A comparative study on adequate anesthesia depth: clinical judgement and the Narcotrend® measurement

Une étude comparative sur la profondeur d’anesthésie appropriée : jugement clinique vs mesure du moniteur Narcotrend®

Wolfgang F. Puchner, MD · Martin W. Dünser, MD · Patrick Paulus, MD · Markus P. Neuner, MD · Charlotte L. Mayer, MD · Irmgard M. Pomberger, MD · Ruth Hackl, MD · Jens M. Meier, MD

Received: 20 June 2019 / Revised: 26 November 2019 / Accepted: 4 December 2019 / Published online: 3 March 2020 © The Author(s) 2020

Abstract

Purpose To compare the clinical judgement of electroencephalogram (EEG)-naïve anesthesiologists with an EEG-based measurement of anesthetic depth (AD) using the Narcotrend® monitor.

Methods In this prospective cohort study including 600 patients, AD during stable anesthesia was assessed by clinical judgement of the attending, EEG-blinded anesthesiologist (using a scale staging the AD as mid-adequate, adequate but fairly deep, or adequate but fairly light) and by simultaneously recorded Narcotrend measurements.

Results In 42% of patients (n = 250), the anesthesiologist’s clinical judgement was in agreement with anesthetic levels as measured by the Narcotrend monitor. In 46% of patients (n = 274), the anesthesiologist’s judgement and the Narcotrend monitor differed by one AD level (minor discordance). Major discordance was observed in 76 (13%) measurements (judged deeper than measured, n = 29 [5%]; judged lighter than measured, n = 47 [8%]). In 7% of patients (n = 44), the Narcotrend index was outside the limits of adequate AD (too deep, n = 28 [5%]; too superficial, n = 16 [3%]). The overall level of agreement between the anesthesiologist’s judgement and the Narcotrend monitor was not statistically significant (Cohen’s kappa, −0.039; P = 0.17). Using a random forests algorithm, age, mean blood pressure, the American Society of Anesthesiologists classification, body mass index, and frailty were the variables with the highest relative feature importance to predict the level of agreement.

Conclusion These results suggest that clinical judgement of AD during stable anesthesia was not in agreement with EEG-based assessment of anesthetic depth in 58% of cases. Nevertheless, this finding could be influenced by the lack of validated scales to clinically judge AD.

Trial registration www.clinicaltrials.gov (NCT02766894); registered 10 May, 2016.

Résumé

Objectif Notre objectif était de comparer le jugement clinique d’anesthésiologistes n’ayant pas accès à un électroencéphalogramme (EEG) à une mesure de la profondeur anesthésique (PA) fondée sur l’EEG utilisant le moniteur Narcotrend®.

Méthode Dans cette étude de cohorte prospective de 600 patients, la PA a été évaluée pendant la phase de maintien stable de l’anesthésie selon le jugement clinique de l’anesthésiologue traitant, qui n’avait pas accès à l’EEG (sur une échelle évaluant la PA comme étant adéquate, adéquate mais relativement profonde ou adéquate mais relativement légère) et par des mesures simultanément enregistrées par le Narcotrend.

Résultats Chez 42 % des patients (n = 250), le jugement clinique de l’anesthésiologue concordait aux niveaux anesthésiques tels que mesurés par le moniteur Narcotrend. Chez 46 % des patients (n = 274), le
jugement de l’anesthésiologiste et le moniteur Narcotrend différaient d’un niveau de PA (discordance mineure). Une discordance majeure a été observée dans 76 (13 %) mesures (jugées plus profondes que mesurées, n = 29 [5 %], jugées plus légères que mesurées, n = 47 [8 %]). Chez 7 % des patients (n = 44), l’indice Narcotrend était situé au-delà des limites d’une PA adéquate (trop profond, n = 28 [5 %]; trop superficiel, n = 16 [3 %]). Le niveau global de concordance entre le jugement de l’anesthésiologiste et le moniteur Narcotrend n’était pas significatif d’un point de vue statistique (kappa de Cohen, -0,039; P = 0,17). En se fondant sur un algorithme de forêt d’arbres décisionnels (random forests algorithm), l’âge, la tension artérielle moyenne, la classification selon l’American Society of Anesthesiologists, l’indice de masse corporelle et l’indice de fragilité ont été identifiés comme les variables ayant la plus grande importance relative pour prédire le niveau de concordance.

Conclusion Ces résultats suggèrent que, dans 58 % des cas, le jugement clinique de la PA ne concordait pas à l’évaluation par EEG de la profondeur anesthésique pendant une phase de maintien stable de l’anesthésie. Toutefois, ces résultats pourraient être influencés par l’absence d’échelles validées pour juger la PA d’un point de vue clinique.

Enregistrement de l’étude www.clinicaltrials.gov (NCT02766894); enregistrée le 10 mai 2016.

Adequate anesthetic depth (AD) is a fundamental responsibility of the anesthesiologist during general anesthesia. Too deep or too light levels of anesthesia are disadvantageous for patients. While a too light AD can result in intraoperative awareness and major psychological sequelae, a too deep AD carries the risk of hemodynamic instability, prolonged awakening, and postoperative delirium.1–6 Clinical evaluation is still considered to play the leading role in AD assessment during general anesthesia.

Despite their well-known limitations,7 electroencephalography (EEG)-based monitors are increasingly used to support AD assessment.2,8,9 It has been shown that using an EEG monitor can reduce anesthetic consumption,10,11 recovery time,10,12 postoperative nausea and vomiting,10,13 awareness,14 and postoperative delirium.15,16 Taking these data into account, one can assume that clinical judgement is inferior to EEG-derived measurements in determining adequate AD, and leads to inadvertent under- or overdosing of anesthetic agents.

The aim of this prospective observational cohort study was to compare the clinical, EEG-free judgement of anesthesiologists in daily practice with an EEG-based measurement of AD during stable general anesthesia. We hypothesized that there would be no differences in the level of agreement between an EEG-based measurement and clinical judgement of AD.

Methods

This study was approved by the Ethics Committee of Upper Austria (study number C-109-16) and written informed consent was obtained from all subjects participating in the trial. The trial was registered prior to patient enrollment at clinicaltrials.gov (NCT02766894, principal investigator: Wolfgang Puchner, date of registration: May 2016) and conducted in accordance with ethical principles of the Declaration of Helsinki. The manuscript was prepared as suggested by the STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) statement. The Narcotrend® monitor used in this study was provided free of charge by Gepa-Med Mediotechnik (Vienna, Austria). The company played no role in the study design, data collection, statistical analysis, or interpretation of study results. This analysis was designed as an investigator-initiated, single-centre, prospective, observational cohort study. It was performed in the operating theatres of a tertiary university hospital in Linz, Austria from May 2016 to February 2017.

Selection of study patients

Patients aged between one and 110 yr, who were scheduled for elective or non-elective surgery requiring a general anesthetic, were eligible for study enrollment. Duration of surgery < ten minutes, a moribund state, proximity of the EEG electrodes to the operating field, skin lesions over the forehead, known pathologies of the frontal brain lobe, hypoxic brain damage, and hypothermia (< 33°C) were considered exclusion criteria.

Selection of anesthesiologists

Twenty anesthesiologists with no experience in EEG monitoring and EEG-based guidance of anesthesia were invited to take part in this study. Participation was voluntary and decided on a first come-first serve basis. Each physician received detailed information before the study and was expected to participate in 30 individual measurements. This number of measurements was estimated to allow for correction of within-subject correlation. If a participant could not complete 30
measurements within the study period, data were omitted, and the physician was replaced by another colleague, who performed 30 measurements.

Data collection

All measurements were taken under stable general anesthesia at the earliest five minutes following induction, when the surgical intervention already in progress. The general anesthetic regimen (balanced anesthesia with volatile anesthetics or total intravenous anesthesia) and the type of anesthetics were at the discretion of the attending anesthesiologist, except for ketamine, dexmedetomidine, and/or nitrous oxide as these substances modify the EEG signal and are not accurately reflected in the processed EEG index.17–19

The Narcotrend®-Compact M monitor (software version 3.0, MT Monitor Technik, Bad Bramstedt, Germany) was used to objectively measure AD.20 This device processes the raw EEG signal (acquired over an average of 20 sec) to the Narcotrend index, which is updated every second. This index is a dimensionless number from 0 (isoelectric EEG) to 100 (awake). These indices are categorized by the manufacturer into six stages (A [awake] to F [anesthesia with burst suppression]), with the stages C, D, E0,1 reflecting adequate AD (Table 1). Similar to the bispectral index, the Narcotrend monitor is a validated tool for measuring AD.21–23 It has repeatedly been used to measure AD in patients aged one year or older.24,25

The scale by which the anesthesiologist determined the AD was derived from the Narcotrend® index scale and had five categories (Table 1). After surgery had started, the attending anesthesiologist excluded an inadequate AD, and the investigator verified a stable and undisturbed EEG processing and index recording (electrode impedance < 8 kΩ, no interference by electrocautery) for at least one minute. Under these conditions, the attending anesthesiologist, who was blinded to the raw EEG and the Narcotrend index, determined the AD using their clinical judgement and standard perioperative monitoring. According to the AD scale (Table 1), the anesthesiologist could then choose from the remaining three AD categories: mid-adequate, (adequate but) fairly deep, or (adequate but) fairly light. Simultaneously with clinical judgement, the raw EEG trace was recorded and the Narcotrend index documented. Only one measurement was performed per patient. The attending anesthesiologist was not given access to the raw EEG or the index on the Narcotrend monitor during the course of study.

In addition, the following data were documented at the same time: the attending anesthesiologist’s years of practice, and the period of time from induction of anesthesia until measurement. We collected the following patient data: age, sex, body mass index, red/haired status (defined by phenotypic appearance, e.g., red or orange hair, freckles), frailty (based on the Fried criteria),26 the American Society of Anesthesiologists (ASA) physical status classification, history of awareness or postoperative nausea and vomiting, urgency of surgery, pre-medication received, anesthesia regimen, calculated effect-site or plasma concentration of propofol (in case a target controlled infusion was used) or the age-adjusted endtidal minimum alveolar concentration at the time of measurement, neuromuscular blockade during assessment, heart rate and mean arterial blood pressure during assessment, intraoperative movements before measurement, and intraoperative use of catecholamines. Within 24 hr after recovery from anesthesia, study patients

Table 1 Levels of anesthetic depth assessed by clinical judgement and Narcotrend® monitor measurement

| Clinical judgement                | Narcotrend® measurement |
|----------------------------------|-------------------------|
| inadequate too light / no judgement | NI 100 – 95 / stage A |
| adequate but fairly light         | NI 79 – 65 / stage C (C0,1,2) |
| mid-adequate                      | NI 64 – 37 / stage D (D0,1,2) |
| adequate but fairly deep          | NI 36 – 20 / substage E0,1 |
| inadequate too deep / no judgement | NI 19 – 13 / substage E2 |
|                                  | NI 12 – 0 / stage F (F0,1) |

Anesthesiologists blinded to the raw electroencephalogram (EEG) and the Narcotrend index (NI) judged anesthetic depth using clinical skills and standard monitoring as either mid-adequate, adequate but fairly light or fairly deep. Simultaneously, the NI was measured. The NI is processed from the raw EEG and is updated every second. It is a dimensionless number ranging from 100 (awake) to 0 (zero-line EEG) and has been categorized into six stages (A–F). According to the recent manufacturer’s recommendations, stage A (awake), stage B (sedated), stage E2 (anesthesia with incipient appearance of burst suppression patterns), and stage F (anesthesia with burst suppression) are considered indicators of inadequate anesthetic depth during conventional surgical procedures. White area represents levels of adequate AD (i.e., stages C [light anesthesia], D [anesthesia middle], E0,1 [anesthesia with deep hypnosis]). On this basis, levels of anesthetic depth were compared with dis/agreement between clinical judgement and the depth of anesthesia monitor Narcotrend in 600 patients.
were interviewed by one of four trained investigators using
the modified Brice method to identify intraoperative
awareness. This information was not obtained in very
young children, patients with postoperative delirium, or
critically ill subjects.

Study endpoints

The primary endpoint of this study was to determine the
level of agreement between the anesthesiologist’s clinical
judgement and the Narcotrend measurement of AD. The
level of agreement was determined for each measurement
by counting the number of AD levels judged by the
anesthesiologist that were discordant from the Narcotrend
monitor measurements(d-score). For example, if the AD
levels determined by clinical judgement and the
Narcotrend monitor were in agreement, the number of
discordant levels was zero. If the AD level judged by the
anesthesiologist was two levels lower than the level
measured by the Narcotrend monitor, the number of
discordant levels (d-score) was two. We used negative
values to indicate that the Narcotrend monitor categorized
patients into deeper AD and positive values if the
Narcotrend monitor categorized patients into lighter AD
than judged by the anesthesiologist. Discordance by one
AD level was considered minor, whereas discordance by
two or more AD levels was considered major. The
secondary study endpoint was risk factors for discordance
between the anesthesiologists’ judgement and the
Narcotrend measurement using a feature selection
algorithm of a “random forest” algorithm.

Statistical analysis

As no data on the level of agreement between clinical
judgement and an EEG monitor based measurement of AD
existed at the time of study planning, we could not perform
a power analysis to determine which sample size would be
adequate for our study. We knew that 20 anesthesiologists
were willing to participate in the study, and we believed
that 30 measurements per anesthesiologist would be
sufficient to identify participant-specific effects. This
resulted in a sample size of 600, a number which can
also be seen as a lower limit for the number of
measurements necessary for performing a machine
learning algorithm.

The level of agreement (beyond that observed by
chance) between the anesthesiologist’s judgement and
Narcotrend measurements was tested using the Cohen’s
crude kappa method. Furthermore, the number of measurements
where the d-score was −1, 0, or 1 was counted to
determine how often a minor deviation between the
judgement of the anesthesiologist and the measurement
of the Narcotrend was observed.

Additionally the “random forest” algorithm, provided
by Breiman was used to predict the d-score as described
above (R statistical software, package “random forest”).
For this purpose, the data set was split into a training data
set (80% of measurements) and a test data set (20% of
measurements). After training the model on the training
data set, the test data set was used for prediction. This
approach determines the relative feature importance using
the “Boruta” package provided by “The Comprehensive R
Archive Network”. The relative feature importance is then
given in arbitrary units.

All statistical analyses were performed using the open-
source R statistical software package, version 3.5.0 (The R
Foundation for Statistical Computing, Vienna, Austria).
Variables are given as median values with interquartile
ranges or absolute numbers with percentages.

Results

During the study period, complete sets of intraoperative
measurements were taken in 600 patients (Table 2, Fig. 1)
who were anesthetized by 20 anesthesiologists. In all but
41 patients, postoperative data on intraoperative awareness
and dreams could be collected.

Primary study endpoint

In 42% of patients (n = 250), the anesthesiologist’s clinical
judgement of AD agreed with anesthetic level measured by
the Narcotrend monitor. In 46% of patients (n = 274), the
anesthesiologist and Narcotrend monitor differed by one
AD level (minor discordance). Major discordance was
observed in 76 (13%) measurements (judged deeper than
measured, n = 29 [5%]; judged lighter than measured, n =
47 [8%]). In 7% of patients (n = 44), the Narcotrend index
was outside the limits of adequate AD (too deep, n = 28
[5%]; too superficial, n = 16 [3%]). The overall level of
agreement between the anesthesiologist’s clinical
judgement and the Narcotrend monitor was statistically
insignificant (Cohen’s kappa, −0.039; P = 0.17). Figure 2
gives an overview over the levels of dis-/agreement
between the anesthesiologists’ judgement and the
Narcotrend measurement for all patients. None of the 559
study patients who were followed up with interviews
reported signs of intraoperative awareness.

Secondary study endpoint

In 90.8% of patients the difference between the real and the
predicted d-score was −1, 0, or 1 showing a very high
Table 2 Characteristics of all study patients as well as subjects with concordant and discordant judgement of the anesthetic depth

|                  | All patients (n = 600) | Concordance (n = 250) | Discordance Judged deeper than measured (n = 157) | Discordance Judged lighter than measured (n = 193) |
|------------------|------------------------|-----------------------|--------------------------------------------------|--------------------------------------------------|
| Age, yr          | 50 [24–67]             | 38 [22–60]            | 42 [17–62]                                       | 62 [46–74]                                       |
| 1–18             | 107 (18)               | 51 (20)               | 46 (29)                                          | 10 (5)                                           |
| 19–70            | 375 (63)               | 171 (68)              | 88 (56)                                          | 116 (60)                                         |
| 71–92            | 118 (20)               | 28 (11)               | 23 (15)                                          | 67 (35)                                          |
| ASA physical status |                       |                       |                                                  |                                                  |
| I                | 244 (41)               | 120 (48)              | 75 (48)                                          | 49 (25)                                          |
| II               | 240 (40)               | 103 (41)              | 58 (37)                                          | 79 (41)                                          |
| III              | 110 (18)               | 26 (10)               | 24 (15)                                          | 60 (31)                                          |
| IV               | 6 (1)                  | 1 (0.4)               | 0                                                | 5 (3)                                            |
| Sex              |                        |                       |                                                  |                                                  |
| Male             | 365 (61)               | 155 (62)              | 95 (61)                                          | 115 (60)                                         |
| Female           | 235 (39)               | 95 (38)               | 62 (39)                                          | 78 (40)                                          |
| Body mass index, kg·m⁻² |                   |                       |                                                  |                                                  |
| 25 [21–29]       | 25 [21–29]             | 24 [20–27]            | 26 [23–29]                                       |                                                  |
| Frailty          | 3 (0.5)                | 0 (0)                 | 0                                                | 3 (2)                                            |
| Redhead status   | 9 (2)                  | 7 (3)                 | 2 (1)                                            | 0                                                |
| History of awareness |                 |                       |                                                  |                                                  |
| 4 (0.7)          | 2 (0.8)                | 2 (1)                 | 0                                                |                                                  |
| History of PONV  | 44 (7)                 | 19 (8)                | 9 (6)                                            | 16 (8)                                           |
| Induction to measurement, min |         |                       |                                                  |                                                  |
| 38 [26–60]       | 35 [25–57]             | 39 [25–60]            | 43 [28–66]                                       |                                                  |
| Urgency of surgery |                       |                       |                                                  |                                                  |
| Elective         | 581 (97)               | 246 (98)              | 150 (96)                                         | 185 (96)                                         |
| Urgent (< 6 hr)  | 19 (3)                 | 4 (2)                 | 7 (4)                                            | 8 (4)                                            |
| Pre-medication received |            |                       |                                                  |                                                  |
| 561 (94)         | 236 (94)               | 143 (91)              | 182 (94)                                         |                                                  |
| Anesthesiologist’s experience, yr |        |                       |                                                  |                                                  |
| 3 [1–12]         | 4 [1–12]               | 2 [1–12]              | 4 [1–12]                                         |                                                  |
| Anesthesia regimen |                       |                       |                                                  |                                                  |
| Balanced anesthesia |           | 147 (59)              | 97 (62)                                          | 89 (46)                                          |
| Total intravenous anesthesia |        | 103 (41)              | 60 (38)                                          | 104 (54)                                         |
| Neuromuscular blockade |          | 182 (73)              | 96 (61)                                          | 139 (72)                                         |
| Narcotrend index | 45 [35–59]             | 48 [40–56]            | 65 [47–73]                                       | 33 [26–37]                                       |
| AD as per clinical judgement |           |                       |                                                  |                                                  |
| Fairly deep      | 122 (20)               | 23 (9)                | 91 (58)                                          | 8 (4)                                            |
| Mid-adquate      | 387 (65)               | 220 (88)              | 65 (41)                                          | 102 (53)                                         |
| Fairly light     | 91 (15)                | 7 (3)                 | 1 (0.6)                                          | 83 (43)                                          |
| MAP, mmHg        | 71 [65–81]             | 72 [65–80]            | 68 [61–75]                                       | 74 [66–87]                                       |
| Heart rate, beats·min⁻¹ |              |                       |                                                  |                                                  |
| 59 [52–68]       | 60 [52–68]             | 59 [52–71]            | 57 [51–66]                                       |                                                  |
| Propofol (TCI) concentration, µg·mL⁻¹ |        |                       |                                                  |                                                  |
| 1.8 [1.5–2.0]    | 1.8 [1.6–2.2]          | 1.6 [1.5–1.9]         | 1.8 [1.4–2]                                      |                                                  |
| Sevoflurane (endtidal), MAC |        |                       |                                                  |                                                  |
| 0.8 [0.7–0.9]    | 0.8 [0.7–0.9]          | 0.8 [0.7–0.9]         | 0.8 [0.7–0.9]                                    |                                                  |
| Intraoperative movements | 8 (2) | 0 (0)                 | 2 (2)                                            | 6 (3)                                            |
| Intraoperative use of catecholamines | 39 (7) | 12 (5)               | 11 (7)                                           | 16 (8)                                           |
| D-score          |                        |                       |                                                  |                                                  |
| 0                | 250 (42)               | 250 (100)             | 0                                                | 0                                                |
| 1                | 274 (46)               | 0                     | 128 (82)                                         | 146 (76)                                         |
| 2                | 71 (12)                | 0                     | 28 (18)                                          | 43 (22)                                          |
| 3                | 5 (0.8)                | 0                     | 1 (0.6)                                          | 4 (2)                                            |

AD = anesthetic depth; ASA = American Society of Anesthesiologists; D-score = discordance by counting the number of AD levels the anesthesiologists’ judgement was discordant from the measurements of the Narcotrend monitor; MAC = minimum alveolar concentration; MAP = mean arterial blood pressure measured invasively or non-invasively; PONV = postoperative nausea and vomiting; TCI = target controlled infusion

Variables are given as median values [interquartile range] or absolute numbers (%)
Fig. 1 Histogram of measured anesthetic depths (AD). Distribution of AD of all 600 patients during stable conditions of adequate anesthesia from the attending anesthesiologist’s perspective as measured by Narcotrend index (NI) values. Dark-shaded bars represent measurements of inadequate AD with an NI beyond the range of 20–79, as recommended by the manufacturer of Narcotrend to be considered an either too deep or too light anesthesia level.

Fig. 2 Cross-reference of judged and measured anesthetic depths (AD). Depiction of all 600 judgments of AD (left: categories of fairly deep and fairly light anesthesia, right: category of mid-adequate level of anesthesia) compared with the measurements by the Narcotrend index (middle: table of index values in steps of 5). The strength of the line corresponds to the number of patients. Horizontal line means conformity between assessed and measured anesthetic levels, falling or ascending lines indicate discordance. Dark-shaded areas show the range of inadequate anesthesia levels according to the recommendations of the manufacturer of Narcotrend.
accuracy of the underlying model. Performing the feature selection algorithm of the “Boruta” package on the same data set revealed age, mean arterial blood pressure, pediatric surgery, ASA physical status classification, body mass index, and frailty as the variables with highest relative feature importance for predicting the d-score (Fig. 3). All other variables contributed less to predicting the d-score.

Discussion

The main finding of our study is that clinical judgement of AD during stable anesthesia by the anesthesiologist agreed with Narcotrend monitor measurements in 42% of patients. We observed minor discordance between clinical and EEG-based assessments of AD in 46% of subjects and major discordance in 13% of subjects. The Cohen’s kappa coefficient, which was used to statistically report the level of agreement between the two assessments of AD, was low (−0.039) and not significant (P = 0.17), indicating that clinical judgement and EEG-based measurements of AD were not in agreement in our study.

Currently, there is no scientific evidence on the optimal AD in terms of either clinical assessment or EEG-derived indices, including the Narcotrend index. In contrast to another study which evaluated the ability of anesthesiologists to predict the bispectral index, the aim of our study was to evaluate the agreement of the clinical judgement of AD and measurements of the EEG-based Narcotrend monitor. Therefore, we used a three-stage scale to clinically report AD (i.e., fairly light, mid-adequate, fairly deep). This pragmatic approach was chosen to reflect clinical practice in the absence of other validated scales to clinically report AD. Nevertheless, it needs to be kept in mind that this three-stage scale has not been validated, and may have influenced the non-agreement between clinical judgement and EEG-based measurement of AD in our study.

Minor discordance was observed in 46% and major discordance in 13% of study subjects. In one third of the study patients, the attending anesthesiologist assessed AD as either adequate but “fairly deep” or “fairly light”. The rationale behind individual decision-making processes was not determined in our protocol. Thus, it can only be speculated on which variables the anesthesiologists’
decisions were based. Given that age, blood pressure, pediatric surgery, the ASA classification, body mass index, and frailty were the variables with the highest relative feature importance for predicting the level of agreement between the anesthesiologist and the Narcotrend monitor, one could assume that the anesthesiologists in our study mainly relied on hemodynamic indicators and physical appearance of the patients. This practice is, however, not supported by scientific evidence, which shows that hemodynamic variables may be influenced by comorbid conditions and a relevant patient to patient variability in pharmacodynamics of anesthetics.\textsuperscript{30,31} Notably, the anesthesiologists’ judgement of adequate but “fairly deep” or adequate but “fairly light” was rarely confirmed by the Narcotrend index. This appears to be clinically relevant for certain surgical procedures and circumstances (e.g., ophthalmological surgery, neurosurgery, end of surgery) where anesthesiologists intend to achieve fairly deep or fairly light AD. Our data indicate that without EEG-based measurement of the AD, this might not be reliably achieved by the anesthesiologist’s clinical judgement alone.

Age, blood pressure, pediatric surgery, the ASA classification, body mass index, and frailty score were identified as clinically relevant risk factors for discordance between clinical judgement and Narcotrend measurements of AD. Deeper levels of anesthesia than clinically perceived were more frequently observed in old and comorbid patients, whereas lighter levels of anesthesia were predominantly observed in younger patients. This finding is in accordance with the well-known phenomenon of anesthesia-overdosing in older patients.\textsuperscript{32,33} It was even more evident when older age was associated with a higher ASA score, suggesting that managing geriatric anesthesia without EEG monitoring is prone to misjudgement of AD and therefore overdosing of hypnotics. Our results could therefore explain why the use of an EEG monitor reduced anesthetic consumption,\textsuperscript{10,11} shortened recovery time,\textsuperscript{10,12} and decreased the incidence of postoperative delirium\textsuperscript{15,16} in other studies.

In study subjects aged one to 18 yr, lighter than clinically judged levels of AD were observed more frequently than in older patients in our analysis. These results are consistent with reports on increased total EEG power and on higher-frequency bands in children and adolescents during anesthesia resulting in processed EEG values indicating a light AD, which may lead to increased doses of hypnotics being administered unnecessarily.\textsuperscript{34,35} Similarly, high EEG index values were measured with an entropy monitor in adults despite an adequate clinical AD.\textsuperscript{36,37} It was hypothesized that this could be an artefact of the device’s processing algorithm.\textsuperscript{36,37} Furthermore, analgesia-based general anesthetic may result in an AD that is lighter than that assessed by EEG-monitors but still clinically adequate.\textsuperscript{38} Accordingly, we did not observe any indicators of explicit awareness in this population, despite current literature suggesting higher rates of intraoperative awareness in children than in adults.\textsuperscript{39}

Neither the anesthesiologist nor their professional experience was associated with the primary study endpoint in our population. It can be hypothesized that anesthesiologists with limited experience at the time of study participation had been sufficiently trained to assess and achieve an adequate AD. Nevertheless, when interpreting these results, it needs to be kept in mind that young anesthesiologists less frequently cared for study patients with a higher ASA score or for very young or very old patients.

Strengths of our study are its prospective design and the high number of study patients included. Enrollment of 20 anesthesiologists, on the other hand, minimized the effect of inter-individual differences on the study results. Certain limitations need to be considered when interpreting the results of this study. First, measurements were taken only once and at single, arbitrarily chosen time points during general anesthesia. Therefore, we cannot conclude that the values obtained correlate with other phases of anesthesia (e.g., during episodes of instability or intense/minimal surgical stimulation). Second, we only recorded the Narcotrend index and the respective AD levels. Doing so, we did not analyze the original EEG trace, which may give more accurate information about the true AD.

In conclusion, the results of our study suggest that clinical judgement of AD during stable anesthesia did not agree with EEG-based assessment of anesthetic depth in 58% of cases. Nevertheless, this finding could be influenced by the lack of validated scales to clinically judge AD. Age, arterial blood pressure, the ASA classification, body mass index, and frailty appear to determine the level of disagreement between clinical judgement and the Narcotrend monitor.

\textbf{Author contributions} Wolfgang F. Puchner contributed to all aspects of this manuscript, including study conception and design; acquisition, analysis, and interpretation of data; and drafting the article. Martin W. Dünser contributed to analysis and interpretation of data and drafting the article. Patrick Paulus helped write and review the manuscript. Markus P. Neuner, Charlotte L. Mayer, Irmgard M. Pomberger, and Ruth Hackl helped collect data and conduct the study. Jens M. Meier led the project, conceived and designed the study, analysed and interpreted data, and drafted the article.

\textbf{Acknowledgements} Open access funding provided by Kepler Universitätsklinikum Linz. The authors are grateful for the voluntary attendance of the 22 anesthesiologists of the Department of Anesthesiology and Intensive Care, Kepler University Hospital, Med Campus III. and Med Campus IV., Linz, Austria. Each participating anesthesiologist received a detailed personal report of their own examinations. The authors wish to thank Gepa-Med.
Medizintechnik G.m.b.H., Vienna, Austria, for providing the Narcotrend® monitor free of charge for the study period. We would like to thank Dr. Wolfgang Schimetta for his assistance with the statistical analysis of the study.

Conflicts of interest None.

Funding statement None.

Editorial responsibility This submission was handled by Dr. Philip M. Jones, Associate Editor, Canadian Journal of Anesthesia.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc/4.0/.

References

1. Myles PS, Williams DL, Hendrata M, Anderson H, Weeks AM. Patient satisfaction after anaesthesia and surgery: results of a prospective survey of 10,811 patients. Br J Anaesth 2000; 84: 6-10.

2. American Society of Anaesthesiologists Task Force on Intraoperative Awareness. Practice advisory for intraoperative awareness and brain function monitoring: a report by the American Society of Anaesthesiologists Task Force on Intraoperative Awareness. Anaesthesiology 2006; 104: 847-64.

3. Sebel PS, Bowdle TA, Ghoneim MM, et al. The incidence of awareness during anesthesia: a multicenter United States study. Anesth Analg 2004; 99: 833-9.

4. Monk TG, Saini V, Weldon BC, Sigl JC. Anesthetic management and one-year mortality after noncardiac surgery. Anesth Analg 2005; 100: 4-10.

5. Kertai MD, Pal N, Palanca BJ, et al. Association of perioperative risk factors and cumulative duration of low bispectral index with intermediate-term mortality after cardiac surgery in the B-Unaware Trial. Anaesthesiology 2010; 112: 1116-27.

6. Fritz BA, Kalarickal PL, Maybrier HR, et al. Intraoperative electroencephalogram suppression predicts postoperative delirium. Anesth Analg 2016; 122: 234-42.

7. Dahaba AA. Different conditions that could result in the bispectral index indicating an incorrect hypnotic state. Anesth Analg 2005; 101: 765-73.

8. Chechetts MR, Alladi R, Ferguson K, et al. Recommendations for standards of monitoring during anaesthesia and recovery 2015: Association of Anaesthetists of Great Britain and Ireland. Anaesthesia 2016; 71: 85-93.

9. Aldecoa C, Bettelli G, Bilotta F, et al. European Society of Anaesthesiology evidence-based and consensus-based guideline on postoperative delirium. Eur J Anaesthesiol 2017; 34: 192-214.

10. Liu SS. Effects of bispectral index monitoring on ambulatory anesthesia: a meta-analysis of randomized controlled trials and a cost analysis. Anesthesiology 2004; 101: 311-5.

11. Aimé I, Verrout N, Masson-Lefoll C, et al. Does monitoring bispectral index or spectral entropy reduce sevoflurane use? Anesth Analg 2006; 103: 1469-77.

12. Gan TJ, Glass PS, Windsor A, et al. Bispectral index monitoring allows faster emergence and improved recovery from propofol, alfentanil, and nitrous oxide anesthesia. BIS Utility Study Group. Anesthesiology 1997; 87: 808-15.

13. Nelskylä KA, Yli-Hankala AM, Puro PH, Korttila KT. Sevoflurane titration using bispectral index decreases postoperative vomiting in phase II recovery after ambulatory surgery. Anesth Analg 2001; 93: 1165-9.

14. Myles PS, Leslie K, McNeil J, Forbes A, Chan MT. Bispectral index monitoring to prevent awareness during anaesthesia: the B-Aware randomised controlled trial. Lancet 2004; 363: 1757-63.

15. Chan MT, Cheng BC, Lee TM, Gin T; CODA Trial Group. BIS-guided anesthesia decreases postoperative delirium and cognitive decline. J Neurosurg Anesthesiol 2013; 25: 33-42.

16. Radtke FM, Franck M, Lenderj J, Krüger S, Wernecke KD, Spies CD. Monitoring depth of anaesthesia in a randomized trial decreases the rate of postoperative delirium but not postoperative cognitive dysfunction. Br J Anaesth 2013; 110(Suppl 1): 98-105.

17. Hayashi K, Tsuda N, Sawa T, Hagihira S. Ketamine increases the frequency of electroencephalographic bicoherecene peak on the alpha spindle area induced with propofol. Br J Anaesth 2007; 99: 389-95.

18. Foster BL, Liley DT. Nitrous oxide paradoxically modulates slow electroencephalogram oscillations: implications for anesthesiology monitoring. Anesth Analg 2011; 113: 758-65.

19. Akeju O, Pavone KJ, Westover MB, et al. A comparison of propofol- and dexmedetomidine-induced electroencephalogram dynamics using spectral and coherence analysis. Anesthesiology 2014; 121: 978-89.

20. Kreuer S, Wilhelm W. The Narcotrend monitor. Best Pract Res Clin Anaesthesiol 2006; 20: 111-9.

21. Schmidt GN, Bischoff P, Standl T, et al. Comparative evaluation of Narcotrend, Bispectral Index, and classical electroencephalographic variables during induction, maintenance, and emergence of a propofol/remifentanil anesthesia. Anesth Analg 2004; 98: 1346-53.

22. Förder S, Bruhn J, Larsen R, Bialas P, Wilhelm W. Comparability of Narcotrend index and bispectral index during propofol anaesthesia. Br J Anaesth 2004; 93: 235-40.

23. Kreuer S, Bruhn J, Larsen R, Grundmann U, Shafer SL, Wilhelm W. Application of bispectral index and Narcotrend index to the measurement of the electroencephalographic effects of isoflurane with and without burst suppression. Anaesthesiology 2004; 101: 847-54.

24. Münste S, Klockars J, van Gils M, et al. The Narcotrend index indicates age-related changes during propofol induction in children. Anesth Analg 2009; 109: 53-9.

25. Dennhardt N, Ardnt S, Beck C, et al. Effect of age on Narcotrend index monitoring during sevoflurane anesthesia in children below 2 years of age. Paediatr Anaesth 2018; 28: 112-9.

26. Fried LP, Tungen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001; 56: M146-57.

27. Brice DD, Hetherington RR, Utting JE. A simple study of awareness and dreaming during anaesthesia. Br J Anaesth 1970; 42: 535-42.

28. Breiman L. Random forests. Machine Learning 2001; 45: 5-32.

29. Bottros MM, Palanca BJ, Mashour GA, et al. Estimation of the bispectral index by anesthesiologists: an inverse turing test. Anaesthesiology 2011; 114: 1093-101.
30. Schnider TW, Minto CF, Shafer SL, et al. The influence of age on propofol pharmacodynamics. Anesthesiology 1999; 90: 1502-16.

31. Peeters MYM, Bras LJ, DeJongh J, et al. Disease severity is a major determinant for the pharmacodynamics of propofol in critically ill patients. Clin Pharmacol Ther 2008; 83: 443-51.

32. Besch G, Liu N, Samain E, et al. Occurrence of and risk factors for electroencephalogram burst suppression during propofol-remifentanil anaesthesia. Br J Anaesth 2011; 107: 749-56.

33. Purdon PL, Pavone KJ, Akeju O, et al. The ageing brain: age-dependent changes in the electroencephalogram during propofol and sevoflurane general anaesthesia. Br J Anaesth 2015; 115(Suppl 1): i46-57.

34. Akeju O, Pavone KJ, Thum JA, et al. Age-dependency of sevoflurane-induced electroencephalogram dynamics in children. Br J Anaesth 2015; 115(Suppl 1): i66-76.

35. Lee JM, Akeju O, Terzakis K, et al. A prospective study of age-dependent changes in propofol-induced electroencephalogram oscillations in children. Anesthesiology 2017; 127: 293-306.

36. Epstein RH, Maga JM, Mahla ME, Schwenk ES, Bloom MJ. Prevalence of discordant elevations of state entropy and bispectral index in patients at amnestic sevoflurane concentrations: a historical cohort study. Can J Anesth 2018; 65: 512-21.

37. Avidan MS, Graetz TJ. Monitoring the brain strikes a discordant note for anesthesiologists. Can J Anesth 2018; 65: 501-6.

38. Manyam SC, Gupta DK, Johnson KB, et al. When is a bispectral index of 60 too low?: Rational processed electroencephalographic targets are dependent on the sedative-opioid ratio. Anesthesiology 2007; 106: 472-83.

39. Ghoneim MM, Block RI, Haffarun M, Mathews MJ. Awareness during anesthesia: risk factors, causes and sequelae: a review of reported cases in the literature. Anesth Analg 2009; 108: 527-35.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.