armitage trend test for BIS <40. A proportional-odds model with training level scored 1–3 estimated the common OR for a higher BIS category, and separate binary logistic models at each cut-point examined whether the proportional-odds assumption held.

Two sensitivity analyses addressed features of the design that the group-level data could not model directly. First, because each resident managed five patients, outcomes may be correlated within residents. Test statistics were deflated by the square root of the design effect, $1 + (m - 1) \times \text{ICC}$ with $m = 5$, across intraclass correlation coefficients (ICCs) from 0 to 0.50; *t*-based tests were referred to resident-level degrees of freedom (13 for junior-versus-senior comparisons and 14 for seven-semester analyses), and the tipping-point ICC at which each result lost significance was located. Second, because the joint patient-level distribution of BIS category, hypotension, ASA status and sex was not available, a bounds analysis was run over all datasets consistent with the published margins. For the association between BIS <40 and hypotension, every admissible 2×2 table within the junior and senior strata was enumerated and summarized by the attainable range of the OR (with Haldane–Anscombe correction for zero cells), Fisher *p* values and the stratified Mantel–Haenszel OR. For multivariable questions, admissible patient-level datasets were generated by permuting each variable within training level, which preserves every published margin; Firth-penalized

logistic regression was fitted to 4,000 random admissible datasets, and 12-restart hill-climbing searched for each coefficient's minimum and maximum. Wald p values are reported for Firth coefficients. Because random sampling and hill-climbing cannot guarantee that global extremes are found, the reported ranges are the attainable values found by the search (random seed 20261012; 500 swap proposals per restart). Model A regressed BIS <40 on junior level, ASA II and male sex; model B regressed hypotension on BIS <40, junior level, ASA II and male sex. A measure whose attainable range spans the null is reported as not identifiable from these data; a conclusion that holds in every admissible dataset is reported as robust.

All tests were two-sided with $\alpha = 0.05$, and p values are reported to three decimal places. Analyses were performed in Python 3 with SciPy 1.18.1, statsmodels 0.15.0 and NumPy 2.4.4; the data set and code are provided in Supplementary File 2. The four junior-versus-senior secondary outcomes (propofol dose, time to loss of consciousness, BIS <40 and hypotension) were treated as one family, and Holm–Bonferroni-adjusted p values are reported alongside unadjusted values. Because learning curves are typically

nonlinear, curvature of the T2 trajectory was examined post hoc with a quadratic contrast, separate slopes were estimated for semesters I–V and V–VII, and the stability of the linear trend was checked by omitting one semester at a time. Approximate normality of BIS within semesters was checked by confirming that mean ± 2 SD lay within the 0–100 scale for every semester.

3. Results

3.1. Participants

All 21 residents and 105 patients completed the protocol and were analyzed (Figure 1). Residents were 31.3 ± 2.7 years old and 16 of 21 (76.2%) were men; neither age ($p=0.777$) nor sex ($p=0.748$) differed between semesters. Patients were 42.9 ± 11.2 years old, 53 (50.5%) were men, mean body mass index was 24.0 ± 3.0 kg/m², and 70 (66.7%) were ASA physical status I. Age ($p=0.999$), sex ($p=0.977$), body mass index ($p=0.994$) and ASA status ($p=0.991$) were similar across semesters, as detailed in Table 1.

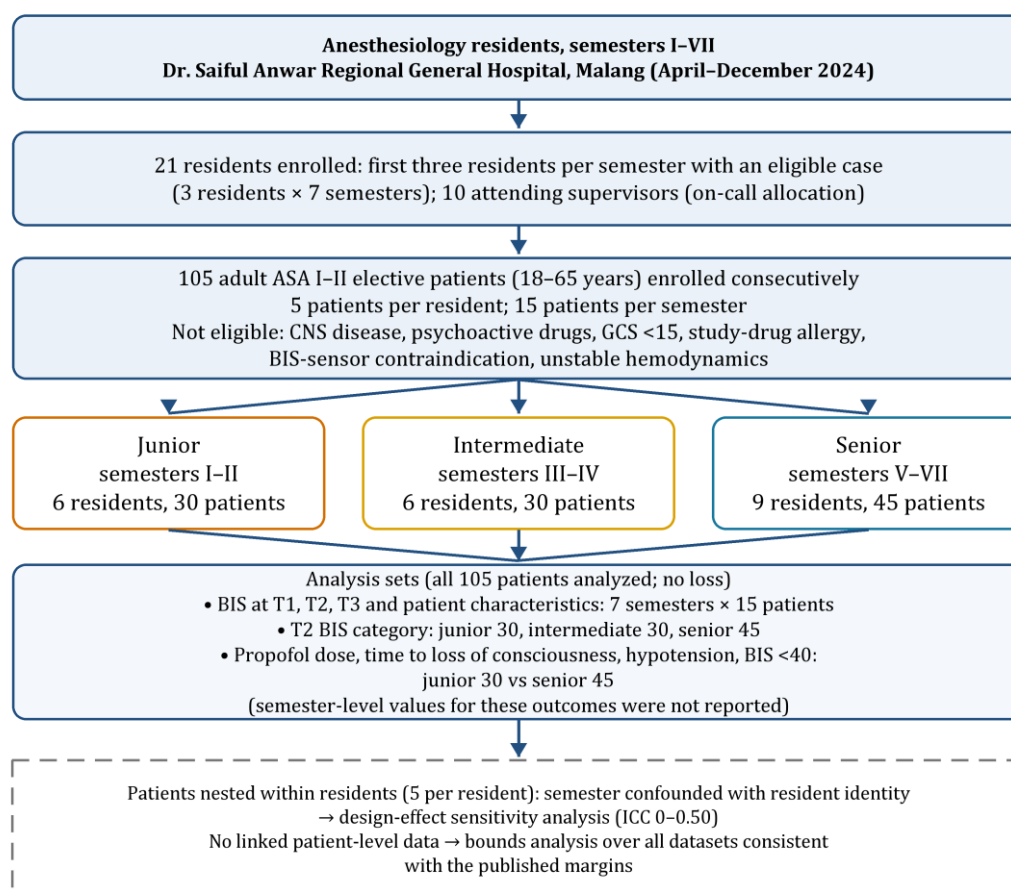


Figure 1. Study flow. Twenty-one residents (three per semester) each managed five consecutive eligible patients; all 105 patients were analyzed. The middle row shows the training-level groups and the box below them the analysis sets. The dashed box summarizes how the nested design and the absence of linked patient-level data were handled.

Table 1. Characteristics of the residents and their patients by residency semester

Characteristic	Sem I	Sem II	Sem III	Sem IV	Sem V	Sem VI	Sem VII	All	p
Residents (n=3 per semester; 21 in total)									
Age, years, mean $\pm$ SD	31.7 $\pm$ 2.9	29.0 $\pm$ 0.0	31.3 $\pm$ 2.1	31.0 $\pm$ 4.4	32.0 $\pm$ 3.5	31.3 $\pm$ 1.2	33.0 $\pm$ 3.6	31.3 $\pm$ 2.7	0.777 ^a

Characteristic	Sem I	Sem II	Sem III	Sem IV	Sem V	Sem VI	Sem VII	All	p
Male, n/3	3/3	3/3	2/3	2/3	2/3	3/3	1/3	16/21 (76.2%)	0.748 ^b
Patients (n=15 per semester; 105 in total)									
Age, years, mean ± SD	42.1 ± 11.2	43.5 ± 10.8	41.9 ± 12.1	44.0 ± 11.5	42.8 ± 10.9	43.2 ± 11.7	42.5 ± 12.3	42.9 ± 11.2	0.999 ^a
Male, n (%)	7 (46.7)	8 (53.3)	7 (46.7)	9 (60.0)	8 (53.3)	6 (40.0)	8 (53.3)	53 (50.5)	0.977 ^b
BMI, kg/m ² , mean ± SD	24.1 ± 3.1	23.8 ± 2.9	24.5 ± 3.3	23.9 ± 3.0	24.2 ± 3.2	24.0 ± 2.8	23.7 ± 3.1	24.0 ± 3.0	0.994 ^a
ASA physical status I/II, n	10/5	9/6	11/4	10/5	9/6	10/5	11/4	70/35	0.991 ^b

Notes: ASA, American Society of Anesthesiologists; BMI, body mass index; Sem, semester. Each semester contributed three residents and 15 patients (5 per resident). ^a One-way ANOVA computed from the semester means and SDs. ^b Fisher–Freeman–Halton exact test (Monte Carlo, 20,000 tables).

3.2. BIS before induction and after incision

The trajectories of BIS through induction and early maintenance are shown in Figure 2 and summarized in Table 2. Baseline BIS differed between semesters ($F(6,98)=7.98$, $p<0.001$; $\eta^2=0.328$, 95% CI 0.146–0.423), with semester means ranging from 85.7 in semester IV to 96.3 in semester III, but the differences showed no linear relation to training (slope 0.12 BIS units per semester, 95% CI –0.43 to 0.67; $p=0.662$), and the pooled baseline means of the junior, intermediate and senior groups were identical (91.0 in each).

BIS fell by 36.3–53.0 units from T1 to T2 in every semester; pooled by training level, the mean fall was 50.5 units in juniors, 46.4 in intermediate residents and 43.0 in seniors, mirroring the T2 gradient (the SD of the change cannot be derived without within-patient correlations). Five minutes after incision, BIS no longer differed between semesters ($F(6,98)=1.30$, $p=0.266$) and showed no trend ($p=0.947$); every semester mean lay within the 40–60 target range (50.3–57.0).

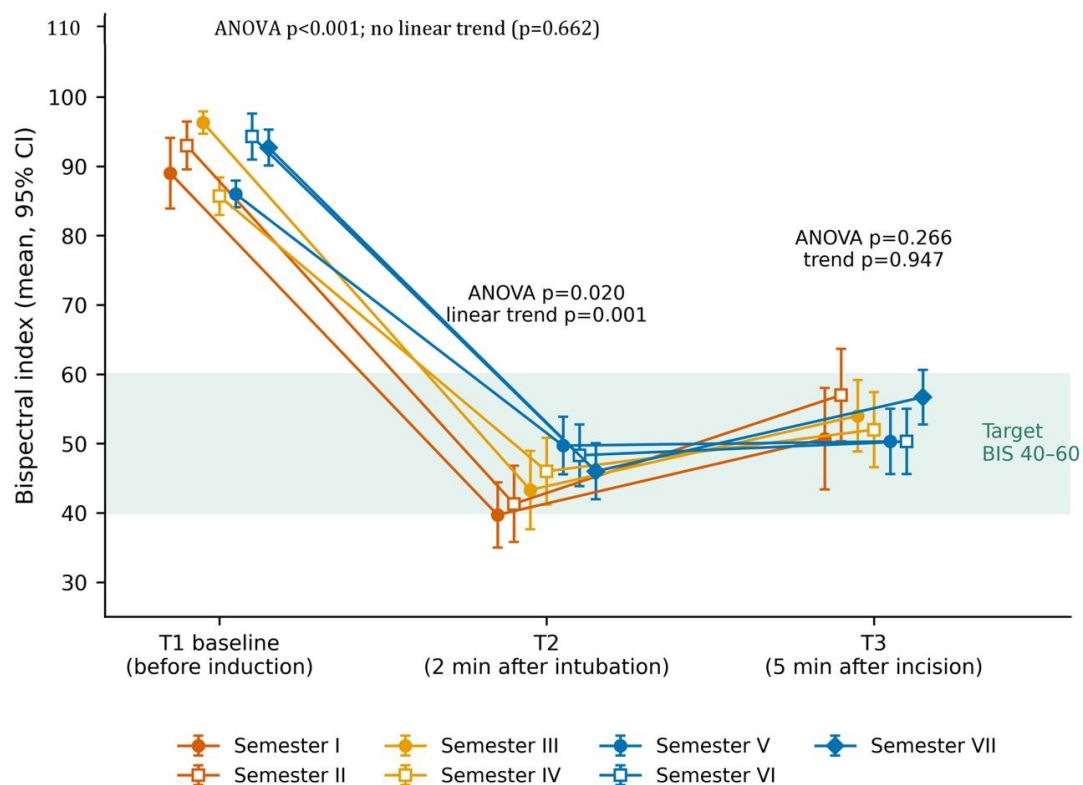


Figure 2. Bispectral index (BIS) before induction (T1), 2 min after intubation (T2) and 5 min after skin incision (T3) by residency semester. Points are means with 95% CIs (n=15 per semester), offset horizontally for legibility. Colors denote training level (vermillion, junior; orange, intermediate; blue, senior) and the shaded band the BIS 40–60 target. ANOVA and linear-trend p values are shown for each time point.

Table 2. Bispectral index before induction, after intubation and after incision by residency semester

Semester (training level)	T1 baseline	T2 2 min after intubation	T3 5 min after incision
I (junior)	89.0 ± 9.2 (83.9–94.1)	39.7 ± 8.5 (35.0–44.4)	50.7 ± 13.2 (43.4–58.0)
II (junior)	93.0 ± 6.2 (89.6–96.4)	41.3 ± 9.9 (35.8–46.8)	57.0 ± 12.0 (50.4–63.6)

Semester (training level)	T1 baseline	T2 2 min after intubation	T3 5 min after incision
III (intermediate)	96.3 ± 2.9 (94.7–97.9)	43.3 ± 10.2 (37.7–48.9)	54.0 ± 9.3 (48.8–59.2)
IV (intermediate)	85.7 ± 4.9 (83.0–88.4)	46.0 ± 8.7 (41.2–50.8)	52.0 ± 9.8 (46.6–57.4)
V (senior)	86.0 ± 3.5 (84.1–87.9)	49.7 ± 7.5 (45.5–53.9)	50.3 ± 8.5 (45.6–55.0)
VI (senior)	94.3 ± 6.0 (91.0–97.6)	48.3 ± 8.0 (43.9–52.7)	50.3 ± 8.5 (45.6–55.0)
VII (senior)	92.7 ± 4.7 (90.1–95.3)	46.0 ± 7.3 (42.0–50.0)	56.7 ± 7.1 (52.8–60.6)
Between-semester comparison			
One-way ANOVA F(6,98); p	7.98; <0.001	2.66; 0.020	1.30; 0.266
Welch ANOVA p	<0.001	0.029	0.210
η^2 (95% CI)	0.328 (0.146–0.423)	0.140 (0.003–0.223)	0.074 (0.000–0.134)
Linear trend, BIS units per semester (95% CI); p	0.12 (–0.43 to 0.67); 0.662	1.40 (0.57–2.24); 0.001	0.03 (–0.93 to 1.00); 0.947
Deviation from linearity, p	<0.001	0.434	0.180
Pooled mean ± SD; junior / intermediate / senior	91.0 ± 8.0 / 91.0 ± 6.7 / 91.0 ± 6.0	40.5 ± 9.1 / 44.7 ± 9.4 / 48.0 ± 7.6	53.9 ± 12.8 / 53.0 ± 9.4 / 52.4 ± 8.4

Notes: Values are mean ± SD (95% CI of the mean), n=15 per semester. ANOVA statistics, η^2 and the linear contrast (semester scores 1–7) were computed from the semester means, SDs and sizes; the linear trend estimate is the least-squares slope across semester means. Pooled training-level means and SDs were derived exactly from the semester statistics (junior n=30, intermediate n=30, senior n=45).

3.3. BIS 2 min after intubation across semesters

Mean BIS at T2 increased from 39.7 ± 8.5 in semester I to 49.7 ± 7.5 in semester V and was 48.3 ± 8.0 and 46.0 ± 7.3 in semesters VI and VII (Figure 3). Between-semester differences were significant by one-way ANOVA ($F(6,98)=2.66$, $p=0.020$) and by Welch's ANOVA ($p=0.029$), with η^2 of 0.140 (95% CI 0.003–0.223) and ω^2 of 0.087. A linear trend accounted for 69% of the between-semester variation, with BIS rising by 1.40 units per semester (95%

CI 0.57–2.24; $p=0.001$), and the residual deviation from linearity was not significant ($p=0.434$). The patient-level Pearson correlation between semester and T2 BIS, recovered exactly from the semester summaries, was $r=0.31$. Pooled by training level, mean T2 BIS was 40.5 ± 9.1 in juniors, 44.7 ± 9.4 in intermediate residents and 48.0 ± 7.6 in seniors ($F(2,102)=6.90$, $p=0.002$); juniors were 7.5 BIS units below seniors (95% CI –11.5 to –3.5; $p<0.001$; Hedges g –0.90). The junior–intermediate (–4.2 units; $p=0.088$) and intermediate–senior (–3.4 units; $p=0.109$) steps were not individually significant.

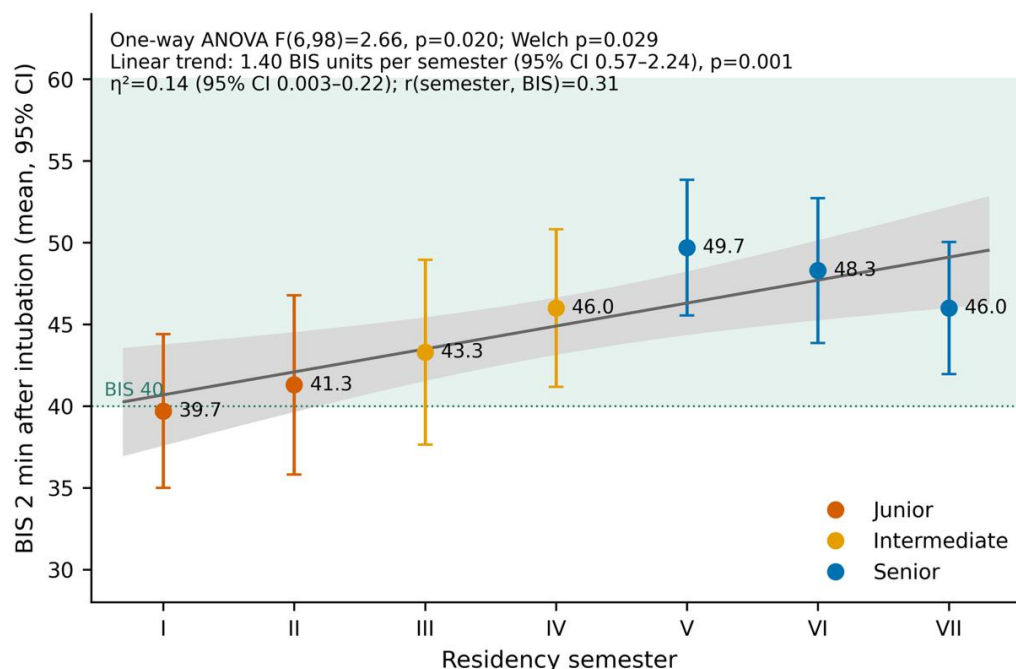


Figure 3. BIS 2 min after intubation by residency semester. Points are means with 95% CIs; the gray line and band are the least-squares linear trend across semester means and its 95% confidence band. The shaded area marks BIS 40–60 and the dotted line BIS 40. All statistics were computed from the semester means and SDs.

Of the 21 pairwise semester comparisons (Table 3 and Figure 4), only semester I versus semester V reached family-wise significance (difference 10.0 BIS units, simultaneous 95% CI 0.7–19.3; Games–Howell $p=0.029$; Tukey–Kramer

$p=0.033$). Semester I versus VI (8.6 units; $p=0.100$) and semester II versus V (8.4 units; $p=0.161$) pointed in the same direction, and no comparison between later and earlier semesters favored the earlier semester by more than 3.7 units.

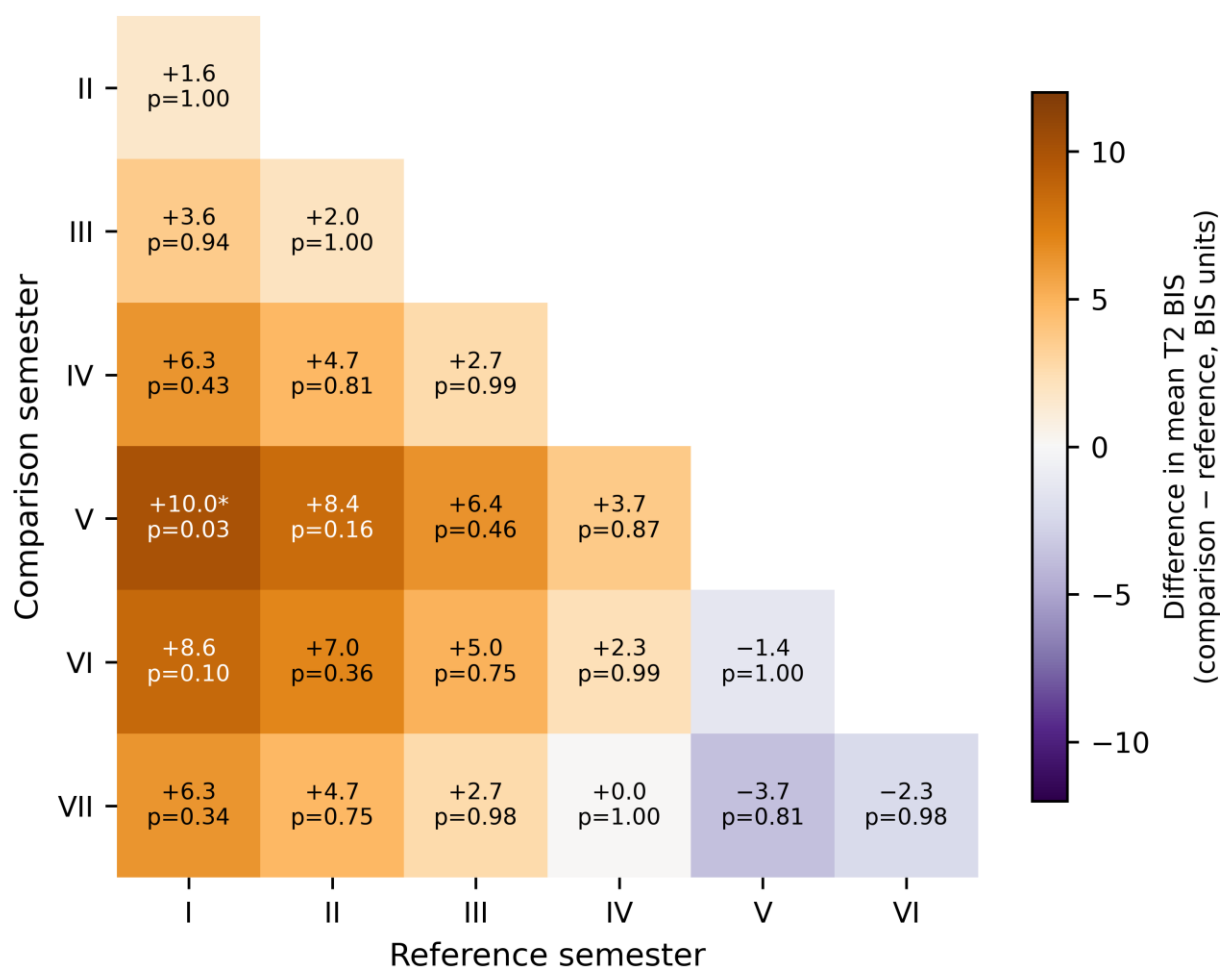


Figure 4. Pairwise differences in mean BIS 2 min after intubation between semesters. Each cell gives the difference (row semester minus column semester, BIS units) and the Games-Howell family-wise adjusted p value; *p<0.05. Warm colors indicate a higher, lighter BIS in the later semester.

Table 3. Pairwise comparisons of bispectral index 2 min after intubation between semesters

Comparison (later vs earlier semester)	Mean difference, BIS units	Simultaneous 95% CI	Games-Howell p	Tukey-Kramer p
II vs I	1.6	-9.1 to 12.3	0.999	0.999
III vs I	3.6	-7.3 to 14.5	0.937	0.914
IV vs I	6.3	-3.7 to 16.3	0.433	0.424
V vs I	10.0	0.7-19.3	0.029	0.033
VI vs I	8.6	-1.0 to 18.2	0.100	0.103
VII vs I	6.3	-2.9 to 15.5	0.339	0.424
III vs II	2.0	-9.6 to 13.6	0.998	0.996
IV vs II	4.7	-6.1 to 15.5	0.807	0.751
V vs II	8.4	-1.8 to 18.6	0.161	0.120
VI vs II	7.0	-3.5 to 17.5	0.365	0.296
VII vs II	4.7	-5.4 to 14.8	0.754	0.751
IV vs III	2.7	-8.3 to 13.7	0.985	0.978
V vs III	6.4	-4.0 to 16.8	0.463	0.405
VI vs III	5.0	-5.7 to 15.7	0.746	0.693
VII vs III	2.7	-7.7 to 13.1	0.979	0.978
V vs IV	3.7	-5.7 to 13.1	0.869	0.903
VI vs IV	2.3	-7.4 to 12.0	0.987	0.991

Comparison (later vs earlier semester)	Mean difference, BIS units	Simultaneous 95% CI	Games-Howell p	Tukey-Kramer p
VII vs IV	0.0	−9.3 to 9.3	1.000	1.000
VI vs V	−1.4	−10.4 to 7.6	0.999	0.999
VII vs V	−3.7	−12.3 to 4.9	0.813	0.903
VII vs VI	−2.3	−11.2 to 6.6	0.980	0.991

Notes: Positive differences indicate a higher (lighter) BIS in the later semester. Games-Howell intervals and p values allow for unequal variances and control the family-wise error rate over all 21 comparisons; Tukey-Kramer values are shown for comparison. Computed from the semester means and SDs (n=15 each).

3.4. Category of BIS after intubation by training level

BIS <40 at T2 followed 14 of 30 junior-led inductions (46.7%, 95% CI 30.2–63.9), 7 or 8 of 30 intermediate-led inductions (23.3–26.7%) and 5 of 45 senior-led inductions (11.1%, 95% CI 4.8–23.5); BIS within the 40–60 target was reached in 43.3%, 63.3–66.7% and 80.0%, respectively, and BIS >60 in 10.0%, 10.0% and 8.9% (Table 4 and Figure 5). Across all 105 patients, 24.8–25.7% had BIS <40 at T2. The distribution differed by training level (Fisher-Freeman-Halton exact p=0.009 under either intermediate

count; Cramér's V 0.25), with a monotonic decline in BIS <40 across levels (Cochran-Armitage $z=-3.45$, $p<0.001$). Each step up in training level multiplied the odds of a higher BIS category by 1.99–2.00 (95% CI 1.21–3.31; $p=0.007$), but the cut-point analysis showed that this common OR reflected the deep tail alone: the OR per level for BIS ≥40 versus <40 was 2.62–2.66 (lower 95% limit ≥1.48), whereas the OR for BIS >60 versus ≤60 was 0.93 (95% CI 0.43–2.03).

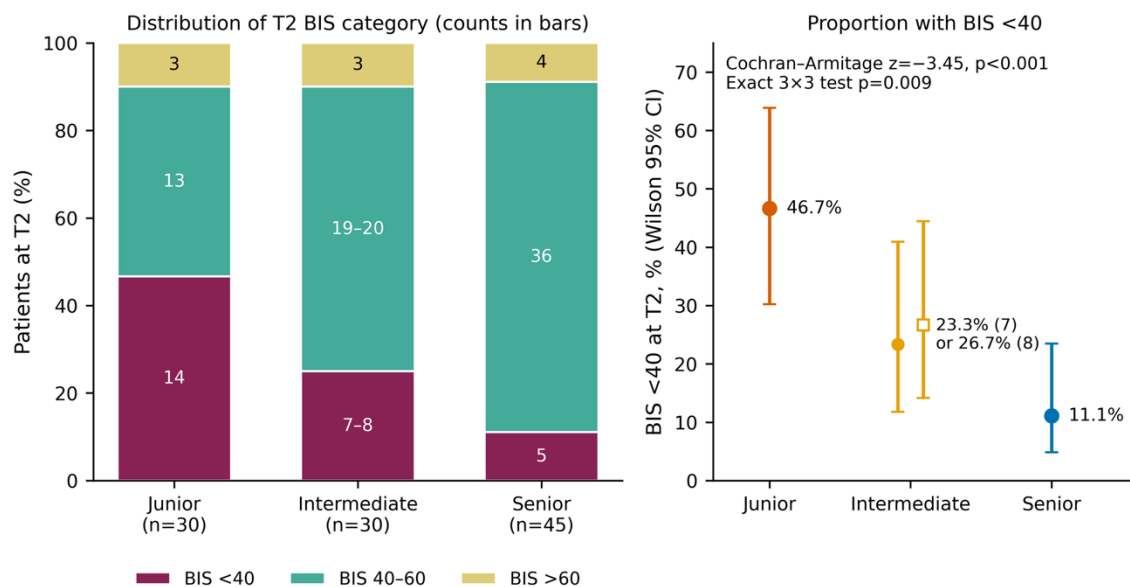


Figure 5. Category of BIS 2 min after intubation by training level. Left panel: distribution of patients across categories, with counts printed in the bars (the intermediate group's count of BIS <40 is 7 or 8). Right panel: proportion with BIS <40 with Wilson 95% CIs; both admissible intermediate values are shown (filled circle, 7 of 30; open square, 8 of 30).

Table 4. Category of bispectral index 2 min after intubation by training level

Training level	BIS <40, n (%) [95% CI]	BIS 40–60, n (%) [95% CI]	BIS >60, n (%) [95% CI]
Junior, semesters I–II (n=30)	14 (46.7) [30.2–63.9]	13 (43.3) [27.4–60.8]	3 (10.0) [3.5–25.6]
Intermediate, semesters III–IV (n=30)	7–8 (23.3–26.7) [11.8–44.4]	20–19 (66.7–63.3) [48.8–78.1]	3 (10.0) [3.5–25.6]
Senior, semesters V–VII (n=45)	5 (11.1) [4.8–23.5]	36 (80.0) [66.2–89.1]	4 (8.9) [3.5–20.7]
All patients (n=105)	26–27 (24.8–25.7)	69–68 (65.7–64.8)	10 (9.5) [5.3–16.6]
Association with training level			
Fisher-Freeman-Halton exact test	p=0.009 (intermediate 7) and p=0.009 (intermediate 8); Pearson $\chi^2(4)=12.65$ –12.95; Cramér's V=0.25–0.25		
Cochran-Armitage trend, BIS <40	z=−3.45 and −3.45; p<0.001 in both		
Proportional-odds model, OR per level toward higher BIS	1.99–2.00 (95% CI 1.21–3.31); p=0.007		
Cut-point-specific ORs per level	BIS ≥40 vs <40: 2.62–2.66 (lower 95% limit ≥1.48); BIS >60 vs ≤60: 0.93 (95% CI 0.43–2.03)		

Notes: 95% CIs are Wilson score intervals. The intermediate group's deep-anaesthesia count is 7 or 8 (the source reported 'approximately 25%' and drew 25.0% of 30); every statistic is given for both admissible values. The exact test enumerated all 24,696 (7/20/3) and 26,376 (8/19/3) tables with the observed margins. Training level was scored 1–3; cut-point ORs come from separate binary logistic models.

3.5. Junior versus senior residents

Junior residents gave 0.50 mg/kg more propofol than senior residents (2.4 ± 0.3 vs 1.9 ± 0.2 mg/kg; 95% CI 0.37–0.63; $p < 0.001$; Hedges g 2.02, 95% CI 1.46–2.59), a dose 20.8% lower in seniors, and reached loss of consciousness 7.0 s sooner (75 ± 15 vs 82 ± 12 s; 95% CI –13.6 to –0.4; $p = 0.037$; g –0.52), as shown in Table 5 and Figure 6. The ratio of mean dose to mean induction time corresponded to roughly 1.9 mg/kg/min in juniors and 1.4 mg/kg/min in seniors. The risk difference for BIS <40 was 35.6 percentage points (95% CI 15.0–53.9; OR 6.79, 95% CI 1.92–28.30; RR 4.20, 95% CI 1.69–10.44; $p < 0.001$), corresponding to one additional over-deep induction for every 2.8 junior-led inductions (95% CI 1.9–6.7; the reciprocal of the risk difference). Juniors reached the target range less often (43.3% vs 80.0%; RD –36.7 percentage points, 95% CI

–55.0 to –14.4; $p = 0.001$), whereas BIS >60 was equally uncommon (10.0% vs 8.9%; $p = 1.000$). Hypotension occurred in 30.0% of junior-led and 17.8% of senior-led inductions (RD 12.2 percentage points, 95% CI –6.8 to 32.0; OR 1.96, 95% CI 0.58–6.85; $p = 0.265$). After Holm–Bonferroni adjustment across the four secondary outcomes, the differences in propofol dose ($p < 0.001$) and BIS <40 ($p = 0.003$) remained significant, whereas the difference in time to loss of consciousness did not (adjusted $p = 0.074$). The CI of the risk difference for hypotension was wide; with 30 and 45 patients, only a junior incidence of at least 48% (a difference of about 30 percentage points) would have been detectable with 80% power. The standardized effects in this section derive from an extreme-groups comparison and overstate those expected across the full training range.

Table 5. Induction outcomes in junior versus senior residents

Outcome	Junior (n=30)	Senior (n=45)	Difference (95% CI)	Effect size (95% CI)	p	Holm p
Propofol induction dose, mg/kg	2.4 ± 0.3	1.9 ± 0.2	0.50 (0.37–0.63)	g 2.02 (1.46–2.59)	<0.001	<0.001
Time to loss of consciousness, s	75 ± 15	82 ± 12	–7.0 (–13.6 to –0.4)	g –0.52 (–0.99 to –0.05)	0.037	0.074
BIS 2 min after intubation	40.5 ± 9.1	48.0 ± 7.6	–7.5 (–11.5 to –3.5)	g –0.90 (–1.39 to –0.42)	<0.001	—
BIS <40, n (%)	14 (46.7)	5 (11.1)	RD 35.6 pp (15.0–53.9)	OR 6.79 (1.92–28.30); RR 4.20 (1.69–10.44)	<0.001	0.003
BIS 40–60, n (%)	13 (43.3)	36 (80.0)	RD –36.7 pp (–55.0 to –14.4)	OR 0.20 (0.06–0.60)	0.001	—
BIS >60, n (%)	3 (10.0)	4 (8.9)	RD 1.1 pp (–12.4 to 17.6)	OR 1.14 (0.15–7.31)	1.000	—
Hypotension (MAP fall >20%), n (%)	9 (30.0)	8 (17.8)	RD 12.2 pp (–6.8 to 32.0)	OR 1.96 (0.57–6.85); RR 1.69 (0.73–3.88)	0.265	0.265

Notes: BIS, bispectral index; g , Hedges g ; MAP, mean arterial pressure; OR, conditional maximum-likelihood odds ratio with exact 95% CI; pp, percentage points; RD, risk difference (Newcombe hybrid score CI); RR, risk ratio. Continuous outcomes are mean $\pm$ SD and were compared with Welch's t test; binary outcomes with Fisher's exact test. Junior BIS values are pooled from semesters I–II and senior values from semesters V–VII; semester-level values for dose, time and hypotension were not reported, so semesters III–IV are not represented in this table. Holm p : Holm–Bonferroni adjustment across the four prespecified secondary outcomes (dose, time to loss of consciousness, BIS <40, hypotension).

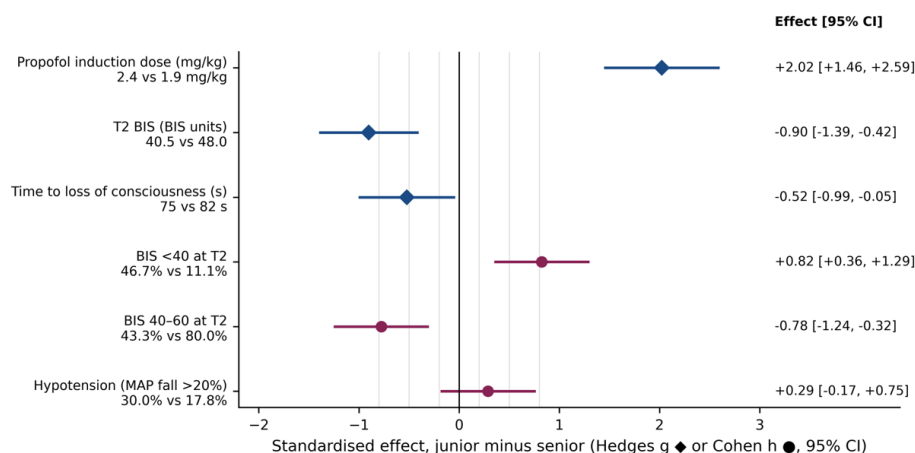


Figure 6. Standardized junior-minus-senior differences in induction outcomes. Diamonds are Hedges g for continuous outcomes and circles Cohen h for proportions, with 95% CIs; group values are given under each label. Grey lines mark conventional small (0.2), medium (0.5) and large (0.8) effects.

3.6. Shape of the T2 trajectory

In post hoc analyses of curvature, BIS at T2 rose by 2.47 units per semester across semesters I–V (95% CI 1.00–3.94; $p = 0.001$) and did not change significantly across semesters V–VII (–1.85 units per semester, 95% CI –4.65 to

0.95; $p = 0.190$); the quadratic contrast over all seven semesters pointed to a concave trajectory ($p = 0.095$). Omitting any single semester left a positive linear trend (1.21–2.03 units per semester; all $p \leq 0.027$). Mean $\pm$ 2 SD of BIS at T2 lay between 21.5 and 64.7 in every semester, within the bounds of the scale, consistent with approximately normal

within-semester distributions. In every senior semester the interval extended below 40, in keeping with the five senior-led inductions that reached BIS <40.

3.7. Sensitivity to clustering of patients within residents

Results of the design-effect analysis are presented in Table 6 and Figure 7. The difference in propofol dose remained significant at every ICC examined, including 0.50

($p < 0.001$). The junior–senior difference in BIS <40 lost significance only at an ICC of 0.53, the junior–senior difference in mean T2 BIS at 0.50 and the linear trend across semesters at 0.35; at an ICC of 0.20 the RD for BIS <40 was 35.6 percentage points with a design-effect-adjusted 95% CI of 7.9 to 60.1. Two results depended on the reference distribution: the seven-group ANOVA for T2 ($p = 0.062$) and the difference in time to loss of consciousness ($p = 0.052$) were not significant when referred to resident-level degrees of freedom even at an ICC of 0. The difference in hypotension was not significant at any ICC.

Table 6. Sensitivity of the main comparisons to clustering of patients within residents

Comparison	ICC 0	ICC 0.10	ICC 0.20	ICC 0.30	ICC 0.50	Tipping-point ICC
Propofol dose, junior vs senior	<0.001	<0.001	<0.001	<0.001	<0.001	>1 (significant at every ICC)
BIS <40, junior vs senior	<0.001	0.003	0.010	0.019	0.045	0.53
T2 BIS, junior vs senior	0.003	0.008	0.016	0.026	0.051	0.50
T2 BIS, linear trend over semesters	0.005	0.014	0.027	0.042	0.075	0.35
T2 BIS, seven-group ANOVA	0.062	0.151	0.256	0.358	0.530	not significant with resident-level df
Time to loss of consciousness, junior vs senior	0.052	0.094	0.135	0.173	0.239	not significant with resident-level df
Hypotension, junior vs senior	0.216	0.295	0.356	0.404	0.475	not significant at any ICC

Notes: ICC, intraclass correlation coefficient of the outcome among patients managed by the same resident. Design effect = $1 + (m - 1) \times \text{ICC}$ with $m = 5$ patients per resident; test statistics were divided by $\sqrt{\text{design effect}}$. t-based tests use resident-level degrees of freedom (13 for junior vs senior; 14 for the seven-semester analyses), which is conservative even at ICC 0; z-based tests (binary outcomes) use the normal reference. Tipping-point ICC: value at which the adjusted p reaches 0.05. Binary outcomes are referred to the normal distribution because their test statistics are score statistics without a residual-variance estimate; continuous outcomes use t distributions with resident-level degrees of freedom, which makes their tipping points conservative.

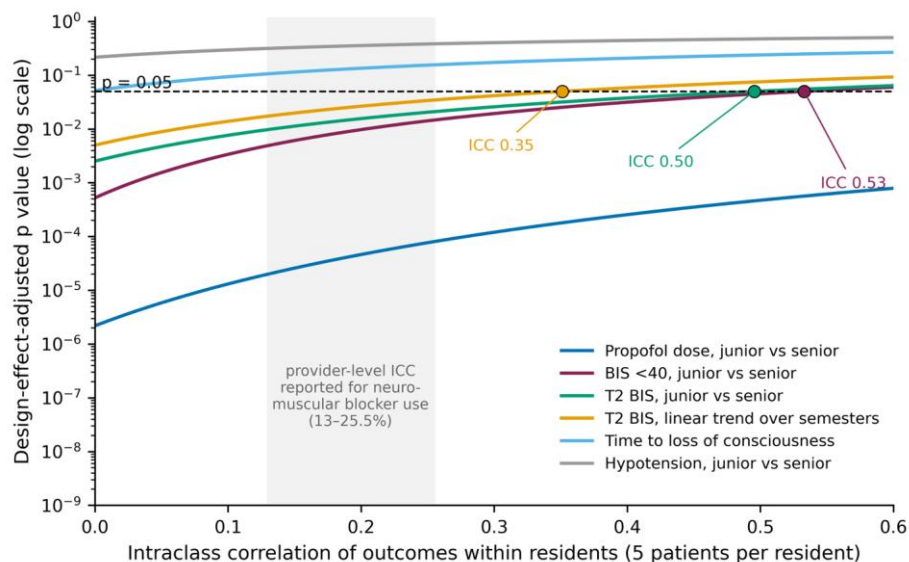


Figure 7. Sensitivity of the main comparisons to clustering of patients within residents. Curves show the p value after design-effect adjustment for five patients per resident as a function of the assumed intraclass correlation (ICC); circles mark the ICC at which p reaches 0.05. The gray band shows the provider-level ICC range reported for intraoperative drug use in a large multicenter cohort.²⁶

3.8. Bounds analysis

The bounds analysis over all datasets consistent with the published margins is summarized in Table 7 and Figure 8. Within the junior stratum, 10 admissible 2×2 tables of BIS <40 by hypotension gave ORs from 0.027 to 57.0, with significant associations in both directions (5 of 10 tables);

within the senior stratum, 6 tables gave ORs from 0.35 to 117.9. Across the 60 stratum combinations the Mantel-Haenszel OR ranged from 0 to infinity, and 21 combinations were significantly positive and 6 significantly negative. The association between BIS <40 and hypotension is therefore not identifiable from these data. In model A, the Firth-adjusted OR for BIS <40 in junior-led versus senior-

led inductions remained at or above 5.99 ($p \leq 0.006$) under every admissible allocation of ASA status and sex, with a median of 6.35 across sampled datasets and significance in all of them; the adjusted ORs for ASA II and male sex were

themselves not identifiable. In model B, neither the adjusted OR of hypotension for BIS <40 (median 0.99; significant in 2.6% of datasets) nor that for junior level (median 1.95; significant in 0.8%) was identifiable.

Table 7. Sensitivity analysis over all datasets consistent with the published margins (bounds analysis)

Analysis / coefficient	Admissible datasets / central OR	Attainable OR range	Datasets with $p < 0.05$	Finding
Bivariate: BIS <40 → hypotension (exact enumeration)				
Junior stratum (30 patients; 14 BIS <40; 9 hypotensive)	10 tables	0.027 to 57.0	5/10 (both directions)	Not identifiable
Senior stratum (45 patients; 5 BIS <40; 8 hypotensive)	6 tables	0.35 to 117.9	3/6 (positive only)	Not identifiable
Stratified, Mantel-Haenszel	60 combinations	0 to ∞	27/60 (21 positive, 6 negative)	Not identifiable
Multivariable Firth logistic regression (4,000 sampled datasets + hill-climbing)				
Model A, BIS <40: junior vs senior (adjusted for ASA II, male sex)	median 6.35 (2.5–97.5th pct 6.11–7.77)	5.99 to 222.7	100.0%	Robust: least favourable OR 5.99 (95% CI 1.91–18.77); $p \leq 0.006$ in every dataset found
Model A: ASA II	median 1.07 (2.5–97.5th pct 0.30–3.02)	→0 to $>10^4$	3.3%	Not identifiable
Model A: male sex	median 1.02 (2.5–97.5th pct 0.32–3.20)	→0 to $>10^4$	3.0%	Not identifiable
Model B, hypotension: BIS <40 (adjusted for level, ASA II, sex)	median 0.99 (2.5–97.5th pct 0.23–3.42)	→0 to $>10^4$	2.5%	Not identifiable
Model B: junior vs senior	median 1.95 (2.5–97.5th pct 1.23–3.16)	→0 to 6728.6	0.8%	Not identifiable

Notes: Patient-level cross-classification of BIS category, hypotension, ASA status and sex was not available. Bivariate bounds enumerate every 2×2 table compatible with the stratum margins (cell a = patients with both BIS <40 and hypotension); ORs use the Haldane-Anscombe correction for zero cells and p values are from Fisher's exact test. Multivariable bounds permute each variable within its training level (preserving every published margin), fit Firth-penalized logistic regression (Wald p) to 4,000 random admissible datasets, and add 12-restart hill-climbing toward each coefficient's minimum and maximum. 'Not identifiable' = the attainable range spans OR = 1.

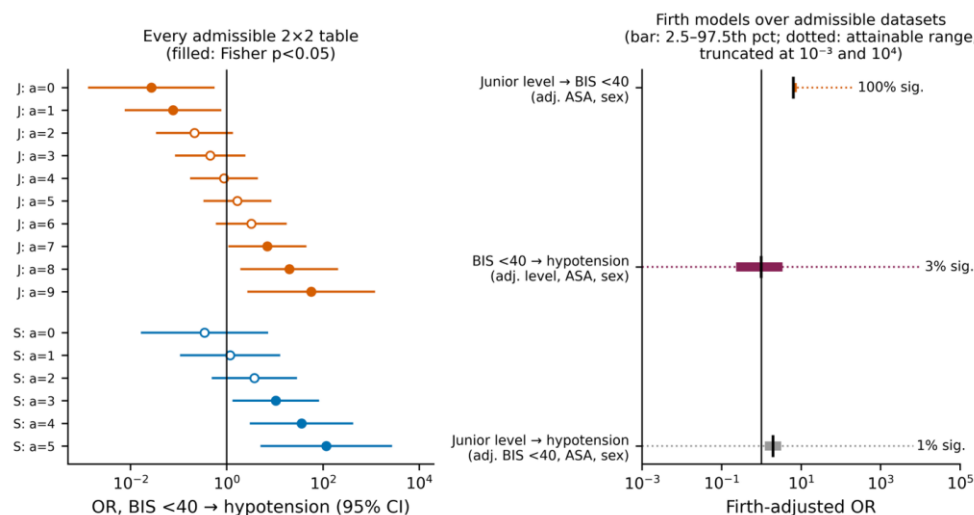


Figure 8. Bounds analysis over all datasets consistent with the published margins. Left panel: odds ratio (95% CI) for hypotension given BIS <40 in each admissible 2×2 table within the junior (J) and senior (S) strata; a is the number of patients with both events, and filled markers indicate Fisher $p < 0.05$. Right panel: Firth-adjusted odds ratios across admissible allocations of ASA status, sex and hypotension; bars span the 2.5th–97.5th percentiles of 4,000 sampled datasets, black ticks mark medians, dotted lines the attainable range found by hill-climbing, and labels the percentage of datasets with $p < 0.05$.

3.9. Precision and sample sizes for future studies

The 95% CI for BIS <40 among juniors spanned 33.6 percentage points, and estimating it to within ± 10 percentage points would require 96 patients (Table 8). A two-group study powered at 80% to confirm the observed difference

in BIS <40 would need 25 patients per group if patients were independent, or 9 residents per group with five patients each at an ICC of 0.20; detecting the observed difference in hypotension would need 191 patients per group, or 69 residents per group at the same ICC (Figure 9).

Table 8. Precision of the present estimates and sample sizes for future cluster-aware studies

Estimate	Observed	Width of Wilson 95% CI	n for ± 10 pp	n for ± 5 pp
Precision of the present estimates				
BIS <40, junior (14/30)	46.7%	33.6 pp	96	383
BIS <40, senior (5/45)	11.1%	18.7 pp	38	152
Hypotension, junior (9/30)	30.0%	31.2 pp	81	323
Hypotension, senior (8/45)	17.8%	22.0 pp	57	225
Residents per group for a future two-group study (5 patients each; 80% power; $\alpha=0.05$)				
Scenario	Patients per group (independent)	ICC 0.10	ICC 0.20	ICC 0.30
BIS <40: 46.7% vs 11.1% (observed)	25	7	9	11
BIS <40: 30% vs 15%	121	34	44	54
Hypotension: 30.0% vs 17.8% (observed)	191	54	69	84
Hypotension: 25% vs 15%	250	70	90	110

Notes: pp, percentage points. Required n for a given half-width uses $n = z^2 p(1-p)/d^2$. Two-group sample sizes use the normal approximation for two proportions and are inflated by the design effect $1 + 4 \times \text{ICC}$, then divided by 5 patients per resident (rounded up).

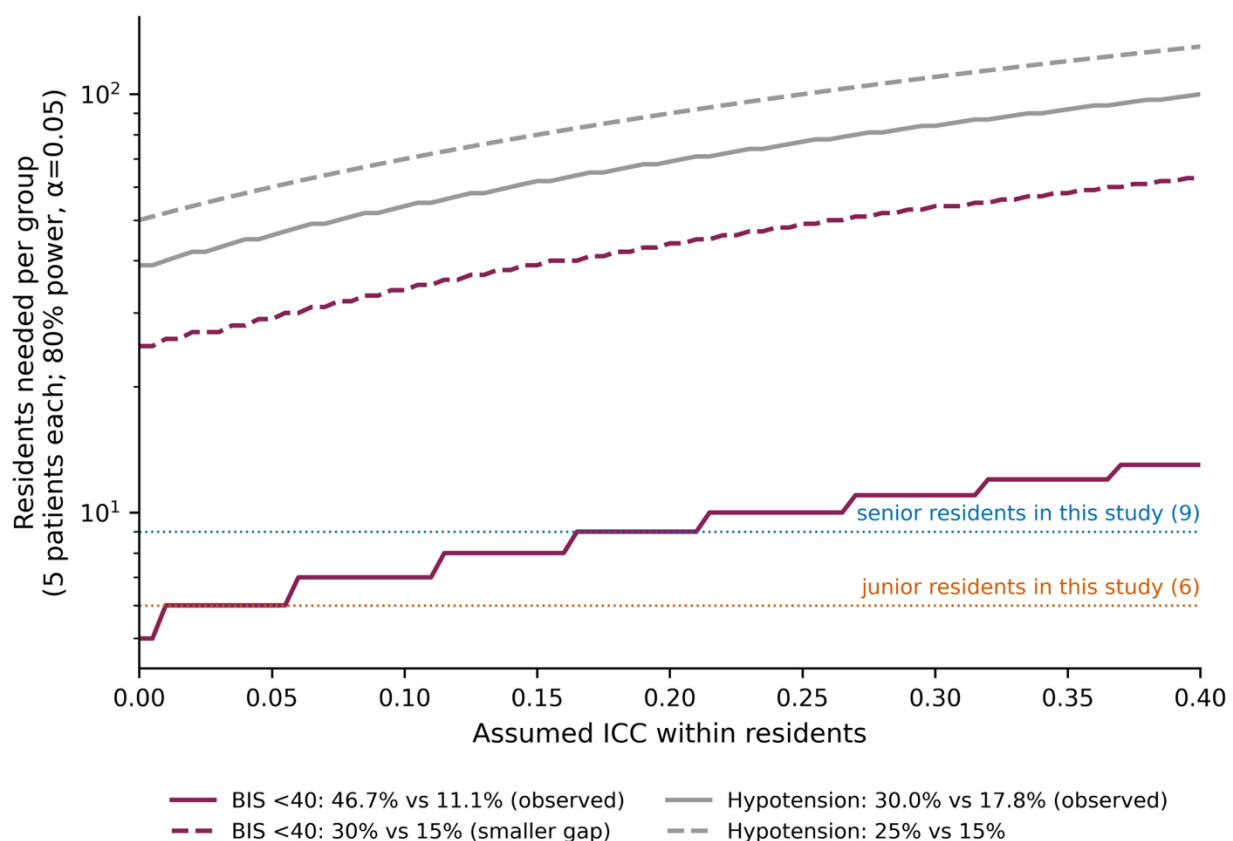


Figure 9. Residents needed per group in a future two-group study with five patients per resident, 80% power and two-sided $\alpha=0.05$, as a function of the intraclass correlation. Solid lines use the differences observed in this study and dashed lines smaller differences. Dotted horizontal lines mark the numbers of junior (6) and senior (9) residents in the present study.

4. Discussion

4.1. Principal findings

In this study at Dr. Saiful Anwar Regional General Hospital, residents in their first year induced markedly deeper hypnosis 2 min after intubation than residents in their final years, despite a standardized drug sequence, the same

monitoring and in-room attending supervision. Nearly half of junior-led inductions produced BIS <40, compared with about one in nine senior-led inductions; mean BIS rose steadily from semester I to semester V and not thereafter; and juniors gave one-quarter more propofol per kilogram. The difference was confined to the deep tail of the distribution — light hypnosis was equally uncommon at all levels — and it was robust to any plausible confounding by ASA

status or sex and to within-resident clustering up to an ICC of 0.53. By 5 min after incision, depth no longer differed between semesters. Whether the deeper inductions caused more hypotension cannot be determined from the available data.

4.2. How deep is too deep after induction?

A BIS below 40 two minutes after intubation is a single-time-point measure and should not be equated with the cumulative exposures studied in outcome trials. The trial that found a benefit of lighter anesthesia compared target values maintained throughout surgery,² and those that found none did so even though EEG guidance halved the duration of suppression;^{3,4} in the BALANCED trial, targeting a BIS of 50 rather than 35 did not significantly reduce 1-year mortality.⁵ In a cohort of 5,927 older patients, cumulative time at BIS <40 during maintenance was not independently associated with 1-year mortality.⁵⁸ Our patients were 18–65 years old and ASA I–II, a population in which brief early suppression is unlikely to have measurable long-term consequences, and depth had converged by T3. The finding is therefore best read as an indicator of titration skill rather than of patient harm. It nevertheless identifies the phase in which inadvertent suppression is generated: the drug-related EEG suppression associated with delirium in meta-analyses is produced by the same propofol-dominated physiology,^{8,9,59} and audits show that sustained excess depth and inadvertent suppression are common in routine practice.^{1,10,11}

4.3. Why novices overshoot

Several mechanisms, summarized in Figure 10, probably combine. The first is dose: after the same fentanyl–midazolam co-induction, juniors gave 2.4 mg/kg and seniors 1.9 mg/kg, a 26% difference in a drug whose hypnotic effect is steeply dose-dependent. The second is speed. Juniors reached loss of consciousness sooner despite the larger dose, which implies faster injection; in a randomized trial in older patients, faster propofol injection increased the induction dose, the incidence of burst suppression and the depth of the BIS nadir.¹³ The third is the endpoint. Loss of the eyelash reflex occurs at BIS values of about 61–65,¹⁴ and because the effect-site concentration continues to rise for about 1–2 min after a bolus, an operator who injects until the reflex is unequivocally gone, or adds a 'safety' increment before giving a neuromuscular blocker, will carry the patient well below 40 by the time the airway is secured. Senior residents may anticipate this lag. Patient age, which lowers propofol requirements and raises BIS at equivalent exposure,^{12,60} cannot explain the gradient because patient age was identical across semesters, and the arousal produced by laryngoscopy, which can raise BIS after intubation,¹⁵ would if anything have narrowed the difference at T2.

Behavioral explanations are consistent with these pharmacological ones. Novices manage a high intrinsic and extraneous cognitive load during induction — drug preparation, airway equipment, an unfamiliar team — which leaves little capacity for continuous titration against a monitor. In simulation, novice anesthesia trainees fall into distinct performance clusters that differ in reported task load,⁶¹ and virtual-reality versus manikin simulation can alter the mental demand that trainees report.⁶² Under such load, a rule-based dose is easier than a titrated one, and fear of awareness during neuromuscular blockade biases errors toward depth. The asymmetry of our result fits this interpretation: juniors overshoot into BIS <40, and there was no

evidence that they more often left patients light, although the CI for the light tail (OR 0.93, 95% CI 0.43–2.03) does not exclude such an effect.

4.4. Post-induction hypotension

The 30.0% incidence of hypotension after junior-led induction and 17.8% after senior-led induction lie within the range of recent reports, which varies substantially with case mix and definition: one prospective cohort reported 26.95%, whereas other cohorts reported 31.1% or 73–82%.^{20–23} Propofol induction was identified as a risk factor for post-induction hypotension in an earlier large cohort,¹⁶ propofol produced more hypotension than ciprofol or remimazolam in randomized comparisons,^{17,18} and age and lower baseline pressure are established risk factors.¹⁹ A dose-dependent increase in hypotension among juniors would therefore be expected. Our data are compatible with a risk difference anywhere from –7 to +32 percentage points, the study could have detected only a difference of about 30 percentage points, and the bounds analysis showed that the published margins are equally compatible with deep inductions having caused all, none or the opposite of the excess hypotension. The correct conclusion is that the question remains open, not that over-deep induction is hemodynamically benign. Because post-induction hypotension was not associated with 30-day mortality in a large cohort, whereas intraoperative hypotension was associated with mortality and organ complications,^{20,21} future studies should measure both.

4.5. The shape of the learning curve

The gradient across semesters resembles the learning curves described for procedural skills, but it differs in kind. CUSUM analyses define competence as a sustained success rate on a binary task — bronchoscopic intubation after 15 procedures, hyperangulated videolaryngoscopy after a median of 12 attempts, caudal block after 9–13, transesophageal echocardiographic views after about 20 — and they have limited accuracy when used for certification.^{33–37} Titration has no pass–fail endpoint, and the relevant unit is not the number of procedures but the accumulation of inductions over months. Mean BIS reached its highest value in semester V and did not rise further, and the post hoc slope analysis showed a rise of about 2.5 units per semester up to semester V and no significant change thereafter, although the formal test of curvature was not conclusive ($p=0.095$). This suggests that most of the learning occurs during the first two years of training, when structured teaching has been shown to shorten procedural learning curves in inexperienced trainees.³⁸

The size of the training-level effect should be judged against the variation already documented between qualified clinicians. Provider-level variation explains 13–26% of the variance in neuromuscular blocker use,²⁶ 44% of the variation in nerve block use,²⁷ variation in vasopressor selection and inotrope use,^{28,29} and reliable differences in hypotension exposure between individual anesthesia professionals.^{24,25} That the junior–senior differences in BIS <40 and in mean T2 BIS survived design effects corresponding to ICCs of 0.53 and 0.50 — although the linear trend did so only up to 0.35 — suggests that the training-level gradient is larger than the between-resident variation one would expect within a training level. The ICC for post-induction BIS itself is unknown, and the benchmark used here comes from neuromuscular blocker use.

4.6. Supervision and the resident-supervisor dyad

Every induction in this study was supervised, so the outcome measured is the performance of the resident-supervisor dyad. Supervisors were asked to intervene only at hemodynamic and oxygenation thresholds, none of which concerned BIS. The deep inductions by juniors therefore occurred within a system that tolerated BIS values well below 40 as long as hemodynamics remained acceptable. Supervision intensity matters for patient outcomes in general — covering more overlapping rooms raises surgical morbidity and mortality⁴⁰ — and novices value direct attending supervision,⁴¹ and postoperative care guidelines require residents to assess recovery, determine readiness for discharge and transfer care safely under supervision.⁴² Just-in-time coaching before a high-stakes procedure improved first-attempt success and reduced cognitive load in inexperienced clinicians.⁴⁹ The same approach applied to the minute before and after a propofol bolus — agreeing a target BIS, an injection rate and a stopping rule — is a natural extension that would make supervisors' expectations explicit. Because supervisors were allocated by the on-call roster and their identity was not recorded, we cannot exclude that senior residents more often worked with particular attendings whose own titration style contributed to the gradient; recording supervisor identity would allow this to be separated from resident learning.

4.7. Implications for training

Post-induction BIS is cheap, automatically recorded, available in routine practice and directly interpretable, which makes it attractive as one component of workplace-based assessment. Current EPA ratings in anesthesiology are only moderately reliable and require many observations,^{30,31} milestone trajectories vary across both residents and programs,³² and EPA-based feedback shifts residents toward technical skills that can be observed.³⁹ An objective, per-case metric such as the proportion of inductions reaching BIS <40 or the time spent below 40 in the first 10 min could complement these ratings and provide the frequent low-stakes data that reliable assessment requires. EEG literacy can be taught quickly⁴³ and incorporated into residency curricula,⁴⁴ and departmental education increases the use of processed EEG.⁴⁵ Decision support may add what education alone cannot: advisory displays reduced BIS <40 and severe hypotension,⁴⁶ Bayesian advisory tools for propofol titration have been proposed specifically as teaching aids,⁴⁷ BIS-guided closed-loop delivery has been associated with lower propofol consumption than manual control,⁴⁸ and an education program without software guidance did not change hypotension exposure.⁵⁰ These observations suggest a package to be tested rather than adopted: an agreed target BIS and stopping rule announced before each induction, injection at a stated rate with a pause to allow for the bolus-effect-site lag, a short post-case review of the BIS trend with the supervisor, and periodic simulation of the lag. Whether such feedback reduces over-deep induction is the question for the trial outlined below.

4.8. Regional context

These findings are relevant beyond a single teaching hospital. Indonesian specialist training in anesthesiology is organized in semesters, so a semester-V resident is in the third year of training and the junior group comprised residents in their first year. In much of Southeast Asia and other low- and middle-income regions, anesthesiologists are scarce,^{51,52} residents carry a large share of the clinical workload, and depth monitoring is far from routine; only

13.7% of anesthesiologists in a three-country Asian survey monitored depth in every patient.⁵³ Where EEG guidance has been studied in the region it reduced anesthetic exposure and burst suppression,⁵⁴ and the BALANCED trial showed no significant difference in 1-year mortality between lighter and deeper BIS-targeted anesthesia.⁵ Competence can be assessed more systematically: Indonesian studies report improvement in clinical reasoning across training levels and better understanding of anesthetic pharmacokinetics after simulation, while the international ANTS framework supports assessment of nontechnical skills^{55–57} — and BIS-based measurement of induction depth fits naturally into this agenda because monitors are already available in teaching hospitals.

4.9. Baseline BIS and the measurement of depth

Baseline BIS differed between semesters by up to 10.6 units, and three semester means were below 90 in awake, unpremedicated adults. Because the differences had no relation to training level and the pooled baseline means of the three levels were identical, they do not explain the T2 gradient, but they deserve comment. Awake BIS values below 90 can arise from electromyographic interference, sensor placement or signal quality, and processed-EEG indices differ in how they respond to these factors and to age.^{60,63} Individual sensitivity to anesthetics also varies widely,⁵⁹ which is one reason why a fixed BIS threshold is an imperfect proxy for pharmacological depth. Future studies should record signal-quality and electromyography indices and the raw EEG alongside the processed number. They should also report the time course of BIS between induction and intubation rather than a single post-intubation value.

4.10. Research agenda

Three studies follow directly. First, a longitudinal cohort should follow the same residents from their first to their sixth semester, record every induction with continuous BIS, dose, injection rate, MAP and supervisor identity, and fit mixed-effects models with resident and supervisor random effects; this would replace the cross-sectional inference used here and estimate the ICC directly. Second, a cluster-randomized trial should allocate junior residents to BIS-anchored coaching and post-case feedback or to usual supervision. With five patients per resident and an ICC of 0.20, 9 residents per arm would detect a reduction in BIS <40 of the size observed here, 44 per arm a reduction from 30% to 15%, and 69 per arm the observed difference in hypotension (Table 8 and Figure 9) — numbers that require a multicenter design in Indonesia. Third, the outcome set should include the time spent below BIS 40 and in burst suppression during the first 10 min, post-induction and intraoperative hypotension defined by both relative and absolute thresholds, and patient-centered outcomes in older patients, in whom the consequences of excess depth are greatest.

4.11. Strengths and limitations

The study used a fully standardized drug sequence, prospective recording by an independent assessor, equal numbers of residents from every semester and the same monitor for every patient. The reports effect sizes with CIs for every comparison, uses exact tests where counts are small, quantifies the influence of clustering instead of ignoring it, and replaces untestable joint claims with a bounds analysis that distinguishes conclusions holding for every dataset compatible with the observations from those the data cannot support.

The limitations are substantial. Linked patient-level data were not available for this analysis, so associations between dose, depth and hypotension within patients could not be estimated and were bounded instead; MAP and heart rate at T1–T3 could not be reported; and the intermediate group's count of BIS <40 is known only to within one patient. The design is cross-sectional across training levels, so differences between semesters reflect both learning and differences between the particular residents enrolled; with three residents per semester, semester and resident identity are confounded. Patients were nested within residents, and the design-effect analysis is an approximation that assumes equal cluster sizes and a common ICC across outcomes. Supervision was not standardized beyond the safety thresholds, supervisor identity was not recorded, and the result therefore describes resident-supervisor dyads. The types of surgery were not available for analysis. Patients were young and ASA I–II, so the clinical consequences of brief deep hypnosis in this population are likely to be small and cannot be extrapolated to older or sicker patients. Finally, the study was not powered for hypotension, and it was conducted at a single Indonesian teaching hospital whose curriculum and supervision culture may differ from those elsewhere. Secondary outcomes compared only the junior and senior extremes, which inflates standardized effects. The bounds search was not exhaustive, and the clustering benchmark derives from a different outcome. Several secondary comparisons were made; the main conclusions survive Holm adjustment, but the analyses of curvature were post hoc and should be regarded as exploratory.

5. Conclusion

At an Indonesian teaching hospital, junior anesthesiology residents induced deeper hypnosis than senior residents under the same protocol and the same in-room supervision. BIS <40 two minutes after intubation followed 46.7% of junior-led and 11.1% of senior-led inductions (risk difference 35.6 percentage points, 95% CI 15.0–53.9), juniors gave 0.50 mg/kg more propofol, and mean BIS rose by about 2.5 units per semester up to the fifth semester and not thereafter. The difference in BIS <40 is robust to plausible clustering of patients within residents and to any confounding by ASA status or sex compatible with the data, whereas the link between deep induction and post-induction hypotension cannot be determined from the available data. Because the patients were young and ASA I–II and depth converged within minutes, the finding indexes titration skill rather than patient harm. Post-induction BIS is a candidate objective metric for formative assessment during the first two years of training. The next step is a multicenter, cluster-randomized trial of BIS-anchored coaching — an agreed target and stopping rule, a stated injection rate and a post-case review — with resident and supervisor identifiers, continuous BIS and patient-level hemodynamic outcomes.

6. Declarations

Ethical approval and consent to participate. The study was approved by the CMHC Research Center Ethics Committee, Indonesia (Approval No. 2024/8854) in 2024, before enrollment began. Written informed consent was obtained from all participating residents and patients. The study was conducted in accordance with the Declaration of Helsinki.

Availability of data and materials. The de-identified dataset and the analysis scripts supporting the findings are available from the corresponding author on reasonable request.

Artificial intelligence (AI) use statement. The authors declare that no generative artificial intelligence tools were used in the design, analysis, or writing of this manuscript.

Author contribution statement. RS: Conceptualization, Methodology, Investigation, Data curation, Writing – original draft. I: Supervision, Methodology, Writing – review & editing. BHL: Methodology, Validation, Writing – review & editing. RML: Supervision, Formal analysis, Writing – review & editing. All authors approved the final version of manuscript.

Funding statement. This study received no specific funding.

Conflict of interest (competing interests) statement. The authors declare no competing interests.

Acknowledgements. The authors thank the residents, supervising anesthesiologists and patients who took part in the study.

7. References

- Dexter F, Epstein RH, Marian AA. The Value of Auditing the Frequency of Inhalational Anesthetics With a Combination of Very Low Bispectral Index and High Fraction of the Age-Adjusted Minimum Alveolar Concentration. *Cureus*. 2024;16(12):e75036. doi: 10.7759/cureus.75036
- Evered LA, Chan MTV, Han R, et al. Anaesthetic depth and delirium after major surgery: a randomised clinical trial. *Br J Anaesth*. 2021;127(5):704–712. doi: 10.1016/j.bja.2021.07.021
- Wildes TS, Mickle AM, Ben Abdallah A, et al. Effect of Electroencephalography-Guided Anesthetic Administration on Postoperative Delirium Among Older Adults Undergoing Major Surgery: The ENGAGES Randomized Clinical Trial. *JAMA*. 2019;321(5):473–483. doi: 10.1001/jama.2018.22005
- Deschamps A, Ben Abdallah A, Jacobsohn E, et al. Electroencephalography-Guided Anesthesia and Delirium in Older Adults After Cardiac Surgery: The ENGAGES-Canada Randomized Clinical Trial. *JAMA*. 2024;332(2):112–123. doi: 10.1001/jama.2024.8144
- Short TG, Campbell D, Frampton C, et al. Anaesthetic depth and complications after major surgery: an international, randomised controlled trial. *Lancet*. 2019;394(10212):1907–1914. doi: 10.1016/S0140-6736(19)32315-3
- Sumner M, Deng C, Evered L, et al. Processed electroencephalography-guided general anaesthesia to reduce postoperative delirium: a systematic review and meta-analysis. *Br J Anaesth*. 2023;130(2):e243–e253. doi: 10.1016/j.bja.2022.01.006
- Cao L, Chen Q, Xiang Y, et al. Effect of processed electroencephalography-guided anesthesia on postoperative delirium: an updated systematic review and meta-analysis. *Int J Surg*. 2025;111(11):8540–8549. doi: 10.1097/JS9.0000000000003046
- Park SK, Han DW, Chang CH, et al. Association between Intraoperative Electroencephalogram Burst Suppression and Postoperative Delirium: A Systematic Review and Meta-analysis. *Anesthesiology*. 2025;142(1):107–120. doi: 10.1097/ALN.0000000000005255
- Xu A, Arnaut B, Scott DA, et al. Intraoperative electroencephalogram-derived measures and their association with postoperative delirium: a systematic review and meta-analysis. *Br J Anaesth*. 2025;135(6):1684–1703. doi: 10.1016/j.bja.2025.08.038
- Lele AV, Furman M, Myers J, et al. Inadvertent Burst Suppression During Total Intravenous Anesthesia in 112 Consecutive Patients Undergoing Spinal Instrumentation Surgery: A Retrospective Observational Quality Improvement Project. *J Neurosurg Anesthesiol*. 2022;34(3):300–305. doi: 10.1097/ANA.0000000000000754
- Fritz BA, Kalarickal PL, Maybrier HR, et al. Intraoperative Electroencephalogram Suppression Predicts Postoperative Delirium. *Anesth Analg*. 2016;122(1):234–242. doi: 10.1213/ane.0000000000000989
- Yang H, Deng HM, Chen HY, et al. The Impact of Age on Propofol Requirement for Inducing Loss of Consciousness in Elderly Surgical Patients. *Front Pharmacol*. 2022;13:739552. doi: 10.3389/fphar.2022.739552
- Li T, Han W, Yang X, et al. Effects of Different Injection Rates of Propofol on Postoperative Cognition in Elderly Patients Undergoing Laparoscopic Inguinal Hernia Repair. *Drug Des Devel Ther*. 2023;17:1741–1752. doi: 10.2147/DDDT.S407905
- Zhuge H, Zhou Y, Qiu Y, et al. Potential increased propofol sensitivity in cognitively impaired elderly: a controlled, double-blind study. *Front Aging Neurosci*. 2024;16:1410181. doi: 10.3389/fnagi.2024.1410181

15. Kramer K, Payne T, Brooks M, et al. Noradrenergic suppression to reduce electroencephalographic arousal after intubation: a randomised, placebo-controlled trial. *BJA Open*. 2025;13:100359. doi: 10.1016/j.bjao.2024.100359
16. Reich DL, Hossain S, Krol M, et al. Predictors of Hypotension After Induction of General Anesthesia. *Anesth Analg*. 2005;101(3):622-628. doi: 10.1213/01.ane.0000175214.38450.91
17. Zou H, Xi F, Fu Y, et al. Bispectral Index-Monitored Anesthesia Induction in Older Adults Undergoing Elective Surgery: Comparing Propofol and Propofol in a Prospective, Single-Center, Double-Blind, Randomized Controlled Study. *Drug Des Devel Ther*. 2024;18:4993-5003. doi: 10.2147/DDDT.S484532
18. Fechner J, El-Boghdady K, Spahn DR, et al. Anaesthetic efficacy and postinduction hypotension with remimazolam compared with propofol: a multicentre randomised controlled trial. *Anaesthesia*. 2024;79(4):410-422. doi: 10.1111/anae.16205
19. Südfeld S, Brechnitz S, Wagner JY, et al. Post-induction hypotension and early intraoperative hypotension associated with general anaesthesia. *Br J Anaesth*. 2017;119(1):57-64. doi: 10.1093/bja/aex127
20. Czajka S, Putowski Z, Krzych ŁJ. Post-induction hypotension and intraoperative hypotension as potential separate risk factors for the adverse outcome: a cohort study. *J Anesth*. 2023;37(3):442-450. doi: 10.1007/s00540-023-03191-7
21. Nakanishi T, Tsuji T, Sento Y, et al. Association between postinduction hypotension and postoperative mortality: a single-centre retrospective cohort study. *Can J Anaesth*. 2024;71(3):343-352. doi: 10.1007/s12630-023-02653-6
22. Wang Y, Wang J, Zhang Y, et al. Body mass index and risk of post-induction hypotension in older adults undergoing non-cardiac surgery: a retrospective cohort study. *BMC Geriatr*. 2025;25(1):939. doi: 10.1186/s12877-025-06691-1
23. Nega MH, Ahmed SA, Tavuyey HY, et al. Incidence and factors associated with post-induction hypotension among adult surgical patients: Prospective follow-up study. *Int J Surg Open*. 2022;49:100565. doi: 10.1016/j.ijsso.2022.100565
24. Christensen AL, Jacobs E, Maheshwari K, et al. Development and Evaluation of a Risk-Adjusted Measure of Intraoperative Hypotension in Patients Having Nonemergent, Noncardiac Surgery. *Anesth Analg*. 2021;133(2):445-454. doi: 10.1213/ANE.0000000000005287
25. Sessler DJ, Khan MZ, Maheshwari K, et al. Blood Pressure Management by Anesthesia Professionals: Evaluating Clinician Skill From Electronic Medical Records. *Anesth Analg*. 2021;132(4):946-956. doi: 10.1213/ANE.00000000000005198
26. Althoff FC, Xu X, Wachtendorf LJ, et al. Provider variability in the intraoperative use of neuromuscular blocking agents: a retrospective multicentre cohort study. *BMJ Open*. 2021;11(4):e048509. doi: 10.1136/bmjopen-2020-048509
27. McGinn R, Talarico R, Hamilton GM, et al. Hospital-, anaesthetist-, and patient-level variation in peripheral nerve block utilisation for hip fracture surgery: a population-based cross-sectional study. *Br J Anaesth*. 2022;128(1):198-206. doi: 10.1016/j.bja.2021.10.011
28. Ma H, Ahrens E, Wachtendorf LJ, et al. Intraoperative Use of Phenylephrine versus Ephedrine and Postoperative Delirium: A Multi-center Retrospective Cohort Study. *Anesthesiology*. 2024;140(4):657-667. doi: 10.1097/aln.0000000000004774
29. Mathis MR, Janda AM, Khetarpal S, et al. Patient-, Clinician-, and Institution-level Variation in Inotropic Use for Cardiac Surgery: A Multicenter Observational Analysis. *Anesthesiology*. 2023;139(2):122-141. doi: 10.1097/aln.00000000000004593
30. Tessmann T, Marty AP, Stricker D, et al. There is no "too small" for frequent workplace-based assessment: Differences between large and small residency programs in anesthesia when using a mobile application to assess EPAs. *GMS J Med Educ*. 2024;41(5):Doc54. doi: 10.3205/zma001709
31. Tanaka P, Park YS, Chen J, et al. Reliability and utility of anesthesiology entrustable professional activities assessed with a mobile web application. *J Clin Anesth*. 2025;106:111922. doi: 10.1016/j.jclinane.2025.111922
32. Tanaka P, Park YS, Roby J, et al. Milestone Learning Trajectories of Residents at Five Anesthesiology Residency Programs. *Teach Learn Med*. 2021;33(3):304-313. doi: 10.1080/10401334.2020.1842210
33. Cai X, Yue M, Liu X, et al. Learning curve for flexible bronchoscope-guided orotracheal intubation for anesthesiology residents: A cumulative sum analysis. *PLoS One*. 2023;18(7):e0288617. doi: 10.1371/journal.pone.0288617
34. Ott S, Müller-Wirtz LM, Bustamante S, et al. Learning tracheal intubation with a hyperangulated videolaryngoscopy blade: sub-analysis of a randomised controlled trial. *Anaesthesia*. 2025;80(4):395-403. doi: 10.1111/anae.16491
35. Ma D, Chen P, Xu J. Learning curve of ultrasound-guided caudal epidural block: a CUSUM pivotal analysis. *Front Med (Lausanne)*. 2025;12:1624205. doi: 10.3389/fmed.2025.1624205
36. Sawasdiwipachai P, Thanasriphakdeekul S, Raksamani K, et al. Learning curve for the acquisition of 20 standard two-dimensional images in advanced perioperative transesophageal echocardiography: a prospective observational study. *BMC Med Educ*. 2022;22(1):412. doi: 10.1186/s12909-022-03280-3
37. de Oliveira Filho GR, Soares Garcia JH. The Accuracy of the Learning-Curve Cumulative Sum Method in Assessing Brachial Plexus Block Competency. *Anesth Analg*. 2024;139(2):281-290. doi: 10.1213/ANE.00000000000006928
38. Scorzoni M, Gonnella GL, Capogna E, et al. Achieving Lumbar Epidural Block Competency in Inexperienced Trainees after a Structured Epidural Teaching Model: A Randomized, Single Blind, Prospective Comparison of CUSUM Learning Curves. *Anesthesiol Res Pract*. 2022;2022:1738783. doi: 10.1155/2022/1738783
39. Berger S, Stalmeijer RE, Marty AP, et al. Exploring the Impact of Entrustable Professional Activities on Feedback Culture: A Qualitative Study of Anesthesiology Residents and Attending. *Acad Med*. 2023;98(7):836-843. doi: 10.1097/ACM.00000000000005188
40. Burns ML, Saager L, Cassidy RB, et al. Association of Anesthesiologist Staffing Ratio With Surgical Patient Morbidity and Mortality. *JAMA Surg*. 2022;157(9):807-815. doi: 10.1001/jamasurg.2022.2804
41. Malgieri C, Kendall MC, Rajae A, et al. The Effectiveness of Direct Supervision by an Attending Compared To a Senior Resident on Quality of Supervision of Novice Anesthesiology Residents: A Randomized Study. *J Educ Perioper Med*. 2023;25(4):E718. doi: 10.46374/volxxv_issue4_Kendall
42. Accreditation Council for Graduate Medical Education. Guidelines for Resident Experience in the Post-Anesthesia Care Unit. Chicago: ACGME; 2007. Available from: https://www.acgme.org/globalassets/PFAssets/ProgramResources/040_GuidelinesforResidentExperienceinthePost-AnesthesiaCareUnit0307.pdf
43. Bombardieri AM, Wildes TS, Stevens T, et al. Practical Training of Anesthesia Clinicians in Electroencephalogram-Based Determination of Hypnotic Depth of General Anesthesia. *Anesth Analg*. 2020;130(3):777-786. doi: 10.1213/ANE.00000000000004537
44. Berger M, Eleswarpu SS, Cooter Wright M, et al. Developing a Real-Time Electroencephalogram-Guided Anesthesia-Management Curriculum for Educating Residents: A Single-Center Randomized Controlled Trial. *Anesth Analg*. 2022;134(1):159-170. doi: 10.1213/ANE.00000000000005677
45. Moaiyeri Z, Mustafa J, Lamperti M, et al. Intraoperative use of processed electroencephalogram in a quaternary center: a quality improvement audit. *J Clin Monit Comput*. 2024;38(6):1263-1268. doi: 10.1007/s10877-024-01189-4
46. Kuizenga MH, Verecke HEM, Absalom AR, et al. Utility of the SmartPilot® View advisory screen to improve anaesthetic drug titration and postoperative outcomes in clinical practice: a two-centre prospective observational trial. *Br J Anaesth*. 2022;128(6):959-970. doi: 10.1016/j.bja.2022.02.033
47. Vellinga R, Intra M, van Amsterdam K, et al. Implementation of a Bayesian based advisory tool for target-controlled infusion of propofol using qCON as control variable. *J Clin Monit Comput*. 2024;38(2):519-529. doi: 10.1007/s10877-023-01106-1
48. Wang D, Song Z, Zhang C, et al. Bispectral index monitoring of the clinical effects of propofol closed-loop target-controlled infusion: Systematic review and meta-analysis of randomized controlled trials. *Medicine (Baltimore)*. 2021;100(4):e23930. doi: 10.1097/md.00000000000023930
49. Flynn SG, Park RS, Jena AB, et al. Coaching inexperienced clinicians before a high stakes medical procedure: randomized clinical trial. *BMJ*. 2024;387:e080924. doi: 10.1136/bmj-2024-080924
50. de Keijzer IN, Vos JJ, Yates D, et al. Impact of clinicians' behavior, an educational intervention with mandated blood pressure and the hypotension prediction index software on intraoperative hypotension: a mixed methods study. *J Clin Monit Comput*. 2024;38(2):325-335. doi: 10.1007/s10877-023-01097-z
51. Law TJ, Lipnick MS, Morriss W, et al. The Global Anesthesia Workforce Survey: Updates and Trends in the Anesthesia Workforce. *Anesth Analg*. 2024;139(1):15-24. doi: 10.1213/ANE.00000000000006836
52. Akavipat P, Suraseranivongse S, Yimrattanabowon P, et al. Anesthesia Workforce Capacity in Thailand. *WHO South East Asia J Public Health*. 2021;10(1):5-11. doi: 10.4103/WHO-SEA/JP.H.WHO-SEA/JP.H.305.20
53. Lee H, Kim J, Lee KY, et al. Awareness and Perspectives among Asian Anesthesiologists on Postoperative Delirium: A Multinational Survey. *J Clin Med*. 2021;10(24):5769. doi: 10.3390/jcm10245769
54. Long MY, Lim EHL, Balanza GA, et al. Sevoflurane requirements during electroencephalogram (EEG)-guided vs standard anesthesia Care in Children: A randomized controlled trial. *J Clin Anesth*. 2022;81:110913. doi: 10.1016/j.jclinane.2022.110913
55. Omega A, Wijaya Ramlan AA, Soenarto RF, et al. Assessing clinical reasoning in airway related cases among anesthesiology fellow residents using Script Concordance Test (SCT). *Med Educ Online*. 2022;27(1):2135421. doi: 10.1080/10872981.2022.2135421
56. Heriwardito A, Ramdhoni L, Ramlan AAW, et al. Comparison of discussion-based and simulation-based learning methods using the Gas Man® on knowledge of the uptake and distribution of inhalation anesthetics of anesthesia resident: A randomized controlled trial. *J Anesthesiol Clin Pharmacol*. 2024;40(4):672-678. doi: 10.4103/joacp.joacp.270_23
57. Fletcher G, Flin R, McGeorge P, et al. Anaesthetists' Non-Technical Skills (ANTS): evaluation of a behavioural marker system. *Br J Anaesth*. 2003;90(5):580-588. doi: 10.1093/bja/aeg112
58. Ma W, Li C, Li Q. Association between low Bispectral Index™ values during anesthesia maintenance and one-year mortality in older patients with different comorbidities: a retrospective cohort study. *Can J Anaesth*. 2025;72(8):1280-1290. doi: 10.1007/s12630-025-02995-3
59. Cooter Wright M, Bunning T, Eleswarpu SS, et al. A Processed Electroencephalogram-Based Brain Anesthetic Resistance Index Is Associated With Postoperative Delirium in Older Adults: A Dual Center Study. *Anesth Analg*. 2022;134(1):149-158. doi: 10.1213/ANE.00000000000005660

60. Obert DP, Schweizer C, Zinn S, et al. The influence of age on EEG-based anaesthesia indices. *J Clin Anesth.* 2021;73:110325. doi: 10.1016/j.jclinane.2021.110325
61. Daly Guris RJ, Miller CR, Schiavi A, et al. Examining novice anaesthesia trainee simulation performance: a tale of two clusters. *BMJ Simul Technol Enhanc Learn.* 2021;7(6):548-554. doi: 10.1136/bmjstel-2020-000812
62. Babus LW, Gurnaney H, Doshi AK, et al. The utility of virtual reality and manikin crisis scenario simulations for anaesthesia trainee education: a randomised crossover pilot study. *Anaesth Rep.* 2024;12(2):e12316. doi: 10.1002/anr3.12316
63. Eagleman SL, Drover CM, Li X, et al. Offline comparison of processed electroencephalogram monitors for anaesthetic-induced electroencephalogram changes in older adults. *Br J Anaesth.* 2021;126(5):975-984. doi: 10.1016/j.bja.2020.12.04