Assessing the predictive value of the bispectral index vs patient state index on clinical assessment of sedation in postoperative cardiac surgery patients

Adebola O. Adesanya MB BS, MPH*, Eric Rosero MD, Christine Wyrick MD, Michael H. Wall MD, Girish P. Joshi MB BS, MD, FFARCSI

Department of Anesthesiology and Pain Management, University of Texas Southwestern Medical Center and the Department of Anesthesiology, Washington University School of Medicine, St. Louis, Missouri

Keywords:
- Sedation;
- Critically ill;
- Bispectral index;
- Patient state index;
- Ramsay sedation scale

Abstract

Purpose: To compare the depth of sedation determined by Ramsay sedation scale (RSS) with electroencephalogram-based bispectral index (BIS) and patient state index (PSI).

Materials and Methods: Fifty mechanically ventilated cardiac surgical patients undergoing propofol and morphine sedation were assessed hourly for up to 6 hours or until tracheal extubation using the BIS, PSI, and RSS. Correlation between RSS, BIS, and PSI was determined, as well as the interrater reliability of RSS, BIS, and PSI. \( \kappa \) statistics was used to further evaluate the agreement between BIS and PSI.

Results: There was positive correlation between BIS and PSI values (\( \rho = 0.731, P < .001 \)). The average weighted \( \kappa \) coefficient was .40 between the BIS and PSI, 0.28 between the RSS and BIS, and 0.16 between the RSS and PSI. Intraclass correlation was consistently higher between the BIS and PSI at all time intervals during the study. Logistic regression modeling over study duration showed that the BIS was consistently better at predicting oversedation (area under the curve, 0.92) than the PSI (area under the curve, 0.78). A comparison of BIS and PSI receiver operating characteristic curves showed that the BIS monitor was a better predictor of oversedation compared with the PSI (\( P = .02 \)).

Conclusions: There is significant positive correlation between the BIS and PSI but poor correlation and poor test agreement between the RSS and BIS as well as RSS and PSI. The BIS is a better predictor of oversedation compared with the PSI. There was no significant difference between the BIS and PSI with respect to the prediction of undersedation.

© 2009 Published by Elsevier Inc.
1. Introduction

Critically ill patients in the intensive care unit (ICU) often require sedation and pain relief to provide comfort, reduce the stress response, and facilitate mechanical ventilation [1-3]. Optimal sedation and analgesia can also reduce the occurrence of agitation and confusion and other untoward consequences such as accidental extubation [1]. One of the important requirements for adequate sedation is appropriate and frequent assessment of the depth of sedation [2]. Currently, several sedation scoring systems are used to guide the administration of the sedative-hypnotic medications in the ICU [3,4]. However, these tools are subjective and can produce varying results depending upon the evaluator [5]. These shortcomings of the clinical sedation scales can result in either under- or oversedation [5].

Electroencephalogram (EEG)-based monitors such as the bispectral index (BIS) [6] and the patient state index (PSI) [7] have been used for assessing the level of hypnosis during sedation and general anesthesia. These monitors provide a value between 0 (unconscious state) and 100 (awake state) using proprietary internal algorithm and are potentially useful as objective measures of sedation in critically ill patients especially when neuromuscular blockers are used. The use of these devices in the ICU setting remains controversial [8]. Furthermore, there are no studies comparing the depth of sedation obtained simultaneously using BIS and PSI monitors in the ICU setting. The goal of this study was to compare the depth of sedation using the BIS and PSI monitors with that determined by the Ramsay sedation scale (RSS) [9].

2. Materials and methods

After institutional review board approval and informed consent, 50 patients undergoing cardiac surgery (coronary artery bypass graft and cardiac valve repair or replacement) who were expected to require at least 6 hours of mechanical ventilation in the ICU were enrolled in this study. Patients with known or suspected hypersensitivity, allergies, or other contraindications to the sedative-hypnotic drugs and those with known neurologic disorder, neurologic injury, or current use of anticonvulsants were excluded from the study. In addition, patients with preoperative abnormal serum creatinine or elevated hepatic transaminases were also excluded from the study. Anesthetic management in the operating room (OR) was standardized and consisted of 2 mg/kg of etomidate, 5 μg/kg of fentanyl, and 0.1 mg/kg of vecuronium for induction. Anesthesia was maintained with 1.0 minimum alveolar concentration (MAC) of inhaled agent, 25 μg/kg of fentanyl, and neuromuscular blockade as needed. The serial “train-of-four” measurement was used to verify recovery from neuromuscular blockade or paralysis on arrival in the ICU. Sedation in the ICU was provided with propofol infusion, which was started in the operating room after successful separation from cardiopulmonary bypass. Propofol infusion was titrated to maintain a sedation goal of MRS of 3 to 4 (Table 1). Analgesia was provided with bolus doses of morphine (maximum, 8 mg intravenously, over study duration) consistent with routine care in the study ICU. Upon arrival in the ICU, the electrodes for BIS (version XP; Aspect Medical Systems, Natick, Mass) and PSI (PSA 4000; Physiometrix Inc, Billerica, Mass) were applied to the forehead consistent with the manufacturers’ recommendations. The BIS sensors (BIS Extend) (Aspect Medical Systems, Natick, MA) were applied closer to the eyebrows, whereas PSI sensors (PSArray-2) (Physiometrix Inc. Billerica, MA) were applied above BIS sensors. Although the electrodes were applied in close proximity, the interference between the 2 monitoring systems is expected to be minimal because they use sophisticated artifact rejection algorithms and amplifiers with medical-grade isolation transformers. The bar readings on the monitors indicating signal quality and electromyogram activity were monitored, as well as raw EEG tracings. Bispectral index and PSI values corresponding to poor signal quality (signal quality index <50) combined with increased electromyogram activity when accompanied by artifacts on the EEG were noted and subsequently excluded from data analysis because of possible artifact signal pollution.

Measurements were obtained once sensor impedance was checked and accepted by both the BIS and PSI monitors (mean BIS sensor impedance, 3.9 ± 0.4 kΩ) followed by determination of RSS in that order, to avoid potential alteration of BIS and PSI values by the patient stimulation and arousal necessary to determine the RSS. The bedside nurse caring for the patients was blinded to the PSI and BIS values. The RSS, BIS, and PSI values were assessed hourly for up to 6 hours by a single observer (J. Wiley), a research nurse with previous training in monitor application and use. The study was terminated at the end of the 6-hour study period or when the patients’ trachea was extubated, whichever occurred earlier. Events during the study such as medication administration were indicated on the PSI monitor using an electronic event marker.

Statistical analysis was performed with SPSS Version 13.0 (SPSS Inc, Chicago, Ill) and SAS V9.2.13 (SAS Institute, Cary, NC). To account for repeated measurements, a mixed model analysis of variance was performed at various intervals for the RSS, PSI, and BIS variables and showed that an interaction effect was present for PSI and BIS and

<table>
<thead>
<tr>
<th>Table 1 Ramsay sedation scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
</tbody>
</table>
marginally for RSS. This indicated that separate comparisons should be performed at each interval. Consequently, analyses were performed at each repeated measure to compare differences at the observed time. The results reported in subsequent analysis reflect these repeated analyses. To adjust the significance level for multiple tests, we used a Bonferroni procedure to determine significance levels in these tests.

The relationship between BIS and PSI values were explored using scatter plots. Furthermore, the ability of BIS and PSI to predict the RSS was analyzed using Spearman correlation coefficients, and a P value less than .05 was considered statistically significant. Intraclass correlation (ICC) coefficient was used to determine the interrater reliability of RSS, BIS, and PSI. κ statistics was also used to check the agreement between BIS and PSI at light, moderate, and deep levels of sedation (light sedation: RSS, 1-2; BIS, >80; PSI, >70; moderate sedation: RSS, 3-4; BIS, 61-80; PSI, 51-70; and deep sedation: RSS, 5-6; BIS, <60; PSI, <50, using manufacturer-suggested sedation values for critically ill patients for BIS and PSI).

To determine the ability of the BIS and PSI to detect undersedation or oversedation, we dichotomized the sample based on RSS. Subjects with RSS lower than 3 were classified as undersedated, whereas subjects with RSS higher than 4 were classified as oversedated because the goal of sedation was RSS 3 to 4 [3]. Logistic regression modeling and receiver operating characteristic (ROC) curves were constructed to determine which monitor better predicted undersedation and oversedation.

### 3. Results

Of the 106 patients screened for the study, 50 patients met the inclusion criteria and were enrolled. Ninety-four percent were male, and the mean age was 62 ± 6 years (range, 55-69 years). The mean duration of observation was 271 ± 38 minutes. The PSI values were recorded at 291 time points, and the values ranged from 22 to 98. The BIS values were recorded at 290 time points, and the values ranged from 29 to 98. The dose of propofol and morphine administered during the period of observation was 448 ± 42 mg (range, 406-490 mg) and 6 ± 2 mg (range, 4-8 mg), respectively. All patients in the study achieved the target sedation goal of RSS 3 to 4.

A scatter plot of BIS and PSI values showed positive correlation between both measurements (Fig. 1). Similarly, analysis using Spearman correlation coefficient showed a positive correlation between BIS and PSI values (ρ = 0.731, P < .001).

To determine the agreement between the 2 monitors, we categorized RSS, BIS, and PSI data into light (RSS, 1-2; BIS, >80; PSI, >70), moderate (RSS, 3-4; BIS, 61-80; PSI, 51-70), and deep sedation (RSS, 5-6; BIS, <60; PSI, <50) using manufacturer-suggested sedation values for the BIS and PSI (Table 2). The range of BIS and PSI values corresponding to RSS are shown in Table 3.

Weighted κ statistics was used to check the proportion of agreement between the different measures of sedation. The average weighted κ statistic between the BIS and the PSI for all periods of measurement was 0.40. The weighted κ statistic for the RSS and BIS was 0.28, and the weighted κ statistic between the RSS and PSI was 0.16. Intraclass correlation coefficient using Shout-Fleiss reliability fixed set was calculated for each pair of measures. The ICC between the PSI and BIS was found to be consistently higher at all time intervals but highest (ICC, 0.80) at 4 hours as shown in Table 4.

### Table 2 Sedation levels and corresponding RSS, BIS, and PSI values

<table>
<thead>
<tr>
<th>Sedation level</th>
<th>RSS</th>
<th>BIS</th>
<th>PSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light sedation</td>
<td>1-2</td>
<td>&gt;80</td>
<td>&gt;70</td>
</tr>
<tr>
<td>Moderate sedation</td>
<td>3-4</td>
<td>60-80</td>
<td>50-70</td>
</tr>
<tr>
<td>Deep sedation</td>
<td>5-6</td>
<td>&lt;60</td>
<td>&lt;50</td>
</tr>
</tbody>
</table>

### Table 3 Ramsay sedation scale and observed corresponding PSI and BIS scores

<table>
<thead>
<tr>
<th>Sedation level</th>
<th>RSS 1-2</th>
<th>RSS 3-4</th>
<th>RSS 5-6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data points (%)</td>
<td>50 (17.1)</td>
<td>18 (6.2)</td>
<td>223 (76.6)</td>
</tr>
<tr>
<td>Median PSI (IQR)</td>
<td>86 (77-91)</td>
<td>81 (60-88)</td>
<td>60 (54-87)</td>
</tr>
<tr>
<td>Median BIS (IQR)</td>
<td>91 (79-97)</td>
<td>79 (74-87)</td>
<td>61 (48-74)</td>
</tr>
</tbody>
</table>

This table shows statistical dispersion of PSI and BIS scores based on RSS grouping. Median BIS and PSI scores as well as interquartile range (IQR) are displayed. The IQR is not affected by outliers or extreme values because IQR shows the distance between the 75th percentile and the 25th percentile and highlights the range of the middle 50% of the data.

**Fig. 1** Scatter plot of BIS vs PSI.
Logistic regression analysis was used to predict under-
semination (RSS, ≤2) and oversedation (RSS, ≥5) using BIS
and PSI individually across time intervals. The BIS
consistently had a higher area under the curve (AUC)
than the PSI. The 3-hour time interval provided the best
predictive value for both undersedation and oversedation.
A comparison of the ROC curves and the AUCs of BIS
and PSI showed that the BIS is a better predictor of
oversedation compared with the PSI (P = .028). There was
no significant difference between the 2 monitors with
regard to the prediction of undersedation (P = .104)
(Figs. 2 and 3).

The sensitivity and specificity of the BIS and PSI
monitor for detection of oversedation and undersedation at
the 3-hour time interval are shown in Table 5. The
sensitivity and specificity as well as the positive and
negative predictive values were consistently higher for the
BIS for detection of both undersedation and oversedation;
however, the specificity for the detection of oversedation
(BIS, 66.7%, vs PSI, 33.3%) and the sensitivity for
detection of undersedation (BIS, 70%, vs PSI, 30%) were
low for both monitors.

4. Discussion

Our findings indicate that there was a positive correlation
between the BIS and PSI based on scatter plots and
statistically significant correlation using Spearman and ICC
analyses. On the other hand, there was poor correlation
between the BIS and the RSS as well as between the PSI and
the RSS. In addition, there was good test agreement between
the BIS and PSI for detection of sedation levels categorized
by RSS groupings. The BIS was found to be a better
predictor of oversedation using ROC curve analysis;
however, the difference between the BIS and PSI with
respect to the prediction of undersedation did not reach
statistical significance.

![Fig. 2](image-url)
Comparison of ROC curves of BIS and PSI to detect undersedation. The AUC for BIS is 0.95 compared with 0.85 for the PSI. The AUC for BIS was not significantly different than the AUC for PSI (P = .104).
The sensitivity and specificity as well as the positive and negative predictive values were consistently higher for the BIS for detection of both undersedation and oversedation. It is noteworthy that the probability of detecting oversedation in study subjects who were not oversedated was low in both the BIS and PSI (specificity, 66.7% vs 33.3%), and the probability of detecting undersedation in study subjects who were truly undersedated was also low for both monitors (sensitivity, 70% vs 30%), although the BIS performed better than the PSI in both instances. Positive and negative predictive values were also found to be higher for the prediction of both undersedation and oversedation by the BIS compared with the PSI.

Given that our data consisted of repeated measurements of sedation levels, the data were analyzed across all time intervals and adjustments made to account for repeated measurements. The 3-hour time interval was found to be most predictive of undersedation and oversedation by logistic regression modeling.

Previous studies evaluating the use of BIS in the ICU with various sedatives have found inconsistent correlation between the BIS and clinical sedation scores [10-17]. A study of mixed surgical and medical ICU patients [10] found suboptimal and inconsistent correlation between BIS and sedation agitation scores (SAS), a measure of consciousness as compared with the SAS. They concluded that BIS is neither reliable nor valid for routine monitoring of the level of consciousness in the critically ill patients. Another study of surgical intensive care patients [11] receiving morphine and midazolam found a wide range of BIS scores in deeply sedated patients with a Ramsay score of 6. There was no correlation between average BIS and duration of sedation or dosages of morphine and midazolam. These findings are consistent with our observations using the BIS monitor in deeply sedated patients.

![Fig. 3](image-url) Comparison of ROC curves of BIS and PSI to detect oversedation. The AUC for BIS is 0.92 compared with 0.78 for the PSI. The AUC for BIS was significantly different than the AUC for PSI (P = .028).

### Table 5

<table>
<thead>
<tr>
<th></th>
<th>BIS</th>
<th>PSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oversedation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>91.4</td>
<td>82.9</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>66.7</td>
<td>33.3</td>
</tr>
<tr>
<td>Positive predictive value (%)</td>
<td>86.48</td>
<td>74.3</td>
</tr>
<tr>
<td>Negative predictive value (%)</td>
<td>76.92</td>
<td>45.45</td>
</tr>
<tr>
<td>Undersedation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>70.0</td>
<td>30.0</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>92.5</td>
<td>90.0</td>
</tr>
<tr>
<td>Positive predictive value (%)</td>
<td>70.0</td>
<td>42.8</td>
</tr>
<tr>
<td>Negative predictive value (%)</td>
<td>92.5</td>
<td>83.7</td>
</tr>
</tbody>
</table>

The BIS has a higher sensitivity and specificity for prediction of both oversedation and undersedation compared with the PSI. Negative and positive predictive values were also higher for the BIS for detection of both undersedation and oversedation.
patients. Conversely, a study of medical intensive care

Patients. Conversely, a study of medical intensive care

Fewer studies are available regarding the use of PSI in the

Fewer studies are available regarding the use of PSI in the

A prospective blinded study of mixed ICU patients by Ramsay et al. [19] also found a strong correlation between the PSI and the RSS. Similarly, another study [20] investigating the relationship between PSI and the sedation/agitation level measured by Richmond Agitation-Sedation Scale score found significant associations between PSI and Richmond Agitation-Sedation Scale to support the validity of the PSI as a tool to monitor the level of sedation in the ICU. Of note, most of these abstracts have not been published in peer reviewed journals since their presentation in the past 2 to 3 years.

Chen et al. [21] simultaneously evaluated the BIS and PSI in patients receiving general anesthesia and found a good correlation between the BIS and the PSI during general anesthesia. They concluded that the PSI was a suitable alternative to the BIS for assessing level of consciousness during general anesthesia. Chisholm et al. [22] compared the Ramsay scale and the Observer’s Assessment of Alertness/Sedation scale with the BIS and the PSI in 50 patients requiring mild to moderate sedation for surgical procedures. The Ramsay and Observer’s Assessment of Alertness/Sedation scores correlated strongly with each other and with both the BIS and the PSI values only at the extremes of sedation. Great variation occurred between BIS and PSI values of 61 and 80. They concluded that both the BIS and PSI cannot reliably distinguish between light and deep sedation.

For this study, patients with neurologic disorders or injury, current anticonvulsant use, or renal or hepatic dysfunction were excluded from the study because of the potential for these conditions to alter processed EEG readings at baseline and potentially decrease the correlation between processed EEG and clinical sedation scales. To our knowledge, this is the first study to simultaneously compare the BIS and PSI for assessment of the depth of sedation in mechanically ventilated patients in the ICU setting where medium- to long-term sedation may be needed.

There is increasing awareness that the EEG-based monitors can be influenced by both patient and external factors [23]. Electrical signals generated from ventilators, chest tube drainage systems, and other electrical noise arising from ICU monitors [16,23] as well as pacemakers and catecholamine infusions [24,25] have been reported to increase the BIS values. Patient-related factors such as eye movements and muscle activity were also found to increase BIS values [26]. Because the PSI is also based on EEG monitoring, it is reasonable to assume that these external and patient related factors also apply to the PSI. We were unable to determine if any of these factors affected our results.

The absence of a “goal standard” to assess the level of sedation in the ICU makes the determination of the adequacy of the BIS and PSI monitors for sedation in the critically ill patients challenging. One of the limitations of this study relates to our use of the RSS to compare the BIS and PSI values. The RSS is a validated and widely used monitor of the depth of sedation in the ICU because of its simplicity and ease of use. However, similar to all clinical sedation scores, the RSS is subject to interobserver variability and relies on the ability of the patient to respond to a stimulus. Therefore, patients in deep coma and patients treated with neuromuscular blockade cannot be assessed using the RSS. In addition, the RSS does not allow for precise reporting of different degrees of light or heavy sedation. Therefore, RSS cannot be considered a “gold standard” with which to compare other monitors of sedation. We chose RSS 3 to 4 as our sedation goal because we feel that it represents a balance between undersedation and oversedation, although a range of RSS values (i.e., 2-6) have been considered ideal in previous studies [3,17]. In addition, given that this patient group was assessed in the immediate postoperative period, the effect of residual anesthetic agents may have been a confounding factor. However, the monitors should still accurately indicate the level of sedation without much variation. The fact that the study was conducted in predominantly male patients may also limit its generalization to female patients. Although the BIS appeared to be a better predictor of both oversedation and undersedation, our inability to find statistical significance with respect to the prediction of undersedation may be related to a type II error.

The etiology of the wide variability in BIS and PSI values observed at deeper levels of sedation in this study merits further investigation by analysis of the BIS and PSI algorithm, both of which are presently proprietary. Further research is needed to identify a monitor that best predicts adequate sedation.

5. Conclusion

Given the importance of sedative administration and its impact on ventilator weaning, ICU length of stay, and cost of hospitalization, there is a need to accurately monitor sedation levels using either clinical scores such as the RSS or EEG-based monitoring such as the BIS and PSI. We found significant positive correlation between the BIS and PSI. In addition, the correlation between the RSS and the BIS, as well as the correlation between the RSS and PSI, was weak. Overall, the BIS was found to be a better predictor of both oversedation and undersedation; however, we could only demonstrate statistical significance with respect to the prediction of oversedation.
Based on this study, we are unable to recommend EEG-based monitoring to ICU practitioners needing to decide whether to monitor depth of sedation using clinical scores or EEG-based monitoring given the weak relationship between both types of monitoring found in this study. The clinical assessment of sedation, although imperfect, remains the gold standard of overall sedation evaluation.

Acknowledgments

The authors would like to thank Jackie Wiley, RN, for help with data collection and data entry.

References