SLEEPJ, 2026, 49, 1–10
doi:10.1093/sleep/zsaf229
Advance Access publication 7 August 2025
Original Article

Original Article

Health-oriented sleep states: making sleep states reflect health conditions

Haoqi Sun [1], Wolfgang Ganglberger [1], Samaneh Nasiri[1,2], Thijs E. Nassi [1], Erik-Jan Meulenbrugge[1], Alice D. Lam [3], Sahar Zafar[3], Aditya Gupta [3], Manohar Ghanta[3], Valdery F. Moura Junior [3], Chol Shin[4], Rhoda Au[5], Sydney S. Cash[3], Robert J. Thomas [6,†] and M. Brandon Westover[1,*,†]

[1]Department of Neurology, Beth Israel Deaconess Medical Center, Boston, MA, United States, [2]Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA, United States, [3]Department of Neurology, Massachusetts General Hospital, Boston, MA, United States, [4]Department of Internal Medicine, Biomedical Research Center, Korea University Ansan Hospital, Ansan, Gyeonggi, South Korea, [5]Department of Anatomy & Neurobiology, Chobanian & Avedisian School of Medicine, Boston University, Boston, MA, United States and [6]Division of Pulmonary, Critical Care and Sleep, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, United States

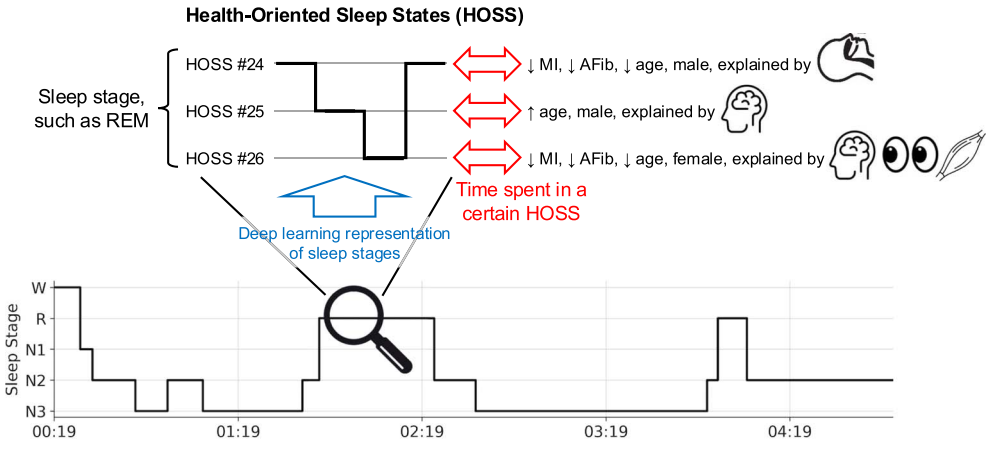
*Corresponding author. M. Brandon Westover, Department of Neurology, Beth Israel Deaconess Medical Center, Harvard Medical School, KS-444, 330 Brookline Avenue, Boston, MA 02215, United States. Email: bwestove@bidmc.harvard.edu.
†These authors contributed equally to this paper.

Abstract

Study Objectives: The rich information in sleep offers insights into brain function and overall health. The current guidelines for sleep staging by the American Academy of Sleep Medicine rely on relatively broad categorizations. These traditional sleep stages are not optimized to reflect health status. Here, we propose health-oriented sleep states to better associate with pre-existing health conditions.
Methods: This observational retrospective cohort study involved 8673 participants from the Massachusetts General Hospital sleep laboratory. We examined seven pre-existing conditions: mild cognitive impairment, ischemic stroke, atrial fibrillation, myocardial infarction, type 2 diabetes, hypertension, and depression. We clustered a sleep staging model's hidden layer within each stage, where clusters represent sleep states. The number of sleep states was selected to maximize the average association with the health conditions, using the average area under the receiver operating characteristic curve across outcomes based on time spent in these states. We also assessed the area under the precision-recall curve.
Results: We identified three states within N3, 14 in N2, 6 in N1, 3 in R, and 9 in W. Average area under the receiver operating characteristic curve ranged from 0.608 to 0.723 across the seven outcomes, and area under the precision-recall curve from 0.064 to 0.524. Among these outcomes, mild cognitive impairment/dementia, atrial fibrillation, myocardial infarction, and hypertension demonstrated significantly stronger associations with the health conditions compared to conventional American Academy of Sleep Medicine sleep stages.
Conclusions: Novel sleep states are linked to health conditions. A better understanding of the physiology behind these sleep states may further enhance the concept of using sleep as a window into overall health.

Key words: sleep/wake physiology; sleep staging; neurophysiology; biological rhythms; machine learning

Graphical Abstract



Submitted: 30 March, 2025; **Revised:** 11 July, 2025; **Accepted:** 2 August, 2025

Statement of Significance

The conventional sleep staging describes sleep physiology rather than indicating health conditions. In contrast to the macrostructure (i.e. sleep stages), the microstructure of sleep, as reflected in multi-organ physiological signals during sleep, contains profound information about health. It would be a conceptual innovation to summarize the multi-organ microstructure of sleep into novel sleep states that better reflect health conditions than the current sleep stages. These sleep states should still align with the conventional sleep stages. We propose health-oriented sleep states, which are data-driven states optimized to associate with health conditions. This approach directly links health to sleep states and interprets them similarly to sleep stages, marking a significant step toward a more comprehensive understanding of the clinical relevance of sleep.

Introduction

Sleep provides a window into brain function and overall health. For example, total sleep time and mortality show a U-shaped relationship [1, 2], where both excessively short and long sleep durations correlate with a higher mortality rate. Similar patterns have been noted with cardiovascular disease [2, 3]. Wake after sleep onset acts as a measure of sleep fragmentation on a "macro" level and is associated with global cognition [4]. These examples are derived from the macrostructure of sleep, captured in the hypnogram. It may be possible to infer cognitive impairment based on sleep architecture, such as reduced deep sleep. However, conventional sleep staging [5] has been designed to describe sleep physiology rather than to indicate health conditions.

In contrast to macrostructure, the microstructure of sleep, as reflected in physiological signals recorded during sleep, contains more explicit information about health. Sleep is a continuous process characterized by complex network interactions, oscillations, and coherence across multiple timescales. For instance, spindle density measured using electroencephalography (EEG) during sleep is linked to memory consolidation [6]. Additionally, heart rate is lower during deep sleep, while non-dipping behavior of nocturnal heart rate correlates with hypertension and cardiovascular disease [7, 8]. Furthermore, it is possible to directly predict brain age [9] and various incident outcomes [10] from sleep EEG using artificial intelligence (AI).

It would be a conceptual innovation to summarize the microstructure of sleep from multiple interacting components (electrocortical activity, respiration, autonomic responses, muscle activation, etc.) into novel sleep stages or states that more accurately reflect health conditions than the current sleep stages. A starting point is to define finer sleep states within each American Academy of Sleep Medicine (AASM) sleep stage [5]. Finer sleep states within each stage have been proposed. About 40 years ago, Santamaria and Chiappa investigated the EEG patterns of drowsiness (N1) in healthy adults [11]. They described various EEG patterns categorized into transitional, post-transitional, and arousal from drowsiness. The Hori 9 stages of sleep onset is another scheme [12]. Additionally, there is an approach for assessing sleep depth through spectral analysis and the odds ratio product (ORP) [13]. However, most of these sleep staging methods rely on visual pattern recognition, and none are specifically optimized for associations with health conditions.

Here, we propose health-oriented sleep states (HOSSs), which define data-driven states within each AASM sleep stage, associating their occurrence with health conditions. Instead of adopting an entirely new staging paradigm, we chose to constrain HOSS within the familiar AASM framework due to practical reasons: ease of use and backward compatibility, so that new insights can be incrementally added on top of conventional staging. This approach connects outcomes to sleep states and interprets them similarly to sleep stages. We compare the time spent in each sleep state with the time spent in each AASM sleep stage across seven health outcomes. We also interpret the HOSSs using various methodologies.

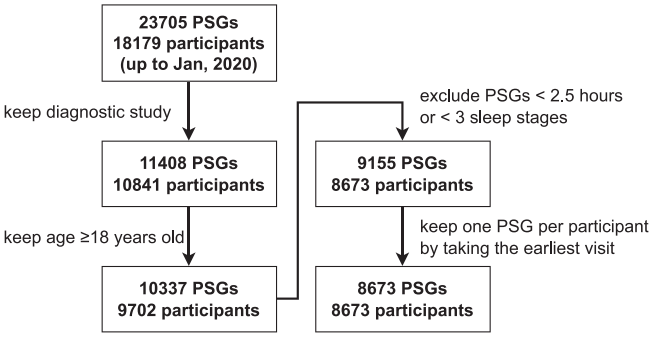


Figure 1. Cohort flowchart.

Materials and Methods

Study design

This is an observational retrospective cohort study, with data collected at the Massachusetts General Hospital sleep clinics. The inclusion criteria were (1) ≥18 years or older at the time of the sleep study (polysomnography [PSG]); and (2) the study type was diagnostic, meaning no CPAP treatment was provided during the PSG. The exclusion criteria included (1) PSG duration shorter than 2.5 h; and (2) fewer than three sleep stages throughout the night. We used the earliest PSG for each patient when multiple PSGs were performed. The Mass General Brigham and Beth Israel Deaconess Medical Center Institutional Review Boards (MGB IRB# 2024P000804, BIDMC IRB# 2022P000417) approved the analysis of PSG data with a waiver of informed consent for this retrospective study. A CONSORT flowchart is shown in Figure 1.

Outcomes

We assessed seven health conditions as outcomes: mild cognitive impairment (MCI) or dementia, ischemic stroke, atrial fibrillation, myocardial infarction, type 2 diabetes, hypertension, and depression.

The ascertainment of MCI or dementia has been described in our prior work [14]. In brief, dementia was defined using the Massachusetts General Hospital (MGH) electronic health record if at least one of the following conditions was met: having an International Classification of Diseases (ICD) code containing at least one dementia key phrase ("Alzheimer" or "dementia"), a problem list containing at least one dementia key phrase, an Montreal Cognitive Assessment (MoCA) score ≤ 19, an Mini-Mental State Examination (MMSE) score ≤ 25, or the use of at least one dementia-related medication. MCI was defined as not qualifying as dementia per the criteria above, and requiring the patient's

problem list to contain an MCI key phrase ("MCI," "mild cognitive impairment," or "minimal cognitive impairment") or that there be a recorded MoCA score between 20 and 25. The rules were verified by a neurologist specializing in memory disorders (author A.D.L.). We combined MCI and dementia to simplify the analysis and increase the sample size.

The ascertainment of other outcomes has been described in our previous work [10]. Briefly, ischemic stroke was ascertained using ICD codes (Table S1) and by detecting specific word patterns (regular expressions) in brain imaging reports (Table S2). For atrial fibrillation, myocardial infarction, type 2 diabetes, hypertension, and depression, we used medications (generic and brand names) and ICD codes (Table S1).

Derivation of HOSSs

We used six channels of EEG, including F3-M2, F4-M1, C3-M2, C4-M1, O1-M2, and O2-M1; one channel of electrooculography (EOG); one channel of chin electromyography (EMG); one channel of electrocardiography (ECG); and two channels of respiratory inductance plethysmography (RIP), including abdomen and chest belts [15, 16]. All signals were notch-filtered at 60 Hz. The EEG and EOG were band-pass filtered from 0.3 to 35 Hz. The ECG was band-pass filtered from 0.3 to 70 Hz. The EMG was band-pass filtered from 10 to 100 Hz. Respiratory effort signals were band-pass filtered from 0.1 to 15 Hz. All signals were resampled to 100 Hz and scaled to have zero median and unit interquartile range. The signals were then segmented into non-overlapping 30-s epochs. Thus, we used all major components of the polysomnogram in addition to EEG.

Epochs were fed into a deep-learning model for sleep staging, ProductGraphSleepNet [15]. This model was chosen because it employs multiple signal modalities. The sleep staging performance on this dataset is shown in Figure S1. We extracted the neural activation vector from the last hidden layer as a representation of the latent sleep information, which has 20 480 dimensions. To facilitate clustering in the next step, we performed principal component analysis (PCA) to reduce the dimensions to 69 while retaining 90 per cent of the variance.

To derive the new sleep states, we performed hierarchical clustering in the PCA-reduced hidden activation space, employing Euclidean distance and Ward linkage, which generated a dendrogram. The clustering was conducted using epochs from the five predicted sleep stages (referred to as "model-predicted sleep stages") separately, resulting in five dendrograms constrained by the AASM sleep stages. Each cluster represents a distinct sleep state.

To make sleep states oriented toward health outcomes, we varied the cutoff values of the five dendrograms and combined the time spent in each resulting sleep state across all sleep stages into a feature vector to classify disease status. For each cutoff value and the resulting sleep states, we performed binary disease outcome classification with 10-fold cross-validation, where 9/10 of the participants were used for training and the remaining 1/10 was used for validation. This process was repeated with different splits until all participants had been used for validation. We assessed classification performance using the average area under the receiver operating characteristic curve (AUROC) for all outcomes, by comparing the pooled disease classification from all validation folds versus the actual disease condition. The sleep states within each conventional sleep stage that achieved the highest average cross-validated AUROC across all outcomes were defined as the HOSSs.

The HOSSs were sorted in descending order based on the model-predicted probability of their respective sleep stages (Figure S2). Within each sleep stage, states with lower numbers show a higher average probability for that stage and are, therefore, more "typical" for that sleep stage.

Interpretations of HOSSs

We interpreted the resulting HOSSs in various ways. First, since the HOSSs are constrained to lie within each model-predicted sleep stage, we calculated the percentage of human-annotated sleep stages in each HOSS and noted the HOSSs that exhibited discrepancies between human annotation and model prediction. Second, some HOSSs dominate during stage transitions, which may indicate a transitional state. We calculated the percentage of HOSSs across all model-predicted sleep stage transitions and identified the dominant ones. Third, we assessed the univariate correlation of HOSS with age and sex, as well as with each outcome after adjusting for these factors. Next, we quantified the organ-specific physiology of HOSS based on the average drop in the probability of being a certain HOSS when the waveforms of a specific signal group were randomly shuffled in time, thereby removing temporal patterns in a specific organ system, while preserving the distribution. The signal–organ system pairs were: EEG for brain, EOG for eye, chin EMG for muscle, ECG for heart, and RIP effort belts for respiration. We will only interpret importance values more than 1 per cent. Lastly, we computed the likelihood of containing at least one obstructive apnea, central apnea, or hypopnea, and their total within a 30-s epoch for each HOSS. To determine whether the likelihood is excessively high, we shuffled the HOSSs to establish a baseline likelihood for each apnea category. In the Supplementary Material, we present the average EEG spectra for each HOSS across all participants (Figures S4–S8) and a dimension-reduced visualization of the latent space vectors used for clustering (Figure S9).

Results

Cohort characteristics

There were 8673 participants. As shown in Table 1, the mean age was 51.0 years, with 50.6 per cent identifying as female and 76.7 per cent as white. The median body mass index was 28.9 kg/m^2. The median Apnea-Hypopnea Index (AHI) was 5.5/h. Antidepressants and benzodiazepines were the most commonly used medications. Among the disease outcomes, hypertension and depression had the highest prevalence.

Health outcome classification performances using sleep states

In Figure 2, the AUROC for classifying different outcomes based on time spent in HOSSs ranges from 0.608 to 0.723. The AUROC for HOSS is significantly higher than that for the five AASM sleep stages across four outcomes: MCI/dementia, atrial fibrillation, myocardial infarction, and hypertension. The area under the precision-recall curve (AUPRC) ranges from 0.064 to 0.524. The AUPRC for HOSS is significantly greater than that for the five conventional sleep stages in three outcomes: MCI/dementia, myocardial infarction, and hypertension.

Sleep state interpretation: association with sleep stage and transition

We present the distribution of human-annotated AASM sleep stages for each sleep state in Table 2. In the "Note" column, we provide qualitative assessments for instances where the dominant human-annotated sleep stage differs from the model-predicted stage. For instance, N2 and N3 form a continuum, whereas the

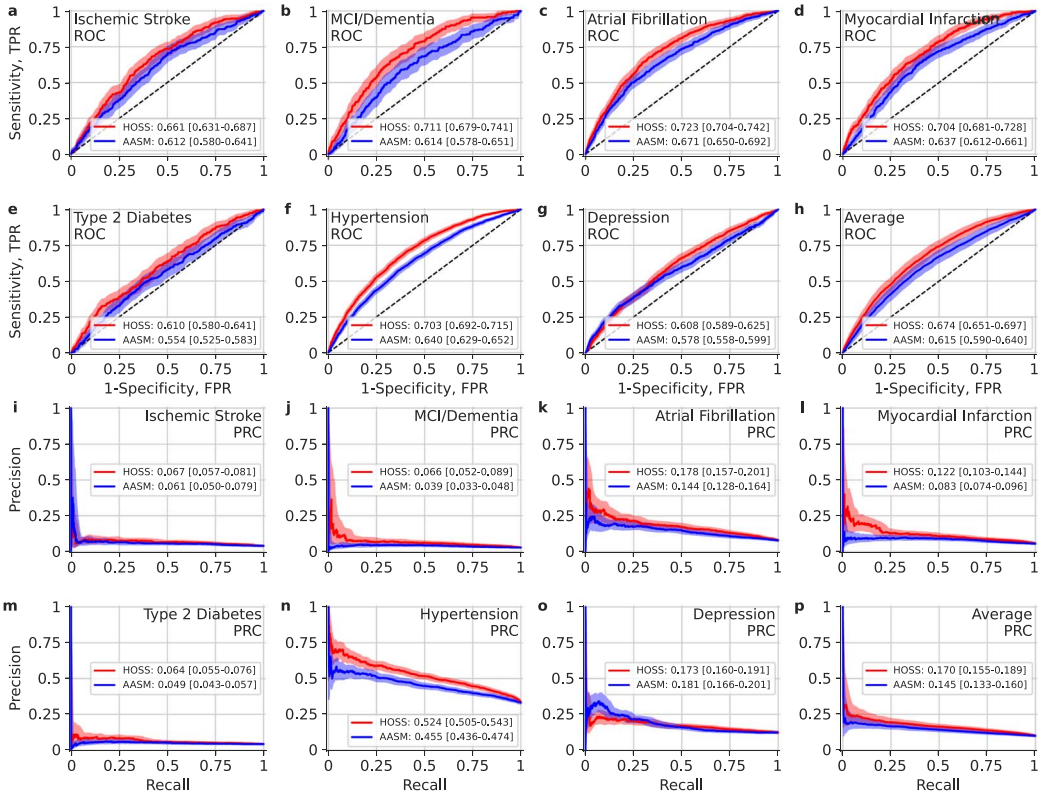


Figure 2. Receiver–operator curve (ROC, top two rows) and precision-recall curve (PRC, bottom two rows) of each outcome.

Table 1. Cohort characteristics

Name	Value
Age, mean (SD), year	51.0 (16.2)
Sex, *n* (%)	
Female	4388 (50.6%)
Male	4285 (49.4%)
Race and ethnicity, *n* (%)	
American-Native	17 (0.2%)
Asian	332 (3.8%)
Black	527 (6.1%)
Hispanic	414 (4.8%)
Middle-Eastern	1 (0.0%)
White	6651 (76.7%)
Other	730 (8.4%)
Body mass index (BMI), median (IQR), kg/m^2	28.9 (25.0–33.9)
AHI, median (IQR), /h	5.5 (1.6–12.8)
Medication, *n* (%)	
Benzodiazepine	744 (8.6%)
Antidepressant	807 (9.3%)
Sedative	426 (4.9%)
Anti-eplipetic	450 (5.2%)
Stimulant	142 (1.6%)
Disease condition, *n* (%)	
Hypertension	2891 (33.3%)
Type 2 diabetes	365 (4.2%)
Ischemic stroke	341 (3.9%)
MCI or dementia	242 (2.8%)
Atrial fibrillation	671 (7.7%)
Myocardial infarction	478 (5.5%)
Depression	1067 (12.3%)

AASM's definition of the boundary between N2 and N3 is arbitrary. Indeed, HOSSs 4, 7, 12, and 14 within the model-predicted N2 are mapped to a relatively high number of human-annotated N3 stages. Additionally, N1 is considered a transitional stage. Here, HOSSs within the model-predicted N1 correspond to distributions with many epochs of human-annotated W and Rapid Eye Movement (REM) stages. Some HOSSs align well with human-annotated sleep stages; for example, HOSSs 27, 28, 29, 31, and 32 at W show strong agreement with human-annotated W stages. Further analysis of their EEG spectra in Figure S8 suggests that HOSSs 27, 28, and 29 likely represent low-power artifacts, as the model accurately and confidently labels their patterns as W.

Some HOSSs may indicate a transitional state. In Table 3, we present the dominant HOSSs immediately following a transition in the model-predicted sleep stages. Some HOSSs correspond with those in Table 2. For example, HOSS 6 is dominant when the model-predicted sleep stage shifts from R to N2, illustrating a transitional state. Meanwhile, HOSS 6 is also identified as a sleep state characterized by confusion between R and N2. It is reasonable to hypothesize that HOSS 6 signifies a transitional state. Likewise, HOSS 17 represents a transitional state between W and N2, while HOSS 21 indicates a transitional state between W and N1. In contrast, HOSS 24 and 26 are dominant sleep states associated with the transition into R, but they do not demonstrate confusion with other sleep stages, as shown in Table 2.

Sleep state interpretation: association with health outcomes

As shown in Figure 3, all three HOSSs in N3 are associated with a reduced likelihood of poor health outcomes, especially HOSS 1, which correlates with multiple outcomes. The HOSSs in N2 show mixed health associations, with 7, 12, and 16 related to a lower likelihood of outcomes in general, while 17 is associated with a higher likelihood of poor health outcomes overall. Referring to Table 2, HOSS 17 displays more W-like characteristics. HOSSs 19 and 23 in N1 are associated with a lower likelihood of outcomes overall, aligning more with R-like patterns as per Table 2. Two HOSSs, 24 and 26 in R, are linked to a lower likelihood of atrial

Table 2. Percent of human-annotated sleep stage in each health-oriented sleep state

Model-predicted stage	Health-oriented sleep state (HOSS)	N3%	N2%	N1%	R%	W%	Note
N3	1	90	10	0	0	0	
	2	91	8	0	0	1	
	3	77	22	0	0	0	
N2	4	50	47	1	1	1	Moderately similar to N3
	5	27	70	1	1	1	
	6	0	30	12	53	5	Similar to R
	7	40	59	0	0	1	Moderately similar to N3
	8	0	16	18	11	56	Similar to W
	9	0	51	23	6	19	
	10	4	81	8	3	4	
	11	6	82	4	7	1	
	12	53	46	1	0	0	Similar to N3
	13	1	60	9	27	4	
	14	80	20	0	0	0	Similar to N3
	15	8	58	11	8	16	
	16	8	89	1	0	1	
	17	5	43	12	1	39	Moderately similar to W
N1	18	0	12	38	31	19	Moderately similar to R
	19	0	13	30	48	8	Similar to R
	20	0	15	29	15	40	Similar to W
	21	0	16	44	9	31	Moderately similar to W
	22	0	16	37	19	28	Moderately similar to W
	23	0	11	20	65	4	Similar to R
R	24	0	13	8	77	2	
	25	0	27	13	52	7	Slightly similar to N2
	26	0	11	10	76	2	
W	27	0	0	0	0	100	
	28	0	2	9	1	88	
	29	0	0	0	0	100	
	30	0	3	23	3	71	
	31	0	1	2	1	96	
	32	0	1	1	0	98	
	33	0	7	23	13	56	Slightly similar to N1
	34	1	3	11	1	85	
	35	1	10	15	4	70	

Table 3. Dominant sleep states immediately after a predicted sleep stage transition, where the dominance is defined as a proportion > 40%

Transition in the predicted sleep stages that occurs more than 1,000 times	Dominant sleep states immediately after the transition (% among all HOSSs)
N2→N3	3 (76%)
R→N2	6 (50%)
W→N2	17 (46%)
N2/W→N1	21 (45%–61%)
N2/N1/W→R	24 (44%–66%)
N1→R	26 (45%)
R→W	33 (41%)
N3/N2/N1/R→W	35 (47%–96%)

fibrillation and myocardial infarction, which are the dominant sleep states while transitioning into R, without being confused with other sleep stages, as shown in Tables 2 and 3. In W, while the results are mixed, hypertension emerges as the outcome with a higher likelihood overall. All HOSSs in W are associated with older age. To further make sense of the health outcomes based on HOSS, we performed hierarchical clustering of health outcomes based on HOSS signatures in Figure 3. As shown in Figure S3, the cardiovascular diseases were clustered together, MCI/dementia and depression were close, and ischemic stroke was by its own. Therefore, disorders of the related organ system shared similar traits (i.e. HOSS signatures) that set them apart from other disorders.

Sleep state interpretation: association with signal organ modality

As shown in Figure 4, the brain represented by EEG had the most widespread importance, with 19/35 HOSS having more than 1 per cent importance (average drop in the probability of being that

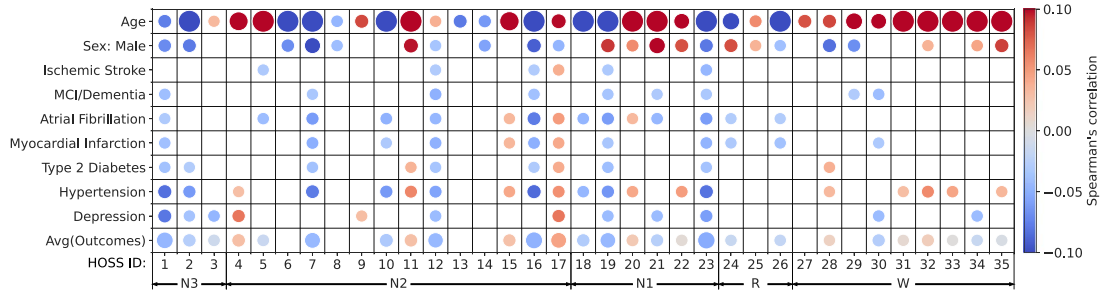


Figure 3. Univariate Speaman's correlation between HOSSs versus age and male sex, and the partial Spearman's correlation between HOSS versus health outcomes after adjusting for age and sex (The color indicates the correlation; while the size is proportional to the minus log *p*-value, i.e. the smaller the *p*-value, the bigger the circle size).

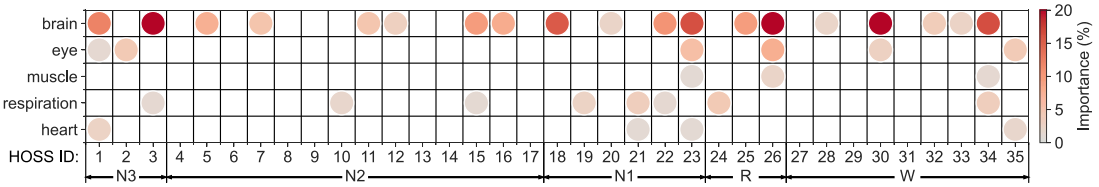


Figure 4. The importance of the organ-specific physiology in each HOSS, which was defined as the average drop in the probability (%) of being a certain HOSS when the waveforms of a specific signal group were randomly shuffled in time, thereby removing temporal patterns in a specific organ system, while preserving the distribution (Brain: F3-M2, F4-M1, C3-M2, C4-M1, O1-M2, and O2-M1; eye: E1-M2; muscle: chin EMG; respiration: Abdomen and chest respiratory inductance plethysmography; heart: ECG).

HOSS if the waveforms in that signal modality were shuffled), followed by respiration represented by RIP effort belts in 8/35 HOSS, eye movement represented by EOG in 6/35 HOSS, cardial activity represented by ECG in 4/35 HOSS, and muscle tone represented by chin EMG in 3/35 HOSS. However, it is important to note that the importance values lacked directionality, i.e. having either healthy or disease-related signal patterns would result in deviation from the typical patterns in a specific HOSS.

Sleep state interpretation: association with apnea and hypopnea

As shown in Table 4, all three HOSSs under R, all six HOSSs under N1, and 7 out of 14 HOSSs under N2 exhibit a high likelihood of apnea and hypopnea.

Example hypnograms

In Figures 5 and 6, we show the results for two example participants of similar age, both female. These results include the human-annotated hypnogram, the model-predicted hypnogram, the HOSS hypnogram (with the background color indicating the model-predicted sleep stage), and the EEG spectrogram from the central channel. The participant in Figure 5 is a 65-year-old female with no medical diagnoses other than hypertension, a common condition in our dataset (33.3 per cent). This participant was predicted to be free from any of the target outcomes, including ischemic stroke (probability 1.8 per cent), MCI or dementia (probability 0.4 per cent), atrial fibrillation (probability 1.4 per cent), myocardial infarction (probability 0.7 per cent), type 2 diabetes (probability 1.6 per cent), hypertension (probability 11.9 per cent), and depression (probability 8 per cent). All probabilities were below the threshold for a positive prediction.

The sleep stage hypnogram and spectrogram show a normal number of cycles, stage transitions, and continuous REM. The EEG spectrogram shows healthy patterns, including strong delta power (red in 1–4 Hz during N3), a clear spindle band at 11–16 Hz, and clear low-power REM.

When examining the HOSS hypnogram, the participant predominantly exhibited HOSS 1 and 2 in N3, which are linked to favorable outcomes, as shown in Figure 3. In N2, the participant primarily displayed HOSSs 7, 12, and 16 during the first half of the night, and HOSSs 9 and 10 during the second half. HOSSs 7, 12, and 16 (along with HOSS 10 to a moderate extent) are associated with generally positive results, as illustrated in Figure 3. In N1, the participant mainly had HOSS 18 and 23, with 23 (and HOSS 18 to a moderate extent) being tied to overall favorable outcomes, as depicted in Figure 3. During REM, the participant mostly showed HOSS 24 and 26, both of which are linked to overall positive outcomes. In W, the participant primarily had HOSSs 30 and 35.

The participant in Figure 6 is a 64-year-old female who had atrial fibrillation, myocardial infarction, hypertension, and depression. The participant was predicted to have ischemic stroke (probability 5.0 per cent), MCI or dementia (probability 5.5 per cent), myocardial infarction (probability 8.3 per cent), type 2 diabetes (probability 9.7 per cent), hypertension (probability 35.9 per cent), and depression (probability 19.9 per cent), but not atrial fibrillation (probability 5.8 per cent).

When examining the sleep stage hypnogram and spectrogram, the EEG spectrogram here, compared to the one in Figure 5, shows similar strong delta power, a clear spindle band at 11–16 Hz, and a normal amount of stage transitions and cycles. The differences include REM with higher power, increased smearing between the spindle band and lower frequencies, and stronger movement artifacts indicated by the vertical bars with high power across all frequencies.

When examining the HOSS in N3, the participant in Figure 6 differs from that in Figure 5, as they spent most of their time in HOSS 3, which may be at the borderline between N2 and N3 according to Table 3, and is associated with only one positive outcome compared to HOSSs 1 and 2. In N2, the participant predominantly occupied HOSSs 5, 6, 10, 11, and 13, but not 12, which is generally linked to good outcomes. In N1, the HOSS durations are brief and mainly involve HOSS 21, which may indicate a transitional state between W and N1. In R, the participant

Table 4. The likelihood of containing obstructive apnea, central apnea, hypopnea, and their total in a 30-s epoch for each HOSS. Bold font indicates the likelihood is higher than the shuffled baseline (see footnote*). We only show the HOSSs with at least one likelihood higher than the baseline. Other HOSSs are not shown

Model-predicted stage	Health-oriented sleep state (HOSS)^	Likelihood of containing apnea in a 30-s epoch (%)			
		Obstructive apnea	Central apnea	Hypopnea	Total
R	24	**2.3**	**1.1**	**5.8**	**9.2**
	25	**2.1**	**0.7**	**6.7**	**9.5**
	26	**1.9**	**0.7**	**4.5**	**7.1**
N1	18	**2.1**	**0.8**	**5.0**	**7.9**
	19	**1.9**	**1.0**	**4.6**	**7.5**
	20	0.3	**1.7**	**3.5**	**5.5**
	21	1.0	**1.0**	2.4	**4.4**
	22	**3.1**	**1.0**	**6.4**	**10.5**
	23	**1.8**	0.3	**4.7**	**6.8**
N2	6	**1.0**	**0.5**	**4.6**	**6.2**
	9	**1.3**	**0.8**	**3.6**	**5.7**
	10	**1.7**	**0.6**	**2.9**	**5.2**
	11	**1.4**	**0.9**	**3.0**	**5.4**
	13	**2.1**	**0.5**	**4.6**	**7.3**
	15	0.5	0.4	**3.2**	**4.1**
	17	0.7	**0.5**	1.9	3.0

*Bold font indicates the likelihood is higher than the shuffled baseline, which is 1% for obstructive apnea, 0.4% for central apnea, 2.6% for hypopnea, and 4% for the sum. ^We only show the HOSSs with at least one likelihood higher than the baseline, while other HOSSs are not shown.

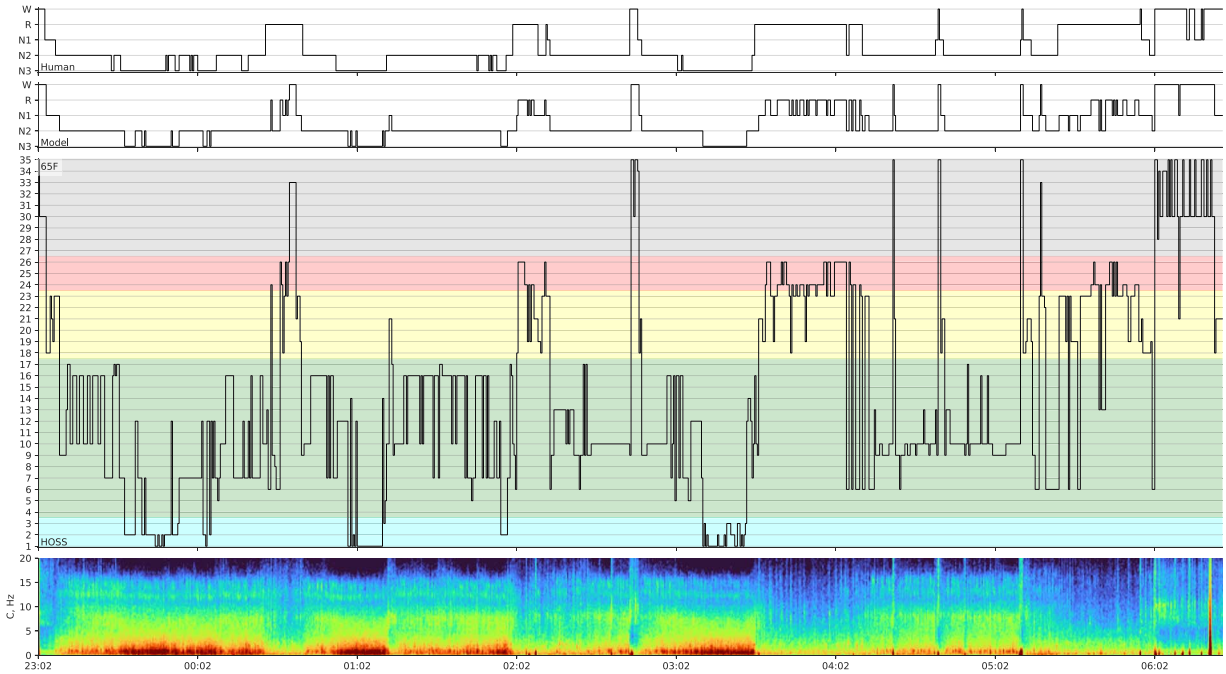


Figure 5. Sleep data from an example participant (65-year-old female) with a low burden of associated disease outcomes, where (A) is a hypnogram based on human annotation; (B) is a hypnogram based on the deep learning sleep staging model; (C) is a hypnogram of the health-oriented sleep states, with background colors indicating different sleep stages; and (D) is an EEG spectrogram of the left–right averaged central channels.

spent all their time in HOSS 24, which is associated with a lower likelihood of two outcomes as shown in Figure 3. In W, the participant primarily engaged with HOSSs 28, 30, 34, and 35, which exhibit mixed associations with the outcomes.

Discussion

A data-driven approach using multiple physiological modalities in PSG identified 35 HOSSs, which offer a more granular view and a closer association with health conditions compared to the AASM sleep stages. The associations of HOSS with health outcomes were stronger than those of AASM sleep stages for MCI/dementia, atrial fibrillation, myocardial infarction, and hypertension. These results reinforce the idea of sleep as a window into overall health: health status is reflected in sleep physiology. The use of a large clinical cohort with a diverse range of health outcomes enhances the generalizability of our findings. Further, HOSS is designed to be compatible with existing scoring and clinical guidelines, as it is not intended to replace traditional staging but rather to refine it by novel data-driven approaches. Removing this constraint may

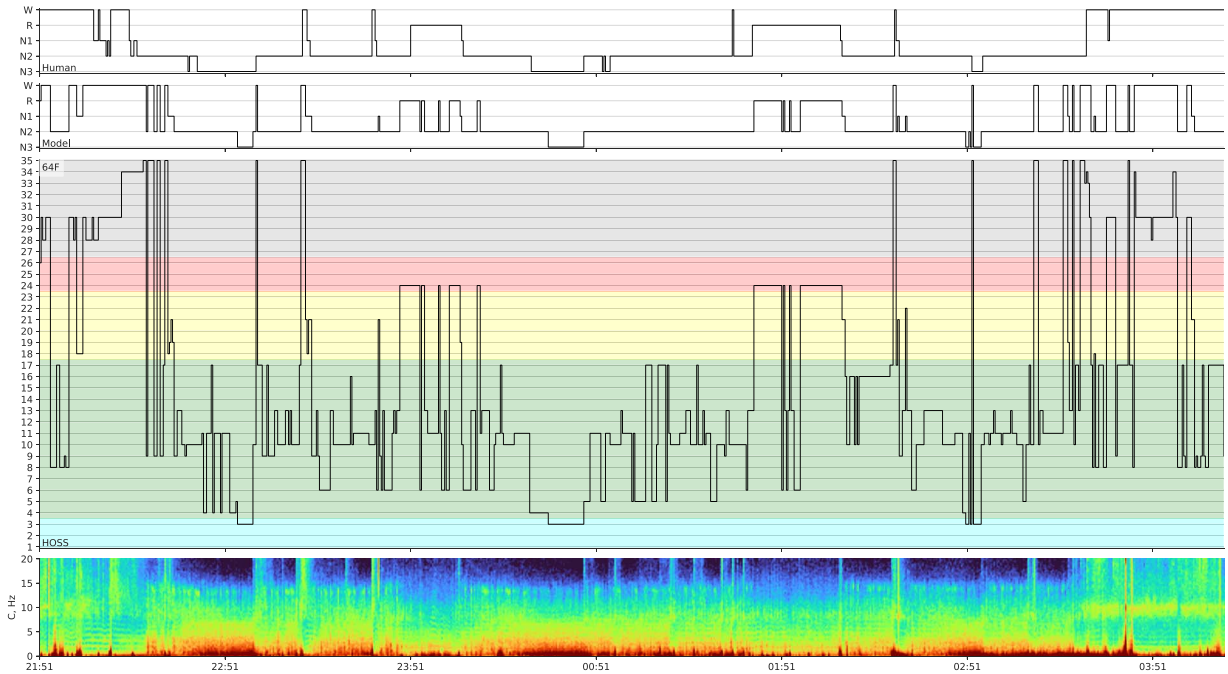


Figure 6. Sleep data from another example participant (64-year-old female) with a high burden of associated disease outcomes, where the format is the same as in Figure 5.

improve outcome prediction performances, but at the cost of reducing its interpretability and applicability.

The findings have potential clinical utility and implications beyond traditional risk assessment. Although the model was optimized to recognize the known history of diseases during training, the disease diagnosis may not always exist when applied to a new individual. HOSS is most helpful when there is a difference in the HOSS-predicted disease vs. the already diagnosed disease. For example, when an individual without a history of myocardial infarction spends much time in certain sleep states associated with myocardial infarction (HOSS 15 and 17), the clinicians could pay attention to subclinical cardiovascular pathology that would otherwise be missed. On top of this, HOSS enables a profile of multiple outcomes based on the time spent in the HOSSs. In this way, HOSS enables clinicians to stratify patients more effectively based on their HOSS hypnogram. Such multiplex screening is difficult to achieve with traditional, disease-specific risk assessments.

Another strength is the use of multi-organ physiology. During sleep, multiple organ systems interact, including the EEG, EOG, EMG, ECG, and respiratory effort signals used by HOSS. For example, sleep apnea in the respiratory system induces oxygen desaturations [17, 18] and an elevated heart rate [19]; heart rate variability is higher in REM [20]; and sleep and breathing are coupled [21]. In individuals with myocardial infarction, there is a high prevalence of central sleep apnea at about 40 per cent [22, 23], where the periodic breathing pattern is measurable from respiratory effort signals; there is also an increase in slow waves, measurable from EEG signals, potentially through an immune response to the elevated inflammation [24]. We innovatively used a deep-learning network using these multi-organ signals, where its embedding space implicitly captures the multi-organ physiology and enables the discovery of sleep states associated with specific health conditions, which is not possible with traditional risk assessment. As shown in Figure 4, 9/35 HOSS had at least two important signal modalities. For example, HOSS 23 in N1 involved multi-organ physiology in brain activity, eye movements, muscle tone, and heart and was associated with favorable health outcomes in Figure 3. HOSS 26 in REM involved multi-organ physiology in brain activity, eye movements, and muscle tone, possibly representing phasic REM [25] and was reasonably associated with cardiovascular outcomes such as atrial fibrillation and myocardial infarction in Figure 3.

Other sleep EEG classification systems have been developed, demonstrating correlations with various health conditions [26]. These include the cyclic alternating pattern (CAP) [27], continuous sleep depth (ORP) [13], alpha-delta sleep [2,8], and hypnodensity graphs [29]. The CAP rate, which indicates the percentage of NREM sleep in CAP, increases in conditions such as insomnia, epilepsy, depression, circadian misalignment, periodic limb movements, and sleep apnea [30]. Hypnodensity graphs reveal a smudging of sleep stages, especially at the edges of REM sleep in narcolepsy [31]. The ORP may be either elevated or low in individuals with approximately similar degrees of sleep apnea [32], with a low ORP indicating deeper sleep, which is associated with better tolerance to continuous positive airway pressure. Furthermore, the proportion of NREM sleep exhibiting an alpha-delta pattern may reflect the severity of chronic fatigue [28]. Scoring schemas for sleep in intensive care units [33] have also been developed, expanding upon standard stages but encompassing fewer categories than those described here.

There are limitations to our study. First, the performance of the sleep staging model could be further improved on this dataset using techniques such as transfer learning [16]. This may help reduce the confusion between N2 and R, and N2 and W. Based on Table 2, we see some HOSSs were confused following the same confusion pattern between the human-annotated and model-predicted sleep stages, such as 6 (N2 and R), 8 (N2 and W), and 25 (N2 and R). They do not have significant associations with health outcomes, as shown in Figure 3. Second, although the HOSSs outperformed traditional sleep stages in predicting

health outcomes, the overall AUROC values were modest. Third, the analyses were still constrained to 30-s epochs, which is due to historical and non-physiological reasons. The optimal epoch length of such analysis is unknown (shorter or longer than 30 s) and may vary in different diseases. Fourth, the health "outcomes" here are cross-sectional associations, not future events, which is the next logical step. In addition, we used one diagnostic PSG per participant, therefore providing a one-night snapshot of sleep, as well as the in-lab effects that confound the findings. Multiple nights of sleep data from the same individual are needed to better characterize the night-to-night variability of HOSSs. There are studies showing lower night-to-night variability of sleep EEG band powers compared to sleep architectural parameters from the hypnogram [34]. Last but not least, further studies are needed to validate HOSS in independent clinical and community cohorts.

In conclusion, this study demonstrates the power of data-driven sleep analysis in refining our understanding and utilization of sleep states as a more detailed aspect of sleep physiology. By orienting the sleep states with health outcomes, we take a significant step toward achieving precision sleep medicine, where sleep staging not only reflects brain states but also provides valuable insights into overall health.

Supplementary material

Supplementary material is available at *SLEEP* online.

Disclosure statement

Financial disclosure: Dr Westover received support from the NIH (RF1NS120947, R01HL161253) and NSF (2014431). Dr Thomas was supported by an AASM Foundation Strategic Research Award.

Non-financial disclosure: Dr Westover is a co-founder, scientific advisor, consultant to, and has personal equity interest in Beacon Biosignals. Other authors have nothing to disclose. Dr Thomas discloses (1) patent and license/royalties from MyCardio, LLC, for the ECG-spectrogram; (2) patent and license/royalties from DeVilbiss-Drive for an auto-CPAP algorithm; (3) consulting for Jazz Pharmaceuticals, Guidepoint Global, and GLG Councils. Other authors declare that they have no conflict of interest.

Data availability

The PSG data can be requested from the Human Sleep Project on the Brain Data Science Platform (https://bdsp.io/content/hsp/2.0/). The outcomes data are included in the Github repo.

Code availability

The code is available on https://github.com/bdsp-core/outcome-oriented-sleep-staging.

References

1. Svensson T, Saito E, Svensson AK, *et al.* Association of sleep duration with all-and major-cause mortality among adults in Japan, China, Singapore, and Korea. *JAMA Netw Open*. 2021;**4**(9):e2122837–e2122837. https://doi.org/10.1001/jamanetworkopen.2021.22837
2. Henríquez-Beltrán M, Dreyse J, Jorquera J, *et al.* The U-shaped association between sleep duration, all-cause mortality and cardiovascular risk in a Hispanic/Latino clinically based cohort. *J Clin Med*. 2023;**12**(15):4961. https://doi.org/10.3390/jcm12154961
3. Yang S, Zhang N, Liang Z, *et al.* Examining the U-shaped relationship of sleep duration and systolic blood pressure with risk of cardiovascular events using a novel recursive gradient scanning model. *Front Cardiovasc Med*. 2023;**10**:1210171
4. Pase MP, Harrison S, Misialek JR, *et al.* Sleep architecture, obstructive sleep apnea, and cognitive function in adults. *JAMA Netw Open*. 2023;**6**(7):e2325152–e2325152. https://doi.org/10.1001/jamanetworkopen.2023.25152
5. Berry RB, Brooks R, Gamaldo CE, *et al.* The AASM manual for the scoring of sleep and associated events. Vol. 176. Darien, Illinois, USA: Rules, Terminology and Technical Specifications, American Academy of Sleep Medicine; 2012:7.
6. Fernandez LM, Lüthi A. Sleep spindles: mechanisms and functions. *Physiol Rev*. 2020;**100**(2):805–868. https://doi.org/10.1152/physrev.00042.2018
7. Sasaki N, Ozono R, Edahiro Y, *et al.* Impact of non-dipping on cardiovascular outcomes in patients with obstructive sleep apnea syndrome. *Clin Exp Hypertens*. 2015;**37**(6):449–453. https://doi.org/10.3109/10641963.2015.1057833
8. Wolf J, Hering D, Narkiewicz K. Non-dipping pattern of hypertension and obstructive sleep apnea syndrome. *Hypertens Res*. 2010;**33**(9):867–871. https://doi.org/10.1038/hr.2010.153
9. Sun H, Paixao L, Oliva JT, *et al.* Brain age from the electroencephalogram of sleep. *Neurobiol Aging*. 2019;**74**:112–120. https://doi.org/10.1016/j.neurobiolaging.2018.10.016
10. Sun H, Adra N, Ayub MA, *et al.* Assessing risk of health outcomes from brain activity in sleep: a retrospective cohort study. *Neurol Clin Pract*. 2024;**14**(1):e200225.
11. Santamaria J, Chiappa KH. The EEG of drowsiness in normal adults. *J Clin Neurophysiol*. 1987;**4**(4):327–382. https://doi.org/10.1097/00004691-198710000-00002
12. Hori T, Hayashi M, Morikawa T. Topographical EEG changes and the hypnagogic experience. In: Ogilvie RD, Harsh JR, eds. *Sleep Onset: Normal and Abnormal Processes*. American Psychological Association; 1994: 237–253.
13. Younes M, Ostrowski M, Soiferman M, *et al.* Odds ratio product of sleep EEG as a continuous measure of sleep state. *Sleep*. 2015;**38**(4):641–654. https://doi.org/10.5665/sleep.4588
14. Ye E, Sun H, Leone MJ, *et al.* Association of sleep electroencephalography-based brain age index with dementia. *JAMA Netw Open*. 2020;**3**(9):e2017357–e2017357. https://doi.org/10.1001/jamanetworkopen.2020.17357
15. Einizade A, Nasiri S, Sardouie SH, Clifford GD. Product-GraphSleepNet: sleep staging using product spatio-temporal graph learning with attentive temporal aggregation. *Neural Netw*. 2023;**164**:667–680. https://doi.org/10.1016/j.neunet.2023.05.016
16. Ganglberger W, Nasiri S, Sun H, *et al.* Refining sleep staging accuracy: transfer learning coupled with scorability models. *Sleep*. 2024;**47**(11). https://doi.org/10.1093/sleep/zsae202
17. Azarbarzin A, Sands SA, Stone KL, *et al.* The hypoxic burden of sleep apnoea predicts cardiovascular disease-related mortality: the osteoporotic fractures in men study and the sleep heart health study. *Eur Heart J*. 2019;**40**(14):1149–1157. https://doi.org/10.1093/eurheartj/ehy624
18. Esmaeili N, Labarca G, Hu WH, *et al.* Hypoxic burden based on automatically identified desaturations is associated with adverse health outcomes. *Ann Am Thorac Soc*. 2023;**20**(11):1633–1641.
19. Azarbarzin A, Sands SA, Younes M, *et al.* The sleep apnea–specific pulse-rate response predicts cardiovascular morbidity and mortality. *Am J Respir Crit Care Med*. 2021;**203**(12):1546–1555. https://doi.org/10.1164/rccm.202010-3900OC

20. Vanoli E, Adamson PB, Ba-Lin, Pinna GD, Lazzara R, Orr WC. Heart rate variability during specific sleep stages: a comparison of healthy subjects with patients after myocardial infarction. *Circulation*. 1995;**91**(7):1918–1922.
21. Ghibaudo V, Juventin M, Buonviso N, Peter-Derex L. The timing of sleep spindles is modulated by the respiratory cycle in humans. *Clin Neurophysiol*. 2024;**166**:252–261. https://doi.org/10.1016/j.clinph.2024.06.014
22. Reshetnik A, Puppe S, Bonnemeier H. Central sleep apnoea and arrhythmogenesis after myocardial infarction—the CESAAR study. *Front Cardiovasc Med*. 2019;**6**:108.
23. Buchner S, Greimel T, Hetzenecker A, *et al*. Natural course of sleep-disordered breathing after acute myocardial infarction. *Eur Respir J*. 2012;**40**(5):1173–1179. https://doi.org/10.1183/09031936.00172211
24. Huynh P, Hoffmann JD, Gerhardt T, *et al*. Myocardial infarction augments sleep to limit cardiac inflammation and damage. *Nature*. 2024;**635**(8037):168–177.
25. Simor P, van der Wijk G, Nobili L, Peigneux P. The microstructure of REM sleep: why phasic and tonic? *Sleep Med Rev*. 2020;**52**:101305. https://doi.org/10.1016/j.smrv.2020.101305
26. Cesari M, Stefani A, Mitterling T, Frauscher B, Schönwald SV, Högl B. Sleep modelled as a continuous and dynamic process predicts healthy ageing better than traditional sleep scoring. *Sleep Med*. 2021;**77**:136–146. https://doi.org/10.1016/j.sleep.2020.11.033
27. Terzano MG, Parrino L, Sherieri A, *et al*. Atlas, r,ules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep. *Sleep Med*. 2001;**2**(6):537–553. https://doi.org/10.1016/S1389-9457(01)00149-6
28. Van Hoof E, De Becker P, De Meirleir K, Cluydts R, Lapp C. Defining the occurrence and influence of alpha-delta sleep in chronic fatigue syndrome. *Am J Med Sci*. 2007;**333**(2):78–84.
29. Perslev M, Darkner S, Kempfner L, Nikolic M, Jennum PJ, Igel C. U-Sleep: resilient high-frequency sleep staging. *NPJ Digit Med*. 2021;**4**(1):72.
30. Parrino L, Ferri R, Bruni O, Terzano MG. Cyclic alternating pattern (CAP): the marker of sleep instability. *Sleep Med Rev*. 2012;**16**(1):27–45. https://doi.org/10.1016/j.smrv.2011.02.003
31. Stephansen JB, Olesen AN, Olsen M, *et al*. Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy. *Nat Commun*. 2018;**9**(1):5229.
32. Younes M, Gerardy B, Pack AI, Kuna ST, Castro-Diehl C, Redline S. Sleep architecture based on sleep depth and propensity: patterns in different demographics and sleep disorders and association with health outcomes. *Sleep*. 2022;**45**(6). https://doi.org/10.1093/sleep/zsac059
33. Ganglberger W, Krishnamurthy PV, Quadri SA, *et al*. Sleep staging in the ICU with heart rate variability and breathing signals. An exploratory cross-sectional study using deep neural networks. *Front Netw Physiol*. 2023;**3**:1120390.
34. LaGoy AD, Cashmere JD, Beckner ME, *et al*. A trait of mind: stability and robustness of sleep across sleep opportunity manipulations during simulated military operational stress. *Sleep*. 2022;**45**(2). https://doi.org/10.1093/sleep/zsab219

SLEEPJ, 2026, **49**, 1–10
https://doi.org/10.1093/sleep/zsaf229
Original Article