CRITICAL CARE

Responsiveness of the frontal EMG for monitoring the sedation state of critically ill patients

T. S. Walsh1*, T. P. Lapinlampi2, P. Ramsay1, M. O. K. Särkelä2, K. Uutela2 and H. E. Viertio-Oja2

1 Anaesthetics, Critical Care and Pain Medicine, General Intensive Care Unit, Edinburgh Royal Infirmary, Little France Crescent, Edinburgh EH16 4SA, UK
2 GE Healthcare Finland Oy, Kuortaneenkatu 2, FI-00510 Helsinki, Finland
* Corresponding author. E-mail: twalsh@staffmail.ed.ac.uk

Editor’s key points

- This study investigated if a responsiveness index (RI) derived from the frontal EMG activity is feasible as a new method of continuously monitoring sedation state in critically ill patients.
- RI is a potentially useful dynamic and continuous measure of patient arousal during intensive care unit management.
- Low RI values might reveal patients at an increased risk of excessive sedation.

Background. Excessive sedation is associated with adverse patient outcomes during critical illness, and a validated monitoring technology could improve care. We developed a novel method, the responsiveness index (RI) of the frontal EMG. We compared RI data with Ramsay clinical sedation assessments in general and cardiac intensive care unit (ICU) patients.

Methods. We developed the algorithm by iterative analysis of detailed observational data in 30 medical–surgical ICU patients and described its performance in this cohort and 15 patients recovering from scheduled cardiac surgery. Continuous EMG data were collected via frontal electrodes and RI data compared with modified Ramsay sedation state assessments recorded regularly by a blinded trained observer. RI performance was compared with Entropy™ across Ramsay categories to assess validity.

Results. RI correlated well with the Ramsay category, especially for the cardiac surgery cohort (general ICU patients \( r = 0.55 \); cardiac surgery patients \( r = 0.85 \), both \( P < 0.0001 \)). Discrimination across all Ramsay categories was reasonable in the general ICU patient cohort \( [P_k = 0.74 \text{ (SEM 0.02)}] \) and excellent in the cardiac surgery cohort \( [P_k = 0.92 \text{ (0.02)}] \). Discrimination between ‘lighter’ vs ‘deeper’ (Ramsay 1–3 vs 4–6) was good for general ICU patients \( [P_k = 0.80 \text{ (0.02)}] \) and excellent for cardiac surgery patients \( [P_k = 0.96 \text{ (0.02)}] \). Performance was significantly better than Entropy™. Examination of individual cases suggested good face validity.

Conclusions. RI of the frontal EMG has promise as a continuous sedation state monitor in critically ill patients. Further investigation to determine its utility in ICU decision-making is warranted.

Keywords: critical care; diagnostic techniques, neurological; electroencephalogram; monitoring, physiologic; sedation, deep

Accepted for publication: 18 June 2011

Most mechanically ventilated critically ill patients require sedation.1 Optimizing sedation is important for patient safety and comfort, but also influences patient outcomes and illness costs.2–6 Excessive sedation, in particular, can prolong mechanical ventilation, intensive care unit (ICU) and hospital stay, and increase complications such as hospital-acquired infections. Recent guidelines recommend using regular clinical sedation scores and daily sedation breaks to assess sedation requirements and guide drug dosing,1 5 6 based on several randomized trials2–4 and cohort studies,7–10 that have shown improvements in a range of patient and economic outcomes when clearly defined sedation management strategies are used.

Despite strong evidence, sedation practices still vary widely between institutions and countries,11–13 and introducing and sustaining protocols into clinical practice is challenging.14 A problem with clinical sedation scales is lack of discrimination within deeper sedation states when the risk of oversedation from drug accumulation is high due to excessive dosing, impaired clearance, or greater patient sensitivity. A device that reliably detects excessive sedation in critically ill patients could improve clinical management by increasing bedside confidence to decrease sedation dose or undertake a sedation hold.

Several devices based on EEG analysis, which were primarily developed as depth of anaesthesia monitors, have been
evaluated as sedation monitors for critically ill patients. The bispectral index (BIS) correlates with clinical sedation state, but does not discriminate well between different levels of sedation and is not widely used in ICUs.\textsuperscript{11} \textsuperscript{15} We recently evaluated Entropy\textsuperscript{TM} and found similarly poor discrimination.\textsuperscript{16} Existing literature examining BIS, and our previous findings with Entropy, suggest that frontal EMG (fEMG) activity is a major confounder for these algorithms due to the frequent arousals that are typical in ICU patients, and the overlap in frequency bands between the EEG and fEMG components of the spectrum.\textsuperscript{17–19}

We hypothesized that an algorithm based on the responsiveness of the fEMG might be a clinically useful method of assessing sedation in critically ill patients. We thought fEMG responsiveness may be useful because it reflects the interaction between a patient’s conscious state and the intensity and frequency of stimulations during treatment. In this study, we used raw data from a study of Entropy in general ICU patients\textsuperscript{16} to derive a new algorithm, the responsiveness index (RI), and then prospectively assessed its validity in sedated patients regaining consciousness after routine cardiac surgery. We assessed the validity and potential clinical utility of the new measure as a method of continuously monitoring sedation state in critically ill patients. We also compared RI with Entropy to establish whether the new algorithm performed better.

**Methods**

We undertook a two-stage development and evaluation process. First, the new algorithm was developed using an existing high-quality clinical data set using raw data from a previously published evaluation of Entropy as a sedation measure in general mixed ICU patients.\textsuperscript{16} Secondly, a data set was prospectively collected in patients recovering after routine cardiac surgery. We chose cardiac surgery patients because we aimed to study patients progressing from deep sedation (on arrival in the cardiac ICU) to full consciousness over a short timescale. This would enable an assessment of the RI over a wide range of clinical sedation states within patients and test the ability of the index to discriminate between different sedation states. We also wanted to exclude the potential confounding effects of encephalopathies from this cohort.

**RI algorithm**

The RI algorithm has been derived to represent fEMG responses in relation to both external (lights, noise, care procedures) and internal (pain, anxiety) stimuli. A value of 0 is intended to indicate a completely non-responsive patient, while a value of 100 is intended to indicate a fully responsive patient. As the amount of external stimulation during intensive care management varies over time with an unpredictable pattern, and the nature, intensity, and source of stimulation also vary, the RI value is intended to be a dynamic indication of the interaction between the current level of stimulation, the sedation and analgesic state of the patient, and incorporate illness-related factors such as encephalopathy that could modify the relation between stimulation and response in individual patients.

The RI algorithm calculates the root mean square (RMS) power from the fEMG in 5 s epochs utilizing the frequency band 50–150 Hz. The mains frequencies and multiples are removed from the signal data with a 10 Hz comb filter. The power value time series is subjected to a morphological filter $F1$, which extracts the steep increases related to patient responses from the signal. The filter is implemented as an FIR filter with a step-shaped impulse response. For constant fEMG activity, the filter output is zero regardless of the baseline level. The length of the filter is eight samples corresponding to 40 s in the fEMG power time series. If the filtered power value at time point $t$ is denoted by $P_f(t)$, the RI value at time $t$ is written as:

$$RI(t) = S \left( \sum_{n=-\Delta}^{t} g(n) \log(P_f(n) + 1 \mu V) \right)$$

where $S$ is a scaling function, $g$ the weighting function and $\Delta$ the number of $P_f$ values included in the summation. Hence, the RI is a weighted average of the logarithm of the $P_f$ values added to 1 $\mu V$ in the given time window. The addition of 1 $\mu V$ is introduced to set the minimum value of the logarithm function to zero. The purpose of the weighting function $g$ is to operate as a low-pass filter and to give more weight to recent fEMG changes compared with those occurring further from the time of assessment. The value of $\Delta$ was set at 720 samples (corresponding to 60 min).

The value of $\Delta$, the length of filter $F1$, and the shape of the weighting function $g$ were determined by a heuristic and iterative analysis of ICU patient data. The purpose of the scaling function $S$ is to scale the RI values to range between 0 and 100 and to enhance the resolution in the lower RI range ($RI < 50$).

At least 30 min of data are required to obtain a reliable RI value. The RI value is updated after 5 s of new data have been recorded, and monitoring is intended to be continuous. Under these circumstances, the value presented is an index of responsiveness over the previous 60 min, with weighting to the more recent period of monitoring.

**Derivation cohort**

We used an iterative approach to derive the RI using data from a cohort of 30 general intensive care patients in whom we had previously compared Entropy with clinical sedation status.\textsuperscript{16} This cohort of patients was studied during routine intensive care management, resulting in a distribution of clinical sedation states within each patient that was sometimes variable (e.g. in patients who were awakened and weaned from ventilation), but sometimes narrow (e.g. deep sedation throughout the study period). Briefly, 30 patients admitted to a general mixed medical–surgical intensive care unit were studied for up to 72 h. Inclusion criteria were a requirement for mechanical ventilation, sedation
using continuous infusions of either midazolam or propofol and concomitant analgesic drugs if clinically indicated. Exclusion criteria were (i) a patient in whom brain injury, namely hypoxic brain injury, traumatic brain injury, and intracranial haemorrhage, was considered present at the time of enrolment to the study, (ii) drug overdose as admission diagnosis, (iii) a patient requiring neuromuscular paralysis at the time of screening for the study, or (iv) status epilepticus. We also excluded patients known to be clinically deaf or who had chronic neuromuscular disorders or brain disease that might interfere with normal clinical sedation assessment. Patients had a mean age of 59 yr (sd 17) and 73% were male. They were studied after a median 88 h after ICU admission (range 2–300 h). Primary diagnosis at ICU admission was pneumonia (7), bowel perforation or infarction (6), major vascular surgery (4), septic shock (3), trauma (3), pancreatitis (2), gastrointestinal bleeding (2), liver transplant (1), cardiac failure (1), and fulminant hepatic failure (1).

Raw EMG data were used to develop the RI algorithm using an iterative process comparing algorithm output with sedation state classified according to the Ramsay scale to progressively improve its discriminatory performance.

Cardiac surgery cohort
We studied a convenience sample of 15 patients after routine cardiac surgery. Any patient undergoing elective cardiac surgery requiring cardiopulmonary bypass was considered eligible unless they were clinically deaf or had neuromuscular disorders or brain disease that might interfere with normal clinical sedation assessment. Patients were studied from the time of return to the ICU from the operating theatre until one of the predefined endpoints was reached (see below).

Study design and protocol
A standard disposable GE Healthcare Entropy sensor (GE Healthcare, Helsinki, Finland) was applied to the forehead symmetrically relative to the midline. This position was modified from the recommended unilateral position for Entropy monitoring in order to acquire bilateral fEMG data. Each sensor comprised a strip that includes one electrode each for the left and right hemisphere, and one central ground electrode. Sensors were changed every 24 h. The monitor performed an automatic impedance test every 10 min to ensure electrical contact fidelity. Periods of poor electrode contact (impedance >5 kΩ) were rejected from the analyses. Once data recording had started, it continued until one of the following endpoints was reached: (i) patient regained consciousness and mechanical ventilation was discontinued; (ii) 72 h had elapsed; (iii) the patient or a relative requested discontinuation of the monitoring, withdrawal from the study, or both; (iv) the patient died.

Observations and management of patients during the protocol
Patients received routine clinical management throughout the study period determined by caring clinicians. Most patients were sedated with propofol as the first-choice sedative; some patients received midazolam. Sedatives were administered by continuous infusion with additional boluses as considered clinically appropriate. Analgesia was provided with alfentanil or morphine infusions. The choice of sedative and analgesic drugs, and the doses prescribed, was determined by medical and nursing staff and was not controlled for the purpose of the study. A single trained observer independent from the clinical team performed and recorded sedation scores up to every 30 min using a modified Ramsay scoring system (Table 1). For the derivation cohort, all observations were performed by a single researcher who was a trained ICU nurse not involved in the patient’s care or sedation decisions. For the cardiac surgery cohort observations, two different research nurses performed the studies, but each study was performed by a single observer. The published Ramsay scoring system was modified slightly to increase the standardization of the stimuli, specifically by including a tetanic stimulus at deep sedation levels. In the derivation data set, the number of observations used depended on the availability of the single trained observer; we did not use data recorded by clinical staff in order to minimize inter-rater variability. In the cardiac surgery validation cohort, the vocal stimulus was further standardized by using headphones and a recorded voice instructing the patient to ‘open your eyes’. The observer also recorded any significant clinical events throughout the study period. RI was calculated off-line after completing the study and the clinical team were blinded from indexes during observations. A laptop-based notation file was used to record all events during periods of observation.

Both studies were approved by the National ethics committee dealing with incapacitated patients (MREC/03/0/62). For the derivation cohort, consent was obtained from the patients’ relative or next of kin. For the validation cohort, all patients gave consent before surgery. Recruitment of patients occurred between 2003 and 2006.

Analysis
Assessment of criterion validity
We chose assessments of clinical sedation state using the modified Ramsay score performed by trained members of the research team as the reference standard. Each Ramsay

<table>
<thead>
<tr>
<th>Score</th>
<th>Clinical state or response to stimulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Patient anxious and/or agitated and/or restless</td>
</tr>
<tr>
<td>2</td>
<td>Spontaneous eye openings</td>
</tr>
<tr>
<td>3</td>
<td>No spontaneous eye opening, response to vocal stimulus</td>
</tr>
<tr>
<td>4</td>
<td>No response to vocal stimulus, response to loud stimulus</td>
</tr>
<tr>
<td>5</td>
<td>No response to loud stimulus, response to tetanic (50 Hz, 40 mA, 0.25 s pulses, duration 4 s) stimulus</td>
</tr>
<tr>
<td>6</td>
<td>No response to tetanic stimulus</td>
</tr>
</tbody>
</table>
score was compared with the RI value 1 min before the assessment. We recognized a priori that this was not an ideal reference standard for two reasons: first, the Ramsay score is a measure of sedation state at a single time point rather than over 60 min of care; secondly, RI is intended to assess responses to all stimuli above the existing background level of stimulation over time, whereas the Ramsay score assesses patient responses to defined stimuli at the time of testing. However, we used this approach because clinical sedation scores are the current standard recommended in guidelines and better reference standards are not available. As a measure of the criterion validity of RI, we examined the correlation between RI and the clinical sedation scores made using the Ramsay scale. We calculated the distribution of RI values corresponding to each modified Ramsay level and presented the results graphically by box-and-whisker plots. Performance of the RI for indicating the depth of sedation as given by the Ramsay score was tested with two statistical approaches. First, we hypothesized that RI should decrease progressively as the Ramsay score increased, but potentially in a non-linear manner. We therefore used a non-parametric rank correlation test (Spearman) to evaluate the strength of the relationship between progressive changes in sedation state and RI. Secondly, the prediction probability, \( P_K \), which is a variant of Kim’s measure of association, was calculated to assess the ability of RI to distinguish different sedation states measured with the Ramsay score. All prediction probabilities and their standard errors were estimated with the jackknife method as described by Smith and colleagues.\(^20\) With this approach, a \( P_K \) value of 0.5 indicates no predictive ability compared with the reference (in this case, clinical sedation level), and a \( P_K \) value of 1 indicates perfect prediction. The calculations were performed with the Matlab software (The Mathworks, Natick, MA, USA). We calculated \( P_K \) values for the ability of RI to distinguish each Ramsay score category from the other categories, treating all the Ramsay assessment as independent observations since the time gap between successive assessments was at least 30 min. The statistical significance of whether the \( P_K \) values differ from 0.5 (no predictive power) was tested with the \( t \)-test as recommended by Smith and colleagues.\(^20\) In all tests, the significance level was set at \( P<0.05 \).

As we were particularly interested in the ability of the RI to distinguish ‘lighter’ sedation ranges from ‘deeper’ sedation states, we also calculated the \( P_K \) value for RI for discriminating patients in the Ramsay range 1–3 from those in the Ramsay range 4–6, and for the Ramsay range 1–4 from those in the Ramsay range 5–6. We calculated \( P_K \) for the derivation general ICU group and for the prospectively collected cardiac surgery group. When used to distinguish two groups, \( P_K \) is equal to the area-under-the-curve of the receiver operating characteristic, and thus summarizes the sensitivity and specificity.

As there was a potential bias from summing repeated measures within individual patients on the overall statistical estimates, we also analysed the individual patient data in the cohorts. We compared mean RI values when the Ramsay score was recorded as 5/6 with those when it was recorded as \( \geq 4 \) within each patient. To minimize the effect of very small numbers of observations within each patient, we only calculated a mean value when at least three observations of Ramsay 5/6 or \( \geq 4 \) were recorded. We described the median (first, third quartiles) for these mean values for each of the patient cohorts.

We also explored the possible importance of encephalopathy using a retrospective analysis. All patient charts for the general ICU cohort were examined focusing on patient status when sedation was reduced. A single clinician (T.S.W.), who was blinded to the responsiveness data, classified patients as ‘high’ vs ‘low’ probability of encephalopathy based on conscious level and the presence of agitation, confusion, or both when patients were assessed without or with minimal sedation. \( P_K \) values were compared for the ‘high’ vs ‘low’ probability groups.

Overall summary values were also compared with the performance of Entropy\textsuperscript{TM} (state entropy, SE; and response entropy, RE) using box-and-whisker plots, because our aim was to derive a parameter that had superior performance to these algorithms.

### Assessment of face validity

We plotted continuous RI data and compared it with intermittent sedation scores and our annotation files. The ability of the index to adequately describe patient sedation status over time, with reference to changing intermittent clinical sedation score, was evaluated visually by the two clinicians (T.S.W. and P.R.). We also examined box-and-whisker plots relating clinical sedation state to RI for each patient within individual cases.

### Results

#### Patient characteristics

**Derivation cohort**

Characteristics for the 30 general intensive care patients have been reported previously for the Entropy evaluation.\(^16\) Data analyses were based on 487 single trained observer sedation assessment performed during a total of 1200 h of EEG/EMG monitoring. Of these, 57 were rejected from the analysis due to poor electrode contact at the time of the assessment, and 64 were rejected because too little data (<30 min) were recorded at the time of the assessment to obtain a reliable RI value. A total of 366 assessments were used to derive the initial RI algorithm. The number of observations made across Ramsay scores 1–6 was 5, 82, 140, 17, 104, and 18, respectively.

**Cardiac surgery cohort**

The mean (range) age of the cardiac surgery patients was 62 (25–82) yr. All patients regained consciousness during the data collection period and were successfully extubated. The median duration of monitoring was 7.3 (5.7, 8.1; 4.1–9.0) (first, third quartile; range) h. Data analyses were based on 96 trained observer sedation score assessments performed.
during a total of 54.7 h of monitoring [median number of scores per patient 12 (8.8, 15.3; 7–17) (first, third quartile; range)]. Seven assessments were rejected because too little data (<30 min) were recorded at the time of the assessment to obtain a reliable RI value leaving 89 Ramsay assessments for the analyses. The number of observations made across Ramsay scores 1–6 was 0, 45, 13, 5, 19, and 7, respectively.

Criterion validity
Pooled values for RI (and Entropy as a comparator) in relation to the Ramsay score for the general ICU patients (derivation cohort) and the post-cardiac surgery patients are shown in Figures 1 and 2. The rank correlation coefficient (r) between the Ramsay score and RI was 0.55 (P<0.0001) in the derivation cohort, but was significantly better for the cohort of post-cardiac surgery patients who tended to progress from deep sedation to the awake state (r=0.85; P<0.0001). The mean P_(k) (SEM) value for discriminating each sedation level from all other levels was 0.74 (0.02) in the derivation cohort and was significantly better for the cardiac surgery cohort [0.92 (0.02)]. The data indicated that RI had predictive power in the derivation cohort [value >0.5, t-test (P<0.00001)], but this was significantly better for patients studied in the cardiac surgery cohort who consistently progressed from deep sedation to the awake state (t-test P<0.00001). Visual inspection of the box-and-whisker plot data from the validation cohort suggested that RI performed particularly well in Ramsay 4–6 categories, with no inappropriate high values, unlike Entropy for which this remained problematic (Fig. 2).

The mean (SEM) P_(k) values for distinguishing Ramsay scores of 1–3 from 4–6 and Ramsay scores of 1–4 from 5–6 for the derivation and cardiac surgery cohorts are shown in Table 2. These data suggested that the RI had useful discriminative value between ‘lighter’ and ‘deeper’ clinical sedation states, which was particularly strong in the post-cardiac surgery validation cohort that progressed from ‘deeper’ to ‘lighter’ sedation during the observation period.

Examination of individual patient data showed variability between patients. In the derivation cohort of general ICU patients, the mean RI values for Ramsay 5/6 were generally lower than those when the Ramsay score was ≥4 (median (first, third quartile) mean value 25 (7, 39) vs 69 (41, 87) for Ramsay 5/6 vs ≥4). This difference was also more marked in the cardiac surgery cohort [1 (0, 3) vs 44 (42, 61) for Ramsay 5/6 vs ≥4].

Using data from patient charts, 17 patients were judged as ‘low’ probability of encephalopathy and 13 as ‘high’ probability. The P_(k) value for distinguishing Ramsay 1–4 from 5–6 was 0.90 (SEM 0.02) for patients judged as ‘low’ probability of encephalopathy and 0.70 (0.04) for patients judged as ‘high’ probability of encephalopathy, suggesting that responsiveness may discriminate clinical sedation states less well in encephalopathic patients.

Comparing RI with Entropy for the pooled patient data, RI had slightly better discriminatory power compared with SE and RE in the general cohort and markedly better discriminatory power in the cardiac surgery cohort (Table 2, P<0.0001).

Face validity
Visual inspection of the majority of continuous plots of the RI over time showed good face validity against the Ramsay scores recorded and clinical annotation notes. This was particularly evident for the cardiac surgery patients who progressed from deeper sedation states to full consciousness during the period of data recording. Examples of different clinical situations are shown in Figure 3.

Discussion
We have shown that responsiveness of the frontal EMG is a potentially useful method for continuously monitoring the sedation state of critically ill patients. The RI had excellent performance in a prospective cohort of patients regaining consciousness after cardiac surgery when compared with clinical sedation assessments, and was better than SE and RE.

Assessment of a new monitoring health technology needs an assessment of validity, preferably against a gold standard. As there is no gold standard for assessing sedation state in critically ill patients, we used clinical sedation assessments, which are a recommended standard of care in most sedation guidelines.156 We chose the Ramsay scale because it was used routinely in our ICU at the time, and is widely used in published studies. We improved its quality as a reference standard by used trained members of research staff to make all assessments independently, which minimized interrater variability and investigator bias. We also modified the Ramsay assessments to standardize the stimuli as much as possible. The detailed event annotation files enabled potential confounders, such as procedures or other clinical events, to be accounted for in the analysis. Despite these precautions, we recognized a priori that there are limitations to the use of clinical sedation assessments as the reference standard for RI. First, inter- and intra-rater variation cannot be completely excluded. Secondly, the RI uses data from the previous 60 min of monitoring, updating over time to create a ‘running average’ trend reflecting patient state over the preceding 30–60 min, whereas clinical assessments were at discrete time points. Unless clinical sedation state remained unaltered over this period, which was unlikely during routine care, the clinical scores related to a much shorter period of care. This may have underestimated discriminatory ability. We could have increased the frequency of clinical scores and averaged these over time, but this would have introduced confounding from the multiple stimulations associated with assessments. Visual inspection of data in conjunction with clinical event and sedation score data suggested that RI had face validity for most patients.

The RI concept is also different from clinical scores using standardized stimulation. For example, if a lightly sedated patient is comfortable, pain free, and receiving little stimulation from other treatments, the RI may be low consistent with the fact that the patient is not in need of sedative

714
Stimulating this patient during a clinical sedation assessment would likely suggest a responsive patient. Under these circumstances, and during natural sleep, the two measures may be discordant. However, we hypothesize that RI may be superior to clinical sedation scoring because it assesses whether the level of sedation is adequate for the patient’s recent clinical status and may be more likely to lead to sedation reduction when clinically appropriate. Further studies are required to test this hypothesis. A limitation to this approach is the reliance on intact neuromuscular function, and RI has no validity during neuromuscular paralysis. However, the few critically ill patients who require paralysis, such as those with raised intracranial pressure, usually require deep sedation.

Performance of RI was superior to SE and RE calculated from the identical raw data, particularly in the cardiac surgery cohort. Our previous study found that Entropy performed poorly in general ICU patients. We identified frontal EMG activations as a major confounder, which specifically resulted in an ‘on–off’ effect in response to intermittent stimulation and arousals that was problematic in the deeper sedation states. Similar problems with frontal EMG have been reported for BIS with healthy volunteers and in the ICU. Our current data confirm this problem and is illustrated by...
the large number of high Entropy values (both SE and RE) associated with Ramsay scores of 4–5 in both the general and cardiac surgery cohorts (Figs 1 and 2). RI potentially reduces this issue in two ways: first, unlike Entropy, the RI was rarely in the ‘awake’ range when the Ramsay score was 4–6; secondly, the use of data over 60 min rather than a short sampling period created a smoothing effect that removed the problematic ‘on–off’ effect previously observed with Entropy. These factors suggest that RI is more likely to act as a useful indicator of deep sedation, and could reliably trigger a sedation reduction or hold in appropriate patients.

**Table 2** Mean (SEM) $P_K$ values for distinguishing a Ramsay score of 1–3 from 4–6, and 1–4 from 5–6 for the general ICU cohort (derivation cohort; $n=30$ patients) and post-cardiac surgery cohort (validation cohort; $n=15$ patients). *The $P_K$ values for RE and SE in the general ICU data were previously reported.**

<table>
<thead>
<tr>
<th>Ramsay Score 1–3 vs 4–6</th>
<th>Data pairs (number)</th>
<th>$P_K$ values [mean (SEM)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>General ICU patients</td>
<td>227</td>
<td>0.80 (0.02)</td>
</tr>
<tr>
<td>Post-cardiac surgery</td>
<td>58</td>
<td>0.96 (0.02)*</td>
</tr>
</tbody>
</table>

*The $P_K$ value for RI was significantly higher than that of RE and SE ($P<0.0001$, paired-data $t$-test)*

<table>
<thead>
<tr>
<th>Ramsay Score 1–4 vs 5–6</th>
<th>Data pairs (number)</th>
<th>$P_K$ values [mean (SEM)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>General ICU patients</td>
<td>244</td>
<td>0.78 (0.02)</td>
</tr>
<tr>
<td>Post-cardiac surgery</td>
<td>63</td>
<td>0.96 (0.02)*</td>
</tr>
</tbody>
</table>

**Fig 3** Examples of continuous RI data plotted over time (continuous curve) with intermittent clinical sedation assessments (modified Ramsay score) (filled circles connected with a dashed line). (A) general ICU patient deeply sedated (derivation cohort); (B) general ICU patient varying sedation state (derivation cohort); (C) cardiac surgery patient during full recovery of consciousness from deep sedation (validation cohort).
We observed better performance in the cohort of cardiac surgery patients than the derivation cohort of general ICU patients, for which there are several possible explanations. First, general ICU patients tended to be studied during more constant levels of sedation, so we were less able to test discriminatory ability within individual cases. For the statistical analyses, we pooled the data and treated all the Ramsay assessments as independent observations, so we were not able to adjust for unequal distributions across the patient population. Some patients had a greater proportion of low Ramsay scores and some higher scores. This was less likely in the cardiac surgery cohort who all transitioned from deeper to lighter sedation states. Secondly, it is possible that delirium and encephalopathy, which are more prevalent in general than cardiac ICU populations, acted as confounders to the relation between clinical sedation assessment and RI. For example, a patient with encephalopathy or reduced conscious level associated with delirium might have facial EMG responses to stimulation, but not respond normally to voice or physical stimulation as part of clinical sedation scoring. Our retrospective comparison of patients considered at ‘high’ vs ‘low’ probability of encephalopathy indicated better performance in those considered at ‘low’ probability, which supports this conjecture. Prospective studies are needed to further explore this issue. Thirdly, discrimination between levels of clinical sedation scales is more difficult in complex general ICU patients than cardiac surgery, such that the consistency of reference standard assessments may have been less in the general patients. We intentionally chose the cardiac surgery patients to reduce these potential confounding issues, and the excellent performance of RI in this cohort was encouraging, especially the consistently low RI values associated with Ramsay levels 4–6. Further studies are required to assess the validity of RI for general ICU cohorts and to evaluate the clinical utility of the monitor as an adjunct to clinical decision-making.

In conclusion, our data suggest that RI is a potentially useful dynamic and continuous measure of patient arousal during ICU management. Specifically, low RI values indicate a patient at increased risk of excessive sedation, which could prompt a reduction of sedative drug doses or a sedation hold. The continuous objective nature of the monitoring could reduce nursing workload and improve sedation practice, which might be both clinically and cost-effective, given, the clear associations between over-sedation and adverse patient outcomes.2 3 22 Future work needs to confirm the promising performance demonstrated in this study and explore whether RI monitoring can usefully modify nurse decision-making in relation to sedation to result in clinically relevant improvements in patient outcomes.

Acknowledgements
The authors wish to thank the medical and nursing staff of the Royal Infirmary of Edinburgh Intensive Care Unit for their support during the study.

Conflict of interest
T.S.W. and P.R. were in receipt of a grant from GE Healthcare to support the research study, which was managed by the institution in which the clinical research was undertaken, but neither have received personal payment or have any financial interest in GE Healthcare. T.P.L., M.O.K.S., K.U., and H.E.V.-O. are all employees of GE Healthcare, who are the company developing the responsiveness index technology. GE Healthcare holds patents relating to responsiveness index.

Funding
This work was funded by GE Healthcare.

References


21 Bruhn J, Bouillon TW, Shafer SL. Electromyographic activity falsely elevates the bispectral index. Anesthesiology 2000; 92: 1485–7