NEJM AI

NEJM AI 2026;3(3)
DOI: 10.1056/AIoa2500487

ORIGINAL ARTICLE

# Brain Health from Sleep EEG: A Multicohort, Deep Learning Biomarker for Cognition, Disease, and Mortality

Wolfgang Ganglberger, Ph.D.,[1,2,3] Haoqi Sun, Ph.D.,[1,2,3] Niels Turley, B.S.,[1,2] Ayush Tripathi, Ph.D.,[1,2] Peter Hadar, M.D., M.S.,[4] Aditya Gupta, M.S.,[1,2,3] Kaileigh Gallagher, B.S.,[1,3] Ryan Tesh, M.S.,[1,3] Soriul Kim, Ph.D., M.P.H.,[5,6] Samaneh Nasiri, Ph.D.,[7] Yue Leng, Ph.D.,[8] Stephanie Harrison, Ph.D.,[9] Katie L. Stone, Ph.D.,[9,10] Timothy Hughes, Ph.D.,[11] Susan Redline, M.D., M.P.H.,[2,12] Rhoda Au, Ph.D.,[13,14] Dara S. Manoach, Ph.D.,[15] Hans-Peter Landolt, Ph.D.,[16,17] Reto Huber, Ph.D.,[17,18] Emmanuel Mignot, M.D., Ph.D.,[19] Chol Shin, M.D., Ph.D.,[6,20] Sydney S. Cash, M.D., Ph.D.,[4] Robert J. Thomas, M.D.,[2,21] and M. Brandon Westover, M.D., Ph.D.[1,2,3]

Received: May 6, 2025; Revised: January 8, 2026; Accepted: January 9, 2026; Published: February 26, 2026

## Abstract

**BACKGROUND** Sleep underpins cognition, disease prevention, and overall brain health, yet objective, integrative biomarkers of brain health remain lacking. We hypothesized that overnight sleep electroencephalography (EEG) could provide a substrate for such a biomarker. We asked whether a newly developed, end-to-end, data-driven deep learning framework for sleep EEG can learn a latent representation of brain health and distill it into a single score relevant to cognition, disease status, and mortality.

**METHODS** We analyzed 36,000 polysomnography recordings from 27,000 subjects from six cohorts. EEG data were represented as one-dimensional time series or a two-dimensional time-frequency spectrogram. A multitask deep neural network, trained end-to-end without expert-defined features, learned a 1024-dimensional brain health latent space and jointly predicted cognitive performance, disease status, and sleep metrics. The latent representation was additionally distilled into a single brain health score. We compared performance with demographic baselines, conventional EEG metrics (e.g., rapid eye movement fraction, spindle density), and classic multivariate machine learning approaches.

**RESULTS** The deep learning–derived brain health scores consistently surpassed demographic and expert-defined EEG feature models. For cognitive outcomes, correlations (r) rose from small (demographic-only) to moderate (up to r=0.40), while disease classification areas under the receiver operator curve improved from 0.50–0.55 at baseline to 0.65–0.75. In age-adjusted Cox models, a one-standard-deviation increase in the brain health score was associated with a 31%–35% reduced risk of mortality (hazard ratio 0.65 to 0.69; P<0.0001), topping conventional EEG metrics. Gains over classic machine learning, plus latent space visualization, indicated that both established physiological markers and novel EEG features drove enhanced performance.

**CONCLUSIONS** A multitask, end-to-end deep learning approach generated an interpretable, sleep-derived brain health biomarker. By modeling cognition, disease, and mortality,

*Sydney S. Cash, Robert J. Thomas, and M. Brandon Westover contributed equally to this article.*

*The author affiliations are listed at the end of the article.*

*Wolfgang Ganglberger can be contacted at wganglbe@bidmc.harvard.edu.*

this framework provides a robust index of brain health and may be extended to additional modalities, further enhancing its clinical utility. (Funded by the National Institutes of Health and others.)

## Introduction

Sleep is a critical determinant of brain and systemic health, influencing cognition, disease risk, and long-term neurologic integrity. Multiple neurobiological processes are implicated, including synaptic homeostasis during slow-wave sleep,[1,2] memory consolidation, and glymphatic clearance of metabolic waste.[3,4] Sleep also regulates endocrine, metabolic, autonomic, and immune function,[5] with slow-wave sleep characterized by parasympathetic predominance and rapid eye movement (REM) sleep by distinct autonomic and physiological dynamics.[6,7] Accordingly, disrupted or insufficient sleep is consistently linked to elevated risks of neurologic, cardiovascular, and psychiatric disorders.[8]

Large epidemiologic studies have documented associations between sleep characteristics, cognition, and adverse health outcomes, including dementia, mood disorders, cardiovascular disease, and mortality.[9-22] However, much of this literature relies on univariate or narrowly multivariate analyses of individual sleep metrics, which likely underestimate the multidimensional information encoded in sleep physiology. Sleep electroencephalography (EEG), which directly captures brain activity across the night, represents a particularly rich but underexploited source of information about brain health. Prior work shows that alterations in sleep EEG features — such as reduced spindle activity or disrupted slow-wave patterns — can precede cognitive decline and neurodegenerative disease.[23-25]

Recent advances in artificial intelligence (AI), particularly deep learning architectures such as convolutional neural networks[26] and transformers,[27] offer new opportunities to model complex physiological signals like EEG. These methods enable hierarchical feature learning and capture long-range temporal dependencies, allowing the extraction of rich latent representations directly from raw data rather than relying on expert-defined features. Such approaches may therefore better reflect the integrated, system-level nature of sleep and brain health.

Here, we assembled a large multicohort dataset and developed an end-to-end, multitask deep learning framework that maps overnight sleep EEG directly to brain health–related outcomes, including cognition, disease status, and mortality. The model learns a data-driven 1024-dimensional latent representation of brain health, from which we derive a single, clinically interpretable brain health score. By jointly optimizing multiple outcomes, this framework minimizes reliance on predefined EEG features and instead leverages fully data-driven representation learning.

We hypothesized that this latent representation and its distilled score would outperform conventional sleep metrics and classic multivariate machine learning approaches in predicting cognitive performance, disease risk, and survival. Accordingly, this study aims to (1) leverage the largest multicohort sleep EEG dataset assembled to date to examine associative and predictive relationships between sleep EEG and brain health, and (2) introduce an AI-based methodology for synthesizing complex neural signals into individualized, clinically actionable biomarkers.

## Methods

### DATA COLLECTION

We compiled overnight polysomnography (PSG) and cognitive or disease-related data across multiple independent cohorts: Massachusetts General Hospital (MGH) patients with neuropsychological testing (MGH-COG)[28] and Mini-Mental State Examinations (MGH-MMSE)[29]; Framingham Heart Study (FHS),[30-32] Multi-Ethnic Study of Atherosclerosis (MESA),[33,34] Osteoporotic Fracture in Men (MrOS),[35] Study of Osteoporotic Fractures (SOF),[36] and Korean Genome and Epidemiology Study (KoGES)[37]; and patients from MGH with electronic health record information but no standardized cognitive testing[38-40] (details in Supplementary Appendix, Cohort Description section). These datasets include large community-based cohorts with home sleep studies (FHS, MESA, MrOS, SOF, and KoGES) and clinical in-laboratory PSG cohorts from MGH representing patients evaluated for sleep, neurologic, psychiatric, and cardiometabolic conditions.

All analyses were approved by an institutional review board, with consent waived where applicable and external data obtained under local approvals and data use agreements.[41]

The temporal proximity of sleep and cognitive assessments varied by cohort: within 30 days in MGH-COG and KoGES; mean (± standard deviation [SD]) delays of 2.1 (±2.1) years in MGH-MMSE and 0.7 (±1.0) years in FHS; and, typically, within 2 years in MESA, MrOS, and SOF. All disease diagnoses reflected participant status at the time of sleep recording.

## DISEASE CATEGORIZATION AND MATCHING PROCEDURE

Disease diagnoses were extracted from electronic health records for the MGH cohort as previously described.[38] In SOF and MrOS, depression was assessed using the Geriatric Depression Scale, and dementia diagnoses in SOF were obtained via standardized self-report questionnaires. Subjects with disease were matched to controls based on joint age–sex distributions, which were verified to be comparable using two-sample Kolmogorov–Smirnov tests.

## COGNITIVE VARIABLES HARMONIZATION

Cognitive measures (Table 1) were harmonized by within-cohort z-standardization. Primary outcomes were composite fluid, crystallized, and total cognition scores derived from established neuropsychological constructs[42];

**Table 1. Baseline Data Description for Cognition.***

| Cognition | MGH-COG | MESA | FHS | SOF | MrOS | KoGES | MGH-MMSE |
|---|---|---|---|---|---|---|---|
| N subjects | 189 | 1751 | 781 | 426 | 2809 | 3652 | 158 |
| N sleep recordings | 189 | 1751 | 1319 | 426 | 3770 | 6224 | 158 |
| Age (years) | 51 (17) | 68 (9) | 60 (10) | 83 (3) | 77 (6) | 62 (7) | 66 (11) |
| Age (years) — 95% central range | 21 to 77 | 55 to 87 | 43 to 78 | 76 to 91 | 69 to 89 | 51 to 78 | 42 to 84 |
| Sex (% female) | 60 | 54 | 53 | 100 | 0 | 50 | 47 |
| N hours of sleep | 6.2 (1.3) | 6.0 (1.4) | 6.4 (1.0) | 5.8 (1.3) | 5.9 (1.2) | 5.7 (1.4) | 5.8 (1.3) |
| N1 % | 10 (8) | 14 (9) | 5 (3) | 5 (4) | 8 (6) | 1 (1) | 17 (13) |
| N2 % | 59 (13) | 58 (11) | 56 (11) | 56 (13) | 63 (11) | 78 (9) | 55 (14) |
| N3 % | 16 (11) | 10 (9) | 18 (11) | 20 (12) | 10 (9) | 8 (7) | 14 (12) |
| R % | 15 (7) | 18 (7) | 21 (6) | 18 (7) | 19 (7) | 13 (7) | 13 (8) |
| SFI | 4 (2) | 10 (5) | 4 (2) | 5 (3) | 6 (4) | 5 (4) | 5 (3) |
| AHI | 6 (10) | 37 (23) | 31 (22) | 27 (18) | 32 (19) | N/A | 10 (11) |
| LMI | 31 (40) | 19 (29) | | 37 (38) | 81 (81) | N/A | 40 (80) |
| Arousal index | N/A | 27 (15) | 20 (10) | 23 (13) | 28 (14) | N/A | 10 (15) |
| Cognitive test 1 | Fluid cognition: 106 (16) | Digit symbol coding: 52 (19) | Logical memory — story recall: 1.0 (0.0) | MMSE total: 28 (2) | Trails B: 123 (57) | Trails A: 37 (16) | MMSE total: 27 (3) |
| Cognitive test 2 | Crystallized cognition: 112 (10) | CASI (global cognition): 88 (8) | Logical memory — DR: 10 (3) | Trails B: 139 (65) | Teng — naming: 4.9 (0.3) | Visual reproductions — DR: 7 (3) | |
| Cognitive test 3 | Total cognition: 110 (12) | Digit span F: 10 (3) | MMSE — Orientation (time rec): 1.0 (0.1) | Short MMSE: 25 (1) | Digit vigilance score: 553 (183) | Digit symbol — coding: 58 (18) | |
| Cognitive test 4 | Picture sequence memory: 101 (15) | Digit span B: 6 (2) | Trail making A: 0.6 (0.3) | | Teng 3MS: 93 (5) | Stroop — word reading: 105 (22) | |
| Cognitive test 5 | Pattern comp. process. speed: 109 (22) | | | | | | |
| Cognitive test 6 | Dimensional change card sort: 107 (11) | | | | | | |
| Cognitive test 7 | Oral reading recognition: 111 (8) | | | | | | |

*All variables are reported as mean (± standard deviation) unless otherwise specified. 3MS denotes Modified Mini-Mental State Examination; AHI, apnea–hypopnea index; B, backward; CASI, Cognitive Abilities Screening Instrument; F, forward; FHS, Framingham Heart Study; KoGES, Korean Genome and Epidemiology Study; LMI, limb movement index; MESA, Multi-Ethnic Study of Atherosclerosis; MGH-COG, Massachusetts General Hospital patients with neuropsychological testing; MGH-MMSE, Massachusetts General Hospital patients with Mini-Mental State Examinations; MMSE, Mini-Mental State Examination; MrOS, Osteoporotic Fracture in Men study; N/A, data not collected for that cohort; SFI, sleep fragmentation index; and SOF, Study of Osteoporotic Fractures.

individual test–level analyses were treated as exploratory (see Supplementary Appendix, Cognitive test groupings section).

## MODEL DEVELOPMENT

We developed two modeling approaches: (1) baseline models using hand-engineered features with conventional machine learning, and (2) a deep learning model.

### *Sleep Feature Computation*

EEG features were computed from a central channel (C4-M1) in all PSGs; home studies used two central channels, and in-laboratory MGH studies incorporated six channels (F3-M2, F4-M1, C3-M2, C4-M1, O3-M2, and O4-M1). We extracted 535 time-, spectral-, and complexity-based EEG features,[29,43] 97 spindle and slow-oscillation features using LUNA,[44] and 23 macro sleep features (e.g., stage durations, apnea–hypopnea index, hypoxic burden, arousal and limb-movement indices) per American Academy of Sleep Medicine guidelines. All features were z-standardized across subjects.

### *Classic Machine Learning Models*

We benchmarked Elastic Net, Random Forest, and XGBoost models within each cohort using three predictor sets: (1) demographics, (2) demographics plus macro sleep and respiratory variables, and (3) demographics plus expert-defined EEG features. As performance did not differ significantly across methods, Elastic Net is reported as the representative classic baseline.

### *Deep Learning — Sleep EEG Brain Health Latent Space*

A single multitask model was trained on the pooled multicohort dataset (no cohort-specific models). Only the C4-M1 EEG channel was used. The signals were band-pass filtered (0.3–58 Hz), resampled to 200 Hz, and zero-padded to 11 hours per recording. Three input formats were used (i) one-dimensional time series, (ii) multitaper spectrograms,[45] and (iii) wavelet spectrograms,[46,47] each with 1-second resolution and 100 frequency bins.

Two end-to-end architectures (spectrogram- and time series–based) encoded overnight EEG into a shared 1024-dimensional latent representation ("sleep EEG brain health latent space"). Both used a shared encoder–decoder design with a three-layer depthwise separable convolutional neural network stem, followed by stacked MaxViT (multiaxis vision transformer)[48] and Transformer blocks capturing long-range time–frequency dependencies.

From the latent space, the model predicted (1) cognitive performance via linear regression heads trained with Huber loss and (2) disease status via logistic regression heads trained with binary cross-entropy, operating flexibly on labels available for each recording.

Auxiliary supervision was added via prediction of full-night sleep physiology metrics and 30-second sleep staging via a U-Net decoder, providing dense targets that stabilized optimization; removing these tasks caused latent-space collapse (Fig. S15).

During end-to-end training — meaning the model learns directly from raw EEG to all outputs in a single automated pipeline — the encoder and all task-specific heads (i.e., cognition, disease, auxiliary physiology, and sleep staging) were optimized jointly using a weighted multitask objective:

- $L_{cog}$=mean Huber loss across all cognitive tests available for that recording.
- $L_{dx}$=weighted sum of binary cross-entropy losses across disease labels available for that recording.
- $L_{aux}$=mean Huber loss across auxiliary physiological regression targets for that recording.
- $L_{stage}$=categorical cross-entropy loss for the recording's sleep-stage sequence.

$$L_{total} = \omega_{cog} L_{cog} + \omega_{dx} L_{dx} + \omega_{aux} L_{aux} + \omega_{stage} L_{stage} \quad (1)$$

$$\text{with } \omega_{cog} = 0.55, \omega_{dx} = 0.30, \omega_{aux} = 0.11, \omega_{stage} = 0.04, \text{ and } \sum_{t \in \{cog,dx,aux,stage\}} \omega_t = 1.$$

A final linear head derives a single brain health score (BHS) by linearly combining the 1024 latent features, trained post hoc on the trained latent representation. Its objective integrates task losses so that higher scores denote better cognition (mean across all N cognitive test score[s] available per subject) and lower disease risk (mean across all M disease categories available per subject):

$$\bar{L}_{cognition} = \frac{1}{N} \sum_{i=1}^{N} \text{Huber}\left(\mathbf{BHS}, \text{cognitive score } i\right) \quad (2)$$

$$\bar{L}_{disease} = \frac{1}{M} \sum_{i=1}^{M} \text{CE}\left(-\mathbf{BHS}, \text{disease } i\right) \quad (3)$$

$$\text{total loss } \mathbf{BHS} = \alpha \times \bar{L}_{cognition} + (1-\alpha) \times \bar{L}_{disease} \quad (4)$$

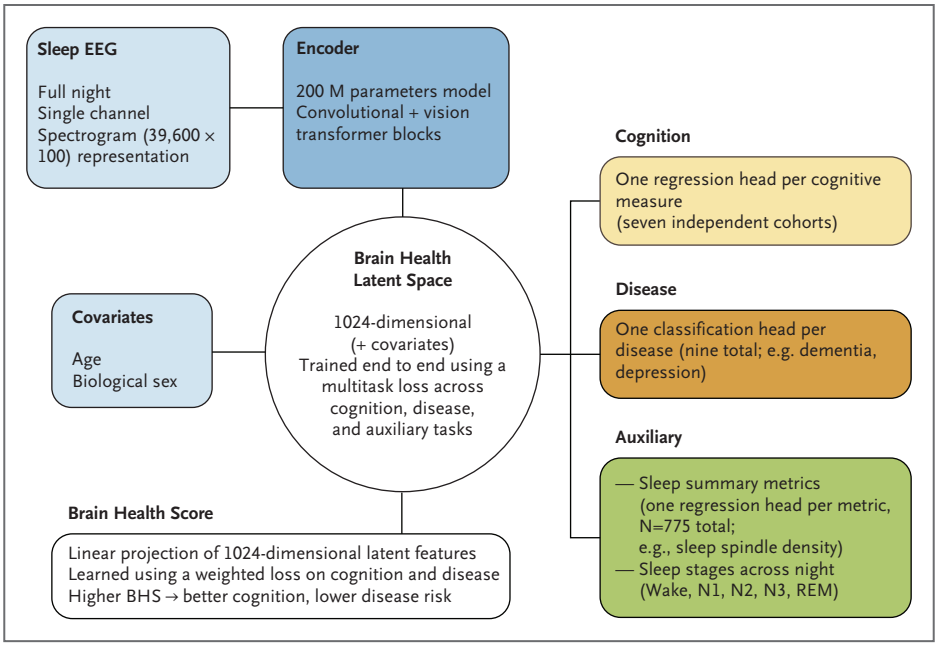


Figure 1. Conceptual Overview of the Multitask Deep Learning Framework for Deriving a Brain Health Latent Space from Sleep EEG.

Full-night, single-channel electroencephalography (EEG) spectrograms feed into an encoder composed of convolutional and transformer blocks, producing a 1024-dimensional latent representation. Demographic covariates (age, biological sex) are appended to this vector. The latent space is trained jointly across cognition, disease, and auxiliary sleep tasks through a shared multitask loss, encouraging the representation to capture EEG features relevant to all downstream outcomes. Task-specific heads then infer cognitive performance, disease status, and summary sleep metrics, and a linear head projects the latent features into a single brain health score. A U-Net–style decoder (not shown) reconstructs sleep stages at 30-second resolution using skip connections from the encoder. The entire framework is optimized end to end across all cohorts, yielding a unified, generalizable EEG-derived brain health embedding. BHS denotes brain health score; EEG, electroencephalography; and REM, rapid eye movement.

The total loss BHS balances cognitive and disease-related objectives, with $\alpha$ controlling their relative weights; in this work, we set alpha ($\alpha$) equal to 0.7. A sensitivity analysis sweeping the cognition–disease weighting parameter $\alpha$ (0.1–0.9) in the BHS optimization showed a smooth trade-off between cognition and disease performance (Fig. S11). Final BHSs were z-scored. Figure 1 provides a conceptual overview of the model design; the Supplementary Appendix provides a conceptual overview of the full architecture, loss functions, parameter details, ablation experiments, and software used.[49,50]

## LABEL AVAILABILITY AND TRAINING LOGIC

All recordings were passed through the full model, but losses were computed only for labels present for each recording. Sleep staging and auxiliary physiology losses were always active, while cognitive and disease losses were masked when absent, avoiding data exclusion or imputation.

## CROSS-VALIDATION

Five-fold cross-validation was performed with participant-level splitting, outcome stratification, and cohort balancing to prevent information leakage and ensure independent evaluation (Fig. S3).

## EVALUATION METRICS FOR PREDICTIVE MODELS

Cognition was assessed using Pearson's correlation (r) and median absolute error; disease classification used the area under the receiver operator curve (AUROC) and the area under the precision recall curve (AUPRC), with bootstrap confidence intervals (10,000 resamples). Disease performance was additionally evaluated at fixed-specificity operating points (50%, 75%, and 90%). Model comparisons, stratified analyses, Integrated Gradients, and uniform manifold approximation and projection (UMAP) visualization are detailed in the Methods section of the Supplementary Appendix.

**Table 2. Baseline Data Description for Disease.***

| Disease | Depression (MrOS) | Depression (SOF) | Dementia (SOF) | Dementia | MCI | Symptomatic Memory | Atrial Fibrillation | Myocardial Infarction | Type 2 Diabetes | Hyper-tension | Bipolar Disorder | Depression |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| N subjects (equal N sleep recordings) | 166 | 36 | 33 | 387 | 585 | 1923 | 625 | 430 | 364 | 2669 | 233 | 1067 |
| Age (years) | 78 (6) | 83 (4) | 84 (4) | 71 (10) | 69 (9) | 65 (9) | 66 (13) | 65 (12) | 59 (15) | 60 (13) | 46 (13) | 50 (15) |
| Age (years) — 95% CI | 69 to 92 | 79 to 91 | 78 to 92 | 52 to 87 | 51 to 84 | 51 to 84 | 32 to 86 | 36 to 84 | 26 to 83 | 31 to 83 | 20 to 71 | 21 to 80 |
| Sex (% female) | 0 | 100 | 100 | 41 | 36 | 46 | 36 | 38 | 55 | 50 | 64 | 65 |
| N hours of sleep | 5.7 (1.5) | 5.1 (1.7) | 5.1 (1.6) | 5.8 (1.3) | 5.8 (1.3) | 6.0 (1.3) | 5.5 (1.4) | 5.7 (1.3) | 5.8 (1.3) | 5.8 (1.3) | 6.2 (1.4) | 6.2 (1.3) |
| N1 % | 8 (6) | 6 (5) | 7 (5) | 21 (18) | 18 (14) | 17 (14) | 20 (15) | 19 (14) | 15 (12) | 18 (14) | 13 (13) | 14 (12) |
| N2 % | 65 (12) | 55 (14) | 54 (16) | 53 (17) | 55 (15) | 54 (15) | 55 (14) | 55 (14) | 55 (14) | 54 (13) | 56 (14) | 56 (14) |
| N3 % | 10 (10) | 21 (14) | 25 (16) | 13 (12) | 13 (12) | 15 (12) | 13 (11) | 13 (11) | 16 (12) | 14 (11) | 17 (12) | 16 (11) |
| R % | 17 (8) | 18 (8) | 15 (8) | 13 (9) | 13 (9) | 14 (8) | 12 (7) | 13 (8) | 14 (8) | 13 (8) | 14 (9) | 14 (9) |
| SFI | 6 (4) | 5 (3) | 4 (3) | 5 (3) | 5 (3) | 5 (3) | 5 (3) | 5 (3) | 4 (3) | 5 (3) | 4 (3) | 4 (3) |
| AHI | 33 (21) | 26 (23) | 32 (23) | 15 (13) | 14 (12) | 13 (13) | 13 (12) | 12 (11) | 11 (12) | 12 (11) | 6 (7) | 8 (10) |
| LMI | 97 (102) | 33 (35) | 31 (37) | 50 (70) | 45 (58) | 35 (54) | 50 (84) | 43 (59) | 38 (57) | 34 (62) | 17 (22) | 29 (62) |
| Arousal index | 28 (16) | 24 (16) | 25 (20) | 13 (16) | 10 (14) | 11 (15) | 12 (16) | 12 (16) | 10 (15) | 11 (16) | 8 (12) | 9 (13) |

*All variables are reported as mean (± standard deviation) unless otherwise specified. AHI denotes apnea–hypopnea index; CI, confidence interval; DR, delayed recall; LMI, limb movement index; MCI, mild cognitive impairment; MrOS, Osteoporotic Fracture in Men study; SFI, sleep fragmentation index; and SOF, Study of Osteoporotic Fractures. N/A indicates data not collected for that cohort.

### SURVIVAL ANALYSIS

Age-adjusted Cox models evaluated mortality prediction across three internally validated cohorts (MrOS, FHS, and MGH) and one independent external cohort (Beth Israel Deaconess Medical Center [BIDMC]). Performance metrics included hazard ratios (per SD), Harrell's C-index, time-dependent discrimination, model fit, and risk reclassification; absolute 5-year mortality risk was estimated by BHS quartiles.

### COMPARISON WITH THE EEG-BASED BRAIN AGE INDEX

The BHS was compared with the EEG-based Brain Age Index (BAI)[51] in the independent BIDMC cohort using nested age-adjusted Cox models; model fit was assessed via likelihood-ratio tests and information criteria (Supplementary Appendix, Comparison with BAI section).

### SLEEP — BRAIN HEALTH ASSOCIATIONS

Univariate sleep–cognition associations used Pearson's and partial Pearson's r (adjusted for age and sex); sleep–disease associations used Cohen's d and two-sample t-tests (Supplementary Appendix, Methods section).

### COMPUTATIONAL RESOURCES

Models (~199 M parameters) were trained in FP32 using PyTorch/PyTorch Lightning on high-memory graphics processing units (RTX 8000, 48 GB VRAM), with batch size 1 for full-night EEG inputs and around 42 hours total training time.

## Results

### DATA COLLECTION

The demographic, sleep, and cognitive baseline variables for all cohorts are presented in Tables 1 and 2. Overall, we collected 36,520 sleep recordings from 27,359 subjects. Most cohorts contributed a single PSG night per subject, whereas a subset (FHS, MrOS, and KoGES) included repeated nights for some participants, typically two nights in FHS and MrOS, and up to three in KoGES. Cognitive data were available for 9766 subjects, some of whom — specifically from the SOF and MrOS cohorts — also had disease diagnoses and are included in the disease-labeled cohort. In total, disease labels were available for 7576 MGH subjects and additional subjects from SOF and MrOS were already included in the cognitive cohort. An additional 11,777 MGH subjects had no associated cognitive or disease data and contributed only to train the auxiliary tasks.

### SLEEP — COGNITION AND DISEASE ASSOCIATIONS

Figure 2 summarizes the age- and sex-adjusted associations between sleep features and both cognitive scores and disease status. Sleep duration correlated positively with

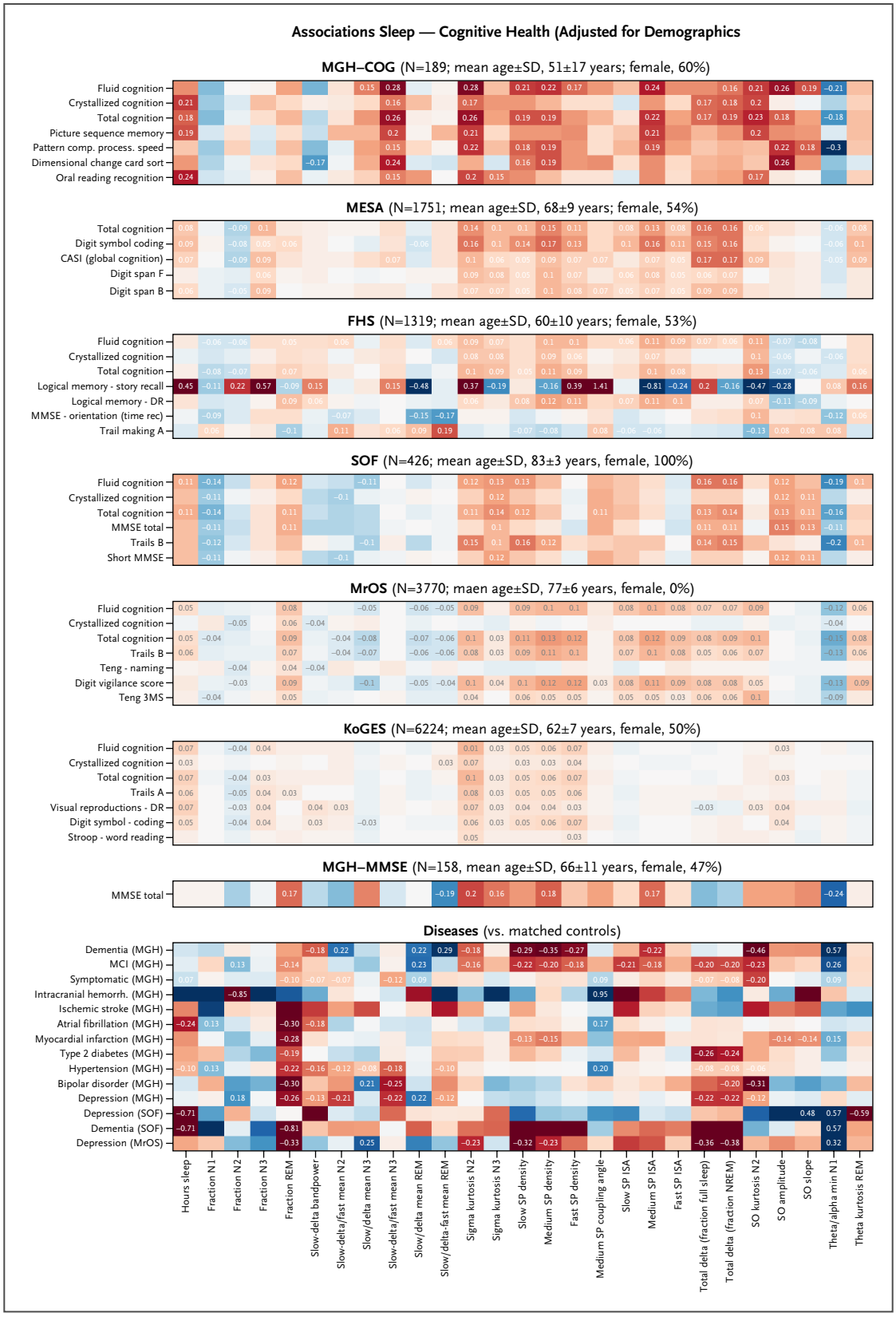


Figure 2. (*Continued*)

cognition across cohorts (partial Pearson's r up to 0.45 for Logical Memory in FHS) and negatively with disease (i.e., positively with health; Cohen's d=−0.71 for depression/dementia in SOF). Among stage fractions, REM was the most consistent marker of brain health — modestly linked to cognition (maximum r of 0.12) but more strongly linked to disease (Cohen's d of −0.81 for dementia in SOF, and −0.19 to −0.30 for atrial fibrillation, hypertension, depression, bipolar disorder). The EEG features most strongly tied to brain health included (1) the fraction of 0.3-to-4-Hz versus 8-to-50-Hz power in N3 (positive), (2) the 0.3-to-1-Hz versus 1-to-4-Hz fraction in REM (negative), (3) sigma power kurtosis and spindle densities (positive), (4) slow-oscillation kurtosis in N2 (positive), and (5) minimal theta to alpha ratio in N1 (negative). Unadjusted associations are shown in Figure S4, where correlations with age were highest for sleep duration, N1 fraction, spindle metrics, spindle–slow oscillation coupling angle, and slow-oscillation kurtosis.

## SLEEP EEG PREDICTS BRAIN HEALTH MORE ACCURATELY THAN DEMOGRAPHIC AND SLEEP MACRO METRICS

Table S2A and Figure 3 show predictive results for (1) demographics alone and (2) our sleep EEG brain health latent space. Figure 3 adds the traditional sleep macro model and visualizes representative outcomes across the three key model classes; full quantitative results for all outcomes, and the complete four-model comparison, including classic EEG-based machine learning, are provided in Tables S2 and S3 and Figure S5.

Across cohorts, the sleep EEG–based deep learning outperformed all baseline models using demographic or sleep macro predictors. Fluid cognition — already partly age driven (demographic r=0.61, MGH; 0.46, KoGES; 0.31, MrOS; 0.29, FHS; 0.14, SOF) — gained modestly ($\Delta r \approx +0.01$–$0.12$). Crystallized cognition, poorly captured by demographics (r=0.09, MrOS; 0.21, KoGES; not significant for other cohorts), improved more in MGH, FHS, and SOF ($\Delta r \approx +0.09$–$0.22$) but not in MrOS or KoGES. In MESA, total cognitive performance improved significantly, with r increasing from 0.30 to 0.39. Improvements were consistent across cohorts for both fluid and crystallized composite scores, with increases in correlation typically in the modest-to-moderate range over demographic baselines. Individual test results are provided in Table 2. For disease prediction, AUROC rose from around 0.50–0.55 to 0.65–0.75 for dementia, mild cognitive impairment (MCI), subjective decline, atrial fibrillation, myocardial infarction, hypertension, bipolar disorder, and depression, underscoring that overnight EEG captures pathology-related signatures far beyond what age conveys.

## SLEEP EEG BRAIN HEALTH LATENT SPACE OUTPERFORMS CLASSIC MACHINE LEARNING

Compared with classic machine learning using expert-based EEG features (Table S3), the sleep EEG brain health latent space generally performed similarly or better. It was significantly superior for crystallized cognition in MrOS and for MCI, depression, atrial fibrillation, and myocardial infarction in MGH. The spectrogram-based deep learning variant outperformed the time series model and was selected as the primary architecture. Wavelet-based spectrograms as inputs led to slightly better performance than multitaper spectrograms (see Supplementary Appendix for detailed results).

## SLEEP EEG BRAIN HEALTH LATENT SPACE: SLEEP STAGING AND TRADITIONAL METRICS

The model accurately performed sleep staging and predicted standard sleep metrics, serving primarily as a

Figure 2. Associations between Sleep EEG Features, Cognition, and Disease.

Age- and sex-adjusted correlations reveal how overnight sleep EEG metrics — spanning spectral bandpowers, slow oscillations, and spindle events — track with cognitive performance and disease status across multiple cohorts. In brief, longer total sleep, higher N3 slow-wave and delta activity, and more sleep spindle activity are associated with better cognition and lower disease prevalence, whereas reduced rapid-eye-movement (REM) sleep fraction aligns with increased risk of dementia, depression, hypertension, and other disorders. Notably, the strength of associations can differ between cognition and disease — for example, REM fraction shows a stronger association with diseases — and notable variability is observed across cohorts with cognitive data. These findings underscore the pivotal role that diverse sleep oscillations play in maintaining — or undermining — brain health. CASI denotes Cognitive Abilities Screening Instrument; DR, delayed recall; EEG, electroencephalography; FHS, Framingham Heart Study; KoGES, Korean Genome and Epidemiology Study; MCI, mild cognitive impairment; MESA, Multi-Ethnic Study of Atherosclerosis; MGH-COG, Massachusetts General Hospital patients with neuropsychological testing; MGH-MMSE, Massachusetts General Hospital patients with Mini-Mental State Examinations; MMSE, Mini-Mental State Examination; MrOS, Osteoporotic Fracture in Men study; NREM, non–rapid eye movement; REM, rapid eye movement; SO, slow oscillations; SOF, Study of Osteoporotic Fractures; SP, spindle event; and 3MS, the Modified Mini-Mental State Examination.

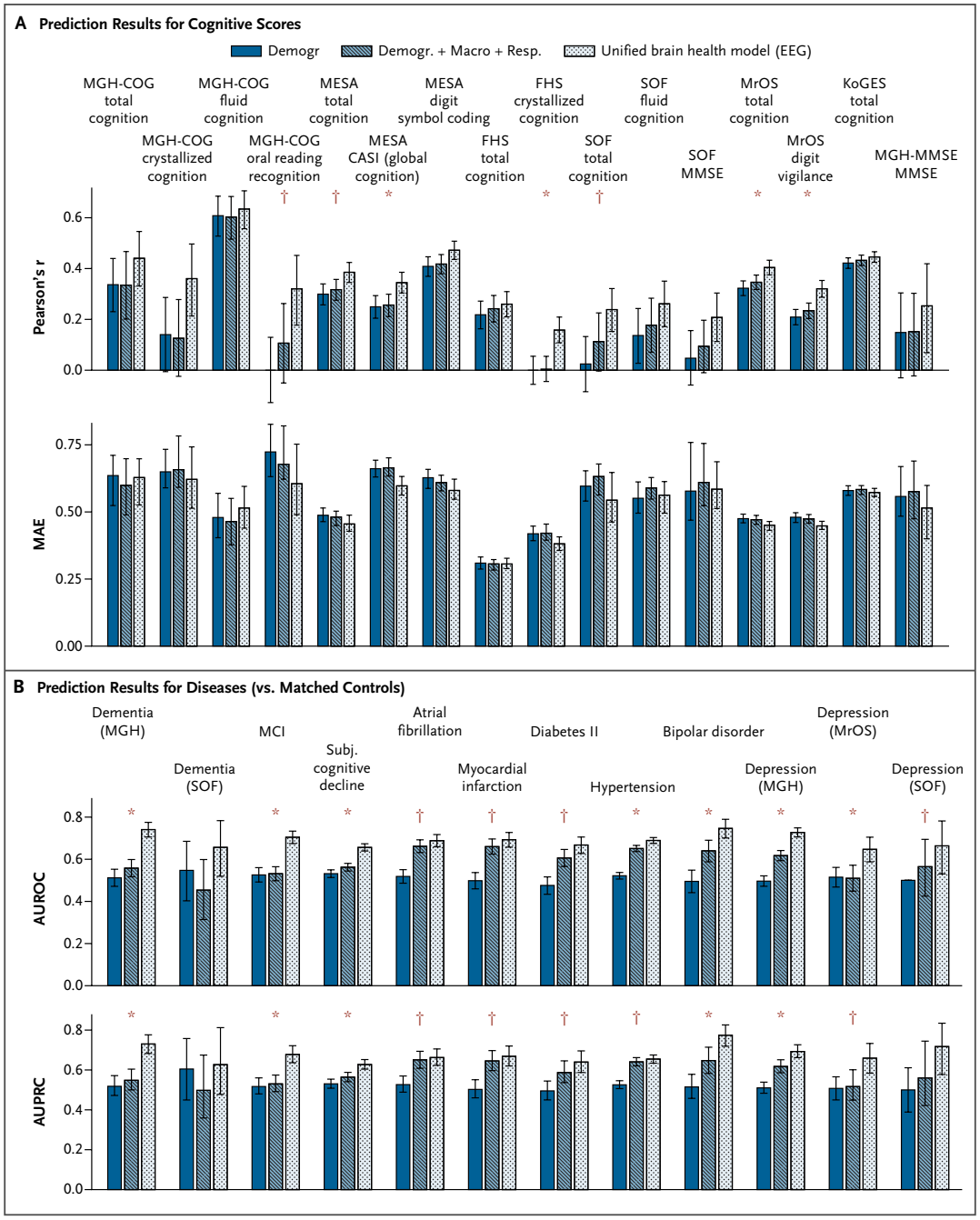


Figure 3. Predicting Cognition and Disease from Sleep EEG with Deep Learning.

Comparison of predictive performance (correlation and median absolute error for cognitive scores shown in Panel A; areas under the receiver operator curves and precision recall curves for disease classification shown in Panel B) between (1) demographic (age and sex) baselines, (2) multivariate models with traditional sleep-stage hypnogram and sleep-disordered breathing respiratory variables, and (3) our multitask deep learning brain health latent space using sleep electroencephalography. The brain health latent space consistently outperforms baseline and the traditional sleep information–based model in inferring levels of cognitive functioning (e.g., reading fluency, sustained attention and psychomotor speed, aggregate cognitive measures) and identifying disease risks (e.g., dementia, depression). *These results show significant improvement over all comparison models. †These results show significant improvement over the demographic model. EEG denotes electroencephalography; FHS, Framingham Heart Study; KoGES, Korean Genome and Epidemiology Study; MAE, median absolute error; MCI, mild cognitive impairment; MESA, Multi-Ethnic Study of Atherosclerosis; MGH, Massachusetts General Hospital; MGH-COG, Massachusetts General Hospital patients with neuropsychological testing; MGH-MMSE, Massachusetts General Hospital patients with Mini-Mental State Examinations; MMSE, Mini-Mental State Examination; MrOS, Osteoporotic Fracture in Men; and SOF, Study of Osteoporotic Fractures.

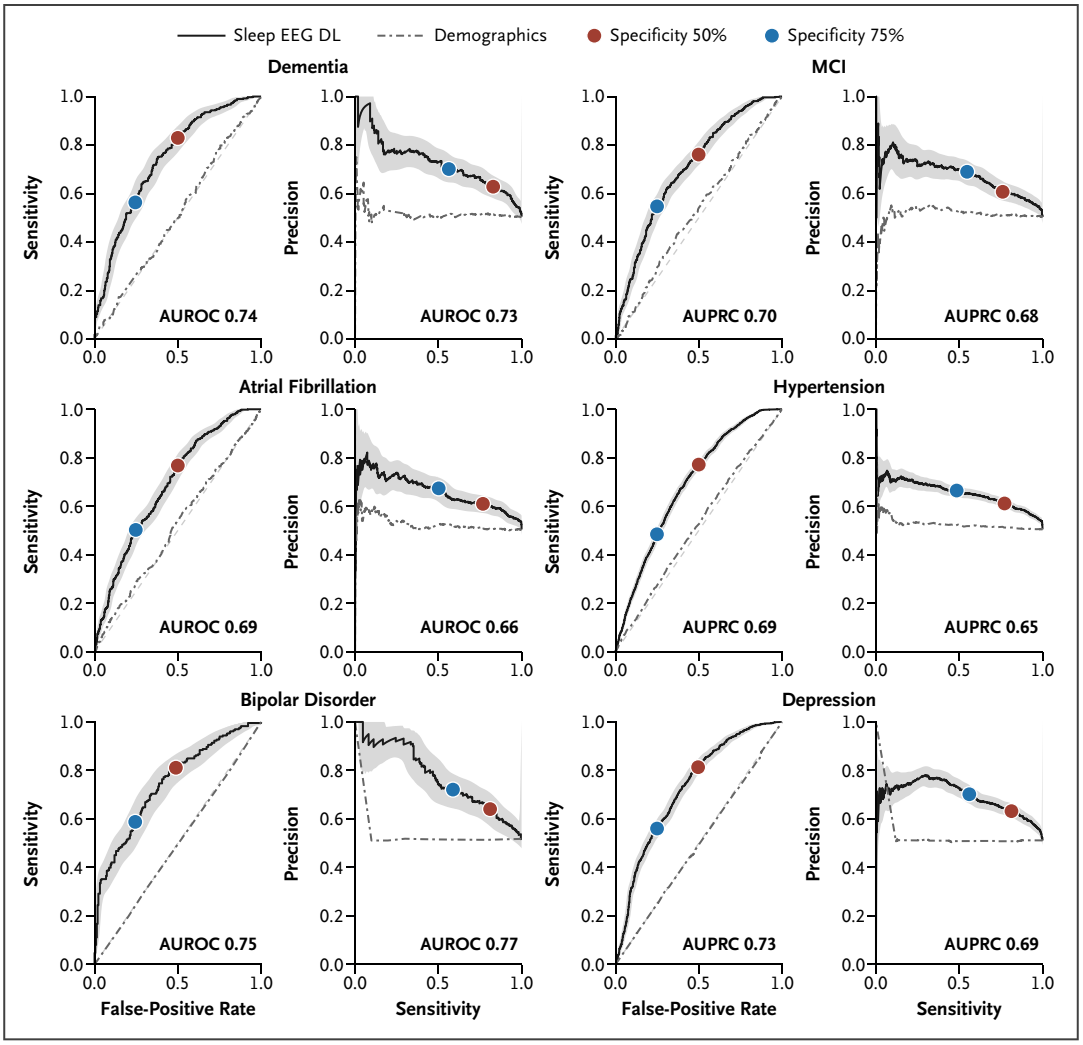


Figure 4. Disease Prediction Performance of the EEG-Based Deep Learning Model at Clinically Relevant Operating Points.

Receiver operator (left) and precision recall (right) curves are shown for six representative diseases, comparing the sleep electroencephalography (EEG) model (black) with a demographic baseline (gray, age+sex). Colored markers denote fixed specificity thresholds (red, 50%; blue, 75%). At 50% specificity, sensitivities typically fall in the 0.70–0.80 range, while higher specificity increases precision at the expected cost of sensitivity. These plots illustrate how the proposed EEG-based biomarker framework performs across screening-oriented and more conservative decision settings. AUPRC denotes area under the precision recall curve; AUROC, area under the receiver operator curve; DL, deep learning; EEG, electroencephalography; and MCI, mild cognitive impairment.

regularization and self-supervised learning component. It achieved state-of-the-art sleep staging performance with a Cohen's kappa of 0.75 across all cohorts (confusion matrices are shown in Fig. S14). Predicted sleep features closely matched conventional estimates: spindle density (r=0.78), spindle–slow oscillation overlap (r=0.66), slow-oscillation count (r=0.85), REM latency (r=0.55), wake-after-sleep onset (r=0.84), and relative delta power in non-REM (NREM) (r=0.89). Notably, despite relying solely on EEG input, the model also estimated respiratory and limb movement indices with moderate to substantial accuracy: apnea–hypopnea index (r=0.58), hypoxic burden (r=0.51), and limb movement index (r=0.53).

## CLINICAL OPERATING POINTS: SENSITIVITY–SPECIFICITY TRADE-OFFS

Across nine diseases, the deep learning model achieved AUROCs of 0.66–0.75, consistently exceeding demographic baselines (0.50–0.62), with clear sensitivity–specificity trade-offs across operating points (Fig. 4; Figs. S6–S8).

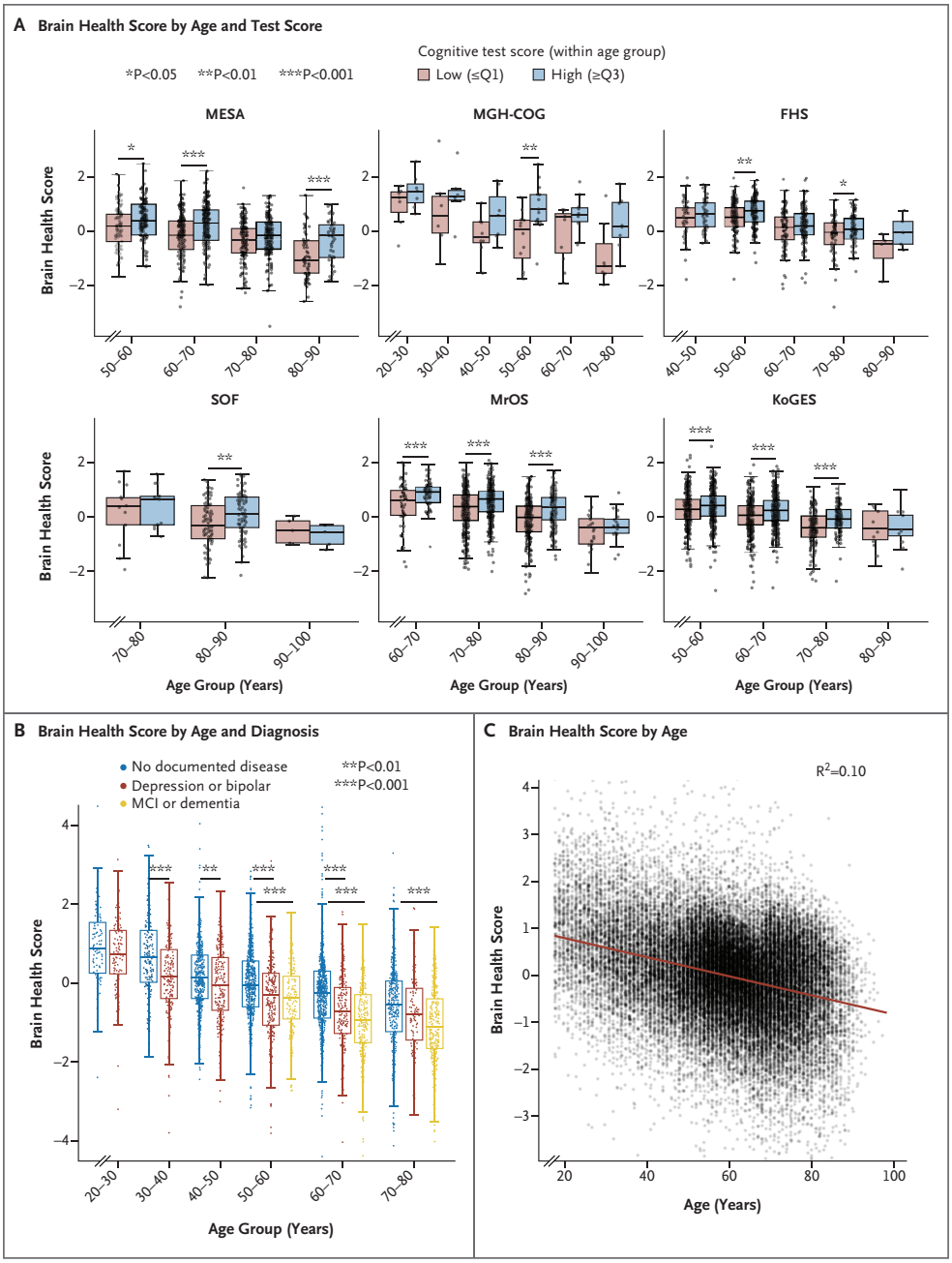


Figure 5. Brain Health Scores Reveal Cognitive and Clinical Differences within Age Bins across Multiple Cohorts.

Panel A shows the distribution of the deep learning–derived brain health score across age groups in six independent cohorts (Multi-Ethnic Study of Atherosclerosis, Massachusetts General Hospital patients with neuropsychological testing, Framingham Heart Study, Study of Osteoporotic Fractures, Osteoporotic Fracture in Men study, and Korean Genome and Epidemiology Study). Participants within each age bracket are stratified by cognitive performance (lowest vs. highest quartile). Higher scores are consistently associated with better cognitive outcomes, confirming that the brain health score robustly reflects cognitive status independent of age. Panel B presents scores further stratified by clinical diagnoses, which show progressively lower values in participants with mood disorders (depression, bipolar disorder) and neurodegeneration (mild cognitive impairment, dementia), relative to disease-free controls. Panel C shows brain health scores decline modestly with age ($r=-0.32$, $R^2=0.10$), capturing age-related neurophysiologic deterioration via sleep EEG. Substantial interindividual variability underscores the clinical value of personalized brain health assessment over chronologic age alone. FHS denotes Framingham Heart Study; KoGES, Korean Genome and Epidemiology Study; MCI, mild cognitive impairment; MESA, Multi-Ethnic Study of Atherosclerosis; MGH-COG, Massachusetts General Hospital patients with neuropsychological testing; MrOS, Osteoporotic Fracture in Men study; Q, quartile; r, Pearson's coefficient; and SOF, Study of Osteoporotic Fractures.

At 50% specificity, sensitivities reached around 0.70–0.80 for depression, MCI, and dementia; at 75% specificity, sensitivities remained moderate (∼0.55) with higher precision (∼0.68–0.72); and at 90% specificity, sensitivity declined (∼0.25–0.34) while precision increased (∼0.73–0.82). These results demonstrate that EEG-derived biomarkers can be tuned for screening-oriented versus conservative clinical use; full metrics are provided in the Supplementary Appendix.

## BRAIN HEALTH SCORES REVEAL COGNITIVE AND CLINICAL DIFFERENCES WITHIN AGE BINS ACROSS MULTIPLE COHORTS

Across cohorts, BHSs negatively correlated with age (combined r=−0.32; Fig. 5C and Figs. S9 and S10). When participants were stratified by observed cognitive scores (top vs. bottom quartiles), the top quartile consistently showed higher BHSs, often reaching statistical significance within age subgroups (Fig. 5A). Similarly, subjects with mood disorders or neurocognitive impairment had lower scores than those without such diagnoses (Fig. 5B), with neurocognitive conditions showing the lowest scores overall.

## INDIVIDUAL-LEVEL INSPECTION

Figure S1 (cognition) and Figure S12 (dementia) illustrate sample sleep spectrograms and salience maps for representative individuals, highlighting how the model processes EEG features and where misclassifications can occur.

Notably, salience maps emphasize two distinct frequency bands during NREM sleep stages (N2 and N3): Frequencies equal to or below 1 Hz predominantly show negative salience, suggesting that increased slow oscillatory activity may predict lower cognitive scores, whereas positive salience appears within the 1-to-4-Hz band, associating increased delta activity with higher predicted cognition. Salience within the spindle frequency range (11–16 Hz) is less clearly discernible, potentially due to methodologic limitations of these salience maps. Fragmented sleep episodes (frequent transitions to N1 or Wake) show increased salience in the theta–alpha range. REM sleep periods demonstrate more dispersed salience, consistent with REM's characteristically heterogeneous EEG patterns.

## BRAIN HEALTH SCORE PREDICTS ALL-CAUSE MORTALITY

In age-adjusted Cox models, a one-SD increase in BHS predicted lower mortality in MrOS (hazard ratio, 0.69; P<0.0001; C-index, 0.691) and MGH (hazard ratio, 0.65; P<0.0001; C-index, 0.806), outperforming conventional EEG metrics including REM fraction and spindle density (MrOS P<0.001; MGH P=0.04 for REM fraction); Figure 6. In FHS, hazard ratios were not significant (hazard ratio, 0.85, P=0.28), but discrimination remained high (C-index, 0.755), consistent with limited event rates. Adding BHS to age improved discrimination (ΔC-index≈+0.02), model fit, and risk stratification, halving 5-year mortality risk between the lowest and highest BHS quartiles and substantially improving individual risk assignment.

In an independent BIDMC cohort (n=6483), BHS again predicted mortality (hazard ratio, 0.76; P<0.0001; C-index, 0.62), exceeding REM fraction (% REM hazard ratio, 0.91; P=0.02) and spindle density (hazard ratio, 0.80; P<0.0001). Discrimination improved from 0.622 (age only) to 0.646 (age+BHS), with better model fit and a 30% net improvement in risk reclassification, confirming external generalization.

## COMPARISON WITH THE EEG-BASED BRAIN AGE INDEX

In the independent BIDMC cohort, both the BHS and EEG-based BAI predicted mortality in age-adjusted Cox models (BHS hazard ratio, 0.76 per SD; P<0.0001; BAI hazard ratio, 1.08; P=0.049), but only BHS provided incremental value beyond the other: Adding BHS to an age+BAI model significantly improved fit (P<0.0001), whereas adding BAI to age+BHS did not (P=0.42). BHS also achieved the highest discrimination, including the top C-index and strongest 5-year time-dependent performance. Across MrOS, MGH, and FHS, both indices showed partial complementarity (Supplementary Appendix), indicating that while both carry prognostic signals, BHS captures additional mortality-relevant EEG features beyond age regression–based approaches. For completeness, cognition and disease comparisons are provided in Table S4.

## MULTIDIMENSIONAL LATENT SPACE

UMAP projections of the latent space (Fig. S2) reveal structured gradients and clusters consistent with quantitative results. Individuals with higher BHSs cluster in regions characterized by longer, less fragmented sleep, higher REM fraction, and greater delta and spindle activity across a broad age range, whereas lower scores align with shorter, fragmented sleep, reduced REM and spindle

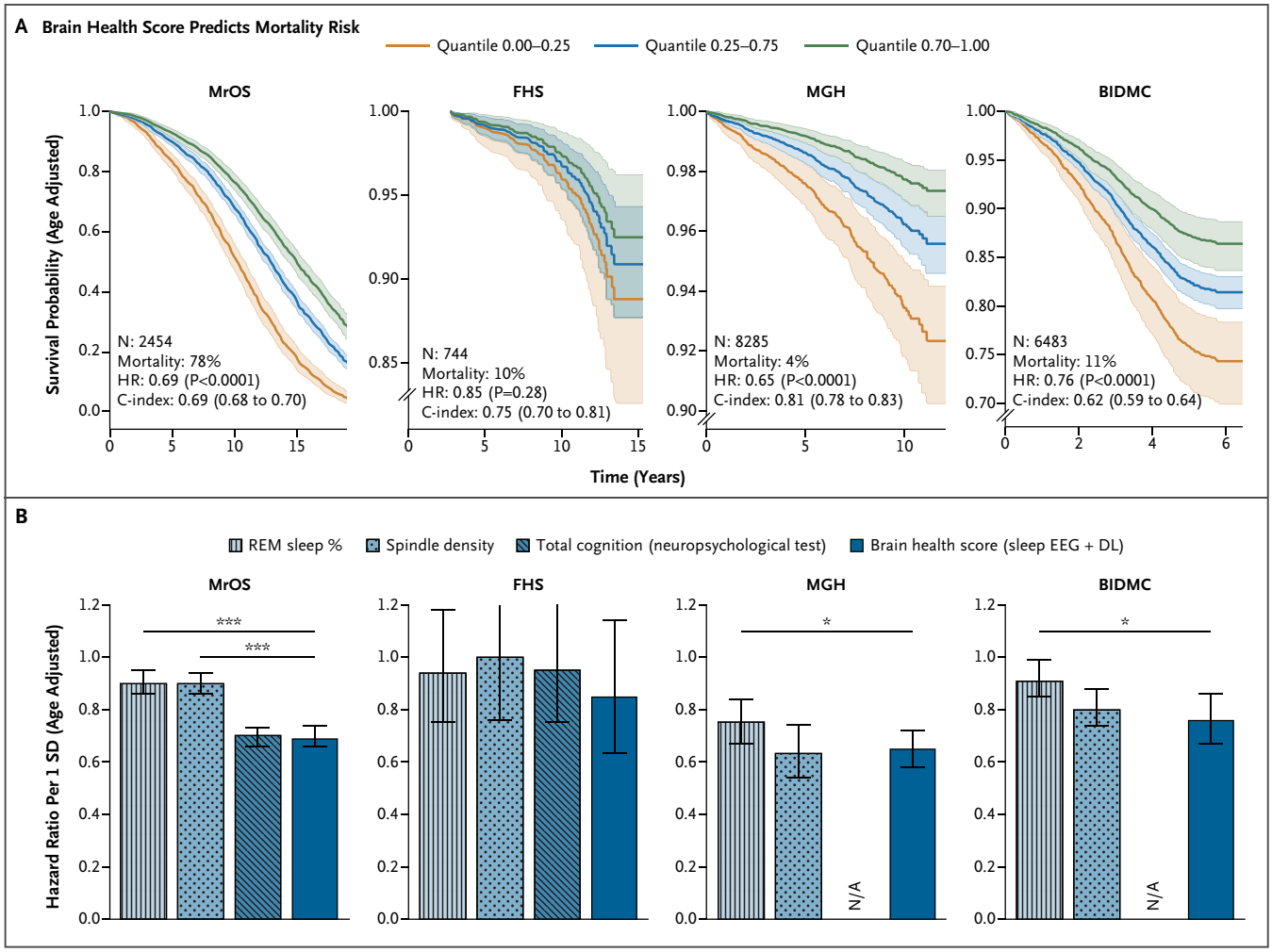


Figure 6. The Deep Learning–Derived Brain Health Score Stratifies Mortality Risk across Cohorts.

Panel A shows survival curves for the Osteoporotic Fracture in Men study (MrOS), Framingham Heart Study (FHS), Massachusetts General Hospital (MGH), and Beth Israel Deaconess Medical Center (BIDMC) cohorts, grouped by quantiles of the brain health score. Age-adjusted Cox models show significant associations between higher brain health scores and reduced mortality in MrOS (hazard ratio, 0.69; P<0.0001; C-index, 0.69), MGH (hazard ratio, 0.65; P<0.0001; C-index, 0.81), and BIDMC (hazard ratio, 0.76; P<0.0001; C-index, 0.62). In FHS, the hazard ratio did not reach significance (hazard ratio, 0.85; P=0.28), likely reflecting the cohort's low event rate and healthier profile; however, discrimination remained high (C-index, 0.75). Survival curves were generated by fitting a Cox model within each cohort using age and brain health score as covariates, followed by stratification into quantile-based risk groups. Panel B presents comparisons against traditional electroencephalography metrics (rapid-eye-movement fraction, spindle density) and neuropsychological test scores, which show that the brain health score provides stronger prognostic information. Error bars represent 95% confidence intervals. BIDMC denotes Beth Israel Deaconess Medical Center; DL, deep learning; EEG, electroencephalography; N/A, data not collected for that cohort; REM, rapid eye movement; and SD, standard deviation.

activity, lower delta power, and older age. Beyond simple gradients, the embedding captures multidimensional phenotypes — for example, separating individuals with similar delta power by spindle density and fragmentation — providing internal validation that the latent space integrates multiple EEG features into clinically meaningful representations.

## Discussion

In this study, we developed a multicohort deep learning framework that learns a brain health latent space and a composite BHS directly from overnight sleep EEG. This representation robustly predicted cognitive performance,

disease status, and all-cause mortality across multiple cohorts. Compared with demographic baselines and classic machine learning models using expert-defined EEG features, the BHS consistently improved predictive performance. Correlations with cognition increased from near zero under demographic-only models to moderate values (up to $r \approx 0.40$), disease classification improved from chance-level areas under the curve ($\sim$0.50–0.55) to around 0.70–0.75, and survival analyses showed substantially reduced mortality risk (hazard ratios as low as 0.65 per SD), outperforming conventional sleep EEG metrics. Together, these findings support the central hypothesis that data-driven analysis of sleep EEG yields a more informative and clinically relevant biomarker of brain health than traditional sleep measures.

A major strength of this work is the scale and diversity of the assembled dataset, which — to our knowledge — represents the largest multicohort resource linking full-night EEG to cognition, disease, and mortality. This breadth enabled a comprehensive assessment of both associative and predictive relationships between sleep EEG and brain health outcomes. The results underscore that integrated representations of sleep physiology capture more variance in brain health than individual sleep metrics considered in isolation, supporting a shift from univariate markers toward multidimensional EEG-derived phenotypes.

Predictive gains differed across outcome domains. Improvements over demographics were modest for fluid cognition but more pronounced for crystallized cognition and disease outcomes. This pattern likely reflects differences in biologic stability and measurement alignment. Fluid cognitive abilities fluctuate and are influenced by acute state factors, while both EEG and cognitive testing represent single, potentially asynchronous snapshots. In contrast, crystallized abilities accumulate over decades and may align more closely with long-term neurophysiological integrity encoded in sleep oscillations. Disease end points, particularly chronic conditions, are likely to imprint more stable and discriminative signatures onto sleep EEG, explaining the stronger classification performance observed for these outcomes.

Operating-point analyses highlighted the clinical flexibility of the framework. Lower specificity thresholds yielded higher sensitivity, consistent with screening-oriented use, whereas stricter thresholds improved precision, appropriate for confirmatory settings. These trade-offs demonstrate that EEG-based biomarkers derived from this framework can be tuned to different clinical contexts, although such applications remain exploratory.

Saliency and latent-space analyses provided insight into the learned representations. The model differentially weighted adjacent frequency bands, assigning negative salience to very slow oscillations ($\leq$1 Hz) and positive salience to delta power (1–4 Hz), echoing recent findings that these features may relate differently to cognitive health and disease risk. Visualization of the latent space revealed coherent organization by known physiological markers such as sleep-stage proportions and spindle density, while performance gains suggest additional information not captured by conventional features. Although interpretability remains limited, these results indicate that the latent space reflects physiologically meaningful structure rather than arbitrary patterns.

Methodologically, this work bridges conventional supervised EEG modeling and emerging self-supervised approaches. Traditional EEG machine learning relies on dense outcome labels, whereas self-supervised models learn general physiological structure from unlabeled data. Our approach combines these paradigms by jointly optimizing clinically meaningful outcomes (cognition and disease) alongside auxiliary physiological tasks available across cohorts. This hybrid strategy stabilizes training, encourages physiologically grounded representations, and directs the latent space toward health-relevant dimensions. The scale and heterogeneity of contemporary sleep cohorts now make such joint optimization feasible.

Within the broader literature on multivariate sleep EEG biomarkers, brain age modeling has received particular attention. Brain age indices — derived by predicting chronologic age from EEG and examining deviations — have been associated with dementia and mortality risk.[40,51,52] Prior work has also explored two-stage deep learning approaches[53] and classic machine learning using handcrafted sleep features, with mixed gains over demographic baselines.[28,29,38] In contrast, our framework directly optimizes clinically meaningful outcomes rather than age; operates end-to-end on raw overnight EEG; and jointly models cognition, disease, and mortality. Our results indicate that the BHS performs favorably relative to brain age indices while capturing partially complementary information, suggesting that outcome-optimized and age regression approaches reflect overlapping but distinct physiological dimensions.

External validation in an independent cohort further supported the robustness of the approach. Despite differences in population characteristics and acquisition protocols, the BHS reproduced mortality associations, discrimination metrics, and risk stratification patterns. This finding is encouraging given the susceptibility of deep learning

models to covariate shift, although broader external validation remains essential.

Several limitations warrant consideration. It remains unclear how much information about long-term brain health is fundamentally encoded in a single-night sleep EEG recording, and the moderate effect sizes observed likely reflect both biological limits and practical constraints of current datasets. Most analyses relied on single-night EEG despite known night-to-night variability, while cognitive and disease outcomes were often measured months or years apart, introducing noise from heterogeneous tests and electronic health records. Interpretability also remains incomplete, and aside from one external cohort, most evaluations relied on internal cross-validation. Together, these factors suggest that the reported results represent lower-bound estimates of the predictive value of sleep EEG.

Finally, this framework should be viewed as a foundation rather than an end point. Although sleep EEG provides a uniquely information-dense and scalable signal, the latent-space architecture is inherently compatible with multimodal expansion, including neuroimaging, blood-based biomarkers, genomics, and wearable or behavioral data. Integrating such modalities within a shared representation may yield more accurate, robust, and mechanistically interpretable models of brain health.

In summary, we present an end-to-end, data-driven framework that transforms overnight sleep EEG into a multidimensional representation of brain health and a clinically interpretable composite score. This approach outperforms conventional sleep metrics and classic machine learning models in predicting cognition, disease, and mortality, while generalizing across cohorts. More broadly, the study demonstrates that sleep EEG contains rich, distributed markers of brain health and illustrates how modern AI methods can convert complex neurophysiological signals into integrative health measures with translational potential.

## Disclosures

Author disclosures and other supplementary materials are available at ai.nejm.org.

The analysis code and the trained sleep electroencephalography (EEG) brain health latent-space model are available at https://github.com/bdsp-core/philosophers-stone and https://bdsp.io/content/wbsrgm4ol-jc1bpksbvro. A convenient application programming interface allows users to run the model on their own sleep EEG data, returning both the 1024-dimensional latent representation and the single brain health score. To facilitate secondary analyses and replication, we also provide a spreadsheet containing all brain health scores generated for the recordings used in this study (for cohorts with data that are publicly available or shareable under their data use agreements). These scores can be downloaded directly and used as ready-made inputs for research without needing to rerun the model.

Funding for this study was provided by National Institutes of Health (NIH) grants RF1NS120947, R01HL161253 and R01AG073410, RF1AG064312, R01NS126282, R01AG073598, R01NS131347, and R01NS130119. The Framingham Heart Study has been funded by NIH through core contracts administered by Boston University (AG062109, AG008122, AG016495, AG068753). The Multi-Ethnic Study of Atherosclerosis received funding from the National Heart, Lung, and Blood Institute under contracts N01-HC-95159 through N01-HC-95169. The Osteoporotic Fractures in Men Study was supported by NIH grants U01 AG027810, U01 AG042124, U01 AG042139, U01 AG042140, U01 AG042143, U01 AG042145, U01 AG042168, U01 AR066160, UL1 TR000128, R01 HL071194, R01 HL070848, R01 HL070847, R01 HL070842, R01 HL070841, R01 HL070837, R01 HL070838, and R01 HL070839. The Study of Osteoporotic Fractures received support from the National Institute on Aging under grant numbers R01 AG005407, R01 AR35582, R01 AR35583, R01 AR35584, R01 AG005394, R01 AG027574, and R01 AG027576. The Korean Genome and Epidemiology Study was supported by the National Human Genome Research Institute.

## Author Affiliations

[1] Department of Neurology, Beth Israel Deaconess Medical Center, Boston, MA, USA

[2] Harvard Medical School, Boston, MA, USA

[3] McCance Center for Brain Health, Massachusetts General Hospital, Boston, MA, USA

[4] Department of Neurology, Massachusetts General Hospital, Boston, MA, USA

[5] Department of Preventive Medicine, College of Medicine, Kangwon National University, Chuncheon, Gangwon, Republic of Korea

[6] Institute of Human Genomic Study, College of Medicine, Korea University, Seoul, Republic of Korea

[7] Emory University School of Medicine, Atlanta, GA, USA

[8] Departments of Psychiatry, Neurology, and Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, USA

[9] California Pacific Medical Center Research Institute, San Francisco, CA, USA

[10] Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, USA

[11] Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA

[12] Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, MA, USA

[13] Framingham Heart Study, Boston University School of Medicine, Boston, MA, USA

[14] Departments of Anatomy and Neurobiology, Neurology, and Medicine and Epidemiology, Chobanian and Avedisian School of Medicine and School of Public Health, Boston University, Boston, MA, USA

[15] Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

[16] Institute of Pharmacology and Toxicology, University of Zurich, Zurich, Switzerland

[17] Sleep and Health Zurich, University of Zurich, Zurich, Switzerland

[18] University Children's Hospital Zurich, University of Zurich, Zurich, Switzerland

[19] Department of Psychiatry and Behavioral Sciences, Stanford University, Palo Alto, CA, USA

[20] Biomedical Research Center, Korea University Ansan Hospital, Ansan, Republic of Korea

[21] Division of Pulmonary Critical Care and Sleep Medicine, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA

## References

1. Tononi G, Cirelli C. Sleep and the price of plasticity: from synaptic and cellular homeostasis to memory consolidation and integration. Neuron 2014;81:12-34. DOI: 10.1016/j.neuron.2013.12.025.

2. Cirelli C, Tononi G. Effects of sleep and waking on the synaptic ultrastructure. Philos Trans R Soc Lond B Biol Sci 2020;375:20190235. DOI: 10.1098/rstb.2019.0235.

3. Xie L, Kang H, Xu Q, et al. Sleep drives metabolite clearance from the adult brain. Science 2013;342:373-377. DOI: 10.1126/science.1241224.

4. Fultz NE, Bonmassar G, Setsompop K, et al. Coupled electrophysiological, hemodynamic, and cerebrospinal fluid oscillations in human sleep. Science 2019;366:628-631. DOI: 10.1126/science.aax5440.

5. Spiegel K, Leproult R, Van Cauter E. Impact of sleep debt on metabolic and endocrine function. Lancet 1999;354:1435-1439. DOI: 10.1016/S0140-6736(99)01376-8.

6. Somers VK, Dyken ME, Mark AL, Abboud FM. Sympathetic-nerve activity during sleep in normal subjects. N Engl J Med 1993;328:303-307. DOI: 10.1056/NEJM199302043280502.

7. Peever J, Fuller PM. The biology of REM sleep. Curr Biol 2017;27:R1237-48. DOI: 10.1016/j.cub.2017.10.026.

8. Irwin MR. Why sleep is important for health: a psychoneuroimmunology perspective. Annu Rev Psychol 2015;66:143-172. DOI: 10.1146/annurev-psych-010213-115205.

9. Djonlagic I, Aeschbach D, Harrison SL, et al. Associations between quantitative sleep EEG and subsequent cognitive decline in older women. J Sleep Res 2019;28:e12666. DOI: 10.1111/jsr.12666.

10. Djonlagic I, Mariani S, Fitzpatrick AL, et al. Macro and micro sleep architecture and cognitive performance in older adults. Nat Hum Behav 2021;5:123-145. DOI: 10.1038/s41562-020-00964-y.

11. Dykierek P, Stadtmüller G, Schramm P, et al. The value of REM sleep parameters in differentiating Alzheimer's disease from old-age depression and normal aging. J Psychiatr Res 1998;32:1-9. DOI: 10.1016/s0022-3956(97)00049-6.

12. Geiger A, Huber R, Kurth S, Ringli M, Jenni OG, Achermann P. The sleep EEG as a marker of intellectual ability in school age children. Sleep 2011;34:181-189. DOI: 10.1093/sleep/34.2.181.

13. Helfrich RF, Mander BA, Jagust WJ, Knight RT, Walker MP. Old brains come uncoupled in sleep — slow wave-spindle synchrony, brain atrophy and forgetting. Neuron 2018;97:221-230.e4. DOI: 10.1016/j.neuron.2017.11.020.

14. Himali JJ, Baril A-A, Cavuoto MG, et al. Association between slow-wave sleep loss and incident dementia. JAMA Neurol 2023;80:1326-1333. DOI: 10.1001/jamaneurol.2023.3889.

15. Lustenberger C, Wehrle F, Tüshaus L, Achermann P, Huber R. The multidimensional aspects of sleep spindles and their relationship to word-pair memory consolidation. Sleep 2015;38:1093-1103. DOI: 10.5665/sleep.4820.

16. Malhotra A, Loscalzo J. Sleep and cardiovascular disease: an overview. Prog Cardiovasc Dis 2009;51:279-284. DOI: 10.1016/j.pcad.2008.10.004.

17. Manoach DS, Pan JQ, Purcell SM, Stickgold R. Reduced sleep spindles in schizophrenia: a treatable endophenotype that links risk genes to impaired cognition? Biol Psychiatry 2016;80:599-608. DOI: 10.1016/j.biopsych.2015.10.003.

18. Pase MP, Himali JJ, Grima NA, et al. Sleep architecture and the risk of incident dementia in the community. Neurology 2017;89:1244-1250. DOI: 10.1212/WNL.0000000000004373.

19. Walker MP, Brakefield T, Morgan A, Hobson JA, Stickgold R. Practice with sleep makes perfect: sleep-dependent motor skill learning. Neuron 2002;35:205-211. DOI: 10.1016/s0896-6273(02)00746-8.

20. Wallace ML, Buysse DJ, Redline S, et al. Multidimensional sleep and mortality in older adults: a machine-learning comparison with other risk factors. J Gerontol Ser A 2019;74:1903-1909. DOI: 10.1093/gerona/glz044.

21. Ye EM, Sun H, Krishnamurthy PV, et al. Dementia detection from brain activity during sleep. Sleep 2023;46:zsac286. DOI: 10.1093/sleep/zsac286.

22. Younes M, Redline S, Peters K, et al. Normalized EEG power: a trait with increased risk of dementia. Sleep 2023;46:zsad195. DOI: 10.1093/sleep/zsad195.

23. Fernandez LMJ, Lüthi A. Sleep spindles: mechanisms and functions. Physiol Rev 2020;100:805-868. DOI: 10.1152/physrev.00042.2018.

24. Lucey BP, Wisch J, Boerwinkle AH, et al. Sleep and longitudinal cognitive performance in preclinical and early symptomatic Alzheimer's disease. Brain 2021;144:2852-2862. DOI: 10.1093/brain/awab272.

25. Winer JR, Mander BA, Helfrich RF, et al. Sleep as a potential biomarker of tau and β-amyloid burden in the human brain. J Neurosci 2019;39:6315-6324. DOI: 10.1523/JNEUROSCI.0503-19.2019.

26. LeCun Y, Bengio Y. Convolutional networks for images, speech, and time-series. In: Arbib MA, ed. The handbook of brain theory and neural networks. Cambridge, MA: MIT Press, 1995:255-258.

27. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: Proceedings of the 2017 Conference on Advances in Neural Information Processing Systems. Long Beach, CA: NeurIPS, 2017 (https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html).

28. Adra N, Dümmer LW, Paixao L, et al. Decoding information about cognitive health from the brainwaves of sleep. Sci Rep 2023;13:11448. DOI: 10.1038/s41598-023-37128-7.

29. Wei R, Ganglberger W, Sun H, et al. Linking brain structure, cognition, and sleep: insights from clinical data. Sleep 2024;47:zsad294. DOI: 10.1093/sleep/zsad294.

30. Zhang G-Q, Cui L, Mueller R, et al. The National Sleep Research Resource: towards a sleep data commons. J Am Med Inform Assoc 2018;25:1351-1358. DOI: 10.1093/jamia/ocy064.

31. Quan SF, Howard BV, Iber C, et al. The Sleep Heart Health Study: design, rationale, and methods. Sleep 1997;20:1077-1085 (https://pubmed.ncbi.nlm.nih.gov/9493915/).

32. Thomas RJ, Kim H, Maillard P, et al. Digital sleep measures and white matter health in the Framingham Heart Study. Explor Med 2021;2:253-267. DOI: 10.37349/emed.2021.00045.

33. Burke G, Lima J, Wong ND, Narula J. The Multiethnic Study of Atherosclerosis. Glob Heart 2016;11:267-268. DOI: 10.1016/j.gheart.2016.09.001.

34. Chen X, Wang R, Zee P, et al. Racial/ethnic differences in sleep disturbances: the Multi-Ethnic Study of Atherosclerosis (MESA). Sleep 2015;38:877-888. DOI: 10.5665/sleep.4732.

35. Stone KL, Blackwell TL, Ancoli-Israel S, et al. Sleep disturbances and risk of falls in older community-dwelling men: the outcomes of sleep disorders in older men (MrOS sleep) study. J Am Geriatr Soc 2014;62:299-305. DOI: 10.1111/jgs.12649.

36. Blackwell T, Yaffe K, Ancoli-Israel S, et al. Poor sleep is associated with impaired cognitive function in older women: the study of osteoporotic fractures. J Gerontol Ser A 2006;61:405-410. DOI: 10.1093/gerona/61.4.405.

37. Kim H, Thomas RJ, Yun C-H, et al. Association of mild obstructive sleep apnea with cognitive performance, excessive daytime sleepiness, and quality of life in the general population: the Korean Genome and Epidemiology Study (KoGES). Sleep 2017;40:zsx012. DOI: 10.1093/sleep/zsx012.

38. 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:e200225. DOI: 10.1212/CPJ.0000000000200225.

39. Sun H, Ganglberger W, Nasiri S, et al. The human sleep project (version 2.0). Brain Data Science Platform, November 1, 2023 (https://bdsp.io/content/hsp/2.0/).

40. Ye E, Sun H, Leone MJ, et al. Association of sleep electroencephalography-based brain age index with dementia. JAMA Netw Open 2020;3:e2017357. DOI: 10.1001/jamanetworkopen.2020.17357.

41. Orwoll E, Blank JB, Barrett-Connor E, et al. Design and baseline characteristics of the Osteoporotic Fractures in Men (MrOS) study — a large observational study of the determinants of fracture in older men. Contemp Clin Trials 2005;26:569-585. DOI: 10.1016/j.cct.2005.05.006.

42. Cattell RB. Theory of fluid and crystallized intelligence: a critical experiment. J Educ Psychol 1963;54:1-22. DOI: 10.1037/h0046743.

43. Sun H, Jia J, Goparaju B, et al. Large-scale automated sleep staging. Sleep 2017;40:zsx139. DOI: 10.1093/sleep/zsx139.

44. Luna. Luna: software for the analysis of sleep signal data. March 7, 2025 (https://zzz.bwh.harvard.edu/luna/).

45. Thomson DJ. Spectrum estimation and harmonic analysis. Proc of IEEE 1982;70:1055-1096. DOI: 10.1109/PROC.1982.12433.

46. Daubechies I. The wavelet transform, time-frequency localization and signal analysis. IEEE Trans Inf Theory 1990;36:961-1005. DOI: 10.1109/18.57199.

47. GitHub. OverLordGoldDragon/ssqueezepy: synchrosqueezing, wavelet transforms, and time-frequency analysis in Python. August 2, 2025 (https://github.com/OverLordGoldDragon/ssqueezepy).

48. Tu Z, Talebi H, Zhang H, et al. MaxViT: multi-axis vision transformer. September 9, 2022 (https://arxiv.org/abs/2204.01697). Preprint.

49. Paszke A, Gross S, Chintala S, et al. Automatic differentiation in PyTorch. NIPS 2017 Autodiff Workshop, October 28, 2017 (https://openreview.net/forum?id=BJJsrmfCZ).

50. Wightman R. PyTorch image models. New York, NY: Hugging Face, January 6, 2025 (https://github.com/huggingface/pytorch-image-models).

51. Sun H, Paixao L, Oliva JT, et al. Brain age from the electroencephalogram of sleep. Neurobiol Aging 2019;74:112-120. DOI: 10.1016/j.neurobiolaging.2018.10.016.

52. Paixao L, Sikka P, Sun H, et al. Excess brain age in the sleep electroencephalogram predicts reduced life expectancy. Neurobiol Aging 2020;88:150-155. DOI: 10.1016/j.neurobiolaging.2019.12.015.

53. Brink-Kjaer A, Leary EB, Sun H, et al. Age estimation from sleep studies using deep learning predicts life expectancy. NPJ Digit Med 2022;5:1-10. DOI: 10.1038/s41746-022-00630-9.