

Original Article

Better sleep now, better cognition later? Predicting cognitive function using a machine learning-based sleep EEG brain health score

Francesca R. Marino^{1,*}, Wolfgang Ganglberger^{1,2,3,4}, Haoqi Sun^{1,2,3,4}, Yulin Liu^{1,5}, Huitong Ding^{1,5}, Robert J. Thomas^{1,5,6}, Rhoda Au^{1,5,7,8,9,10} and M. Brandon Westover^{1,2,3,4}

¹Department of Anatomy and Neurobiology, Boston University Chobanian and Avedisian School of Medicine, Boston, MA, United States, ²Department of Neurology, Beth Israel Deaconess Medical Center, Boston, MA, United States, ³Harvard Medical School, Boston, MA, United States, ⁴Department of Neurology, Massachusetts General Hospital, Boston, MA, United States, ⁵The Framingham Heart Study, Boston University Chobanian and Avedisian School of Medicine, Boston, MA, United States, ⁶Department of Medicine, Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA, United States, ⁷Department of Medicine, Boston University Chobanian and Avedisian School of Medicine, Boston, MA, United States, ⁸Department of Epidemiology, Boston University School of Public Health, Boston, MA, United States, ⁹Department of Neurology, Boston University Chobanian and Avedisian School of Medicine, Boston, MA, United States and ¹⁰Sloan Epidemiology Center, Boston University Chobanian and Avedisian School of Medicine, Boston, MA, United States

*Corresponding author: Francesca R. Marino, Department of Anatomy and Neurobiology, Boston University Chobanian and Avedisian School of Medicine, 72 E Concord Street, Suite 912-C, Boston, MA 02118, United States. Email: fmarino1@bu.edu.

Abstract

Study Objectives: Sleep state electrocortical activity measured using electroencephalograms (EEG) is linked with cognitive function and dementia risk. The rich information in overnight EEG can be condensed into integrated scores using machine learning. The Brain Health Score (BHS) was derived in prior work using an end-to-end deep learning model trained to jointly optimize higher cognitive function and lower disease risk from raw EEG. However, it is unknown whether the BHS predicts future neuropsychological (NP) performance.

Methods: This study included 426 Framingham Heart Study (FHS) Generation 2 or Omni 1 participants with BHS values from in-home polysomnography in mid-to-late life, and subsequent digital clock drawing (dCDT) and NP testing an average of 12.6 years later. Linear regression models estimated associations between BHS and dCDT, memory, language, or executive function scores, adjusting for age, sex, race, education, smoking, body mass index, and FHS cohort. To enable comparisons across outcomes, the independent variables were centered and rescaled, and then the model was refitted to generate standardized estimates.

Results: Participants were on average 56 years at sleep assessment, 55% female, and 86% non-Hispanic White. Each 1-SD higher BHS was associated with higher dCDT, memory, language, and executive function scores (dCDT: $\beta = 0.16$, 95% CI = 0.06–0.26; memory: $\beta = 0.13$, 95% CI = 0.03–0.23; language: $\beta = 0.13$, 95% CI = 0.03–0.23; executive: $\beta = 0.10$, 95% CI = 0.01–0.19).

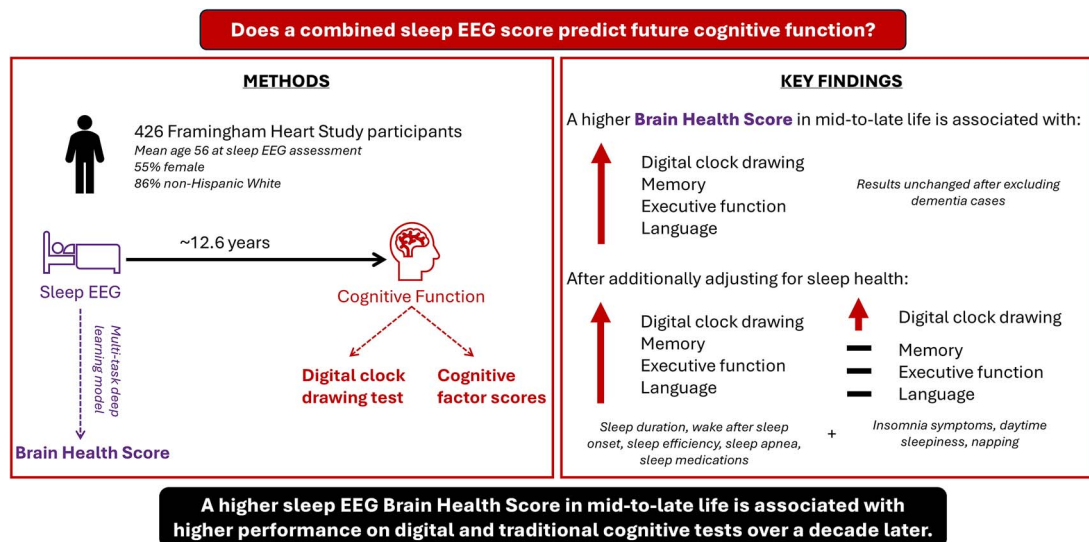
Conclusions: Higher BHS in mid-to-late life was associated with better digital and traditional NP performance more than a decade later. These findings support the potential of EEG-derived, data-driven scores as a biomarker of future cognitive health.

Key words: sleep; EEG; cognition; digital; risk prediction

Submitted: 24 October, 2025; Revised: 26 February, 2026; Accepted: 5 April, 2026

© The Author(s) 2026. Published by Oxford University Press on behalf of Sleep Research Society. All rights reserved. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Graphical Abstract



Statement of Significance

This study is among the first to demonstrate associations between a sleep-derived brain health score in mid-to-late life and subsequent performance on both digital and traditional cognitive tests in late-life. These results suggest the potential utility of using this data-driven sleep score as a biomarker of future cognitive function. Moreover, the use of sleep electroencephalogram data, which can be assessed in the free-living environment, adds to the feasibility of using this sleep score for remote monitoring. Future studies with longer follow-up periods in more diverse samples are needed to determine whether this sleep score predicts the rate of cognitive decline or risk of dementia, as well as to evaluate whether sleep interventions can mitigate cognitive risk.

Introduction

Sleep is important for maintaining cognitive function into later-life. Indeed, poor sleep quality is associated with higher burden of amyloid- β [1, 2] and tau [2], smaller brain volumes [3, 4], and faster rates of cognitive decline [5, 6]. Sleep is also prospectively linked with risk of cognitive impairment and dementia [2, 6, 7]. Studies to date often rely on either self-reported or actigraphy-assessed sleep. However, electroencephalograms (EEGs), which measure the electrical activity of the brain during sleep, provide additional information on sleep physiology and pathology. These EEG-based sleep measures have been linked with neuropathology [8] and early changes in cognitive function [9], raising the possibility of using these measures as biomarkers for early dementia detection.

With technological advancements, accurate assessment of sleep in the free-living environment is now possible. For example, in-home polysomnography (PSG) allows for remote measurement of sleep EEGs [10]. Beyond individual EEG features, data-driven approaches can distill the rich information in overnight EEG into integrated scores of brain health. Earlier work using classical machine learning showed such scores relate to cognitive function, though with limited accuracy [11]. More recently, a multi-cohort study applied an end-to-end deep learning framework directly to raw EEG to derive the Brain Health Score (BHS), jointly optimized for higher cognitive function and lower disease risk [12]. However, it remains unknown whether the BHS is associated with future cognitive performance.

Digital cognitive assessments have also become increasingly common. The digital clock drawing test (dCDT) is a digital version of the traditional assessment that has been linked with Alzheimer's disease (AD) genetic risk [13] and neuropathology

[14, 15], neuropsychological (NP) test performance [14], and mild cognitive impairment (MCI) status [16]. dCDT performance also accurately differentiates individuals living with MCI or mild dementia from cognitively intact controls [14], as well as amyloid- β status among cognitively normal populations [14, 15]. However, it has yet to be determined whether the BHS is associated with dCDT performance. To understand its potential utility as a biomarker of early cognitive changes, there is a need to evaluate associations between the BHS and sensitive indicators of cognitive function, such as the dCDT.

This study aimed to evaluate whether the BHS in mid-to-late life was associated with cognitive function on average 12.6 years later. A secondary aim was to compare the strength of associations between this score and either digital (dCDT) or traditional (NP) cognitive performance. We hypothesized that higher mid-to-late life BHS, a score trained to reflect better cognition and lower disease risk, would relate to superior late-life performance across both digital (dCDT) and traditional (NP) cognitive domains, thereby testing the generalizability of a sleep-EEG brain-health phenotype to future cognition.

Materials and methods

Study population

The Framingham Heart Study (FHS) aims to understand risk factors for cardiovascular and cerebrovascular disease. Beginning in 1948, 5209 adults living in Framingham, Massachusetts, were enrolled in the Original Cohort [17]. In 1971, the offspring and offspring spouses of the Original Cohort were recruited to join the Generation 2 cohort (Gen 2, $n = 5124$) [18]. Then, in 1994, FHS

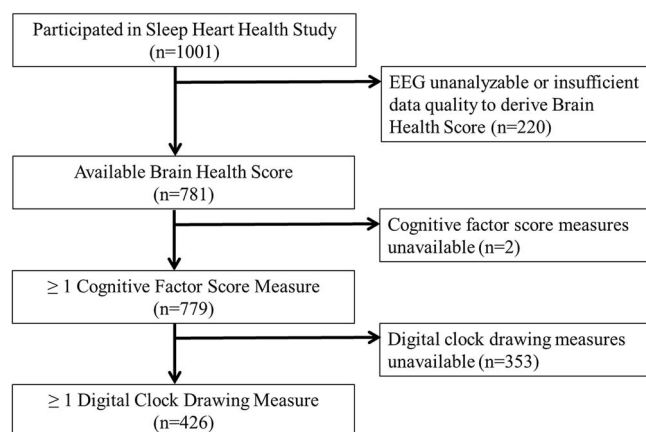


Figure 1. Sample flow chart.

expanded enrollment to include participants from more racially and ethnically diverse populations to form the Omni 1 cohort ($n = 506$). Participants in the Gen 2 and Omni 1 cohorts complete examinations on average every 4 years.

Beginning in 1994, a subset of FHS participants enrolled in the multi-site Sleep Heart Health Study (SHHS), which aimed to evaluate whether obstructive sleep apnea or sleep-disordered breathing were risk factors for cardiovascular disease [10]. Participants from six ongoing cohort studies, including FHS, completed in-home PSG and were assessed for cardiovascular events. Participants were excluded if they had sleep apnea treated with continuous positive airway pressure or oral devices, used home oxygen therapy, or had a tracheostomy [10]. Visit 1 occurred from 1995–2003 and Visit 2 from 2001–2003. The current study includes FHS participants who attended at least one visit ($n = 1001$), had analyzable EEG and sufficient data quality to derive BHS ($n = 781$), and data on both NP factor scores ($n = 779$) and the dCDT ($n = 426$) (Figure 1). All procedures were approved by the Boston University Medical Center Institutional Review Board. Written informed consent was obtained from all participants.

Sleep EEG-derived Brain Health Score

Sleep EEGs were collected during in-home PSG as part of SHHS visits [10]. In a prior study, a multi-task deep learning model (“Philosopher’s Stone”) was trained on full-night sleep EEG from a single central derivation (C4-M1) to jointly predict cognitive performance, disease status, and auxiliary sleep-related targets across multiple cohorts. Each overnight EEG was processed through a shared encoder yielding a 1024-dimensional latent representation supporting multiple supervised prediction heads [12].

Cognitive outcomes and disease labels were predicted directly from this representation using task-specific regression and classification heads. A composite BHS was additionally learned via a linear projection from the latent space, trained with a composite objective favoring higher cognitive performance and lower disease burden. Domain-specific cognitive scores for Total Cognition (TC), Fluid Cognition (FC), and Crystallized Cognition (CC) were obtained from corresponding prediction heads. Higher values indicate better brain health (BHS: higher cognition, lower disease risk; TC: higher total cognition; FC: higher fluid cognition; CC: higher crystallized cognition) [12]. For the present analysis, the primary exposure was the composite BHS, and the domain-specific sub-scores were examined as secondary exposures. When multiple SHHS visits were available, BHS values were averaged

across visits. Full details for the derivation of the BHS and domain-specific sub-scores are described elsewhere [12].

Digital clock drawing test

Starting in 2011, FHS began administering the dCDT. In brief, participants use a digital pen to complete command and copy test conditions on a paper test form, consistent with the traditional clock drawing test [19]. For the command condition, participants are asked to “draw the face of a clock, put in all numbers, and set the hands to 10 after 11.” For the copy condition, participants are asked to replicate a clock showing the time as 10 after 11. During both conditions, the digital pen records multiple features of drawing performance [13, 14]. The primary outcome for this analysis was the dCDT summary score, which ranges from 0 to 100 with higher scores reflecting better performance. dCDT performance is scored using the DCTclock™ algorithm (Linus Health). This algorithm uses machine learning methods to integrate all test performance information from the digital pen into one summary score. The full details on this algorithm are described elsewhere [14].

Neuropsychological factor scores

We also examined performance on traditional cognitive assessments, including memory, executive function, and language factor scores. Participants completed an NP test battery administered by trained examiners every 2–3 years between FHS core study exams. For the current analysis, we selected NP factor scores measured at the same visit as the dCDT. In a previous analysis focused on cross-study harmonization, cognitive domain factor scores were derived across multiple FHS exams [20]. Each NP test was assigned to a single cognitive domain, then factor scores for each domain were estimated using separate bi-factor confirmatory factor analysis models. NP tests calibrated from other studies as part of the cross-study harmonization were included as fixed item parameters, and NP tests only administered in FHS were freely estimated. Full methods for generating the FHS cognitive domain factor scores are described elsewhere [21].

Covariates

Potential confounders were measured at the time of sleep assessment and included age (years), sex (female vs. male), race (White vs. Black vs. other), education (some high school or high school graduate vs. some college vs. college graduate), body mass index (BMI; kg/m^2), current smoking status (yes vs. no), cohort (Gen 2 vs. Omni 1), sleep duration (< 6 vs. $6-9$ vs. ≥ 9 hours), wake after sleep onset (minutes), sleep efficiency (%), sleep apnea (< 10 apnea/hypopnea events/hour vs. $10-19$ events/hour vs. $20-29$ events/hour vs. ≥ 30 events/hour), self-reported use of “sleep pills” or benzodiazepines (yes vs. no), self-reported insomnia symptoms (never, rarely, or sometimes vs. often or almost always), self-reported daytime sleepiness (never, rarely, or sometimes vs. often or almost always), and self-reported number of naps ≥ 5 minutes per month (0 vs. ≥ 1). Due to small sample sizes, participants who self-reported as Asian, American Indian, Native Hawaiian, Pacific Islander, or preferred not to answer were combined into the other race category. Participants who endorsed often or almost always, for either waking up too early in the morning and unable to resume sleep or waking up during the night and having difficulty resuming sleep, were classified as having insomnia symptoms. Sleep duration, wake after sleep onset, sleep efficiency, and sleep apnea were measured from PSG data. Potential mediators were accounted for in a sensitivity analysis and included: self-reported cardiovascular diseases, including hypertension, diabetes, dyslipidemia, coronary heart disease, stroke, and atrial fibrillation, and

depressive symptoms, defined as scores >16 on the Center for Epidemiologic Studies Depression Scale [22]. Dementia status was ascertained by expert consensus according to the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders [23]. Participants with dementia were excluded in a sensitivity analysis.

Statistical analysis

Sample characteristics at the time of sleep assessment were summarized as means and SDs or as counts and proportions for continuous or categorical variables, respectively. Differences in sample characteristics were evaluated overall and within each FHS cohort.

Separate linear regression models estimated associations between the BHS and each cognitive outcome: (1) dCDT score; (2) memory factor score; (3) executive function factor score; or (4) language factor score. Cognitive outcomes were assessed on average 12.6 years after sleep assessment. Models were adjusted for age, sex, race, education, cohort, BMI, and smoking status. To enable comparison across outcomes, the independent variables were centered and rescaled, and then the model was refitted to generate standardized estimates. These standardized estimates were interpreted as the SD change in each cognitive outcome associated with each 1-SD higher BHS. Confidence intervals were bootstrapped over 1000 iterations.

In sensitivity analyses, we compared associations after excluding $n=8$ participants with dementia between sleep and cognitive assessments to determine whether individuals with greater cognitive impairment were driving the results. We also additionally adjusted for several covariates. First, we accounted for the difference in time between sleep and cognitive assessments. Second, we adjusted for NP factor scores at the time of sleep assessment. We could not account for dCDT performance at the time of sleep assessment because these data were not collected until 2011. Third, we evaluated associations after accounting for potential mediating variables, including depressive symptoms and cardiovascular diseases. Finally, we additionally adjusted for potential confounders of sleep health at the time of sleep assessment, including sleep duration, wake after sleep onset, sleep efficiency, sleep apnea, and use of sleep medications. We also adjusted for self-reported insomnia symptoms, daytime sleepiness, and napping among the $n=414$ participants with available data.

We also conducted several secondary analyses. First, we changed the primary exposure from the average BHS to the score at visit 1 or visit 2. We included the scores at visit 1 and visit 2 in separate models, and then together in the same model. Then, we tested whether including both timepoints significantly contributed to the model using likelihood ratio tests for nested models. Second, we explored associations using TC, FC, or CC as the primary exposure.

Statistical significance was defined as $\alpha < 0.05$. All analyses were performed using Stata version 18.0 (StataCorp LLC, College Station, Texas). Figures were generated using R version 4.4.1 (<https://CRAN.R-project.org>).

Results

Sample characteristics

Of the 426 participants in this study, the mean age at the time of sleep assessment was 55.9 (SD: 8.2) years, 55% were female, and 86% were non-Hispanic White. Participants in the Gen 2 cohort were older, more likely to be non-Hispanic White, had lower

educational attainment, and higher cognitive function than those in the Omni 1 cohort (Table 1).

Sleep EEG-derived BHS and cognitive function

A higher average BHS was associated with better cognitive function 12.6 (SD: 2.6) years later. After full covariate adjustment, 1-SD higher BHS was associated with 0.16-SD higher dCDT, 0.14-SD higher memory, 0.13-SD higher language, and 0.11-SD higher executive function scores (dCDT: β : 0.16, 95% CI: 0.06-0.26; memory: β : 0.14, 95% CI: 0.04-0.24; language: β : 0.13, 95% CI: 0.03-0.23; executive function: β : 0.11, 95% CI: 0.02-0.20). The magnitude of the associations was similar for all cognitive outcomes, but highest for the dCDT (Figure 2).

BHS and cognitive function, excluding dementia cases

Results were robust when excluding $n=8$ participants with dementia during the study period. The magnitude of the association was slightly attenuated for each cognitive outcome, but findings remained statistically significant (Table 2).

BHS and cognitive function, using BHS from different visits

We then examined associations using the BHS at the first SHHS visit, last SHHS visit, or both SHHS visits as the primary exposures. The average time between the first or last SHHS visits and cognitive testing was 16.6 (SD: 1.9) or 11.7 (SD: 1.5) years, respectively. The first visit BHS was associated with higher dCDT, memory, executive function, and language scores. The magnitude of the associations was slightly lower compared to the average BHS across available visits. Of the 332 participants with sleep assessments at both visits, higher BHS at the last visit was associated with higher memory and language factor scores. There were no statistically significant associations between the BHS and any cognitive outcome when the BHS from both the first and last visits were included in the same model. Including the BHS from both visits only contributed to model fit for the dCDT, compared to using the last visit BHS alone (Table 3).

BHS and cognitive function, adjusting for time lag, cognitive function at sleep assessment, or depressive symptoms and cardiovascular diseases

First, results were robust when adjusting for the time lag between sleep and cognitive assessments (Supplemental Table 1). Second, findings with memory were robust when adjusting for memory factor scores at the time of sleep assessment. However, findings with executive function and language were attenuated when accounting for baseline performance (Supplemental Table 2). Finally, results were robust when adjusting for potential mediators, including depressive symptoms, hypertension, diabetes, dyslipidemia, coronary heart disease, stroke, and atrial fibrillation (Supplemental Table 3).

BHS and cognitive function, adjusting for sleep health

117 participants in our sample had obstructive sleep apnea as defined by >10 apnea/hypopnea events/hour, of which 12% ($n=50$) had mild sleep apnea (10-19 events/hour), 5% ($n=20$) had moderate sleep apnea (20-29 events/hour), and 11% ($n=47$) had severe sleep apnea (≥ 30 events/hour). Results were robust after additionally adjusting for sleep duration, wake after sleep onset, sleep efficiency, sleep apnea, and sleep medications, although the magnitude of the coefficients were attenuated. Among the $n=414$ participants with available data, results with dCDT were robust after additionally accounting for insomnia symptoms, daytime

Table 1. Sample characteristics overall and by FHS cohort

	Overall (n = 426)	Generation 2 (n = 351)	Omni 1 (n = 75)
Age (years) at sleep assessment, mean (SD)	55.9 (8.2)	56.8 (7.9)	51.5 (8.1)
Age (years) at cognitive assessment, mean (SD)	70.4 (8.1)	71.3 (7.8)	66.1 (8.1)
Female, N (%)	233 (55%)	190 (54%)	43 (57%)
Non-Hispanic White, N (%)	367 (86%)	347 (99%)	20 (27%)
College education, N (%)	174 (41%)	130 (37%)	44 (59%)
Body mass index (kg/m ²), mean (SD)	27.7 (5.0)	27.8 (4.9)	27.5 (5.5)
Current smoker, N (%)	50 (12%)	44 (13%)	6 (8%)
Framingham risk score category, N (%)			
Low (< 10%)	256 (60%)	203 (58%)	53 (71%)
Moderate (10-19%)	107 (25%)	91 (26%)	16 (21%)
High (≥ 20%)	63 (15%)	57 (16%)	6 (8%)
BHS, mean (SD)*	0.39 (0.57)	0.38 (0.57)	0.42 (0.56)
Digital clock drawing score, mean (SD)	70.2 (23.4)	71.0 (23.6)	66.7 (22.3)
Memory factor score, mean (SD)	0.43 (0.53)	0.44 (0.53)	0.39 (0.52)
Executive function factor score, mean (SD)	0.08 (0.61)	0.11 (0.61)	-0.04 (0.59)
Language factor score, mean (SD)	0.40 (0.60)	0.42 (0.61)	0.28 (0.56)
Time lag (years), mean (SD) [†]	12.6 (2.6)	12.4 (2.6)	13.3 (2.2)

*Average sleep-derived BHS across two sleep assessments [†]Difference in time (years) between sleep assessment and cognitive assessment

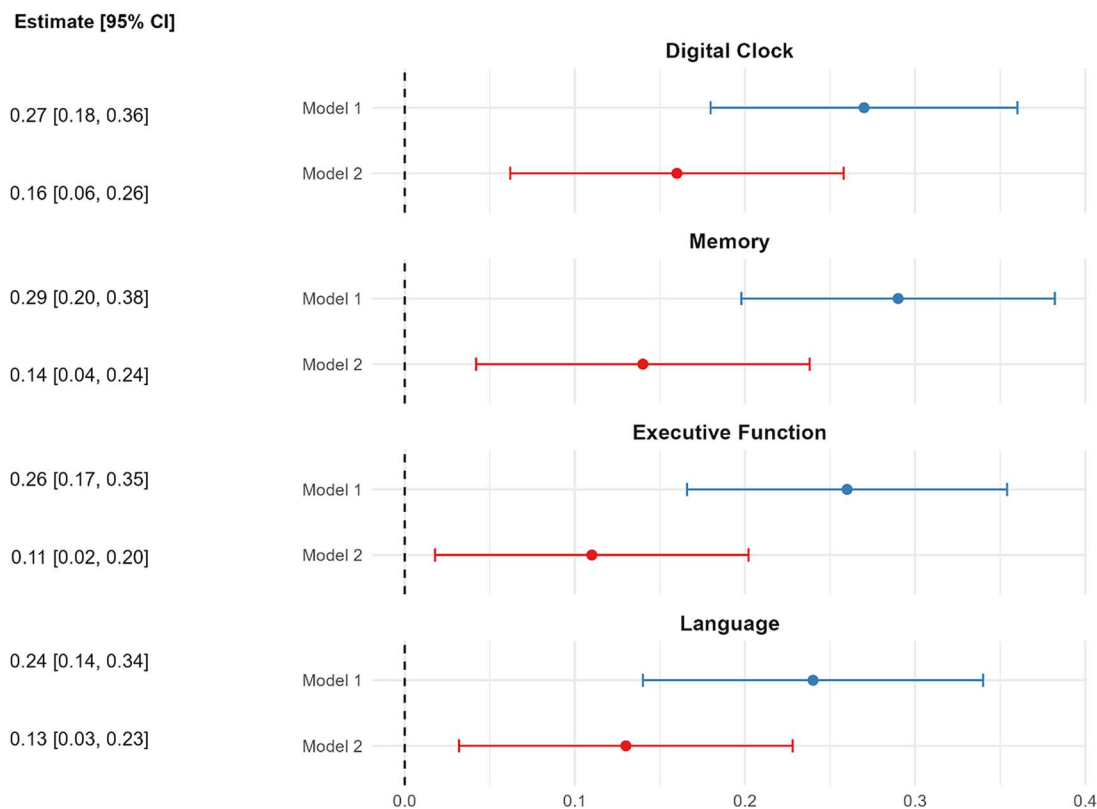


Figure 2. Associations between average sleep-derived BHS and cognitive function. Higher sleep-derived BHS was associated with better cognitive performance more than a decade later. In 426 FHS participants, each 1-SD higher mid-to-late life BHS predicted higher digital clock drawing (dCDT), memory, executive function, and language scores at follow-up (~12.6 years later). Associations remained after adjustment for age, sex, race, education, body mass index, smoking, and cohort (model 2). Estimates are standardized β 's with 95% confidence intervals from 1000 bootstrap iterations.

Table 2. Associations between average sleep-derived BHS and cognitive function, excluding $n = 8$ prevalent dementia cases

	Digital clock β (95% CI)*	Memory β (95% CI)*	Executive function β (95% CI)*	Language β (95% CI)*
Model 1	0.25 (0.16, 0.34)	0.27 (0.18, 0.36)	0.25 (0.16, 0.34)	0.23 (0.13, 0.33)
Model 2	0.15 (0.05, 0.25)	0.11 (0.02, 0.20)	0.09 (0.004, 0.18)	0.11 (0.02, 0.20)

BHS averaged across all available visits *Estimates are standardized β 's and corresponding confidence intervals, which were bootstrapped over 1000 iterations
Model 1: unadjusted Model 2: adjusted for age, sex, race, education, body mass index, smoking status, and cohort.

Table 3. Associations between sleep-derived BHS by visit and cognitive function

	Digital clock β (95% CI)*	Memory β (95% CI)*	Executive function β (95% CI)*	Language β (95% CI)*
Average of visits 1 and 2	0.16 (0.06, 0.26)	0.14 (0.04, 0.24)	0.11 (0.02, 0.20)	0.13 (0.03, 0.23)
Visits 1 and 2 together†				
Visit 1	0.12 (−0.02, 0.26)	0.10 (−0.02, 0.22)	0.09 (−0.03, 0.20)	0.10 (−0.04, 0.24)
Visit 2	0.02 (−0.11, 0.15)	0.09 (−0.04, 0.21)	0.02 (−0.09, 0.14)	0.07 (−0.04, 0.18)
Visit 1 alone	0.15 (0.05, 0.25)	0.11 (0.01, 0.21)	0.10 (0.01, 0.19)	0.11 (0.01, 0.21)
p-value‡	0.77	0.16	0.68	0.25
Visit 2 alone†	0.07 (−0.05, 0.19)	0.12 (0.01, 0.23)	0.06 (−0.05, 0.16)	0.11 (0.01, 0.21)
p-value‡	0.05	0.10	0.14	0.11

BHS averaged across all available visits. Cognitive function was measured on average 16.6 or 11.7 years after visits 1 or 2, respectively *Estimates are standardized β 's and corresponding confidence intervals, which were bootstrapped over 1000 iterations †n = 332 participants with both visits ‡p-value for likelihood ratio test for whether including both visits 1 and 2 together significantly contributes to the model, compared to the models with either visit 1 or visit 2 alone. p-values <0.05 indicate that the BHS from both visits contributes to the model. Model adjusted for age, sex, race, education, body mass index, smoking, and cohort.

sleepiness, and napping, but findings with memory, executive function, and language were attenuated (Supplemental Table 4).

BHS sub-scores for total, fluid, or crystallized cognition and cognitive function

Finally, we examined associations using TC, FC, or CC sub-scores as the primary exposure. As expected, the TC and FC scores were associated with higher memory performance (TC: β : 0.12, 95% CI: 0.003-0.24; FC: β : 0.15, 95% CI: 0.01-0.29). 1-SD higher FC was also associated with 0.14-SD higher language factor score (95% CI: 0.01-0.27). There were no associations with CC and any cognitive outcome (Supplemental Table 5).

Discussion

In this study, we found that higher sleep-derived BHS in mid-to-late life were associated with higher performance on both digital and traditional cognitive assessments on average 12.6 years later. Indeed, a higher BHS was related to better dCDT, memory, executive function, and language scores independent of potential confounding or mediating variables. These associations were also robust when excluding participants with dementia, suggesting this sleep score could potentially serve as a biomarker of future cognitive function.

These findings add to the growing body of evidence linking sleep EEG microarchitecture features and cognitive function. Individual sleep EEG features, such as theta power and spindle density, are associated with cognitive function [24] and differ by amyloid- β status [25]. Among middle aged and older men, sleep microarchitecture parameters (e.g. lower non-rapid eye movement (NREM) and REM delta power, higher NREM and REM theta or alpha power) are associated with the cross-sectional level [26], but not subsequent performance [27], on executive function, attention, processing speed tests. Further, compared to older women who remain cognitively intact, those who transition to MCI have higher baseline alpha power, theta power, and slow wave activity during NREM sleep, but do not differ in traditional sleep measures (e.g. total sleep time, wake after sleep onset, sleep efficiency, etc.) [28]. Similarly, individuals living with AD have higher theta power but lower sigma power, delta power, and spindle density during NREM sleep than cognitively normal controls [29]. NREM sigma power and REM delta or beta power also correlate with cortical thickness in several regions impacted by AD-related neurodegeneration [30]. Prior work has shown that combinations of sleep EEG features can predict cognition [11] or differentiate MCI and dementia [31, 32],

and, in some cases, produce continuous indices reflecting cognitive domains [11]. Yet, these approaches generally rely on hand-crafted features and limited cohorts, whereas the current framework generates an integrated brain-health score from raw EEG, optimized across cognition and disease in multiple large cohorts [12]. The composite score used in this study may also explain differences between previous studies that have not found associations between individual EEG features and subsequent cognitive function [27]. Taken together, the current findings suggest that such an integrated sleep EEG score may relate to future performance on both digital and traditional cognitive assessments.

There are several biological mechanisms through which sleep may impact cognitive function. At the simplest level, good sleep is generally necessary for adequate attention, vigilance, and working function. For example, blood pressure dipping during sleep requires stable sleep [33]. However, sleep becomes more fragmented and decreases in quality with older age [2], and this is further exacerbated among individuals living with dementia [34]. Indeed, sleep fragmentation is a whole-body stressor including increasing sympathetic activation [35] and dysglycemia [36]. Sleep is critical for removing waste via the glymphatic system [2, 37], regulating the production and clearance of amyloid- β [2, 37] and tau [2], and modulating glia cells, which are involved in synaptic homeostasis and neuroinflammation [2]. Within specific cognitive domains, there is extensive evidence linking sleep and memory. Indeed, memory consolidation depends on REM and non-REM sleep [2, 31], sharp wave ripples in the hippocampus [2], and sleep spindle-slow oscillation coupling [24, 31]. Similar mechanisms are also associated with language functions, such as learning word order rules [38], generalizing grammatical rules [39], and word consolidation [40, 41]. In addition to the mechanisms listed above [2], sleep and executive function may be linked through better reorganization of large-scale cortical networks during slow wave sleep [42], larger brain volumes in the frontal, temporal, or parietal cortices [4], or better brain connectivity [43]. Importantly, the sleep EEG score used in this study remained associated with cognitive function after additionally controlling for measures of sleep health, including sleep duration, wake after sleep onset, sleep efficiency, sleep apnea, use of sleep medications, daytime sleepiness, and napping. These findings support further evaluation of the importance of sleep microarchitecture, above and beyond summary measures of sleep duration, quality, and disorders, for cognitive function. Additional studies are also needed to evaluate the specific mechanisms through which sleep physiology impacts specific cognitive domains.

Unlike earlier work limited to associations between hand-crafted EEG features and cognition, this study shows that a data-driven, end-to-end sleep EEG score predicts future cognitive performance, including on a digital assessment (dCDT). The dCDT has become increasingly popular as a screening tool for early cognitive impairment. Indeed, cognitively intact individuals with amyloid- β positivity have lower dCDT scores compared to those without amyloid- β [14, 15]. The dCDT also accurately differentiates MCI sub-types from AD or each other [44, 45], as well as between older adults living with MCI or dementia and cognitively intact controls [14, 16]. These findings support the dCDT for detecting subtle differences in cognitive function relating to early stages of dementia. Further, this test is short and simple to administer, leading to improved feasibility over traditional NP batteries [16]. Given the utility of this test, it is critical to understand how potential biomarkers relate to both traditional NP and dCDT scores. Indeed, this study found the BHS was associated with future dCDT scores, with slightly stronger associations with the dCDT than NP factor scores. The BHS also remained associated with dCDT after accounting for sleep health (e.g. insomnia symptoms, sleep apnea), while findings with traditional NP factor scores were attenuated. Findings were also robust when excluding individuals who transitioned to dementia over the study period, highlighting potential utility for the BHS as a biomarker of cognitive function.

When evaluating the BHS sub-scores trained on different cognitive domains, the total cognition (TC) sub-score was associated with memory, while the fluid cognition (FC) sub-score was related to both memory and language. The magnitudes of these estimates were also comparable to the overall BHS score, which was trained to jointly optimize cognition and disease. Additionally, a higher BHS during each visit was associated with better cognitive function, although the magnitudes of the associations were strongest when using the average across available visits compared to individual visits. Taken together with these sub-score findings, this study demonstrates that a data-driven, end-to-end EEG score of brain health, originally trained on cross-sectional cognition and disease, also predicts future cognitive performance more than a decade later. The associations extended across both digital and traditional tests, suggesting that the BHS captures a durable, generalizable signal of neurophysiological integrity. Importantly, these findings underscore that a single overnight EEG can contain information relevant to long-term brain health, raising the possibility of scalable risk stratification and remote monitoring. Future work should determine whether tailoring model objectives (e.g. cognition-specific vs. multi-task) or leveraging repeated assessments can further enhance predictive value and refine the clinical use of EEG-derived BHSs.

Limitations and strengths

It is important to evaluate these results in the context of the study's limitations. First, FHS Gen 2 participants are predominantly of White European descent, which may limit generalizability. We attempted to address this limitation by including participants from the Omni 1 cohort, which has greater racial and ethnic diversity. Still, replication of findings in larger, more diverse cohorts is needed. Second, there was a gap between sleep EEG and cognitive assessments (mean: 12.6 years). We were also unable to account for baseline cognitive performance since the dCDT was not measured until 2011. However, findings were robust when additionally adjusting for the time between measures and when adjusting for memory, but not executive function or language, factor scores at the time of sleep

assessment. Future longitudinal studies are needed to evaluate predictive value and mechanisms between the BHS and cognitive function. Third, despite accounting for confounders and adjusting for potential mediating variables in sensitivity analyses, residual confounding is possible. Fourth, this analysis was dependent on single night of polysomnography, a brief snapshot of an evolving process. As multi-night sleep EEG recordings become more convenient and practical, the stability, evolution, and inter- vs. intra-individual variability domains of analysis will add new layers of complexity and hopefully improve prediction.

This study also has several strengths, including use of the BHS and its sub-scores, which were derived from sleep EEG using data-driven methods [12]. We also examined cognitive function using both traditional and digital assessments, as the dCDT has been found to detect subtle differences in cognitive function possibly relating to early stages of dementia [14, 16, 44, 45].

Conclusions

This study found that participants with a higher sleep-derived BHS in mid-to-late life had better dCDT performance, memory, executive function, and language scores on average 12.6 years later. These findings add to the growing body of evidence linking sleep EEG features with cognitive function and extend this work to suggest that integrated sleep EEG scores are also associated with future performance on both digital and traditional NP tests.

Acknowledgments

The authors would like to acknowledge the Framingham Heart Study participants, staff, and study personnel.

Author contributions

F.R.M. made substantial contributions to conception or design of the work; R.J.T., R.A., and M.B.W. to the acquisition of data for this work; and F.R.M. and W.G. to the analysis or interpretation of data for the work. All authors made substantial contributions to drafting of the work and reviewing it critically for important intellectual content. All authors gave final approval of the version to be published. F.R.M. had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Supplementary material

[Supplementary material](#) is available at *SLEEP* online.

Funding

This study was supported by the Framingham Heart Study of the National Heart, Lung, and Blood Institute of the National Institutes of Health in collaboration with Boston University (contract numbers: 75N92025D00012, N01-HC-25195, and HHSN2692015000011). The Framingham Heart Study is also supported by the National Institute on Aging of the National Institutes of Health (grant numbers U19-AG068753 and R01-AG072654). Funding to support the Omni cohort recruitment, retention and examination was provided by the National Heart, Lung, and Blood Institute of the National Institutes of Health (contract numbers: N01-HC-25195, HHSN2682015000011, and 75N92019D00031; grant numbers: R01-HL070100, R01-HL076784, R01-HL49869, and U01-HL053941). The authors were supported by the National Institute of Neurological Disorders and Stroke,

National Institutes of Health (grant numbers R01-NS120947, RF1-NS120947 and R01-HL161253). The funder was not involved in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Disclosure statement

Financial disclosure: R.J.T. discloses (1) general sleep medicine consulting for Guidepoint and GLG Councils; and (2) episodic consulting for Jazz Pharmaceuticals and Apnimed. RA is a scientific advisor to Signant Health, GSK, and NovoNordisk. M.B.W. is a co-founder, scientific advisor, and consultant to Beacon Biosignals and has a personal equity interest in the company. No other authors have conflicts to disclose.

Non-financial disclosure: R.J.T. discloses (1) licensed patent for cardiopulmonary coupling sleep spectrograms (sleep apnea and sleep quality), to MyCardio, LLC, through the Beth Israel Deaconess Medical Center, Boston; (2) licensed patents for respiratory self-similarity and Enhanced Expiratory Rebreathing Space (EERS) for diagnosis and treatment of high loop gain sleep apnea; and (3) licensed auto-CPAP algorithm (DeVilbiss-Drive). No other authors have conflicts to disclose.

Data availability

The datasets generated and analyzed for this study could be shared on reasonable request after approval of a Framingham Heart Study research application.

References

1. Nguyen Ho PT, Hoepel SJW, Rodriguez-Ayllon M, Luik AI, Verhoeven MW, Neitzel J. Sleep, 24-hour activity rhythms, and subsequent amyloid- β pathology. *JAMA Neurol.* 2024;**81**(8):824–834. <https://doi.org/10.1001/jamaneurol.2024.1755>
2. Chen J, Peng G, Sun B. Alzheimer's disease and sleep disorders: a bidirectional relationship. *Neuroscience.* 2024;**557**:12–23. <https://doi.org/10.1016/j.neuroscience.2024.08.008>
3. Thomas RJ, Kim H, Maillard P, et al. Digital sleep measures and white matter health in the Framingham Heart Study. *Explor Med.* 2021;**2**(3):253–267. <https://doi.org/10.37349/emed.2021.00045>
4. Tai XY, Chen C, Manohar S, Husain M. Impact of sleep duration on executive function and brain structure. *Commun Biol.* 2022;**5**(1):201. <https://doi.org/10.1038/s42003-022-03123-3>
5. Ma Y, Liang L, Zheng F, Shi L, Zhong B, Xie W. Association between sleep duration and cognitive decline. *JAMA Netw Open.* 2020;**3**(9):e2013573. <https://doi.org/10.1001/jamanetworkopen.2020.13573>
6. Chen J, Espeland MA, Brunner RL, et al. Sleep duration, cognitive decline, and dementia risk in older women. *Alzheimers Dement.* 2016;**12**(1):21–33. <https://doi.org/10.1016/j.jalz.2015.03.004>
7. Keil SA, Schindler AG, Wang MX, et al. Longitudinal sleep patterns and cognitive impairment in older adults. *JAMA Netw Open.* 2023;**6**(12):e2346006. <https://doi.org/10.1001/jamanetworkopen.2023.46006>
8. 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**(32):6315–6324. <https://doi.org/10.1523/JNEUROSCI.0503-19.2019>
9. Lucey BP, Wisch J, Boerwinkle AH, et al. Sleep and longitudinal cognitive performance in preclinical and early symptomatic Alzheimer's disease. *Brain.* 2021;**144**(9):2852–2862. <https://doi.org/10.1093/brain/awab272>
10. Quan SF, Howard BV, Iber C, et al. The Sleep Heart Health Study: design, rationale, and methods. *Sleep.* 1997;**20**(12):1077–1085. <https://doi.org/10.1093/sleep/20.12.1077>
11. Adra N, Dümmer LW, Paixao L, et al. Decoding information about cognitive health from the brainwaves of sleep. *Sci Rep.* 2023;**13**(1):11448. <https://doi.org/10.1038/s41598-023-37128-7>
12. Ganglberger W, Sun H, Turley N, et al. Brain health from sleep EEG: a multicohort, deep learning biomarker for cognition, disease, and mortality. *NEJM AI.* 2026;**3**(3):1–17. <https://doi.org/10.1056/AIoa2500487>
13. Thompson LI, Cummings M, Emrani S, et al. Digital clock drawing as an Alzheimer's disease susceptibility biomarker: associations with genetic risk score and APOE in older adults. *J Prev Alzheimers Dis.* 2024;**11**(1):79–87. <https://doi.org/10.14283/jpad.2023.48>
14. Rentz DM, Papp KV, Mayblyum DV, et al. Association of digital clock drawing with PET amyloid and tau pathology in normal older adults. *Neurology.* 2021;**96**(14):e1844–e1854. <https://doi.org/10.1212/WNL.00000000000011697>
15. Ding H, Lyu C, Karjadi C, et al. Association of the digital clock drawing test with amyloid and tau PET biomarkers in low age risk adults. *Sci Rep.* 2025;**15**(1):11104. <https://doi.org/10.1038/s41598-025-95852-8>
16. Yuan J, Libon DJ, Karjadi C, et al. Association between the digital clock drawing test and neuropsychological test performance: large community-based prospective cohort (Framingham Heart Study). *J Med Internet Res.* 2021;**23**(6):e27407. <https://doi.org/10.2196/27407>
17. Dawber TR, Kannel WB. An epidemiologic study of heart disease: the Framingham study. *Nutr Rev.* 1958;**16**(1):1–4. <https://doi.org/10.1111/j.1753-4887.1958.tb00605.x>
18. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham offspring study. *Design and preliminary data.* *Prev Med.* 1975;**4**(4):518–525. [https://doi.org/10.1016/0091-7435\(75\)90037-7](https://doi.org/10.1016/0091-7435(75)90037-7)
19. Schramm U, Berger G, Müller R, Kratzsch T, Peters J, Frölich L. Psychometric properties of clock drawing test and MMSE or short performance test (SKT) in dementia screening in a memory clinic population. *Int J Geriatr Psychiatry.* 2002;**17**(3):254–260. <https://doi.org/10.1002/gps.585>
20. Mukherjee S, Choi SE, Lee ML, et al. Cognitive domain harmonization and cocalibration in studies of older adults. *Neuropsychology.* 2023;**37**(4):409–423. <https://doi.org/10.1037/neu0000835>
21. Scollard P, Choi SE, Lee ML, et al. Ceiling effects and differential measurement precision across calibrated cognitive scores in the Framingham Study. *Neuropsychology.* 2023;**37**(4):383–397. <https://doi.org/10.1037/neu0000828>
22. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Measur.* 1977;**1**(3):385–401. <https://doi.org/10.1177/014662167700100306>
23. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders 4th Edition.* Washington DC: American Psychiatric Association; 1994.
24. Suen C, Ma J, Lin X, et al. Linking sleep microstructure, blood markers of inflammation and metabolism, and cognition: mediation analysis in the osteoporotic fractures in men study. *Geroscience.* 2025;**48**(1):1277–1289. <https://doi.org/10.1007/s11357-025-01730-6>
25. Pulver RL, Kronberg E, Medenblik LM, et al. Mapping sleep's oscillatory events as a biomarker of Alzheimer's disease. *Alzheimers Dement.* 2024;**20**(1):301–315. <https://doi.org/10.1002/alz.13420>

26. Parker JL, Appleton SL, Melaku YA, et al. The association between sleep microarchitecture and cognitive function in middle-aged and older men: a community-based cohort study. *J Clin Sleep Med*. 2022;**18**(6):1593–1608. <https://doi.org/10.5664/jcsm.9934>
27. Parker JL, Vakulin A, Melaku YA, et al. Associations of baseline sleep microarchitecture with cognitive function after 8 years in middle-aged and older men from a community-based cohort study. *Nat Sci Sleep*. 2023;**15**:389–406. <https://doi.org/10.2147/NSS.S401655>
28. 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**(3):e12666. <https://doi.org/10.1111/jsr.12666>
29. Sun H, Ye E, Paixao L, et al. The sleep and wake electroencephalogram over the lifespan. *Neurobiol Aging*. 2023;**124**:60–70. <https://doi.org/10.1016/j.neurobiolaging.2023.01.006>
30. D'Atri A, Gorgoni M, Scarpelli S, et al. Relationship between cortical thickness and EEG alterations during sleep in the Alzheimer's disease. *Brain Sci*. 2021;**11**(9):1174. <https://doi.org/10.3390/brainsci11091174>
31. Geng D, Wang C, Fu Z, Zhang Y, Yang K, An H. Sleep EEG-based approach to detect mild cognitive impairment. *Front Aging Neurosci*. 2022;**14**:865558. <https://doi.org/10.3389/fnagi.2022.865558>
32. Ye EM, Sun H, Krishnamurthy PV, et al. Dementia detection from brain activity during sleep. *Sleep*. 2023;**46**(3). <https://doi.org/10.1093/sleep/zsac286>
33. Seravalle G, Mancina G, Grassi G. Sympathetic nervous system, sleep, and hypertension. *Curr Hypertens Rep*. 2018;**20**(9):74. <https://doi.org/10.1007/s11906-018-0874-y>
34. Spira AP, Chen-Edinboro LP, Wu MN, Yaffe K. Impact of sleep on the risk of cognitive decline and dementia. *Curr Opin Psychiatry*. 2014;**27**(6):478–483. <https://doi.org/10.1097/YCO.0000000000000106>
35. Taylor KS, Murai H, Millar PJ, et al. Arousal from sleep and sympathetic excitation during wakefulness. *Hypertension*. 2016;**68**(6):1467–1474. <https://doi.org/10.1161/HYPERTENSIONAHA.116.08212>
36. Stamatakis KA, Punjabi NM. Effects of sleep fragmentation on glucose metabolism in normal subjects. *Chest*. 2010;**137**(1):95–101. <https://doi.org/10.1378/chest.09-0791>
37. Mendelsohn AR, Larrick JW. Sleep facilitates clearance of metabolites from the brain: Glymphatic function in aging and neurodegenerative diseases. *Rejuvenation Res*. 2013;**16**(6):518–523. <https://doi.org/10.1089/rej.2013.1530>
38. Cross ZR, Helfrich RF, Corcoran AW, et al. Slow oscillation–spindle coupling predicts sequence-based language learning. *J Neurosci*. 2025;**45**(3):e2193232024. <https://doi.org/10.1523/JNEUROSCI.2193-23.2024>
39. Batterink LJ, Paller KA. Sleep-based memory processing facilitates grammatical generalization: evidence from targeted memory reactivation. *Brain Lang*. 2017;**167**:83–93. <https://doi.org/10.1016/j.bandl.2015.09.003>
40. Bakker I, Takashima A, van Hell JG, Janzen G, McQueen JM. Changes in theta and beta oscillations as signatures of novel word consolidation. *J Cogn Neurosci*. 2015;**27**(7):1286–1297. https://doi.org/10.1162/jocn_a_00801
41. Batterink LJ, Westerberg CE, Paller KA. Vocabulary learning benefits from REM after slow-wave sleep. *Neurobiol Learn Mem*. 2017;**144**:102–113. <https://doi.org/10.1016/j.nlm.2017.07.001>
42. Diep C, Ftouni S, Manousakis JE, Nicholas CL, Drummond SPA, Anderson C. Acoustic slow wave sleep enhancement via a novel, automated device improves executive function in middle-aged men. *Sleep*. 2019;**43**(1). <https://doi.org/10.1093/sleep/zsz197>
43. Sen A, Tai XY. Sleep duration and executive function in adults. *Curr Neurol Neurosci Rep*. 2023;**23**(11):801–813. <https://doi.org/10.1007/s11910-023-01309-8>
44. Binaco R, Calzaretto N, Epifano J, et al. Machine learning analysis of digital clock drawing test performance for differential classification of mild cognitive impairment subtypes versus Alzheimer's disease. *J Int Neuropsychol Soc*. 2020;**26**(7):690–700. <https://doi.org/10.1017/S1355617720000144>
45. Müller S, Herde L, Preische O, et al. Diagnostic value of digital clock drawing test in comparison with CERAD neuropsychological battery total score for discrimination of patients in the early course of Alzheimer's disease from healthy individuals. *Sci Rep*. 2019;**9**(1):3543. <https://doi.org/10.1038/s41598-019-40010-0>