Assessment of Sleep in Ventilator-Supported Critically Ill Patients

Cristina Ambrogio, MD[1,2]; Jeffrey Koebnick, RPSGT[1]; Stuart F. Quan, MD[1,3]; V. Marco Ranieri, MD[4]; Sairam Parthasarathy, MD[1]

[1]Section of Pulmonary, Critical Care, and Sleep Medicine, Southern Arizona VA Health Care System and the Department of Medicine, Arizona Respiratory Center, University of Arizona, Tucson, AZ; [2]Visiting research fellow from University of Torino, Turin, Italy; [3]Division of Sleep Medicine, Harvard Medical School, Boston, MA; [4]Department of Anesthesiology and Critical Care, University of Torino, Turin, Italy

Objectives: In critically ill patients, sleep derangements are reported to be severe using Rechtschaffen and Kales (R&K) methodology; however, whether such methodology can reliably assess sleep during critical illness is unknown. We set out to determine the reproducibility of 4 different sleep-assessment methods (3 manual and 1 computer-based) for ventilator-supported critically ill patients and also to quantify the extent to which the reproducibility of the manual methods for measuring sleep differed between critically ill and ambulatory (control) patients.

Design: Observational methodologic study.

Setting: Academic center.

Patients: Critically ill patients receiving mechanical ventilation and age-matched controls underwent polysomnography.

Interventions: None.

Measurements and Results: Reproducibility for the computer-based method (spectral analysis of electroencephalography [EEG]) was better than that for the manual methods: R&K methodology and sleep-wakefulness organization pattern (P = 0.03). In critically ill patients, the proportion of misclassifications for measurements using spectral analysis, sleep-wakefulness organization pattern, and R&K methodology were 0%, 36%, and 53%, respectively (P < 0.0001). The EEG pattern of burst suppression was not observed. Interobserver and intraobserver reliability of the manual sleep-assessment methods for critically ill patients (κ = 0.52 ± 0.23) was worse than that for control patients (κ = 0.89 ± 0.13; P = 0.03). In critically ill patients, the overall reliability of the R&K methodology was relatively low for assessing sleep (κ = 0.19), but detection of rapid eye movement sleep revealed good agreement (κ = 0.70).

Conclusions: Reproducibility for spectral analysis of EEG was better than that for the manual methods: R&K methodology and sleep-wakefulness organization pattern. For assessment of sleep in critically ill patients, the use of spectral analysis, sleep-wakefulness organization state, or rapid eye movement sleep alone may be preferred over the R&K methodology.

Keywords: Critical illness, polysomnography, physiologic monitoring, artificial respiration, reproducibility of results

Citation: Ambrogio C; Koebnick J; Quan SF; Ranieri VM; Parthasarathy S. Assessment of sleep in ventilator-supported critically ill patients. *SLEEP* 2008;31(11):1559-1568.

SLEEP DERANGEMENTS CAN LEAD TO DELETERIOUS CONSEQUENCES IN AMBULATORY PATIENTS.[1-5] IN CRITICALLY ILL PATIENTS, SLEEP DERANGEMENTS have been reported to be more severe than in ambulatory patients, but there is *large variability* in the nature and severity of sleep derangements reported.[6] For example, the time spent in slow wave sleep (SWS) has been assessed at 50% of sleep time by some investigators,[7] whereas others have observed little (3%-9%) or no SWS.[8-10] Such large variability in the assessment of sleep in critically ill patients may, in part, be due to difficulty analyzing electroencephalography (EEG); secondary to the confounding effects of sedative medications,[11-13] underlying illnesses such as sepsis,[9,14] measurement artifacts in the intensive care unit (ICU) environment; and the result of other factors.[6,12,15] A knowledge gap exists in that a systematic assessment of various methods—manual or computer-based—for assessment of sleep in critically ill patients has not been performed. A reliable method to assess sleep may further the study of sleep during critical illness.

In ambulatory patients, sleep is usually assessed by the Rechtschaffen and Kales (R&K) method, with good to excellent interobserver reliability for assessing sleep in the same population (Cohen κ range; 0.68 to 0.82).[16-18] But, whether the R&K methodology can facilitate reliable assessment of sleep in critically ill patients is unknown. There are reasons to believe that the R&K method may be less reliable when used in patients with critical illnesses because investigators have found that, in ambulatory patients, both the presence of neurologic diseases and ingestion of psychotropic medications may lessen the reliability of such methodology.[19] It is for such, and other, reasons that automated sleep assessment is being attempted. One such automated (computer-based) method is the spectral analysis of EEG signals with Fast Fourier Transform (FFT).[20] Spectral analysis quantifies activity across the EEG spectrum—from fast frequencies (with greater representation during wakefulness) to slow frequencies (signifying more restorative sleep). Although excellent reliability for such spectral analysis has been demonstrated in normal subjects,[20] such methodology has not been applied to critically ill patients. Two other manual methods are categorization of the EEG of critically ill patients into 5 groups based upon sleep-wakefulness organization pattern and counting the number of burst-suppression episodes in the EEG—a characteristic burst of EEG activity interspersed by short periods of EEG suppression.[21-23] Both such manual measures may help prognosticate neurologic recovery in critically ill patients, but rigorous interobserver and intraobserver reliability measurements of such manual methods are also lacking.[21-23]

The primary aim of this study was to determine the reproducibility of 4 different sleep-assessment methods (3 manual and 1 computer-based) for ventilator-supported critically ill pa-

Submitted for publication November, 2007
Accepted for publication August, 2008

Address correspondence to: Sairam Parthasarathy, MD, Section of Pulmonary, Critical Care and Sleep Medicine, Southern Arizona Veterans Administration Health Care System, 3601 South Sixth Ave, Mail Stop 1-111A, Tucson, AZ, 85723; Tel: 520 792 1450, Ext: 5447; Fax: 520 629 1801; E-mail: spartha@arc.arizona.edu

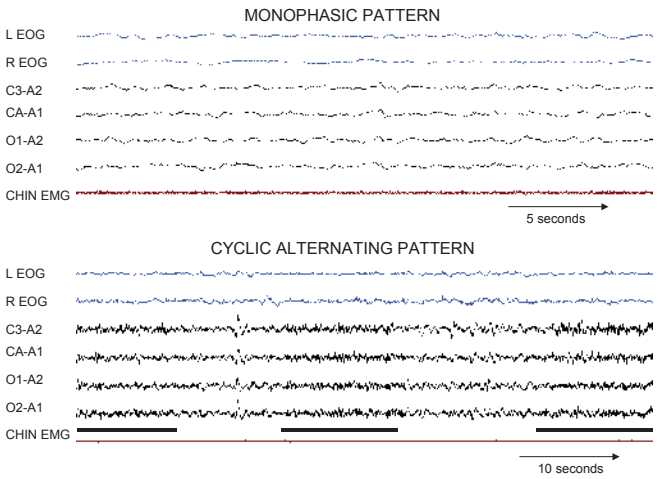

Figure 1—Representative polysomnography tracings of a critically ill patient that show monophasic pattern, characterized by continuous low-voltage theta-delta activity (top panel; 30-second epoch). The bottom panel is a representative tracing from a different critically ill patient that shows cyclic alternating pattern of electroencephalographic activity. Note that this panel is a compressed (60-second) window that reveals low-voltage theta activity alternating (highlighted by horizontal bars) with high-voltage monomorphic delta waves. L EOG refers to left electrooculogram; R EOG, right electrooculogram; C3-A2 and C4-A1, coronal electroencephalograms; O1-A2 and O2-A1, occipital electroencephalograms; EMG, electromyogram.

Table 1—Patient Characteristics

Critically Ill Patients			Diagnosis
Patient No.	Sex	Age, y	
1	M	74	Pneumonia
2	M	55	Necrotizing fasciitis
3	M	74	Septic shock
4	M	79	Septic shock
5	M	66	Dissecting aortic aneurysm
6	M	80	Acute respiratory distress syndrome
7	F	45	Tracheal stenosis resection
8	M	62	Acute exacerbation of COPD
9	M	59	Pneumonia, respiratory failure
10	M	60	Empyema, thoracoplasty, respiratory failure
11	M	66	Acute lung injury
12	M	74	Pneumonia; acute respiratory failure
13	M	67	SIRS; Post-radiofrequency ablation of liver metastasis
14	M	64	Septic shock; necrotizing fasciitis
Ambulatory (control) Patients			**Diagnosis**
Patient No.	Sex	Age, y	
1	M	58	OSA, asthma
2	M	54	OSA
3	M	54	OSA
4	M	50	OSA, PLMD
5	M	53	OSA
6	M	52	OSA
7	M	78	OSA
8	M	57	OSA
9	M	68	OSA
10	M	55	OSA
11	M	59	OSA
12	M	61	OSA
13	M	78	OSA
14	M	84	PLMD
15	M	72	PLMD
16	F	83	OSA
17	M	63	OSA

Abbreviations: M refers to male; F, female; COPD, chronic obstructive pulmonary disease; SIRS, systemic inflammatory response syndrome; OSA, obstructive sleep apnea; PLMD, periodic leg movement disorder; y, years.

tients. We hypothesized that, in critically ill patients receiving mechanical ventilation, the reproducibility of a computer-based system (spectral analysis) for measuring EEG would be better than manual methods. A secondary aim of our study was to quantify the extent to which the reproducibility of the manual methods for measuring sleep differed between critically ill and ambulatory (control) patients. We hypothesized that the reproducibility for manual methods of measuring EEG would be worse in critically ill patients, when compared with age-matched ambulatory patients.

METHODS

Patients

Fourteen critically ill patients receiving mechanical ventilation and 17 age-matched ambulatory (control) patients with sleep disorders were studied at a single ICU (Table 1). The Institutional Review Board of the University of Arizona approved the study, and written informed consent was obtained from surrogates or participants.

For critically ill patients, inclusion criteria included medical patients with acute respiratory failure who were receiving mechanical ventilation for acute exacerbation of chronic obstructive pulmonary disease, pneumonia, acute respiratory distress syndrome, or septic shock. Exclusion criteria included (a) patients who were considered by their primary physician to be too unstable to undergo this investigation; (b) patients with refractory hypotension, defined as systolic blood pressure less than 90 mm Hg despite the use of inotropic agents; (c) comatose patients or patients with severe debilitating neurologic disease such as cerebrovascular accidents, intracranial hemorrhage, subdural hematoma, intracranial primary or secondary cancers, or anoxic-hypoxic encephalopathy; (d) patients suffering from metastatic or terminal cancer and patients with do-not-resuscitate orders; (e) pregnant women; and (f) patients who were expected to be extubated within 24 hours.

The ambulatory patients were undergoing split-night studies (n = 13), positive airway pressure therapy (n = 2), or diagnostic polysomnographic studies (n = 2).

Polysomnography

Critically ill patients underwent a video-assisted polysomnogram for a 24-hour period that included EEG (C4-A1, C3-

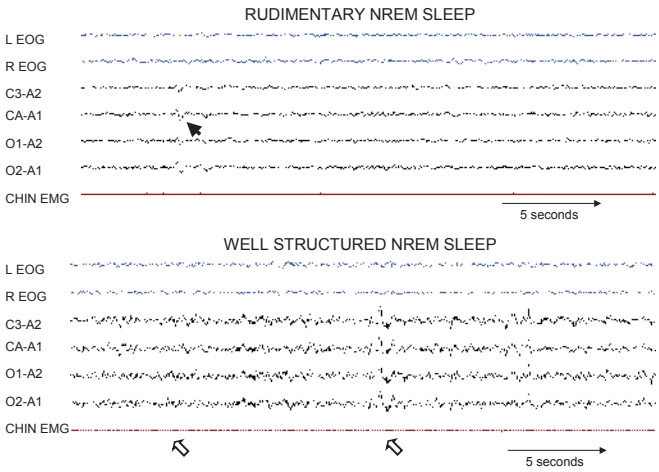

Figure 2—Representative polysomnography tracings of rudimentary sleep pattern in a critically ill patient as previously described by Valente and colleagues[21] (upper panel). The arrow points to a rudimentary K complex. Representative polysomnographic tracings of a patient with well-formed elements of non-rapid eye movement (NREM) sleep. Note the well-structured K-complex and classic spindles depicted by open arrows (lower panel). Other abbreviations are the same as for Figure 1.

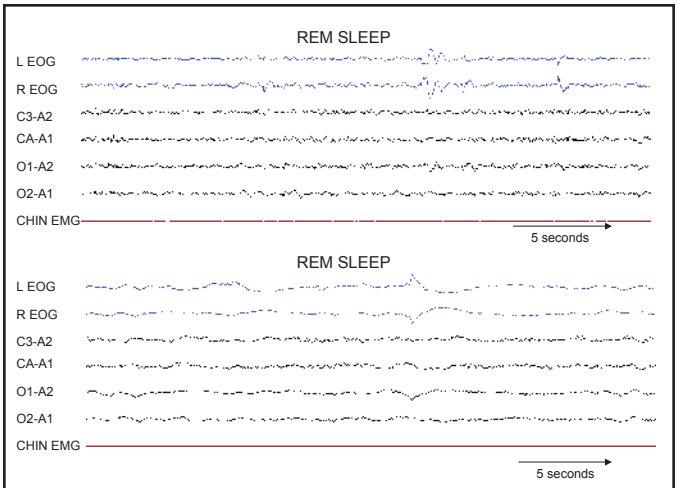

Figure 3—Representative polysomnography tracings of sleep pattern in a critically ill patient that shows rapid-eye movement (REM) sleep (upper and lower panels). Such classic REM sleep alternating with non-REM sleep in critically ill patients were classified as the best sleep organization pattern as previously described by Valente and colleagues.[21] Other abbreviations are the same as for Figure 1.

A2, O1-A2, and O2-A1), left and right electrooculograms, submental electromyogram, chest and abdominal movement by inductance plethysmography (Ambulatory Monitoring Inc., Ardsley, NY), leg movements by bilateral anterior tibialis electromyograms, and finger pulse oximetry (Sandman, Ontario, CA). Ambulatory patients underwent polysomnography for approximately 8 hours.

Ventilator Settings

All patients were receiving mechanical ventilation in the assist-control mode of ventilation to prevent the effects of mode of ventilation and ventilator setting on the sleep of critically ill patients.[8,24] Inspiratory flow rate, positive end-expiratory pressure, and fractional inspired oxygen concentration were kept at the same settings as before the study. All studies were done for approximately 24 hours and, in the event of patient-care–related activity that required transport (computerized tomography, surgery, etc.), the study was prematurely concluded short of the 24-hour period.

Sedatives

Continuous sedative and analgesic infusions (midazolam, fentanyl, or propofol) were administered per clinical sedation protocols that target the Ramsay sedation scale.[25]

Sleep Assessment

Sleep scoring was performed according to 3 manual methods (R&K, sleep-wakefulness organization pattern, and visual detection of burst suppression) and 1 computer-based (spectral analysis of EEG signals with FFT). For the R&K methodology and spectral analysis, for each patient, approximately 200 epochs of polysomnography recordings that were artifact free was randomly selected for analysis.[26]

(i) Using the **R&K methodology**, sleep was manually staged in 30-second epochs according to standard criteria.[16,17]

(ii) For sleep staging with the **sleep-wakefulness organization pattern**, patients were classified into 1 of 5 groups according to the type of sleep-wakefulness organization pattern[21]:
   (a) Monophasic: continuous low-voltage theta-delta activity (Figure 1; upper panel).
   (b) Cyclic alternating pattern[27]: absence of sleep elements in the presence of recurrent abnormal spontaneous arousal activity. Epochs characterized by low-voltage theta activity alternate with epochs of high-voltage monomorphic delta waves (Figure 1; lower panel).
   (c) Rudimentary sleep: presence of rudimentary non-rapid eye movement sleep (NREM) sleep elements (K-complexes and/or spindles)(Figure 2; upper panel).
   (d) Well-structured elements of NREM (Figure 2; lower panel).
   (e) Rapid eye movement (REM) sleep elements (rapid eye movements and saw-tooth waves) alternating with NREM sleep (Figure 3).

(iii) **Spectral analysis** that employed EEG derived from the C3/A2 channel was partitioned into 5-second record lengths to obtain a frequency resolution of 0.2 Hz. Power spectral analysis of EEG was performed using discrete FFT. FFT analysis yielded the classic measure of power in μV^2 x second. The resulting power in the following 4 different EEG spectral bandwidths was then computed for each 5-second record: δ (0.8-4.0 Hz), θ (4.1-8.0 Hz), α (8.1-13.0 Hz), and β (13.1-20.0 Hz). Relative proportions of δ, θ, α, and β power were expressed as a percentage of average total power. Both sleep and wakefulness periods contained in the 200-epoch fragment per patient were included in

the analysis. This is slightly different than the "2-step process" used by other investigators.[20] In the 2-step process, an initial manual determination of sleep-wakefulness state (NREM, REM, or wakefulness) is made by an observer and is then followed by spectral analysis based upon such sleep-wakefulness state determination. Such state-dependent spectral-analysis method is therefore likely dependent on the prior manual adjudication of the sleep-wakefulness state using the R&K methodology. We chose the single-step process, ie, to include all of the 200 epochs regardless of sleep-wakefulness state to avoid the overlap of methodologies (R&K and FFT), as in the method used by others.[20] Although the measurements were made by the computerized software, the data input (raw EEG-derived voltage [μV]) was performed independently by 2 blinded observers and therefore was referred to as an observer-based rather than computer-based observation. Measurements were performed by computerized software module (Sandman, Ontario, CA and Excel, Microsoft Corp., Seattle, WA). The FFT Export Module of the sleep diagnostic software (Sandman) was used to export the FFT data of 5-second record lengths. The exported FFT file from the Sandman software was then imported into Excel, wherein the relative proportions of δ, θ, α, and β power were expressed as a percentage of average total power for each 5-second record length, and the group average for the entire 200-second period was derived.

(iv) **Number of burst suppressions**: The depth of sedation, or pharmacologically induced sleep, has been correlated with the number of burst suppressions of EEG that occur per hour.[28] Burst suppression can be defined as an EEG pattern characterized by bursts of EEG activity (sharp and slow waves) periodically interrupted by episodes of suppression (activity < 10 mV). Typically, the episodes of suppression are longer (usually 5-10 seconds) than the bursts of activity (usually 1-3 seconds). This pattern is not specific to any etiology but represents severe diffuse encephalopathy due to deep sedation (barbiturates, benzodiazepines, or propofol), trauma, cerebrovascular accidents, hypothermia, and anoxic brain injury.

Two observers (28 years total experience for analyzing EEG) analyzed each study twice while blinded to the first assessment, the other observer's scores, and the group assignment (ie, critically ill versus control patients). For the R&K methodology and spectral analysis, for each patient, approximately 200 epochs of polysomnography recordings that were artifact free were randomly selected for analysis.[26] In the body of literature that addresses interscorer reliability of sleep, the choice of number of epochs has ranged from 90 to 962 epochs per patient.[17,19,26] We chose 200 epochs because preliminary analyses (in 5 patients) showed that scoring more than 200 epochs increased the intensity of the analyses without materially changing the results. Moreover, the current standards for accreditation of sleep centers advocated by the American Academy of Sleep Medicine require that interscorer reliability in scoring sleep stages be determined on 200 consecutive epochs.[29] Such a choice of 200 epochs falls well within the range of epoch choices (90-962 epochs) of the published literature. For assessing the sleep-wakefulness organization pattern and burst suppression, approximately 8 hours of polysomnography were used, as opposed to the 200 epochs for spectral analysis and R&K methodology. For sleep-wakefulness organization pattern and burst suppression, a greater sample duration was necessary by nature of the measurement and in keeping with the duration analyzed in prior such reports.[21,12] Moreover, in an initial sample of patients (n = 5), because we did not find evidence for burst suppression in 200 epochs, review of 8 hours of polysomnography was warranted.

Blinding

Studies were analyzed in a random manner with at least a 2-week interval between each assessment. An investigator—who was not involved in the scoring of the study—selected 200 epochs that had been recorded between 22:00 and 04:00 that were free of artifacts (constituted 7% ± 1% of the total epochs). The starting point of the 200-epoch period was selected by randomly clicking on the unscored hypnogram window (Sandman) to ensure that the starting point fell somewhere between 22:00 and 04:00. This investigator then codified the electronic file and named it differently (using codes [random alphanumeric]) for each observation of each of the 2 observers. Multiple copies of the files were created so that there was no observer recognition of the coded file name for the respective repeat observations. These files were randomly provided to the observers. The research work flow for each observer included data from both critically ill and ambulatory patients in a random order that was provided over regular intervals while making sure that the repeat measurements were at least 2 weeks apart.

Statistical Analysis

Interobserver and intraobserver agreement were quantified as Cohen κ statistic (SPSS v12.0, Chicago, IL).[30] By definition, κ was equal to 1.0 for complete agreement, and κ was equal to 0.0 for agreement no better than chance alone. κ values of 0.21 to 0.4 indicate fair agreement; 0.41 to 0.6, moderate agreement; 0.61 to 0.8, substantial agreement; and 0.81 to 0.99, almost perfect agreement. Results are reported as mean ± SD unless otherwise specified. Paired or unpaired t-tests or equivalent nonparametric tests were used when appropriate, and χ^2 was used for categorical variables. Friedman 1-way analysis of variance was used for comparing nonparametrically distributed continuous variables with repeated measures.

RESULTS

The critically ill and ambulatory (control) patients were matched for age: 66 ± 10 years and 64 ± 11 years, respectively (P = 0.5; t-test). The APACHE II score for the critically ill patients was 17 ± 5.

R&K Methodology

In critically ill patients, when sleep was scored according to R&K methodology with 5 sleep-wakefulness states—stage 1 and 2 NREM sleep, SWS, REM sleep, and wakefulness —the interobserver reliability was poor (κ = 0.19; 4640 epochs; Table

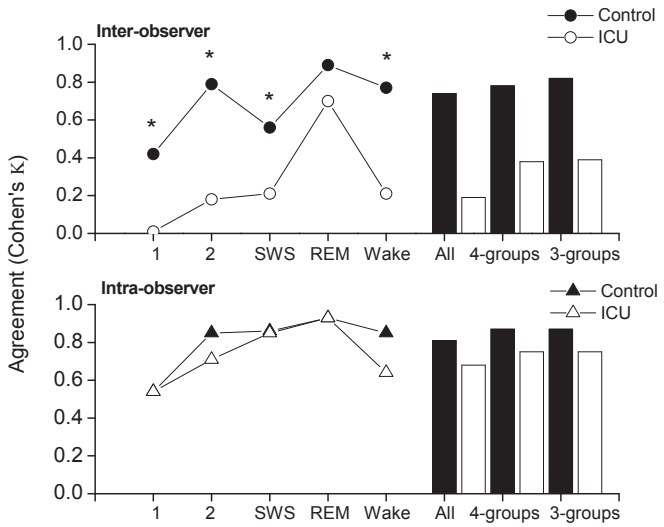

Figure 4—Interobserver (top panel) and intraobserver (bottom panel) agreement for sleep staging per Rechtschaffen and Kales methodology in critically ill (ICU; open symbols) and ambulatory patients (closed symbols). Columns depict agreements (Cohen κ agreement value) for critically ill (white columns) and control (black columns) patients for all sleep-wakefulness stages by Rechtschaffen and Kales methodology (all), for 4 collapsed groups (light non-rapid eye movement sleep [NREM] stage 1 and 2; slow wave sleep [SWS]; rapid eye movement [REM]; and wakefulness), 3 collapsed groups (NREM, REM, wakefulness). Note that interobserver reliability is worse for critically ill than control patients (top panel; analysis of variance $P < 0.0001$). Interobserver reliability was worse in critically ill patients than controls for all sleep stages except REM sleep (Neuman-Keuls; $*P < 0.05$). Also note that the reliability for assessing REM sleep is good in critically ill patients and is not statistically different than that in controls (open symbols; top panel).

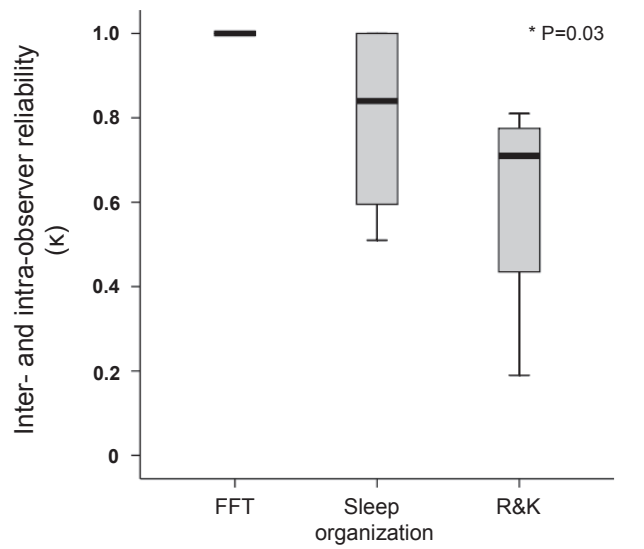

Figure 5—Box and whisker plots of the interobserver and intraobserver reliability (Cohen κ agreement value) for assessing sleep in critically ill patients while using spectral analysis by Fast Fourier Transform (FFT), sleep-wakefulness organization pattern (*Sleep organization*), or Rechtschaffen and Kales (*R&K*) methodology. The reliability was best for spectral analysis (FFT) when compared with the other methods (Friedman analysis of variance; $P = 0.03$). Post-hoc (Neuman-Keuls) comparisons revealed a tendency for spectral analysis to be better than sleep-wakefulness organization pattern ($P = 0.10$) and R&K methodology ($P = 0.06$).

2). In contrast, in control patients, the interobserver reliability was good (κ = 0.74; 7123 epochs; Table 2).

To determine the source of disagreements in the assessment of sleep in critically ill patients, interobserver agreement for each of the different sleep stages was calculated. Such calculations for critically ill patients revealed poor interobserver agreement for NREM stage 1 and 2 sleep (κ = 0.01 and 0.18), fair agreement for SWS and wakefulness (both κ = 0.21), and good agreement for REM sleep (κ = 0.7; Figure 4; top panel). Similarly, stage-specific interobserver reliability testing for controls revealed best agreement for REM sleep (κ = 0.89; Figure 4; top panel). The interobserver agreement for individual sleep stages was worse for critically ill (median κ = 0.21; interquartile range, 0.10, 0.46) than for ambulatory patients (median κ = 0.77; interquartile range, 0.49, 0.84; $P = 0.03$; Mann-Whitney U test).

The intraobserver agreement for individual sleep stages was not different between critically ill and ambulatory patients ($P = 0.8$; Mann-Whitney U test) (Figure 4; bottom panel). The overall intraobserver reliability for all sleep stages combined was good for critically ill patients and excellent for control patients (Table 4).

To focus on the more clinically relevant agreements, we first collapsed the sleep-wakefulness data into 4 groups: light NREM sleep (NREM stage 1 and 2), SWS, REM, and wakefulness. Such collapsing only slightly improved the interobserver agreement from poor to fair in critically ill patients (κ = 0.38; Table 2). A further collapse into 3 groups (REM, NREM, wakefulness) did not improve interobserver agreement in critically ill patients. Small improvements in interobserver and intraobserver reliability were observed for ambulatory patients in response to such categorization.

Sleep-Wakefulness Organization Pattern

In critically ill patients, sleep-wakefulness organization patterns demonstrated moderate interobserver agreement, whereas intraobserver agreement was good (Table 2). Both intraobserver and interobserver agreements were perfect for the control group (Table 2). In critically ill patients, the majority of discrepancies were between monophasic and rudimentary sleep groups for both intraobserver and interobserver reliability assessments (Table 3a,b). However, in the best sleep group, characterized by the presence of REM alternating with NREM sleep, perfect agreement was reached (κ = 1.0; Table 3a, b).

Spectral Analysis

Relative proportions of δ, θ, α, and β power, expressed as a percentage of average total power, for critically ill patients and ambulatory patients are available in Table 4. Bland-Altman plots for both interobserver and intraobserver measurements using spectral analysis revealed bias and precision errors of 0 (not shown). To facilitate comparisons across the 4 methods of sleep assessment, Cohen κ values were derived after categorizing the

Table 2—Interobserver and Intraobserver Agreement of Sleep Scoring*

	Interobserver agreement		Intraobserver agreement	
	Critically ill patients	Ambulatory patients	Critically ill patients	Ambulatory patients
Spectral analysis by fast Fourier transform	1.0	1.0	1.0	1.0
Sleep-wakefulness organization	0.51	1.0	0.68	1.0
Rechtschaffen & Kales Methodology				
Five groups				
(NREM 1, NREM 2, SWS, REM, & Wakefulness)	0.19	0.74	0.68	0.81
Four groups (LNREM, SWS, REM, & Wakefulness)	0.38	0.78	0.75	0.87
Three groups (NREM, REM, & Wakefulness)	0.39	0.82	0.75	0.87

Results are expressed as Cohen kappa [κ] value. NREM refers to non-rapid eye movement sleep; REM, rapid eye movement sleep; SWS, slow wave sleep; LNREM, light NREM sleep (stage 1 and stage 2 NREM).

FFT values into quartiles. The intraobserver and interobserver agreement for spectral analysis of relative proportions of δ, θ, α, and β power were perfect (κ = 1.0) for both critically ill and control patients (Table 2). Reanalysis of the data using a different software (MATLAB, Mathworks Inc., Natick, MA) for spectral analysis did not change the results when comparisons were made between the 2 statistical programs (κ =1.0)

Number of Burst Suppressions

None of our patients showed an EEG pattern of burst suppression.

Comparisons Across Sleep-Assessment Methods

Interobserver and intraobserver reliability for the computer-based method was better than that for the manual methods: R&K methodology and sleep-wakefulness organization pattern (Friedman test, P = 0.03; Figure 5). In critically ill patients, for interobserver reliability testing, the proportion of misclassifications between observations for spectral analysis, sleep organization, and R&K methodology were 0%, 36%, and 53%, respectively (χ^2; P < 0.0001; Figure 6). In critically ill patients, for intraobserver reliability testing, the proportion of misclassifications between observations for spectral analysis, sleep-wakefulness organization, and R&K methodology were 0%, 21%, and 20%, respectively (χ^2; P < 0.0001; Figure 6).

For both of the manual methods combined (R&K methodology and sleep-wakefulness organization pattern), the interobserver and intraobserver reliability for critically ill patients (κ = 0.52 ± 0.23) was worse than that for control patients (κ = 0.89 ± 0.13; t-test, P = 0.03).

DISCUSSION

Certain general observations can be made. First, the reliability for sleep assessment in critically ill patients receiving mechanical ventilation was better with computer-based (spectral analysis) than with manual methods. Second, in critically ill patients, the overall reliability of R&K methodology was poor for assessing sleep, but the detection of REM sleep revealed good agreement for both interobserver and intraobserver determinations. In contrast, in ambulatory (control) patients, the overall reliability of R&K methodology was good or excellent. Third, an alternative manual method for assessing sleep in critically ill patients—sleep-wakefulness organization pattern—revealed moderate to good interobserver and intraobserver agreement. Fourth, we did not observe burst suppression in our critically ill patients. Lastly, the reliability of both manual methods (R&K and sleep-wakefulness organization pattern) for assessing sleep in ambulatory patients was better than that in the critically ill patients.

Prior reports have suggested that the analysis of sleep was difficult or not possible in 23% to 42% of critically ill patients who were studied.[9,31] Most such prior studies relied on visual scoring by experts, but, to our knowledge, this is the first study to systematically evaluate the interobserver and intraobserver agreement of different methods for assessing sleep in critically ill patients.

In the absence of a gold standard for assessing sleep in critically ill patients, we performed reliability measurements of currently available methods to assess sleep. We found that, in critically ill patients, the overall reliability of conventionally used R&K methodology for assessing sleep was poor. To verify that the 2 observers in the study were indeed reliable, these observers also assessed sleep in ambulatory (control) patients. As anticipated, in the control patients, the interobserver and intraobserver agreement for sleep assessment using R&K methodology was good to excellent: κ = 0.74 and 0.81, respectively. Such levels of agreement fell within a previously described range for interobserver (κ range from 0.68-0.82) and intraobserver reliability (κ range from 0.79-0.87) of ambulatory patients.[17-19]

Although the overall reliability for assessing sleep of critically ill patients using the R&K methodology was poor, the reliability for assessing REM sleep was preserved (Figure 4; top panel). Moreover, when a different methodology was used—sleep-wakefulness organization pattern—2 observers were in perfect agreement while classifying the group that was characterized by REM sleep (Table 3). Such findings are also in agreement with those of prior investigators who demonstrated best reproducibility for scoring REM sleep in ambulatory patients.[17] In ambulatory patients, the assessment of features of NREM stage 1 and 2 sleep (K complexes [κ = 0.5], spindles [κ = 0.7], and vertex waves)—when compared with REM sleep—are known to be less reproducible.[32-35]

In our study, the reliability of manual methods (R&K methodology and sleep-wakefulness organization pattern) for assessing sleep was better in ambulatory patients than in critically

Table 3a—Interobserver Agreement for Classifying Sleep-Wakefulness Organization Pattern (in Critically Ill Patients)

	Observer #2					Total
	Monophasic	**CAP**	**Rudimentary**	**NREM**	**REM+NREM**	
Observer #1						
Monophasic	**3**	0	7	0	0	10
CAP	0	**1**	3	0	0	4
Rudimentary	0	0	**8**	0	0	8
NREM	0	0	0	**0**	0	0
REM+NREM	0	0	0	0	**6**	6
Total	3	1	18	0	6	28

Table 3b—Intraobserver Agreement for Classifying Sleep-Wakefulness Organization Pattern (in Critically Ill Patients)

	Second observation					Total
	Monophasic	**CAP**	**Rudimentary**	**NREM**	**REM+NREM**	
First Observation						
Monophasic	**4**	0	2	0	0	6
CAP	0	**2**	1	0	0	3
Rudimentary	3	0	**10**	0	0	13
NREM	0	0	0	**0**	0	0
REM+NREM	0	0	0	0	**6**	6
Total	7	2	13	0	6	28

Abbreviations: Monophasic, continuous low-voltage theta-delta activity, CAP, cyclic alternating pattern, Rudimentary sleep, NREM, presence of well-structured elements of non-rapid eye movement sleep [NREM], and rapid eye movement sleep (REM)+NREM, REM sleep elements alternating with NREM sleep. Numbers in bold type signify observations wherein agreement was achieved.

ill patients. Some of the reasons for the lower reliability in critically ill patients may be the effect of sedating medications,[11-13] the presence of underlying illnesses such as sepsis,[9,14] artifacts in the ICU,[6,12,15,35] and other factors. Such variability in assessing the sleep of critically ill patients may hinder the progress of research in this area, and hence more reproducible methods may be desirable.

Other techniques for assessing sleep in critically ill patients (sleep-wakefulness organization pattern) have been described[20,21] but have not been subjected to systematic assessment of reliability in critically ill patients. In our study, as hypothesized, we found that spectral analysis was more reliable (κ=1.0) than manual methods for assessing sleep in critically ill patients, suggesting that the variability induced by the human visual analysis can be minimized. Indeed, one could argue that a computerized analysis is, of course, expected to achieve perfect agreement. However, by obviating the manual part of a 2-step process of determining sleep stage by R&K methodology before subjecting the EEG signals to FFT, we have removed the human element from such analysis. In doing so, we not only prevented "overlap" with R&K methodology (a comparator), but, in the process, made the spectral-analysis technique more reliable. Therefore it comes as no surprise that perfect agreement achieved by spectral analysis was better than the excellent agreement reported previously.[20] But this is the very purpose of our methodological study, i.e., to identify a reliable technique to analyze sleep-wakefulness EEG in critically ill patients. It is important to note that, unlike the wealth of data that support the physiologic and clinical significance of various sleep stages of the R&K methodology, less is known of the biological relevance of data derived from spectral analysis of EEG—relative proportions of δ, θ, α, and β power. Currently, based upon the study of ambulatory patients, we know that an increase of δ and θ power in the sleep-EEG spectrum signifies higher sleep intensity and deeper sleep, whereas an increase of α activity may be presumed to be a sign of increased arousal and nonrestorative sleep.[35] Nevertheless, we believe that the study of sleep in critically ill patients is still at its infancy and that this would be the time to establish a reliable scoring technique upon which future studies could build.

Valente and colleagues previously described the biological (clinical) relevance of the sleep-wakefulness organization pattern.[21] These investigators classified sleep-wakefulness organization patterns into 5 groups in patients with posttraumatic head-injury coma. They found that the sleep-wakefulness organization was a better prognosticator for neurologic recovery than were other commonly used indexes such as Glasgow Coma Scale, neuroradiologic findings, and age.[21] These investigators, however, had not performed a rigorous reliability testing for their methodology. In our study, however, we found that their method for assessing sleep in critically ill patients has moderate interobserver and good intraobserver reliability (Table 2). In our study, group-specific reliability was best for the group with elements of well-organized REM sleep alternating with NREM sleep, and, notably, the patients in the same group had the best prognosis for functional recovery from head injury. Future studies should probably examine whether the presence of REM sleep alone (a reliable measure) carries a good prognosis.

Despite reviewing 24-hour polysomnograms of 14 critically ill patients (approximately 330 hours of data), we did not find

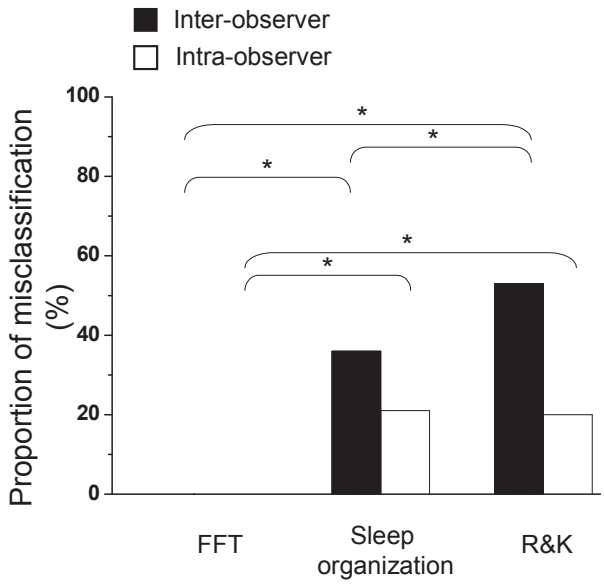

Figure 6—Proportion of observations (patients or epochs) that were misclassified while assessing sleep in critically ill patients with either spectral analysis by Fast Fourier Transform (FFT), sleep-wakefulness organization pattern (*Sleep organization*), or Rechtschaffen and Kales (*R&K*) methodology. Proportion of misclassifications for both interobserver (*black columns*) and intraobserver (*white columns*) are shown. For both interobserver and intraobserver reliability testing, the proportion of misclassifications was highest for R&K methodology and least for FFT (2 x 3 χ^2 comparisons; $P < 0.0001$). Significant post-hoc comparisons (2 x 2 χ^2 comparisons with Bonferroni correction) are also shown (*$P < 0.001$).

Table 4—Results of Spectral Analysis

Spectral Bandwidth, Hz	Ambulatory patients (n = 17)	Critically ill patients (n = 14)
Relative proportion of power, %		
δ [0.8-4.0 Hz]	61.0 ± 5.6	66.1 ± 7.1[a]
θ [4.1-8.0 Hz]	21.1 ± 2.2	20.2 ± 3.6[a]
α [8.1-13.0 Hz]	12.4 ± 2.7	9.60 ± 3.1[a]
β [13.1-20 Hz]	5.60 ± 1.7	4.20 ± 1.4[a]
Ratio between bandwidths		
δ/α Ratio	5.2 + 1.4	7.8 + 3.1[b]
δ/β Ratio	12.0 + 3.7	18.1 + 7.4[b]

Results are expressed as mean ± standard deviation. Hz, Hertz. Relative proportion of power was expressed as percentage of average total power.
[a]$P < 0.05$ when compared with that of ambulatory patients.
[b]$P < 0.01$ when compared with that of ambulatory patients.

evidence for burst suppression. Conceivably, lighter sedation level and exclusion of patients with primary neurologic disease or neurologic catastrophes may have been responsible for such an observation (see exclusion criteria). Prior studies have shown that heavy sedation, underlying hypoxic encephalopathy, and other neurologic catastrophes are associated with the presence of burst suppression.[12]

In conclusion, in ventilator-supported critically ill patients, the interobserver and intraobserver reliability of spectral analysis for assessing sleep was better than manual methods: R&K methodology and sleep-wakefulness organization pattern. Moreover, it could be stated that, considering the relatively low reliability of the R&K methodology in assessing sleep of critically ill patients, other methods, such as spectral analysis, sleep-wakefulness organization pattern, and detection of REM sleep alone, may be preferable. However, future studies need to address the biological significance of such findings in critically ill patients.

Limitations

There are limitations to our study. First, we chose different recording time periods for ambulatory and critically ill patients (8 versus 24 hours, respectively) because the main sleep episode of ambulatory patients is at nighttime but the sleep episodes of critically ill patients are scattered throughout a 24-hour period.[31] Although such rationale and study design ensured adequate sleep recordings in critically ill patients, the sleep measurement (sleep stages or environmental effects) may have been influenced by the time of day. Nevertheless, considering that circadian rhythm is either absent or significantly diminished in critically ill patients,[38] this may be less of a factor. Moreover, such differences may be a limitation in a biological or outcome study but may be less of a factor in our methodological (*reliability*) study. Second, the modest sample size is a limitation. Third, our failure to measure frontal EEG derivatives (F4,F3)—which are known to better measure K complexes and delta waves than do other derivatives[17]—may have deleteriously influenced the reliability of adjudicating NREM sleep in both critically ill and ambulatory patients. However, in ambulatory patients, our failure to collect frontal derivatives did not appear to influence the reliability of scoring NREM sleep because our κ values were well within the range reported in the published literature.[17] It seems unlikely that such a lack of frontal derivatives would have influenced sleep-staging reliability in critically ill patients but not in ambulatory patients. Nevertheless, we are unable to speculate whether such a lack of frontal EEG derivative may have influenced our study findings in critically ill patients other than to recognize this as a limitation.

Fourth, despite our efforts, the observer may have still distinguished the sleep architecture of an ICU subject as different from that of an ambulatory patient. However, we suspect that such a bias was less likely for the following reasons. If, indeed, the observers were biased toward demonstrating poor agreement for the data derived from critically ill patients, then it is unlikely for them to systematically demonstrate good interobserver agreement for REM sleep epochs alone ($\kappa = 0.7$; Figure 4; top panel, open symbol, REM sleep). Moreover, if the observers were indeed biased toward poor interobserver agreement—by scoring sleep in an erratic fashion for subjects that they believed to be critically ill patients—then it is unlikely that their intraobserver agreement (based upon measurements that were > 2 weeks apart) would still demonstrate moderate agreement ($\kappa = 0.68$; Figure 4; bottom panel, open symbols; also see Table 2).

Fifth, although FFT analysis could help improve the reliability of the sleep assessment, other valuable biological infor-

mation that was afforded by the manual methodology may be lost. Such biological information may pertain to the ability to prognosticate functional recovery after blunt head trauma (cyclic alternating pattern events; sleep-wakefulness organization pattern[21]); assess the relationship among sleep, learning, and memory (selective REM sleep deprivation derived from R&K methodology [39]); or prognosticate outcome following coma (burst suppression[23]). Further research is required to determine whether FFT analysis is comparable with, inferior to, or superior to other available methods of assessing sleep and wakefulness in critically ill patients.

ACKNOWLEDGMENTS

The authors are indebted to the participants in this study. The authors are grateful to Drs. Ralph Fregosi and Richard Bootzin for their critical review of the manuscript.

DISCLOSURE STATEMENT

This was not an industry supported study. Dr. Parthasarathy has received research support from Respironics and Takeda. Dr. Quan has received research support from Respironics and has participated in speaking engagements for Takeda. The other authors have indicated no financial conflicts of interest.

REFERENCES

1. Arzt M, Young T, Finn L, Skatrud JB, Bradley TD. Association of sleep-disordered breathing and the occurrence of stroke. Am J Respir Crit Care Med 2005;172:1447-51.
2. Gottlieb DJ, Redline S, Nieto FJ, et al. Association of usual sleep duration with hypertension: the Sleep Heart Health Study. Sleep 2006;29:1009-14.
3. Lorenzi-Filho G, Drager LF. Obstructive sleep apnea and atherosclerosis: a new paradigm. Am J Respir Crit Care Med 2007;175:1219-21.
4. Parthasarathy S, Tobin MJ. Sleep in the intensive care unit. Intensive Care Med 2004;30:197-206.
5. Leung RS, Bradley TD. Sleep apnea and cardiovascular disease. Am J Respir Crit Care Med 2001;164:2147-65.
6. Naughton MT. The link between obstructive sleep apnea and heart failure: underappreciated opportunity for treatment. Curr Heart Fail Rep 2006;3:183-8.
7. Hardin KA, Seyal M, Stewart T, Bonekat HW. Sleep in critically ill chemically paralyzed patients requiring mechanical ventilation. Chest 2006;129:1468-77.
8. Parthasarathy S, Tobin MJ. Effect of ventilator mode on sleep quality in critically ill patients. Am J Respir Crit Care Med 2002;166:1423-9.
9. Freedman NS, Gazendam J, Levan L, Pack AI, Schwab RJ. Abnormal sleep/wake cycles and the effect of environmental noise on sleep disruption in the intensive care unit. Am J Respir Crit Care Med 2001;163:451-7.
10. Bosma K, Ferreyra G, Ambrogio C, et al. Patient-ventilator interaction and sleep in mechanically ventilated patients: pressure support versus proportional assist ventilation. Crit Care Med 2007;35:1048-54.
11. Knill RL, Moote CA, Skinner MI, Rose EA. Anesthesia with abdominal surgery leads to intense REM sleep during the first postoperative week. Anesthesiology 1990;73:52-61.
12. Wauquier A. EEG and neuropharmacology. In: Neidermeyer E Da Silva FL, eds. Electroencephalography. Basic Principles, Clinical Applications, and Related Fields. Baltimore, MD: Williams and Wilkins; 1993:619-29.
13. Sebel PS, Bovill JG, Wauquier A, Rog P. Effects of high-dose fentanyl anesthesia on the electroencephalogram. Anesthesiology 1981;55:203-11.
14. Imeri L, Bianchi S, Opp MR. Inhibition of caspase-1 in rat brain reduces spontaneous nonrapid eye movement sleep and nonrapid eye movement sleep enhancement induced by lipopolysaccharide. Am J Physiol Regul Integr Comp Physiol 2006;291:R197-204.
15. Prior PF. EEG monitoring and evoked potentials in brain ischemia. Br J Anaesth 1985;57:63-81.
16. Rechtschaffen A, Kales A. A Manual of Standardized Terminology, Techniques and Scoring System of Sleep Stages of Human Subjects. Los Angeles, CA: Brain Information Service, UCLA; 1968.
17. Silber MH, Ancoli-Israel S, Bonnet MH, et al. The visual scoring of sleep in adults. J Clin Sleep Med 2007;3:121-31.
18. Whitney CW, Gottlieb DJ, Redline S, et al. Reliability of scoring respiratory disturbance index and sleep staging. Sleep 1998;21:749-757.
19. Danker-Hopfe H, Kunz D, Gruber G, et al. Interrater reliability between scorers from eight European sleep laboratories in subjects with different sleep disorders. J Sleep Res 2004;13:63-9.
20. Tan X, Campbell IG, Feinberg I. Internight reliability and benchmark values for computer analyses of non-rapid eye movement (NREM) and REM EEG in normal young adult and elderly subjects. Clin Neurophysiol 2001;112:1540-52.
21. Valente M, Placidi F, Oliveira AJ, et al. Sleep organization pattern as a prognostic marker at the subacute stage of post-traumatic coma Clin Neurophysiol 2002;113:1798-1805.
22. Wijdicks EF, Hijdra A, Young GB, Bassetti CL, Wiebe S; Quality Standards Subcommittee of the American Academy of Neurology. Practice parameter: prediction of outcome in comatose survivors after cardiopulmonary resuscitation (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2006;67:203-10.
23. Young GB, Doig G, Ragazzoni A. Anoxic-ischemic encephalopathy: clinical and electrophysiological associations with outcome. Neurocrit Care 2005;2:159-64.
24. Fanfulla F, Delmastro M, Berardinelli A, Lupo ND, Nava S. Effects of different ventilator settings on sleep and inspiratory effort in patients with neuromuscular disease. Am J Respir Crit Care Med 2005;172:619-24.
25. Jacobi J, Fraser GL, Coursin DB, et al, Task Force of the American College of Critical Care Medicine (ACCM) of the Society of Critical Care Medicine (SCCM), American Society of Health-System Pharmacists (ASHP), American College of Chest Physicians. Crit Care Med 2002;30:119-41.
26. Drinnan MJ, Murray A, Griffiths CJ, Gibson GJ. Interobserver variability in recognizing arousal in respiratory sleep disorders. Am J Respir Crit Care Med 1998;158:358-62.
27. Terzano MG, Mancia D, Salati MR, Costani G, Decembrino A, Parrino L. The cyclic alternating pattern as a physiological component of NREM sleep. Sleep 1985;8:137-45.
28. Witte H, Schelenz C, Specht M et al. Interrelations between EEG frequency components in sedated intensive care patients during burst-suppression period. Neurosci Lett 1999;260:53-6.
29. Standards for Accreditation of Sleep Disorders Centers; page 12. Available at http://www.aasmnet.org/Resources/PDF/CenterStandards.pdf. Accessed April 27, 2008.
30. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
31. Cooper AB, Thornley KS, Young GB, Slutsky AS, Stewart TE, Hanly PJ. Sleep in critically ill patients requiring mechanical ventilation. Chest 2000;117:809-18.
32. Bremer G, Smith JR, Karacan I. Automatic detection of the K-

Complex in sleep electroencephalograms. IEEE Trans Biomed Eng 1970;17:314-23.
33. Campbell K, Kumar A, Hofman W. Human and automatic validation of a phase–locked loop spindle detection system. Electroencephalogr Clin Neurophysiol 1980;48:602-5.
34. Ferri R, Dequae P, Colognola RM, Musumeci SA, Bergonzi P. A simple electronic and computer system for automatic spindle detection. Neurophysiol Clin 1989;19:171-7.
35. Kumar A, Hofman W, Campbell K. An automatic spindle analysis and detection system based on the evaluation of human ratings of the spindle quality. Waking Sleeping 1979;3:325-33.
36. Ambrogio C, Parthasarathy S. Polysomnography during critical illness. J Clin Sleep Med 2007;3:649-50.
37. Schmid DA, Wichniak A, Uhr M, et al. Changes of sleep architecture, spectral composition of sleep EEG, the nocturnal secretion of cortisol, ACTH, GH, prolactin, melatonin, ghrelin, and leptin, and the DEX-CRH test in depressed patients during treatment with mirtazapine. Neuropsychopharmacology 2006;31:832-44.
38. Mundigler G, Delle-Karth G, Koreny M, et al. Impaired circadian rhythm of melatonin secretion in sedated critically ill patients with severe sepsis. Crit Care Med 2002;30:536-40.
39. Ishikawa A, Kanayama Y, Matsumura H, Tsuchimochi H, Ishida Y, Nakamura S. Selective rapid eye movement sleep deprivation impairs the maintenance of long-term potentiation in the rat hippocampus. Eur J Neurosci 2006;24:243-8.