Inter-rater disagreement in manual scoring of intensive care unit sleep data

Reinke, Laurens; van der Heide, Esther M; Fonseca, Pedro; Absalom, Anthony R; Tulleken, Jaap E

Published in:
BMC Research Notes

DOI:
10.1186/s13104-025-07198-z

BMC Research Notes

RESEARCH NOTE

Open Access

Inter-rater disagreement in manual scoring of intensive care unit sleep data

Laurens Reinke[1*], Esther M. van der Heide[2], Pedro Fonseca[2,3], Anthony R. Absalom[4] and Jaap E. Tulleken[1]

Abstract

Objective Severe sleep disruption is common among intensive care unit (ICU) patients. However, the applicability of standard sleep scoring guidelines by the American Academy of Sleep Medicine (AASM) has been questioned, with most polysomnography (PSG) studies in critically ill patients reporting difficulties in setting up and processing and scoring the recordings. The present study explores human inter-rater agreement in sleep stage scoring following the AASM guidelines, within a heterogenous ICU patient cohort.

Results Two human experts independently scored a total of 51,454 epochs in 20 PSG recordings acquired at the ICU. Epoch-per-epoch comparison of scored stages revealed a Cohen's κ coefficient of agreement of 0.36 for standard 5-stage scoring. Highest agreement occurred in Wake (κ = 0.46), while REM showed the lowest (κ = 0.12). Significant correlations were found between inter-rater agreement, and Simplified Acute Physiology Score (SAPS II, r = − 0.506, p = 0.038), and 12-month mortality (r = − 0.524, p = 0.031). Comparison with similar studies underscore challenges in applying AASM criteria to ICU patients. Despite accounting for artifacts, disparities persisted, emphasizing the need for a nuanced exploration of factors influencing scoring inconsistencies in critically ill patients.

Trial registration: Trial was registered as "Sleep and biorhythm in the ICU", in the Centrale Commissie Mensgebonden Onderzoek register, with number NL-OMON43659 (https://onderzoekmetmensen.nl/nl/trial/43659), on registration date august 4th 2015.

Keywords Sleep, Polysomnography, Intensive care, Sleep scoring

Introduction

Sleep is a dynamic, complex physiological process essential for homeostasis, recovery, and survival [1, 2]. Disrupted or delayed sleep is associated with impaired immune function [3], increased susceptibility to infections and impaired wound healing [4, 5], impaired metabolic and endocrine function [6], increased pain perception [7, 8] and impairment of neurophysiologic organization and memory consolidation [9].

Sleep deprivation affects up to 60% of all critically ill patients admitted to an intensive care unit (ICU) [10, 11]. Sleep among these patients is often fragmented by frequent arousals and awakenings which hamper transitions to deeper stages of sleep, reduced duration of sleep, and disturbed distribution of sleep with up to half of the total sleep time occurring during the day [4, 5, 11, 12]. Poor sleep during critical illness is considered to be a major stressor for patients during and after ICU admission. It might be associated with the development of ICU delirium and long-term cognitive decline, and has detrimental effects on recovery, morbidity, and mortality [13–15].

*Correspondence:
Laurens Reinke
l.reinke@umcg.nl
[1] Department of Critical Care, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands
[2] Philips Research, Eindhoven, the Netherlands
[3] Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, the Netherlands
[4] Department of Anaesthesiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands

The ICU is a unique environment where a multitude of intrinsic and environmental factors may hamper sleep [16–22]. Although previous studies have provided new insights into the etiology and possible prevention of disturbed sleep in the ICU, their scope, statistical significance and reliability have thus far been constrained by the logistical challenges of measuring and assessing sleep objectively [2, 4, 20, 22–28].

Electroencephalography (EEG) has historically been the primary tool for objective sleep monitoring [29, 30]. Polysomnography (PSG), combining EEG electromyography (EMG), and electrooculography (EOG) is the technique used to investigate sleep. The visual and manual annotation or scoring of these recordings commonly follows criteria originally set by Kales and Rechtschaffen [31], with additional changes later culminating in the American Academy of Sleep Medicine (AASM) Manual for the Scoring of Sleep [32]. Hundreds or even thousands of 30 s epochs each comprising multiple channels of PSG data are typically processed by a single human expert. Although this method is considered to be the gold standard for routine clinical sleep analysis, most PSG studies in critically ill patients report difficulties in setting up, maintaining, and manually processing and scoring ICU sleep recordings [4, 12, 33–36]. The practical expertise required to apply and maintain the array of electrodes required for human scoring further limits scalability and increases costs. Furthermore, the reliability and repeatability of manual analysis of ICU sleep recordings is lower than for other clinical recordings [9]. While Elliott et al. reported observed 'reasonable' to 'good' agreement between two combinations of 3 human scorers in discerning wake from sleep activity, the agreement on detailed sleep staging was much lower depending on individual sleep stages and the combination of human scorers [23].

The objective of this study is to investigate human inter-rater agreement in sleep staging following the AASM rules for sleep scoring, in a heterogeneous population of ICU patients.

Methods

Study population and patient recruitment

We obtained 70 PSG recordings during an observational study (NL-OMON43659) primarily investigating the influence of disrupted biorhythms on the quantity and quality of sleep among non-sedated patients of the department of Critical Care of the University Medical Center Groningen (UMCG). After approval by the local ethics committee (UMCG METc, registration number 2015/00295), data collection started in September 2015 and finished in September 2018. All adult patients without a history of sleep pathology, an expected ICU stay of at least 48 h, and a Richmond Agitation and Sedation Scale (RASS) above -3 were eligible for inclusion in the study. Informed consent was gained from patients with capacity to do so. For patient lacking capacity, informed consent was first obtained from their legal representatives, followed by consent after they recovered consciousness. Neurosurgical patients, and patients taking melatonin supplements were excluded from participation. We did not consider this supplement as part of critical care. However, we did classify many potentially sleep-altering medications as such, and therefore chose not to exclude them.

Data acquisition

PSG was recorded for a period of 24–72 h depending on patient's tolerance, RASS scores, and ICU length of stay. The recording consisted of six EEG channels (F3, A1, A2, C3, C4, O1), two EOG channels and EMG of the left and right masseter or submental muscles. Ag/AgCl EEG-electrodes were placed according to the international 10–20 system after skin preparation according to standardized techniques. A BrainAmp DC32 amplifier with a BrainVision recorder (Brain Vision Solutions, Montreal, Canada) or an Alice 6 LDx system (Philips Respironics, Murrysville, USA) was used. EEG was recorded with a sample frequency of 256 Hz. Due to technical failure we switched devices. We observed no difference in quality of the respective recordings. Anonymized data were stored for sleep scoring by two experienced human experts.

Sleep analysis

All recorded data were blindly assessed for data quality by a human expert and sets of sufficient quality were then analysed by a human expert scorer (M_1). We randomly selected 20 patients for further analysis by an additional human expert scorer (M_2). Human expert scorers were free to select either the C4-A1 or C3-A2 EEG channel for scoring depending on signal quality.

The scoring of discrete wake and sleep stages (rapid eye-movement sleep, REM; non-REM sleep stages, N1, N2, N3) according to the latest AASM scoring guidelines by the scorers was done by visual interpretation of individual 30 s epochs in the Brain RT software (OSG, Rumst, Belgium).

Statistical analysis

Demographics of the group with valid recording, and the subgroup randomly selected for additional analysis are shown in Table 1. Significance in differences in demographics between groups was established after evaluating all 15 repeated comparisons with a Benjamini–Hochberg procedure [38], using a maximum acceptable false discovery rate of 10%: the comparison

Table 1 Demographics of the group with valid recordings, and the subgroup randomly selected for additional analysis by M_2

Characteristic	Valid recordings (n = 61) N (%)	Analysed by M_2 (n = 17) N (%)	p-value*
Sex	23 (37.7)	8 (47.06)	0.357
ICU admission diagnosis			
Surgical	19 (31.15)	5 (29.41)	0.859
Medical	42 (68.85)	12 (70.59)	0.859
12 month mortality	19 (31.15)	4 (23.53)	0.433
	Median (IQR)	**Median (IQR)**	
BMI	26 (23–29)	28 (24–32)	0.256
Age, years	60 (52–67)	63 (52–67)	0.388
Duration of hospital stay, days	31 (19–55.49)	21.99 (15.99–43)	0.573
Duration of ICU stay, days	10.99 (5.51–25.50)	7.01 (4.01–18.98)	0.014
Duration of recording, hours	47.74 (17.22)	46.66 (43–65.89)	0.248
APACHE IV	69 (47–82)	74 (61.25–81.25)	0.098
SAPS II	40 (31–49)	42 (36–44)	0.214
Mechanical ventilation, days	7 (1–16)	5 (1–14)	0.282
		Mean (SD)	
Medication dose per day			
Benzodiazepines, mg Lorazepam equivalent	1.55 (4.82)	2.95 (8.77)	0.371
Opioids, mg Morphine equivalent	1.80 (3.63)	0.95 (1.92)	0.242
Propofol 2%, ml	1.81 (7.27)	2.96 (10.59)	0.459

* p-values for the comparison between the subgroup analysis by M_2 versus those not analysed by M_2

None of the differences between the groups were statistically significant after a Benjamini–Hochberg procedure for 15 comparisons, with an acceptable false discovery rate of 10%

i with the largest P value still below the critical P(i) was considered significant, as were all comparisons with lower P

Sleep-related parameters were calculated using Matlab (Matlab 2014b, Natick, MA, USA). Statistics were calculated using SPSS 24 (2016, IBM, Armonk, NY, USA). Cohen's Kappa statistic was used to evaluate epoch-per-epoch agreement between human expert scorers for all sleep stages individually and for full 5-stage sleep scoring. Cohen's Kappa is a dimensionless index that corrects for chance agreement due to imbalanced datasets, such as the imbalanced distribution of sleep and wake stages. Scoring agreement statistics for wake and individual sleep classes were calculated using a binary one-versus-rest strategy. For normative interpretation of inter-rater agreement we used the guidelines by Landis and Koch [39]. Spearman's correlation coefficient was used to quantify the correlation between inter-rater agreements, predicted mortality, and mortality. For estimation of statistical significance, an alpha of 0.05 was used. Unless indicated otherwise, results are presented as mean values (standard deviation).

Results

Seventy patients were included in the main study. PSG data from 4 (5.71%) were lost due to undetected technical failure of EEG equipment during measurement. A further 5 (7.14%) recordings were deemed entirely unscorable by the human expert scorers. Of which three cases of low-quality recordings with substantial movement, sweating, and electrode dislocation artifacts that could not be filtered out, one case of continuous biphasic activity despite low sedation and high RASS, one case of intrusion of a large electrical artifact. Of the remaining 61 patient recordings a median of 0.22% of epochs (0.01–0.56% interquartile range, IQR) were entirely excluded due to artifacts, leaving 339,901 30-s epochs (2832.51 h) for further analysis by scorer M_1. In total 20 recordings were randomly selected for classification by a second scorer (M_2), 3 (15%) were rejected entirely due to low signal quality. Of the remaining 17 patient recordings 0.26% of epochs (0.09–0.86% IQR) were entirely excluded due to artifacts, leaving 51,454 epochs (428.78 h) for analysis of inter-rater agreement between two human scorers. Patient characteristics, medication and sedation use for all valid recordings and for the subgroup scored

by M_2 are summarized in Table 1. Recordings randomly selected for additional classification by M_2 were from patients with a no significant difference after Benjamini–Hochberg procedure.

Table 2 indicates the prevalence of each sleep stage (according to M_1 scorings), and agreement between the two human scorers for the 5-class scoring task, as well as for each class versus the rest. Mean κ agreement for 5-classes was 0.36, with the best agreement obtained for Wake, with a κ of 0.46, and worst for REM, with a κ of 0.12. REM was also the least prevalent class, with an average number of 0.00 h per 24-h period.

Per-subject κ agreement between M_1 and M_2 correlated significantly with the Simplified Acute Physiology Score (SAPS II) predictor of mortality ($r = -0.506$, $p = 0.038$), and with recorded 12-month mortality ($r = -0.524$, $p = 0.031$).

Figure 1 illustrates the confusion matrix for the pooled classification of all epochs in the recordings scored by both human experts scorers. Even for the class with the best κ agreement, i.e., Wake, inconsistent scoring was found between the two scorers: M_2 scored a large proportion of M_1-Wake epochs as N2 and to a certain degree, even N3, whereas M_1 scored a larger proportion of M_2-Wake as N1.

Discussion

Human inter-rater agreement in our sample was comparable to that between human scorers in other studies of ICU sleep. Elliott et al. [23] reported a Cohen's kappa of $\kappa = 0.58–0.68$, which they deemed to be 'reasonable' to 'good' agreement [39], for sleep–wake scoring by two combinations of 3 manual/human scorers. Agreement for the results of detailed sleep staging was much lower, with only slight agreement for stage N1 ($\kappa = 0.08–0.12$), moderate agreement for N2 and REM ($\kappa = 0.55–0.58$ and $\kappa = 0.41–0.44$, respectively), and slight to good agreement for slow wave sleep ($\kappa = 0.20–0.76$), depending on the combinations of manual scorers. Similarly, disagreement in our sample was highest for REM and N1, likely due to a general deficit of this stage of sleep in ICU populations. Additional disagreement was found between individual sleep stages and the wake stage, which could be the result of the relatively high amount of EEG and EMG artifacts in this intensive care population being interpreted as proof of wakefulness. The remainder of substantial disagreement exists between the already notoriously difficult to separate N2 and N3 stages.

Ambrogio et al. compared the agreement between two manual scorers for PSG recordings of 14 mechanically ventilated ICU patients and 17 ambulatory control patients [37]. Inter-rater reliability was good ($\kappa = 0.74$) for recordings of ambulatory patients, but there was only slight agreement on the scoring of recordings of ICU patients ($\kappa = 0.19$). Although in our study we observed a slightly higher interrater agreement ($\kappa = 0.36$), we invite caution when comparing this with interrater agreement studies on non-ICU populations. It is tempting to debate the adequacy of the AASM criteria for scoring ICU recorded PSG-data, particularly among the critically ill patients. However, we found that only part of the source of confusion could be attributed to the high amount of EEG and EMG artifacts and this did not fully explain the disparity in scoring between otherwise relatively unambiguous stages, such as Wake and N3, or REM and N2. For these patients, rather than deeming the scoring rules as inadequate, a better understanding of the factors driving this disparity could help shed light on the sleep of this population. We hypothesize that interrater (dis)agreement might be indicative of more fundamental underlying EEG-related phenomena, and advocate for a more fundamental approach to EEG-analysis to help inform the development of potential new scoring systems or criteria and to advance research in this area.

Limitations

PSG is notoriously labour-intensive during set-up, maintenance, and analysis, which limited the sample size of this study a priori. Despite our best efforts, the amount of usable data was further limited by artifacts from

Table 2 Agreement between human scorers (M_1 vs. M_2)

Class	Wake	REM	N1	N2	N3	5 classes
Prevalence based on M_1, hours per day (SD)	13.98 (8.45)	0.00 (0.01)	1.22 (1.34)	6.15 (5.16)	2.66 (4.09)	24
Kappa,—(SD)	0.46 (0.27)	0.12 (0.24)	0.13 (0.13)	0.32 (0.21)	0.26 (0.24)	0.36 (0.21)
Accuracy, % (SD)	80.62 (14.57)	97.89 (5.26)	90.33 (8.24)	80.65 (13.24)	91.54 (9.22)	70.51 (17.5)
Sensitivity, % (SD)	81.3 (21.99)	11.72 (24.78)	15.92 (14.76)	53.02 (24.92)	36.04 (30.97)	–
Specificity, % (SD)	64.86 (26.93)	99.87 (0.33)	96.35 (3.92)	84.89 (13.73)	95.01 (8.96)	–
PPV, % (SD)	79.55 (19.38)	55.56 (39.59)	24.1 (17.91)	44.74 (22.67)	48.63 (36.49)	–

Performance statistics for individual classes were calculated using a binary one-vs-rest strategy

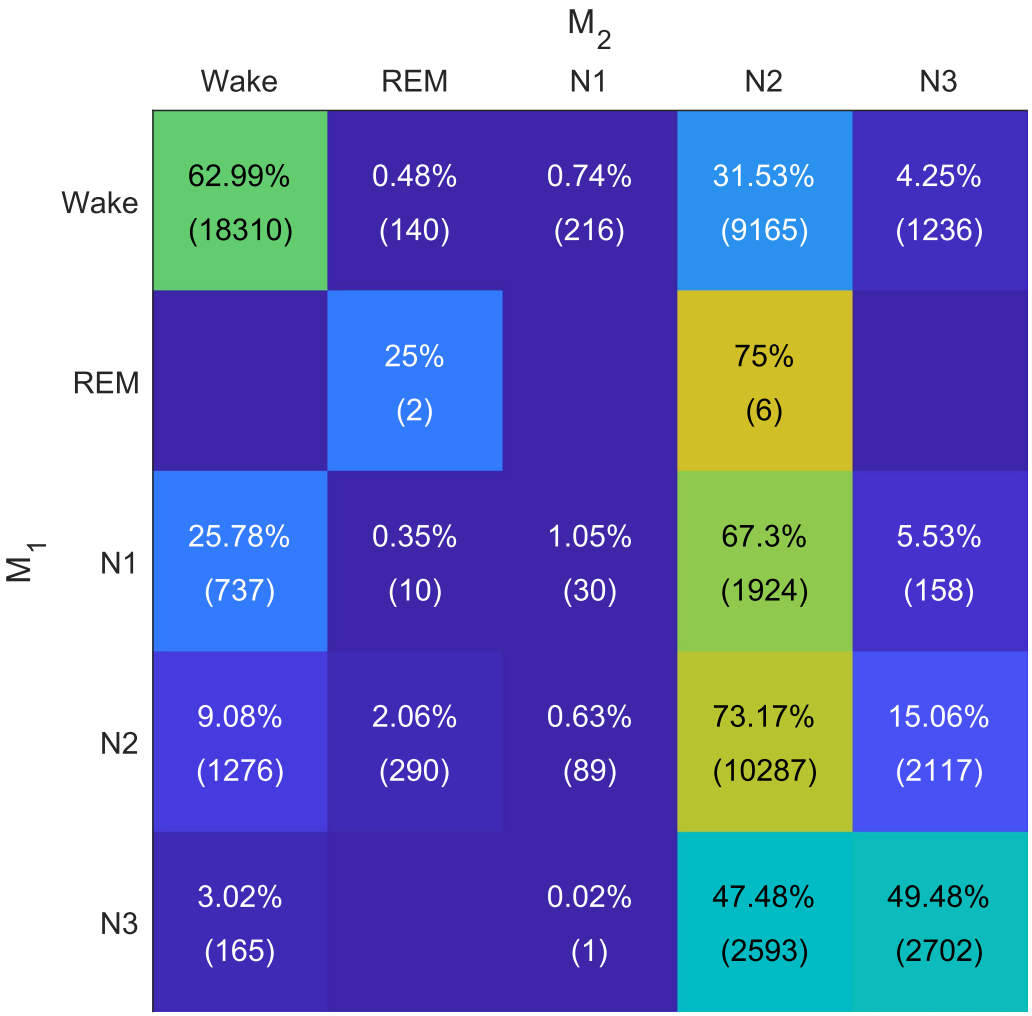

Fig. 1 Confusion matrix for scoring by human scorer M1 versus human scorer M2. Percentages are calculated from class-totals as scored by M1

frequent and intensive care, electromagnetic pollution, motor restlessness, excessive sweating and other technical challenges. Study inclusion and exclusion criteria were chosen to minimize the likelihood of unproductive measurements but may have decreased the already limited generalizability of results from inherently heterogeneous ICU patients.

Study inclusion did not always start immediately after ICU admission and varied in duration due to the unpredictable progression of critical illness. This caused an imbalance in the contribution of individual recordings to aggregated means, which is why all statistics were calculated from per-subject means.

ICU patients could not be relied upon for subjective sleep evaluation, and the neurocognitive state of subjects was not assessed.

The limited practical scalability of polysomnography and human expert sleep scoring has not only restricted the sample size of our comparison but has also limited our ability to do proper consensus scoring or full-sample multi-rater human expert scoring for this investigation. Future efforts to provide more comprehensive investigation of interrater agreements are still encouraged and could benefit from aggregating recordings from previous studies and the adherence to standardized scoring.

Abbreviations

AASM	American Academy of Sleep Medicine
EEG	Electroencephalography
EMG	Electromyography
EOG	Electrooculography
ICU	Intensive care unit
IQR	Interquartile range
PSG	Polysomnography
RASS	Richmond Agitation and Sedation Scale
REM	Rapid eye-movement (sleep)
N1, N2, N3	Non-REM sleep stages 1, 2, 3
SAPS II	Simplified acute physiology score
UMCG	University Medical Center Groningen

Acknowledgements
We thank the entire UMCG ICV research team and student team for their support in setting up and executing this study. We thank dr. J.H. van der Hoeven for analysing PSG data, and Philips Sleep & Respiratory Care for automated sleep analysis. The initiation and success of this study is owed largely to the contribution of Technical Physicians in training.

Author contributions
L.R. drafted the first manuscript, all other authors provided feedback on drafts of the paper. All authors were equally responsible for the conception of the study. L.R. was responsible for implementation of the study and enrolled participants. A.R.A., J.E.T. and L.R. collated the data and analysed results. All authors contributed to, read, and approved the final manuscript.

Funding
No funding was obtained for the purpose of this study. The BrainAmp DC32 amplifier, BrainVision recorder, and disposables were property of the investigating hospital. The Alice 6 LDx system was supplied by Philips Research Eindhoven. LR received partial funding (paid to institution) from Philips Research Eindhoven for a PhD position at the University Medical Center Groningen.

Availability of data and materials
Data are available by contacting the corresponding author.

Declarations

Ethics approval and consent to participate
The study was approved by the local ethics committee of the University Medical Center Groningen ("Medisch Ethische Toetsingscommissie Universitair Medisch Centrum Groningen", registration number 2015/00295). Informed consent was obtained from patients with capacity to do so. Otherwise, informed consent was first obtained from their legal representatives, followed by patients after they recovered consciousness.

Consent for publication
Not applicable.

Competing interests
The BrainAmp DC32 amplifier, BrainVision recorder, and disposables were property of the investigating hospital. The Alice 6 LDx system was supplied by Philips Research Eindhoven. LR received partial funding (paid to institution) from Philips Research Eindhoven for a PhD position at the University Medical Center Groningen. EMH and PF were employed by Philips Research Eindhoven. The remaining authors did not have any conflicts of interest to declare.

Received: 21 February 2024 Accepted: 24 March 2025
Published online: 01 April 2025

References

1. Weinhouse GL, Schwab RJ. Sleep in the critically ill patient. Sleep. 2006;29(5):707–16.
2. Kamdar BB, Needham DM, Collop NA. Sleep deprivation in critical illness: its role in physical and psychological recovery. J Intensive Care Med. 2012;27(2):97–111.
3. Irwin M, McClintick J, Costlow C, Fortner M, White J, Gillin JC. Partial night sleep deprivation reduces natural killer and celhdar immune responses in humans. FASEB J. 1996;10(5):643–53.
4. Cooper AB, Thornley KS, Young GB, Slutsky AS, Stewart TE, Hanly PJ. Sleep in critically ill patients requiring mechanical ventilation. Chest. 2000;117(3):809–18.
5. Friese RS, Diaz-Arrastia R, McBride D, Frankel H, Gentilello LM. Quantity and quality of sleep in the surgical intensive care unit: are our patients sleeping? J Trauma. 2007;63(6):1210–4.
6. Spiegel K, Leproult R, Van Cauter E. Impact of sleep debt on metabolic and endocrine function. Lancet. 1999;354(9188):1435–9.
7. Lautenbacher S, Kundermann B, Krieg JC. Sleep deprivation and pain perception. Sleep Med Rev. 2006;10(5):357–69.
8. Onen SH, Alloui A, Gross A, Eschallier A, Dubray C. The effects of total sleep deprivation, selective sleep interruption and sleep recovery on pain tolerance thresholds in healthy subjects. J Sleep Res. 2001;10(1):35–42.
9. Boyko Y, Ørding H, Jennum P. Sleep disturbances in critically ill patients in ICU: how much do we know? Acta Anaesthesiol Scand. 2012;56(8):950–8.
10. Mistraletti G, Carloni E, Cigada M, Zambrelli E, Taverna M, Sabbatini G, et al. Sleep and delirium in the intensive care unit. Minerva Anestesiol. 2008;74(6):329–33.
11. Freedman NS, Kotzer N, Schwab RJ. Patient perception of sleep quality and etiology of sleep disruption in the intensive care unit. Am J Respir Crit Care Med. 1999;159(4l):1155–62.
12. Bourne RS, Minelli C, Mills GH, Kandler R. Clinical review: sleep measurement in critical care patients: research and clinical implications. Crit Care. 2007;11(4):226.
13. Roche Campo F, Drouot X, Thille AW, Galia F, Cabello B, d'Ortho M-P, et al. Poor sleep quality is associated with late noninvasive ventilation failure in patients with acute hypercapnic respiratory failure. Crit Care Med. 2010;38(2):477–85.
14. McNicoll L, Pisani MA, Zhang Y, Ely EW, Siegel MD, Inouye SK. Delirium in the intensive care unit: occurrence and clinical course in older patients. J Am Geriatr Soc. 2003;51(5):591–8.
15. Ely EW, Speroff T, Gordon SM, Harrell FE, Inouye SK, Bernard GR. Delirium as a predictor of mortality in mechanically ventilated patients in the intensive care unit. JAMA. 2004;291(14):1753–62.
16. Weinhouse GL, Watson PL. Sedation and sleep disturbances in the ICU. Anesthesiol Clin. 2011;29(4):675–85.
17. Van Rompaey B, Elseviers MM, Van Drom W, Fromont V, Jorens PG. The effect of earplugs during the night on the onset of delirium and sleep perception: a randomized controlled trial in intensive care patients. Crit Care. 2012;16(3):R73.
18. Gabor JY, Cooper AB, Crombach SA, Lee B, Kadikar N, Bettger HE, et al. Contribution of the intensive care unit environment to sleep disruption in mechanically ventilated patients and healthy subjects. Am J Respir Crit Care Med. 2003;167(5):708–15.
19. Walder B, Francioli D, Meyer J-J, Lançon M, Romand J-A. Effects of guidelines implementation in a surgical intensive care unit to control nighttime light and noise levels. Crit Care Med. 2000;28(7):2242.
20. Tembo AC, Parker V. Factors that impact on sleep in intensive care patients. Intensive Crit Care Nurs. 2009;25(6):314–22.
21. Bosma KJ, Ranieri VM. Filtering out the noise: evaluating the impact of noise and sound reduction strategies on sleep quality for ICU patients. Crit Care. 2009;13(3):151.
22. Friese RS. Sleep and recovery from critical illness and injury: a review of theory, current practice, and future directions. Crit Care Med. 2008;36(3):697–705.
23. Elliott R, McKinley S, Cistulli P, Fien M. Characterisation of sleep in intensive care using 24-hour polysomnography: an observational study. Crit Care. 2013. https://doi.org/10.1186/cc12565.
24. Figueroa-Ramos MI, Arroyo-Novoa CM, Lee KA, Padilla G, Puntillo KA. Sleep and delirium in ICU patients: a review of mechanisms and manifestations. Intensive Care Med. 2009;35(5):781–95.
25. Weinhouse GL, Schwab RJ, Watson PL, Patil N, Vaccaro B, Pandharipande P, et al. Bench-to-bedside review: delirium in ICU patients—importance of sleep deprivation. Crit Care. 2009;13(6):234.
26. Litton E, Carnegie V, Elliott R, Webb SAR. The efficacy of earplugs as a sleep hygiene strategy for reducing delirium in the ICU. Crit Care Med. 2016. https://doi.org/10.1097/CCM.0000000000001557.
27. Beltrami FG, Nguyen XL, Pichereau C, Maury E, Fleury B, Fagondes S. Sono na unidade de terapia intensiva. J Bras Pneumol. 2015;41(6):539–46.
28. Andersen JH, Boesen HC, Skovgaard OK. Sleep in the Intensive Care Unit measured by polysomnography. Minerva Anestesiol. 2013;79(7):804–15.
29. Loomis AL, Harvey EN, Hobart GA. Cerebral states during sleep, as studied by human brain potentials. J Exp Psychol. 1937;21(2):127–44.
30. Aserinsky E, Kleitman N. Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. Science. 1953;118(3062):273–4.
31. Kales A, Rechtschaffen A. A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects. U.S. Govt. Printing Office, Washington, D.C. Bethesda: U.S. National Institute of

Neurological Diseases and Blindness, Neurological Information Network; 1968.
32. Iber CC, Ancoli-israel S, Chesson AL, Quan SF. Medicine AA of S. The AASM manual for the scoring of sleep and associated events; rules, terminology and technical specifications. 1st ed. Westchester: American Academy of Sleep Medicine; 2007.
33. 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(2):451–7.
34. Drouot X, Roche-Campo F, Thille AW, Cabello B, Galia F, Margarit L, et al. A new classification for sleep analysis in critically ill patients. Sleep Med. 2012;13(1):7–14.
35. Foreman B, Westwood AJ, Claassen J, Bazil CW. Sleep in the neurological intensive care unit. J Clin Neurophysiol. 2015;32(1):66–74.
36. Watson PL. Measuring sleep in critically ill patients: beware the pitfalls. Crit Care. 2007;11(4):159.
37. Ambrogio C, Koebnick J, Quan SF, Ranieri M, Parthasarathy S. Assessment of sleep in ventilator-supported critically Ill patients. Sleep. 2008;31(11):1559–68.
38. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B. 1995;57(1):289–300.
39. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74.

Publisher's Note