Comparison of contemporary EEG derived depth of anesthesia monitors with a 5 step validation process


Abstract: During the last decennium, a growing number of depth of anesthesia monitors, extracting information from the spontaneous electroencephalogram (EEG) have been developed and commercialized. The growing interest in depth of anesthesia monitoring resulted in an intensified technological progress. Innovations on both hardware and mathematical algorithms were introduced for improving the extraction of data. Because of the abundance of monitors now commercially available, it becomes increasingly important to develop a standardized reproducible methodology for comparing depth of anesthesia monitors. In this review, the authors present a strategy to compare monitors of the hypnotic component of anesthesia, based on the available literature and their own experience with validation studies. They also discuss the level of validation of the most commonly used EEG derived depth of anesthesia monitors.

Key words: Anesthesia; depth of anesthesia monitors; hypnotic drug effect; EEG; validation; Bispectral Index; Narcotrend; Spectral Entropy; Cerebral State Index; Patient State Index; Index of Consciousness.

INTRODUCTION

The search for an objective, reproducible and continuous measure of cerebral hypnotic drug effect has lead to the development of monitors that interpret the changes in neurophysiologic endpoints, such as the spontaneous electro-encephalogram (EEG) (19, 44). The major advantage of this technology is the availability of continuous information on the cerebral state even during conditions during which patients have lost all responses to a verbal or painful stimulus.

Although many validation studies have been performed, no consensus on the minimum level of performance is available. Why would it be favorable to have such a consensus on the validation methodology? First, there is no gold standard against which to test EEG derived indices of anesthetic depth because the fundamental effects of anesthetic drugs on consciousness remain unknown. The information obtained by these monitors therefore remains a surrogate for the more fundamental effect of anesthetic drugs on consciousness (43). In this view, EEG derived indices have been called “black boxes”, that might result in dangerous therapeutic choices (43, 45). A good example of such a potential conflict can be found when using EEG derived indices during conditions with high risk for brain ischemia. The EEG changes evoked by ischemia might be reflected on the EEG derived index as an overdosing of cerebral hypnotic drug effect. This triggers the anesthesiologist to erroneously decrease the level of sedation. In order to avoid such therapeutic conflicts a good understanding of the strengths and weaknesses of the algorithms is mandatory.

Despite the lack of a golden standard for consciousness, a consensus on the validation method allows to define a minimal level of performance in several clinical and experimental settings.

An additional reason to standardize the validation process is to provide a more efficient strategy to find the most relevant information out of the vast number of trials in the literature. For some older monitors, e.g. bispectral index (BIS), thousands of
studies have been published which hampers the ability to have a good overview of what is relevant and what is not.

In this review, we propose a five step method for validating new monitors of the hypnotic component of anesthesia based on the available literature and our own experience at the University Ghent (Table 1) (107). Secondly, we performed a literature search to evaluate the level of validation of the currently commercialized devices: Bispectral Index (BIS) (Aspect medical, Newton, MA, USA), Spectral Entropy (RE/SE) (GE Healthcare, Helsinki, Finland), Narcotrend (Schiller AG, Baar, Switzerland), the Patient State Index (PSI) (Hospira, Illinois, USA), the Cerebral State Index (CSI) (Danmeter, Odense, Denmark) and the Index of Consciousness (IoC) (Morpheus, Barcelona, Spain) (Table 2).

**Table 1**

<table>
<thead>
<tr>
<th>Validation step</th>
<th>What is being evaluated ?</th>
<th>Methods and materials</th>
<th>Limitations of this validation step</th>
<th>Advantages of this step</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical sign validation</td>
<td>1) Is the index able to distinguish the presence or absence of predefined clinical signs of cerebral hypnotic drug effect ?  2) Does the interaction between opioids and hypnotics alter sensitivity ?</td>
<td>1) Use of dichotomous or (preferably) scaled clinical signs as independent variable  2) Use of reproducible drug administration e.g. TCI or constant F&lt;sub&gt;e&lt;/sub&gt;  3) Steady-state required  4) Both induction and awakening studied</td>
<td>1) Only limited range of cerebral hypnotic drug effect studied  2) Poor assessment of “deep” anesthesia  3) Unclear relationship between clinical signs and consciousness</td>
<td>1) Direct clinically relevant information  2) Several monitors can be compared in the same setting</td>
</tr>
<tr>
<td>Pharmacological (PKPD) validation</td>
<td>1) What is the correlation between the concentrations of drugs (measured or predicted) and the behavior of the index ?  2) Is response comparable between drugs ?</td>
<td>1) Reproducible drug administration ET% or TCI parameters adequately reported  2) Advanced statistics  3) Large range of drug concentrations tested  4) Non steady state conditions tested</td>
<td>1) Variable relationship between drug concentration and consciousness  2) Dependent on adequacy of the PKPD model used</td>
<td>1) Index can be compared against objective and reproducible measures (concentrations)  2) Assessment possible of “deep” anesthesia  3) Non-steady state evaluation  4) Several monitors can be compared simultaneously</td>
</tr>
<tr>
<td>Clinical utility</td>
<td>1) Performance under clinical conditions (= effect of mechanical and physiological factors)  2) Effects during drug interactions [PK or PD]</td>
<td>1) Specific clinical conditions in which study was conducted  2) Interactions studied in a two-way step-up procedure</td>
<td>1) Low reproducibility  2) Monitor specific results  3) Case reports have little scientific value</td>
<td>1) Many problems are revealed that were not apparent during the first validation phase  2) Advanced PKPD relationships become available in well performed interaction studies</td>
</tr>
<tr>
<td>Outcome improvement</td>
<td>Drug use reduction/recovery time and quality of anesthesia/length of stay/mortality/morbidity/awareness</td>
<td>Large (multicentre) trials with high number of patients  1) Time consuming and expensive  2) Limitations to quantify outcome</td>
<td>1) Reveals clinically relevant conditions for using EEG derived monitors</td>
<td></td>
</tr>
<tr>
<td>Cost-effectiveness</td>
<td>Is the cost of the device in balance with the benefits in outcome ?</td>
<td>Need for adequate outcome studies first  Adequate calculation of all costs and benefits</td>
<td>1) Country and year dependent (US versus Europe, price evolutions)  2) Defining relevant costs and benefits</td>
<td>Pressure on the market to lower costs of new devices</td>
</tr>
</tbody>
</table>

ET = endtidal, TCI = target controlled infusion, PKPD = pharmaco-kinetic-dynamic, PK = pharmaco-kinetic, PD = pharmaco-dynamic.

**FIVE STEPS TO FULL VALIDATION OF AN INDEX OF THE HYPNOTIC COMPONENT OF ANESTHESIA**

*Step 1 : Validation of the index for detecting clinical signs of anesthesia*

Up until today, we still have no direct measure of consciousness available in clinical practice (43). Throughout history of anesthesia, clinical signs have been used to describe anesthetic drug...
effects (29). These clinical endpoints of anesthesia are generally based on the application of a predefined stimulus (name calling, touching of the eyelashes, trapezius muscle squeeze, etc…) and observing the presence or absence of a response (88). The clinical validation of an index of the hypnotic component of anesthesia describes the variability of the index around the transition from one clinical condition to the other. During this validation phase, several questions should be addressed:

### Table 2

**Comparison of different devices**

<table>
<thead>
<tr>
<th>Device</th>
<th>BIS</th>
<th>Spectral Entropy</th>
<th>Narcotrend</th>
<th>PSI</th>
<th>CSI</th>
<th>IoC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturer</td>
<td>Aspect Medical Systems, Inc, Norwood, MA, USA</td>
<td>General Electrics Healthcare, Helsinki, Finland</td>
<td>Schiller AG, Baar, Switzerland</td>
<td>Hospira, Illinois, USA</td>
<td>Danmeter, Odense, Denmark</td>
<td>Morpheus, Barcelona, Spain</td>
</tr>
<tr>
<td>Clinical validation</td>
<td>1) Excellent correlation for both single and scaled endpoints 2) Different thresholds during opioid interactions</td>
<td>1) Excellent correlation for both single and scaled endpoints 2) Different thresholds during opioid interactions</td>
<td>Excellent correlation for both single and scaled endpoints</td>
<td>Only limitedly validated, study with poor methodology</td>
<td>Acceptable correlation for both single and scaled endpoints</td>
<td>Acceptable correlation for single endpoints, OAAS not described</td>
</tr>
<tr>
<td>Pharmacological (PKPD) validation</td>
<td>1) Excellent correlation between BIS and CePROP 2) Little discrimination during burst suppression 3) Good correlation between BIS and stepwise F, change</td>
<td>1) Better baseline variability than BIS 2) Slightly lower correlation with CePROP due to steep decent at LOC 3) Good correlation with burst suppression %</td>
<td>1) Excellent correlation between BIS and CePROP 2) Good correlation between Narcotrend and F, change</td>
<td>Only limitedly validated, study with poor methodology</td>
<td>1) Acceptable correlation with CePROP 2) Less validation for inhalation Anesthetics</td>
<td>Not available</td>
</tr>
<tr>
<td>Clinical utility - adults</td>
<td>Problems with high freq, EMG, electrocoag, ketamine, seizures, inotropic use</td>
<td>1) Same problems as BIS RE has a faster response compared to BIS 2) Value of EMG content in RE questionable 3) Useful as analgesia monitor? 4) Problems with ketamine and N,O</td>
<td>Same as BIS</td>
<td>Little studies available due to limited use of the device</td>
<td>Very vulnerable for EMG activity. Ketamine, inotropies etc… not yet tested</td>
<td>Studies needed</td>
</tr>
<tr>
<td>Clinical utility - peds</td>
<td>1) Might offer advantages 2) Pediatric electrode available</td>
<td>Studies needed</td>
<td>Studies needed</td>
<td>Studies needed</td>
<td>Studies needed</td>
<td>Studies needed</td>
</tr>
<tr>
<td>Outcome improvement</td>
<td>1) Conflicting results, mostly small effect 2) Effects on mortality and morbidity unclear 3) Some scientific support for use as a detector of awareness in high risk patients</td>
<td>Studies needed</td>
<td>Studies needed</td>
<td>Studies needed</td>
<td>Studies needed</td>
<td>Studies needed</td>
</tr>
<tr>
<td>Cost-effectiveness</td>
<td>Poor (drug use, length of stay, awareness), Use only in selected cases</td>
<td>Studies needed</td>
<td>Studies needed</td>
<td>Studies needed</td>
<td>Studies needed</td>
<td>Studies needed</td>
</tr>
</tbody>
</table>

© Acta Anesthesiologica Belgica, 2009, 60, n° 1
Is the index able to detect the transition from presence to absence of a response evoked by a hypnotic drug? What is the sensitivity/specificity of the index for detecting the transition? Are the absolute values of the index comparable when different hypnotic drugs are used? Does the interaction between opioids and hypnotics modify the sensitivity/specificity of the index for detecting the clinical endpoint?

Although this validation phase has direct clinical relevancy, some shortfalls should be kept in mind when evaluating the quality of a clinical validation study (43). A response to a stimulus is dichotomous (it is either present or absent), while the EEG derived index is continuous. The ability of a continuous index to detect a single momentum of transition will partially be determined by the intensity and frequency of the stimuli delivered to the patient. A slow responding index will benefit from wide intervals between clinical observations. In contrast, a fast responding index will profit from the opposite. We will illustrate this with an example. When you determine loss of consciousness by name calling on a 15 seconds interval, and the patient loses consciousness somewhere during that 15 seconds, a fast responding monitor will have a significant lower value at the second observation, whereas a slow responding monitor (with a delay that exceeds 15 seconds) will not have responded on the clinical change yet. In contrast, if you increase the interval of name calling to 1 minute, both monitors will have detected the loss of consciousness at the second observation. Thus, the slow responding monitor has an advantage when being tested over wide intervals.

An additional drawback of this validation is the inability to study the performance of the continuous index during conditions when all clinical responses have disappeared (13). At that level of cerebral hypnotic drug effect, we can not evaluate the performance for discriminating deep levels of anesthesia.

Another drawback is more fundamental. Although applied by many clinicians in daily practice, the correlation of clinical signs of anesthesia and the real cerebral hypnotic drug effect on memory formation or consciousness is not established at all (43). Even if the EEG derived index should have a close correlation between the fundamental cerebral effects on consciousness, it does not guarantee a close correlation with clinical surrogate observations, such as “loss of eyelash reflex”. The latter results from independent neurological pathways, that are not directly involved in the neurological processes providing consciousness.

In order to bypass the problem of describing only one momentum, specific clinical scales have been developed that describe a range of hypnotic effects. The Assessment of Alertness and Sedation scale (OAAS/S) was originally developed by Chernik et al. (22). The OAAS/S uses gradually intensified verbal and tactile stimuli to classify the patients in 6 separate levels of progressively deeper anesthesia. This scale has been used many times as an independent endpoint for validation studies.

Due to the difficulty of detecting the exact momentum of transition from one clinical condition to another, it is recommended that the clinical observations are performed during pharmacokinetic-dynamic (PKPD) steady-state conditions (43, 98). The goal of this validation is to quantify the performance of the monitors to detect clinical hypnotic effects, independently of the amount of drug needed for that effect to occur. By controlling the pharmacokinetic-dynamic differences between patients, it becomes more obvious that some patients loose consciousness at lower drug concentrations than others. This setting allows to test whether the EEG derived index can distinguish both responsive versus unresponsive patients. If we study this in non steady-state conditions, it is impossible to evaluate whether the monitor measures a clinical change or rather a difference in drug concentration. Steady state conditions can be established by target controlled infusion technology for intravenous drugs, or by targeting the end-tidal concentration of inhaled anesthetics.

Finally, it is recommended that the clinical observations are performed both during induction as well as during recovery (84). Due to the large interpatient variability in pharmacokinetic-dynamic characteristics, additional confounding variables are involved for evoking return of consciousness, compared to the more predictable loss of consciousness. Examples of such confounding characteristics are: the context sensitive half time and differences in organ function that provoke a less predictable washout of drug effect. Therefore, a full clinical validation includes a description of measurement accuracy for both induction and recovery.

**Step 2 : Pharmacokinetic-dynamic validation : Quantifying the correlation between the index and the hypnotic drug concentration**

In this phase of validation, the correlation between the EEG derived index and the administered doses of hypnotic medication are investigated. This correlation should be described for both intra-
venous as well as for inhaled anesthetics. By introducing continuously available information on the underlying drug concentration as an independent variable (or “golden standard”), more advanced comparisons between indices become available. For inhaled anesthetics, this “golden standard” is generally the measured end-tidal vapor percentage (20). For intravenously administered hypnotics, this is the predicted effect-site concentration of the studied drug (109). Both these independent variables can be criticized as measures of hypnotic drug effect, as they both have an uncertain correlation with the fundamental mechanisms of consciousness. However, because both endpoints provide continuously available, reproducible, and objective data, they allow a perfect comparison of the performance of several EEG derived monitors.

Additionally, the pharmacokinetic-dynamic validation allows to test the performance of the indices during non-steady state condition, as well as during deep anesthesia, when all clinical responses on a stimulus have disappeared. Consequently, compared with the validation of the indices as detectors of clinical signs, the performance is tested over a wider spectrum of cerebral hypnotic drug effect. Additionally, pharmacodynamic models can be developed, that depict the variability of the measurement within the studied population (42, 109). These models allow extrapolation of the corresponding drug concentration in conditions where only EEG derived information is present. They also provide the necessary information for the further development of closed loop anesthesia controlling systems (96).

Advanced statistical methods can be applied on the datasets obtained with this methodology. These methods include:

1. the baseline variability: How large is the inter- and intraindividual variability in the awake patient? (14)
2. the correlation of the index with the level of burst suppression as a reflection of the ability to detect deep anesthesia (13, 106, 109)
3. the prediction probability (91, 92) or individualized spearman rank correlation (106) which both reflect the all round accuracy to detect a wide spectrum of drug effects
4. the non linear mixed effect modeling (88, 106, 109).

Step 3: Validation of performance under clinical conditions

The devices have to be tested in a clinical setting where additional factors are likely to interfere with the performance compared to a well controlled lab setting. All EEG derived indices have to cope with major interfering factors such as conditions with low impedance of the electrodes, high EMG activity, and suboptimal electrode positioning. Also specific drugs and physiological responses on haemodynamics might alter the correlation between the indices and the cerebral hypnotic drug effect. Finally, a population of patients planned for surgery might have different characteristics compared to healthy volunteers.

Many of these problems only arise once a monitor is implemented in clinical practice. Well known interferences on EEG measurements are the electromyogram (76, 90), the electrocardiogram (114), hot air blanket systems (38), electrocoagulation (34, 59) etc… Physiological conditions, such as age (9, 87), race (70), gender (40, 41, 47), temperature (27, 82), hypercapnia (55, 102, 125), hypoglycemia (113, 123) etc also have a significant effect on raw EEG. Subsequently, all these conditions might alter the behavior of EEG derived indices.

During routine practice, an important number of drugs is administered to the patient. These drugs might interfere with the measurement either through a direct mechanism on the raw EEG (80, 101) or by interacting with the pharmacokinetics of the hypnotic drugs (68, 121). The use of neuromuscular blocking agents has a distinct effect on the electromyogram and the signal quality of the EEG measurement (60, 108).

Finally, the clinical and pharmacokinetic-dynamic validation studies are generally performed on healthy adult volunteers or ASA 1 patients. During the clinical utility validation phase, the performance of the indices in other categories of patients should be determined (100).

Step 4: Improving outcome

Although a measurement might have a high performance in the above test phase, eventually, it must be studied whether the new index improves the outcome of the patient. It should be determined whether the use of a cerebral hypnotic drug effect index has an influence on the amount of anesthetics used, the recovery time or the length of stay in the recovery area or in the hospital. Such endpoints are rather easy to quantify and do not require excessive numbers of patients. It becomes more complicated when endpoints such as morbidity, mortality or awareness are under investigation. Due to the low incidence of these endpoints and the many
covariates that have a potential effect, a very large number of patients will be needed to obtain a sufficiently powered trial (65).

**Step 5: Cost-effectiveness**

It is important to consider that every measurement costs money to the patient and to society. Therefore the costs must be balanced against the benefits. Additionally, to prove a cost/benefit, it may be necessary to determine whether the general population or rather a selection of high-risk patients should be monitored to obtain the desired result. Again, large multicentre trials are needed for this cost-benefit validation (65).

**Level of validation of commercialized EEG derived indices of cerebral hypnotic drug effect**

In this section, we describe the current level of validation of commercialized EEG derived indices based on the five steps of validation. For every monitor and for every validation step, we aim to present some key articles that contributed strongly to a safer implementation of the index of cerebral hypnotic drug effect in clinical practice. In many studies, more than one monitor is included. Consequently, some studies will be mentioned several times for different monitors.

**BISPECTRAL INDEX (BIS) (ASPECT MEDICAL, NEWTON, MA, USA)**

Introduced in the mid nineties, the Bispectral Index (BIS) remains the most extensively studied index of hypnotic drug effect, with thousands of publications. In this view, BIS is often considered to be a “golden standard” of EEG extracted indices of hypnotic drug effect. However, it must be stressed that BIS can not be considered as such. It merely has the advantage of being the first worldwide commercialized system (19).

**Clinical Sign validation**

It has been demonstrated in numerous publications that BIS has an acceptable correlation with the presence or absence of many clinical signs of cerebral hypnotic drug effect. This has been shown both for single endpoints, such as loss of eyelash reflex and loss of response to name calling, as well as for gradual clinical scales, such as the OAAS scale (15, 95, 110). Although BIS detects the subsequent levels of the OAAS scale, the large overlap of BIS between the subsequent OAAS levels, does not allow to define clinical useful thresholds, with sufficiently high sensitivity and specificity, but this appears to be a problem for all commercialized devices (95). Moreover, when opioids are combined with hypnotics the threshold compatible with loss of a response to both verbal and tactile stimuli will be higher compared to anesthesia without opioids (105). In other words, the clinical response (e.g. loss of response to name calling) will appear at higher BIS values and at lower propofol concentrations when opioids are administered first. This finding confirms that BIS only correlates with the hypnotic component of anesthesia and the corresponding hypnotic drug concentration. This interaction complicates the interpretation of a single BIS value a lot, because one BIS represents different clinical and pharmacological conditions, dependent of the presence or absence of opioids. However, this behavior does not decrease the sensitivity/specificity for detecting clinical endpoints of anesthesia. In other words, the change in BIS value will be equally sensitive for the change in cerebral hypnotic drug effect.

In contrast to many other indices, the spectrum of probabilities of response has been described in detail for BIS during propofol and remifentanil interaction (12).

**Pharmacokinetic-dynamic validation**

**STRUYKS et al.** studied the ability of a simultaneously attached BIS, Spectral Entropy and A-line Auditory evoked potential Index (Danmeter, Odense, Denmark) to detect a predicted effect-site concentration of propofol during a non steady state propofol induction (106). This study included several advanced statistical tests to compare the accuracy of all monitors: PK analysis, Individualized Spearman Rank Correlation, baseline stability and the ability to detect deep anesthesia as reflected by the percentage of burst suppression. Moreover, pharmaco-dynamic modeling using non linear mixed effect modeling techniques, depicted the population variability of the indices versus a drug concentration (89). Such studies reveal significant amounts of information that allow a direct comparison between very divers EEG derived indices. In this section we will only discuss the results for BIS. The studies of **STRUYKS et al.** revealed a strong correlation between BIS and the predicted effect-site concentration of propofol, a high performance for detecting increases and decreases in the drug concentration. However, between 0 and 40% of burst...
suppression, the BIS values have little discriminating capacity. When burst suppression levels increase above 40%, a linear decrease in BIS values occurs (13, 16, 106). It is recommended to use the suppression rate (SR) as indicator of cerebral hypnotic drug effect, rather than BIS, to describe very deep anesthesia levels.

For inhalation anesthesia, other methods are mandatory to depict the full spectrum of cerebral hypnotic drug effects. When given as a solitary induction agent, in non steady-state conditions, it remains a challenge to evoke a transition from fully awake to high levels of burst suppression without having respiratory or haemodynamical adverse events (23). Moreover, inhaled anesthetics have the ability to shift the balance between nociception and antinociception at the level of the spinal chord in such a way that the effect of opioids are potentiated during inhaled anesthesia. Therefore, the interpretation of inhaled anesthetics as having pure “hypnotic” cerebral effects is questionable (43, 107). Despite of these pitfalls, several studies using steady state methodology to evoke a transition from fully awake to high levels of burst suppression with negative results for BIS as a monitor of awareness etc. (66). Although these studies of “awareness” have been performed. The B-Aware Trial of MYLES et al. is a multicentered, double blinded, randomized controlled trial that provides some evidence that BIS reduces the incidence of awareness by 87%, in a “high risk” population, including patients scheduled for cardiac surgery, trauma, direct laryngoscopy, caesarian section and a history for awareness etc. (66). Although these results suggest a valuable addition to our anesthesia armamentarium, it does not support a generalized use of BIS in all patients.

In contrast, AVIDAN et al. published a study with negative results for BIS as a monitor of awareness (10). However, this study was fiercely criticized for many reasons (46, 64). A major concern was the potential bias between groups, because the BIS monitored group was compared with a protocol based on the measurement of end-tidal gas concentration. By doing so the group without BIS monitoring was not allowed to receive low concentrations that more recently developed EEG derived indices suffer from comparable problems.

Although the performance of BIS to detect clinical signs of anesthesia is less consistent in children compared to adults, a sufficient number of studies has shown that BIS might offer advantages for a pediatric population (1, 23, 26, 63, 75, 78, 120). In this view smaller pediatric BIS electrodes have been developed.

Outcome validation

Several studies have investigated the effect of using BIS on recovery times, hospital stay or drug consumption. Conflicting results are found, depending on the study methodology and the endpoints of interest. But even those studies that do find a significant reduction in recovery times or drug consumption indicate that these outcome effects are clinically insignificant or too small to accept the additional cost of a BIS electrode (93, 97). The effects on morbidity and mortality are more difficult to study, due to the need of large numbers of patients (65). However, a prospective study of Monk et al. suggests that excessive anesthesia, as (blindly) measured by BIS, might correlate with an increased one year mortality (62). This finding needs to be confirmed, before drastic changes in anesthetic management have to be implemented. Moreover, it was not clear whether titration of anesthesia towards BIS values between 40 and 60 would result in an improved one year mortality (62)?
of sevoflurane below a MAC of 0.7. Therefore, the real value of BIS as a tool to prevent awareness is probably not reflected in this study and more carefully designed trials are still mandatory.

Economic validation

The B-Aware trial calculated that it takes US $2200 to avoid one case of awareness in a high risk population, taking into account that one electrode costs US $16 at that time. The reduction in drug use or hospital stay is mostly insignificant to support a general use of expensive electrodes. Finally, because it is not established yet whether BIS has an effect on morbidity and mortality, the potential effects on costs has not been studied either.

Spectral Entropy (RE/SE) (GE Healthcare, Helsinki, Finland)

The M-Entropy is a module developed by Datex-Ohmeda and was launched in 2003. It converts the entropy content of the EEG signal into an index that measures the level of anesthesia (18). The module shows the index in two separate numerical values (RE and SE) on a scale between 0 and 100. As the RE includes part of the frontal EMG spectrum it must be studied as a separate index compared to SE, that only includes EEG information (104, 111).

Clinical Sign validation

The ability of RE and SE to describe distinct clinical signs of anesthesia has been studied during propofol, thiopental and sevoflurane anesthesia (104). These studies show a very high accuracy for RE and SE, approaching perfect predictions in some cases. However, these results are partially biased due to the clinical endpoints chosen (fully awake versus deeply asleep is much easier to discriminate by a monitor compared to subtle stages of sedation). A more realistic approach can be found in a clinical trial of Vanluchene et al. (105). This study confirmed the ability of both SE and RE to describe the more gradual OAA S scale, while comparing their performance with a simultaneously attached BIS and A-Line monitor (105). Although these devices were all able to describe the gradual transition from awake to deep anesthesia in the individual patient, it remains impossible to find adequate thresholds of RE and SE for every step in the OAA S scale for the population (4). This is due to a large overlap of absolute values of RE and SE between the anesthetic stages (4, 105).

Pharmacokinetic-dynamic validation

During propofol anesthesia, Vanluchene et al showed a better baseline variability for RE and SE compared to BIS indicating that awake patients will have a more constant value compared to BIS (106). This was confirmed for inhaled anesthetics, by Bruhn et al. (14, 18). The correlation between the propofol concentration and both RE and SE is sufficiently high and comparable with BIS. In contrast, Ellerkmann et al. concluded that BIS seems to show a slight advantage in predicting propofol effect-site concentrations compared to spectral entropy as measured by the prediction probability (pK) (30).

Validation of performance under clinical conditions

Feld et al. demonstrated a comparable performance of RE and BIS as a measure of hypnotic drug effect during laparoscopic procedures. He describes a faster response of RE compared to BIS in clinical anesthesia, indicating a potential advantage for entropy when sudden changes in drug effect are expected (31).

Although a part of the EMG spectrum is included in the RE algorithm as a source of information on the cerebral hypnotic drug effect, it remains unclear whether this unique approach has advantages compared to purely EEG derived indices (94). The use of muscle relaxants has clear effect on RE but SE seems to be less affected (11, 56, 108).

Independent of the ability to measure hypnotic drug effects, some studies investigate the ability of entropy to detect analgesia (118). The RE-SE difference has been successfully used as a guide for titrating remifentanil, suggesting the potential use of this measurement as a monitor of the balance between nociception and antinociception (58). These preliminary results have to be considered with much prudence, because many questions remain on the clinical value of RE-SE as a monitor of analgesia (99).

Hans et al. and Verbeeck et al. demonstrated that the entropy monitor is not adequate for monitoring ketamine-induced hypnosis (36, 108). Spectral entropy also has limitations for detecting loss of consciousness induced by N₂O (5).

Outcome and economic validation

Vakkuri et al. showed that there is a decreased consumption of propofol and a shorter recovery...
time in patients with spectral entropy monitoring compared with standard practice (103). A similar sparing effect of sevoflurane was found by Aimé et al. (2). At this time no sufficiently powered studies are available on the ability of entropy monitoring to decrease the incidence of awareness. Möller et al. showed that entropy was a reliable predictor of auditory recall in volunteers (61). But at the other hand, reports can be found of awareness despite low entropy (124).

Narcotrend (Schiller AG, Baar, Switzerland)

The Narcotrend monitor is an EEG monitor that has been developed at the University Medical School of Hanover, Germany and has been commercially available in Europe since 2000. The Narcotrend algorithm classifies the EEG traces into different stages from A (awake) to F (general anaesthesia with increasing burst supression) (49). The newest Narcotrend software version includes a dimensionless Narcotrend index from 100 (awake) to 0 (electrical silence).

Clinical sign validation

Kreuer et al. have performed several studies investigating different steps of the validation process for Narcotrend. Narcotrend is able to detect loss of eyelash reflex during anesthesia with opioids and propofol (52). Both BIS and Narcotrend adequately detected loss and return of consciousness in a non steady state protocol. Schmidt et al. showed high correlation between Narcotrend and the detection of 4 clinical conditions during routine anesthesia: awake, steady state anesthesia, first reaction during recovery and extubation (83). Although, these studies provide some degree of clinical validation, the methodology used can be criticised. Measurements were performed during surgery, which might interfere with the relationship between hypnotic drug concentrations versus the expected EEG response. Secondly the studies were performed during non steady state anaesthesia combining a variety of opioids and hypnotic drug concentrations.

Pharmacokinetic-dynamic validation

Some studies aim to bypass the pharmacological validation by describing the correlation between BIS (as a “golden standard” of hypnotic drug effect) and the Narcotrend (49, 51). However, it is more adequate to study the ability of both indices to detect a hypnotic drug concentration. For Narcotrend, several studies are available that show a high correlation between the index and the effect-site concentration of propofol (35, 54, 86) or the end-tidal percentage of inhaled anesthetics (52, 53, 116). The pharmacological validation indicates that BIS and Narcotrend are more suitable for predicting propofol effect concentrations compared to the less accurate spectral entropy.

Pilge et al. studied the time delay to calculate several EEG derived indices including Narcotrend, cerebral state index and bispectral index. He showed a large variability in the time delay depending on the quality of the signal. Pilge states that this characteristic might limit the use of EEG derived indices for fast detection of awareness and for pharmacodynamic research (73). In order to cope with this problem, future updates of the commercialized algorithms should allow to track the calculation time delay at any time.

Validation of performance under clinical conditions

Panosius et al. found that increased electromyographic activity causes less interference on Narcotrend compared to BIS (72). Narcotrend has successfully been used in children, although no monior can be considered to be superior in a pediatric population (26, 115). No studies can be found on the effects of ketamine on Narcotrend. The effects of neuromuscular blocking agents have not been investigated as thorough as for BIS and spectral entropy.

Outcome and economical validation

Rundshagen et al. and Kreuer et al. found conflicting results on propofol consumption and recovery parameters when comparing standard practice with a Narcotrend guided anesthesia (48, 81). Weber found a reduced propofol consumption in children when Narcotrend is used to titrate propofol (117). Large randomized clinical trials that examine the ability of Narcotrend monitoring to decrease the incidence of awareness are not available. Moreover, Schneider et al. showed that Narcotrend did not differentiate between deliberate return of consciousness during anesthesia and awareness (85).

The patient state index (PSI) (Hospira, Illinois, USA)

The patient state index has been introduced in 2002 into clinical practice. A considerable smaller
number of papers is available compared to BIS and Entropy.

**Clinical sign and pharmacokinetic-dynamic validation**

Xiaoguang Chen et al. designed an observational study to validate the index both clinically as well as pharmacologically (21). The authors concluded that PSI could be a useful alternative to the BIS for assessing level of consciousness during anesthesia. However, this study is performed during surgery, in non steady state conditions while using multiple drugs with a potential impact on the measurement (including neuromuscular blocking agents). Therefore, many questions remain on the comparability of PSI versus the other more adequately validated monitors. Comparable problems occur in the clinical validation study of Prichep et al. (74). In conclusion, for PSI, more standardized validation remains mandatory both on the clinical and pharmacological level.

**Validation of performance under clinical conditions and outcome**

The effects of neuromuscular blocking agents, ketamine, EMG and other potential interfering effects have not been studied yet. Although the fundamental validation of patient state index is limited, Drover et al. reported a reduction in propofol consumption and recovery parameters when using PSI as a guide for drug titration (28). This study was insufficiently powered to conclude anything on awareness.

**Economical validation**

White et al. calculated that the use of PSI is a cost effective alternative for BIS because the costs for electrodes are lower (119). However, due to the limited validation of PSI, it remains doubtful whether both devices are comparable in their performance.

The Cerebral state index (CSI) (Dammeter, Odense, Denmark)

In 2004, the cerebral state monitor (CSM), was launched as a low-cost alternative to the bispectral index, for monitoring depth of hypnosis during anesthesia. It is a handheld device that analyzes a single channel EEG and presents the CSI scaled from 0 to 100, where 0 indicates a flat EEG and 100 indicate EEG activity corresponding to the awake state. The range of adequate anesthesia is designed to be between 40 and 60 (42).

**Clinical sign validation**

Jensen et al. published a first validation study of CSI that combined both clinical and pharmacological endpoints (42). The raw EEG data obtained by means of the A-Line auditory evoked potential monitor (Dannmeter, Odense, Denmark) were reused to validate the cerebral state index algorithm (98, 109). This methodology implies that the commercialized hardware (electrodes, isolation of electrical wiring etc…) of the cerebral state monitor was not tested, but rather the performance of the mathematical algorithm. Despite this limitation it appeared that CSI had a comparable performance to BIS for detecting both clinical endpoints as well as propofol effect-site concentrations. This was clinically confirmed by Zhong et al., who did use new EEG recordings directly obtained by the CSM (126). Despite larger baseline stability for BIS (to measure awake patients), he found comparable overall performance between BIS and CSI (126). Both Hoymork et al. and Anderson et al. report larger overlap in CSI values compared to BIS when discriminating loss of consciousness or OAAS scale (7, 39). The clinical validation of Anderson et al. was performed by using either propofol or inhaled anesthetics to obtain comparable clinical endpoints (7). CSI appears more accurate than BIS to describe deep levels of anesthesia, whereas BIS scored better at intermediate levels of propofol concentrations (24).

**Validation of performance under clinical conditions**

Anderson et al. published that the CSI is not able to measure N2O sedation (3). CSI appears to be rather sensitive for electromyographic interference compared to other EEG derived indices (39, 77) Measuring CSI from either the left or right hemisphere does not appear to have clinical importance. (6) The effects of ketamine have not been investigated yet.

**Outcome validation**

No published literature was found that examined the impact of using the Cerebral State Monitor on the reduction of drug consumption or on the incidence of intraoperative awareness.
Economic validation

By using conventional ECG electrodes instead of proprietary electrodes, CSI remains accurate at a lower cost (8).

The Index of Consciousness (IoC) (Morpheus, Barcelona, Spain)

The most recently commercialized index is the “index of consciousness” (IoC). The IoC was designed to follow the unitless scaling between 0 and 100, with a target for general anesthesia between 40-60. This scale is most familiar to the anesthesiologist. However, due to the incomparable mathematical algorithms used for index calculation, one should be aware of potential differences in behavior.

Due to the recent introduction of IoC, only a limited number of studies have been published. REVUELTA et al. performed a first clinical validation in non steady state conditions using sevoflurane and remifentanil (77). He found good agreement between IoC and BIS. CSI was more vulnerable to EMG interference compared to IoC (77).

According to the proceedings of recent international congresses, more validation studies are being performed. GAMBUS et al. and MAESTRE et al. looked at detectability of clinical hypnotic endpoints evoked by respectively propofol and sevoflurane (32, 57). On a pharmacological level, IoC has been successfully used to model the interaction between propofol and remifentanil (33). For the IoC, much work still needs to be done before a full validation is available.

Conclusion

In this review we propose a five step validation process to quantify the performance of EEG derived indices for clinical use in anesthesia.

Most commercially available EEG derived indices, have not been evaluated as thorough as our schedule suggests. Table 2 gives an overview of the level of validation for all monitors based on the presence or lack of information in the scientific literature.

Before interpreting the information obtained from an EEG derived index, the anesthesiologist must be aware of the differences in level of validation between monitors. Data obtained with one EEG derived algorithm cannot be extrapolated easily to another EEG derived index!
30

B. HEYSE et al.

45. Kalkman C. J., Drummond J. C., Monitors of depth of anaesthesia, quo vadis ?, Anaesthesiology, 96 (4) : 784-7, 2002 Apr.

© Acta Anaesthesiologica Belgica, 2009, 60, n° 1
75. Reeves S. T., Havrich D. E., Tobin D. P., Conscious sedation of children with propofol is anything but conscious, PEDIATRICS, 114 (1) : 74-6, 2004 Jul.


