A Multimodal Sleep Foundation Model Developed with 500K Hours of Sleep Recordings for Disease Predictions

Rahul Thapa[1,2,*], Magnus Ruud Kjær[3,4,5,*], Bryan He[2], Ian Covert[2], Hyatt Moore[3], Umaer Hanif[6], Gauri Ganjoo[3], M. Brandon Westover[8], Poul Jennum[5,7], Andreas Brink-Kjær[4], Emmanuel Mignot[3,+], and James Zou[1,2,+]

[1]Department of Biomedical Data Science, Stanford University, Stanford, CA, USA
[2]Department of Computer Science, Stanford University, Stanford, CA, USA
[3]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA
[4]Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark
[5]Department of Clinical Neurophysiology, Danish Center for Sleep Medicine, Copenhagen University Hospital – Rigshospitalet, Copenhagen, Denmark
[6]Data Science, BioSerenity, Paris, France
[7]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
[8]Department of Neurology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA
[*]Co-first authors
[+]Co-senior authors
[‡]Correspondence: mignot@stanford.edu, jamesz@stanford.edu

ABSTRACT

Sleep is a fundamental biological process with profound implications for physical and mental health, yet our understanding of its complex patterns and their relationships to a broad spectrum of diseases remains limited. While polysomnography (PSG), the gold standard for sleep analysis, captures rich multimodal physiological data, analyzing these measurements has been challenging due to limited flexibility across recording environments, poor generalizability across cohorts, and difficulty in leveraging information from multiple signals simultaneously. To address this gap, we curated over 585,000 hours of high-quality sleep recordings from approximately 65,000 participants across multiple cohorts and developed SleepFM, a multimodal sleep foundation model trained with a novel contrastive learning approach, designed to accommodate any PSG montage. SleepFM produces informative sleep embeddings that enable predictions of future diseases. We systematically demonstrate that SleepFM embeddings can predict 130 future diseases, as modeled by Phecodes, with C-Index and AUROC of at least 0.75 on held-out participants (Bonferroni-corrected $p < 0.01$). This includes accurate predictions for death (C-Index: 0.84 [95% CI: 0.81–0.87]), heart failure (C-Index: 0.80 [95% CI: 0.77–0.83]), chronic kidney disease (C-Index: 0.79 [95% CI: 0.77–0.81]), dementia (C-Index: 0.85 [95% CI: 0.82–0.87]), stroke (C-Index: 0.78 [95% CI: 0.76–0.81]), atrial fibrillation (C-Index: 0.78 [95% CI: 0.75–0.81]), and myocardial infarction (C-Index: 0.81 [95% CI: 0.78–0.84]). The model's generalizability was further validated through strong performance on the Sleep Heart Health Study (SHHS), a dataset unseen during pre-training. Additionally, SleepFM demonstrates strong performance on traditional sleep analysis tasks, achieving competitive results in both sleep staging (mean F1 scores: 0.70–0.78) and sleep apnea diagnosis (AUROC: 0.90–0.94). Beyond these standard applications, our analysis reveals that specific sleep stages and physiological signals carry distinct predictive power for different diseases. This work demonstrates how foundation models can leverage sleep polysomnography data to uncover the extensive relationship between sleep physiology and future disease risk.

1 Introduction

Sleep is a complex process that involves intricate interactions across multiple physiological systems that involve brain, heart, lung, and muscle activity[1]. Polysomnography (PSG), the gold standard for sleep evaluation, captures these interactions through recordings of multiple modalities including brain activity signals (BAS), electrocardiography (ECG), electromyography (EMG), and respiratory signals[2].

Sleep disorders affect millions of people worldwide and are increasingly recognized as indicators of and contributors to various health conditions[3–5]. Sleep disturbances often precede the clinical onset of numerous conditions, notably psychiatric disorders[6], neurodegenerative diseases[7] and cardiovascular disorders[8,9]. These associations highlight the profound role sleep

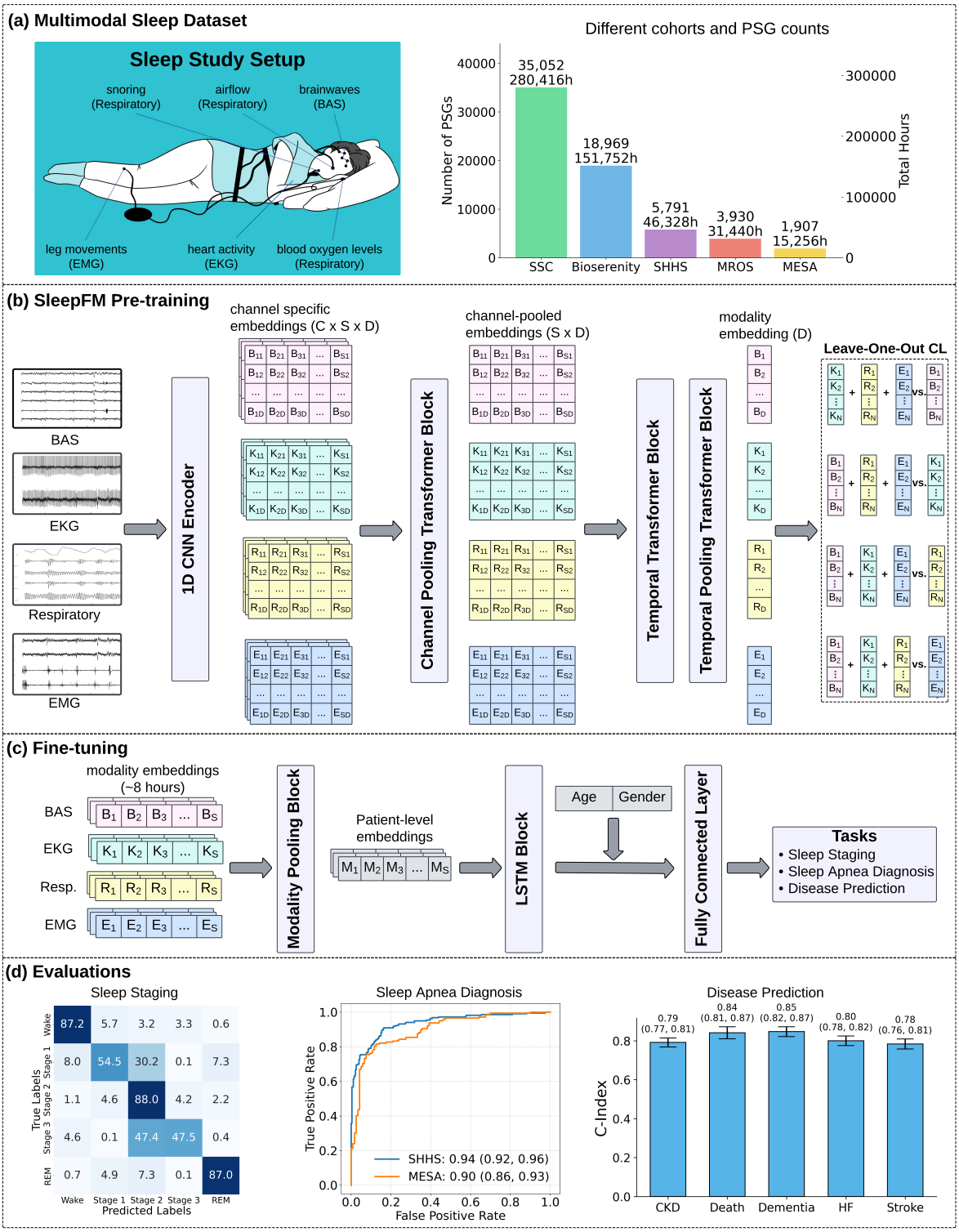


Figure 1. Overview of the SleepFM framework. (a) Polysomnography setup and dataset statistics across multiple sleep centers. (b) Multimodal contrastive learning approach for model pre-training with shared parameters. (c) Fine-tuning pipeline using frozen embeddings from the pre-trained encoder for downstream tasks such as sleep staging, apnea detection, and disease prediction. (d) Evaluation results demonstrating performance across tasks and clinical applications. In the model architecture, C denotes the number of input channels, S represents the sequence length, and D indicates the embedding dimension. For disease prediction results, CKD denotes chronic kidney disease, and HF denotes heart failure.

plays in maintaining overall health and underscore its predictive potential across a wide spectrum of diseases. However, most existing studies have focused on identifying links between sleep and specific diseases using isolated metrics or manual annotations, leaving the full complexity of sleep physiology, as captured in PSG, underutilized.

Recent advances in deep learning methods have enabled effective use of PSG's multimodal data for a range of applications, from traditional sleep analysis tasks (sleep staging, sleep apnea classification, and periodic leg movement counting)[10–13] to specific disease predictions, including atrial fibrillation[14], biological aging[4], and narcolepsy[15]. While these efforts represent significant progress, they face several key limitations. Most approaches target individual diseases or specific outcomes, rely on supervised training requiring expert-annotated data, and are constrained by relatively small datasets ranging from 2,500 to 15,913 recordings[4,12,14–16]. This reliance on labeled data is particularly limiting given that annotation requires sleep experts to manually inspect signals—a time-consuming process prone to inconsistencies across different sites and interpreters[17]. Furthermore, existing models often struggle with limited flexibility across recording environments, poor generalizability across patient cohorts, and inability to leverage the full spectrum of information contained in multimodal sleep recordings. The development of flexible and robust architectures that can leverage full spectrum of multimodal sleep data, along with a systematic investigation of sleep's predictive capabilities across a broad spectrum of diseases, remain unexplored.

Foundation models have emerged as a transformative approach in machine learning, enabling robust representation learning from large-scale, unlabeled data. By leveraging self-supervised learning, these models can be fine-tuned efficiently for diverse applications[18]. In biomedicine, foundation models have demonstrated remarkable capabilities in analyzing complex, heterogeneous datasets, driving advances in disease prediction, patient stratification, and therapeutic discovery[19–22]. Their ability to extract meaningful patterns from large-scale data has addressed many challenges associated with the diverse and high-dimensional nature of clinical datasets.

Despite these successes, their application to sleep—a critical yet underexplored area of medical research—remains limited. Sleep data, particularly from PSG, presents unique challenges due to its complexity and variability, including differences in the number and types of recording channels across clinical cohorts. Most sleep studies have narrowly focused on sleep-specific outcomes, constraining the broader potential of foundation models for disease prediction. In preliminary work, we explored self-supervised learning on PSG data in a smaller cohort of participants[16]. While this effort highlighted the potential of foundation models for analyzing sleep data, it primarily targeted sleep-specific outcomes and lacked the flexibility to accommodate the diverse configurations of PSG recordings. These limitations emphasize the need for models that can generalize across heterogeneous datasets and systematically uncover sleep's role in predicting a wider range of diseases.

In this paper, we present a sleep foundation model, SleepFM, to advances the understanding of sleep's relationship with human health. SleepFM is trained and evaluated on over 585,000 hours of multimodal PSG recordings from more than 65,000 participants through a flexible architecture that accommodates varying PSG configurations across different clinical cohorts. The scale of our dataset represents a significant advance, exceeding previous supervised learning approaches in sleep analysis by 5-25 times[4,12,14,15]. When compared to other large-scale biosignal-based foundation models, SleepFM offers substantially richer data: existing approaches have been limited to 4,611 hours of ECG data[23], 4,246 hours of EEG recordings[24], or 51,120 hours of wearable biosignals[25].

Inspired by phenome-wide association studies (PheWAS), which systematically investigate associations between a single genetic variant or biomarker and thousands of phenotypes[26,27], we comprehensively evaluate SleepFM's predictive power across 1,041 different phenotypes, specifically disease codes recorded in electronic health records (EHRs). Similar to how PheWAS enables discovery of associations between a single factor and multiple outcomes, our approach provides a systematic, data-driven investigation of which health conditions might be predictable from sleep patterns.

Our work makes several key contributions. First, we developed a channel-agnostic foundation model that effectively handles variability in recording channels across different clinical sites, enabling robust processing of sleep data regardless of differences in channel types or numbers. Second, through a systematic PheWAS-inspired analysis of 1,041 disease phenotypes, we identify specific health conditions where sleep recordings demonstrate strong predictive power. Among these, we achieve particularly robust performance for conditions such as death (C-Index: 0.84 [95% CI: 0.81–0.87], AUROC: 0.84 [95% CI: 0.80–0.88]), heart failure (C-Index: 0.80 [95% CI: 0.77–0.83], AUROC: 0.83 [95% CI: 0.79–0.86]), chronic kidney disease (C-Index: 0.79 [95% CI: 0.77–0.81], AUROC: 0.82 [95% CI: 0.79–0.85]), dementia (C-Index: 0.85 [95% CI: 0.82–0.87], AUROC: 0.87 [95% CI: 0.84–0.91]), stroke (C-Index: 0.78 [95% CI: 0.76–0.81], AUROC: 0.81 [95% CI: 0.78–0.85]), atrial fibrillation (C-Index: 0.78 [95% CI: 0.75–0.81], AUROC: 0.81 [95% CI: 0.77–0.84]), myocardial infarction (C-Index: 0.81 [95% CI: 0.78–0.84], AUROC: 0.85 [95% CI: 0.82–0.88]), and atrial flutter (C-Index: 0.82 [95% CI: 0.78–0.85], AUROC: 0.88 [95% CI: 0.84–0.92]). In total, we identify over 130 conditions where the model achieves statistically significant predictive performance (C-Index and AUROC > 0.75, Bonferroni-corrected $p < 0.01$). Third, we show that this single foundation model can effectively process diverse sleep-related signals and generalize across multiple cohorts and recording environments, demonstrating robust transfer learning capabilities even on completely independent datasets. Fourth, while primarily focused on disease prediction, our foundation model achieves competitive performance on standard sleep analysis tasks, including sleep staging (mean F1 scores:

0.70–0.78), sleep apnea diagnosis (AUROC: 0.90–0.94), and age estimation (MAE: 7.33 ± 6.17 years, correlation: 0.88), suggesting that the learned representations capture both disease-relevant and sleep-specific patterns.

2 Results

We provide a detailed description of our dataset and model training procedure in Section 3. Briefly, we used PSG data from four different cohorts: Stanford Sleep Clinic (SSC)[16], BioSerenity[28,29], Multi-Ethnic Study of Atherosclerosis (MESA)[30,31], and Outcomes of Sleep Disorders in Older Men (MrOS)[30,32]. SSC is our internal dataset, comprising approximately 35,052 PSGs collected over two decades (1999–2024) from participants aged 1 to 100 years. The BioSerenity dataset, provided by the medtech company BioSerenity for research and development purposes, contains approximately 18,900 PSGs from participants aged 7 to 90 years. MESA, an NHLBI-sponsored six-center study, includes data from participants initially aged 45–84 at baseline in 2000–2002, with sleep exams conducted on 2,237 participants between 2010 and 2012. MrOS is an ancillary study of the Osteoporotic Fractures in Men Study, which enrolled 3,930 community-dwelling men aged 65 years or older between 2003 and 2005 for comprehensive sleep assessments, including PSG and actigraphy.

Combined, these datasets encompass approximately 65,000 participants, totaling over 585,000 hours of multimodal sleep recordings, forming the basis for pre-training, fine-tuning, and testing. Additionally, we utilized the Sleep Heart Health Study (SHHS)[30,33] as a transfer learning validation dataset. SHHS is a multi-center cohort study implemented by the National Heart, Lung, and Blood Institute to investigate the cardiovascular and other consequences of sleep-disordered breathing. It enrolled 6,441 men and women aged 40 years and older between November 1, 1995, and January 31, 1998. This dataset was used for fine-tuning but excluded during pre-training, allowing us to assess our model's generalizability.

The distribution of our datasets, after filtering for incomplete labels and corrupted data, is shown in Table 1. A demographic summary of the SSC and BioSerenity datasets is presented in Supp. Table 1 and Supp. Table 2, respectively. For the publicly available datasets (SHHS, MrOS, MESA), we refer to their respective publications for demographic details, as described on the National Sleep Research Resource website.

Sleep data analysis poses unique challenges due to its inherently multimodal nature. PSG recordings encompass diverse physiological signals including brain activity signals (BAS), cardiac function (ECG), muscle activity (EMG), and respiratory (flow, effort, oxygen saturation) patterns. This complexity is further compounded by variations in channel availability and recording protocols and filter settings that can vary across different sleep centers, making it difficult to develop models that generalize across clinical settings. To address these challenges, our approach consists of two main phases: channel-agnostic multimodal pre-training and task-specific fine-tuning (Fig. 1). Our key technical innovation lies in combining a channel-agnostic architecture that handles arbitrary numbers and orderings of input channels with a novel leave-one-out contrastive learning strategy, enabling robust pre-training across diverse clinical settings.

Our pre-processing pipeline starts with resampling all signals to 128 Hz to ensure consistency across cohorts and equipment. We segment the resampled signals into 5-second windows, which serve as the basic unit (token) for our model's processing. The model architecture comprises 1D convolutional layers for initial signal processing, followed by channel-agnostic attention pooling to handle both the varying numbers of channels and the different orders in which they are presented across cohorts. A transformer block then captures temporal dependencies across a 5-minute context window. During pre-training, we employed a multimodal contrastive learning (CL) objective to align all modalities into a shared representation space. The model's robustness arises from its channel-agnostic design, which accommodates varying channel availability (missing channels), different numbers of channels across cohorts, and diverse channel types across recording environments. More details about model pre-training are discussed in Section 3.

For downstream tasks, we leverage the pre-trained model's embeddings through lightweight fine-tuning. The token embeddings from different modalities are pooled again and processed by a 2-layer Long Short-Term Memory (LSTM) network before passing through task-specific output heads. For patient-level prediction tasks (e.g., disease prediction), an additional temporal pooling layer before the output layer compresses all token embeddings into a single 128-dimensional embedding. More details about model fine-tuning are explained in Section 3.

To assess model performance across different tasks, we employ task-specific evaluation metrics. For classification tasks such as gender classification and sleep apnea diagnosis, we use AUROC (Area Under the Receiver Operating Characteristic curve) and AUPRC (Area Under the Precision-Recall Curve). For age estimation, we use mean absolute error (MAE) and Pearson correlation coefficient. For sleep staging, we use the F1 score, which is particularly suitable for evaluating performance across imbalanced sleep stages. For disease prediction tasks, we employ two metrics: the Harrell's concordance index (C-Index) and AUROC. Harrell's C-Index is a widely used metric in survival analysis that measures the model's ability to correctly rank relative risks. It is defined as the proportion of all comparable pairs in which the predictions and outcomes are concordant, meaning that the predicted risk aligns with the observed survival times. The C-Index ranges from 0 to 1, where 1 indicates perfect concordance, 0.5 indicates random guessing, and 0 indicates perfect discordance. All metrics range from 0 to 1, with

higher values indicating better performance. All reported metrics include 95% confidence intervals (CIs), calculated using bootstrapping.

2.1 SleepFM Demonstrates Strong Performance on Standard Sleep Analysis Tasks

After pre-training SleepFM, we used its frozen representations for downstream task-specific fine-tuning. We evaluated the quality of these learned representations through four common benchmark tasks: age estimation, gender classification, sleep stage classification, and sleep apnea diagnosis. Even though our model was not explicitly pre-trained for these tasks, a strong foundation model should effectively capture such fundamental sleep characteristics, as demonstrated in our previous work[16]. For all tasks, we trained a simple attention-pooling-based model on top of the frozen multimodal representations derived from entire nights of sleep data.

For age estimation, we assessed the model's ability to predict chronological age from sleep recordings. Overall performance is shown in Supp. Fig. 1, with the model achieving a MAE of 7.33 ± 6.17 years and a correlation coefficient of 0.88 with chronological age. When analyzing performance across five age groups (0–18: pediatric, 19–35: young adult, 36–50: early middle-age, 51–65: late middle-age, and 66+: elderly), we found varying levels of accuracy. The model showed lowest MAE in the pediatric (5.05 ± 6.76 years) and early middle-age (5.26 ± 4.33 years) groups, while demonstrating higher error rates for the elderly group (11.62 ± 7.13 years). This pattern suggests that age prediction is more challenging at the extremes of the age spectrum. For gender classification, considering a binary distinction between male and female, the model achieved an AUROC of 0.86 (95% CI: 0.85–0.87) and an AUPRC of 0.90 (95% CI: 0.89–0.91).

Sleep stage classification, an important task traditionally requiring extensive manual analysis by trained technicians, involves distinguishing five stages: Wake, Stage 1, Stage 2, Stage 3, and REM. While conventional methods rely on 30-second epochs, we reduced this to 5-second windows, enabling more granular classification, which has been shown to improve diagnostic accuracy (e.g., narcolepsy[15]). We fine-tuned a single sleep staging model using annotated data from four cohorts: SSC, MESA, MROS, and SHHS. For comparison, we included performance metrics from U-Sleep[12], a state-of-the-art model trained exclusively for sleep staging. SleepFM achieves competitive performance compared to U-Sleep across all cohorts, as shown in Supp. Table 3. The confusion matrices in Supp. Fig. 2 demonstrate that SleepFM excels in distinguishing Wake, Stage 2, and REM sleep, with some mild confusion between Stage 1 and Stage 2—an expected challenge given Stage 1's role as a transitional state[34].

For sleep apnea diagnosis, we performed patient-level binary classification to distinguish between moderate/severe and no/mild apnea cases, using an apnea-hypopnea index threshold of 15. On the MESA dataset, SleepFM achieved an AUROC of 0.90 (95% CI: 0.86–0.93) and an AUPRC of 0.93 (95% CI: 0.89–0.96). Notably, on the SHHS dataset, which was not included during pre-training, SleepFM demonstrated even higher performance, with an AUROC of 0.94 (95% CI: 0.92–0.96) and an AUPRC of 0.93 (95% CI: 0.91–0.95). These results highlight SleepFM's robust generalizability and its potential for clinical applications.

2.2 SleepFM Enables Comprehensive Disease Prediction from Sleep Data

Building on our model's strong performance on common sleep analysis tasks, we next investigate a novel and more ambitious application: systematically identifying which health conditions can be predicted from sleep data. While our previous evaluations focused on classification tasks using concurrent data, we now tackle the more challenging problem of predicting future disease onset. Similar to a PheWAS, we identify conditions for which sleep data demonstrates high and statistically significant predictive power.

To accomplish this, we pair SSC data with EHR data, extracting all diagnostic information, including ICD-9 and ICD-10 codes along with their associated timestamps. The ICD-9 and ICD-10 codes are translated into phecodes, a hierarchical categorization system of 1,868 disease diagnosis codes designed specifically for PheWAS studies[27,35]. The timestamp for each phecode is determined as the earliest timestamp among the ICD-9 or ICD-10 codes that fall within the corresponding phecode category. Positive cases are composed of patients whose first instance of a phecode occurs more than 7 days after the sleep study. This 7-day offset helps avoid trivial cases where sleep disturbances might be an immediate precursor to diagnosis. From the complete set of phecodes, we filter based on prevalence, excluding those with a prevalence of less than 1.5% in the overall dataset. This minimum prevalence threshold ensures sufficient statistical power for meaningful evaluation of the model's predictive performance. This filtering results in a final set of 1,041 phecodes used to evaluate our foundation model for EHR-based tasks. For model fine-tuning, we use the Cox proportional hazards (CoxPH) loss function. We extend it to a multi-label setting by computing the standard CoxPH loss independently for each label and averaging the losses. More details about the CoxPH loss can be found in Section 3.

Figure 2 illustrates SleepFM's performance across disease categories on the test set. While performance varies across categories, SleepFM demonstrates strong results in several key areas, including neoplasms, pregnancy complications, circulatory conditions, and mental disorders. Overall, 130 future diseases, modeled using Phecodes, achieved a C-Index and AUROC of at least 0.75 on held-out participants (and Bonferroni-corrected $p < 0.01$), as summarized in Supp. Table 5. AUROC was

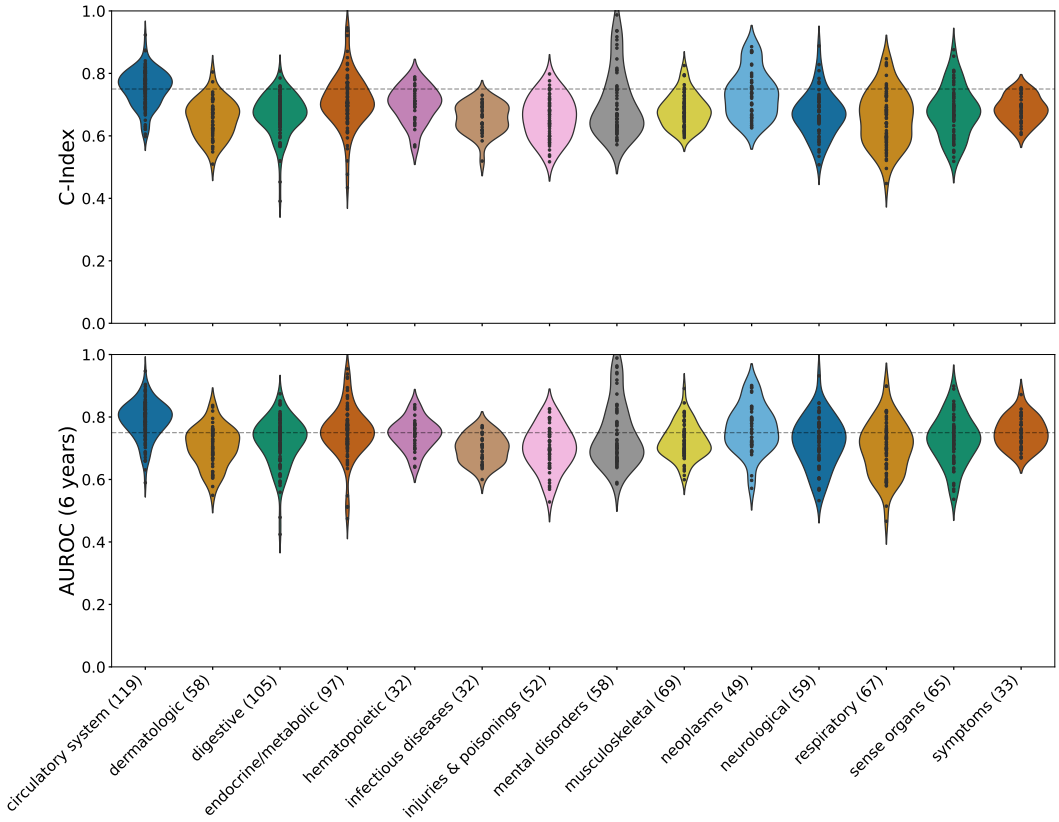


Figure 2. Performance of SleepFM on the held-out test set stratified by disease categories. The results are evaluated using two metrics: the C-Index, which measures the model's ability to rank patient risk accurately, and the 6-year AUROC, which assesses the model's discrimination performance by evaluating its ability to distinguish between patients who experience the event of interest and those who do not within a 6-year prediction window. For reference, the horizontal dotted line indicates a threshold of 0.75.

calculated using a six-year horizon, meaning a condition is considered positive if the patient develops the disease within six years of their PSG study. The six-year horizon for AUROC calculation—where a condition is positive if the patient develops the disease within six years of their PSG study—was chosen to balance performance and account for both long-term and short-term conditions.

In neurological and mental disorders, the model showed high accuracy for mild cognitive impairment (C-Index: 0.81 [95% CI: 0.78–0.84], AUROC: 0.84 [95% CI: 0.80–0.88]), aligning with studies showing sleep disturbances as early markers of cognitive decline[36]. Notably strong performance was observed for Parkinson's disease (C-Index: 0.89 [95% CI: 0.85–0.92], AUROC: 0.93 [95% CI: 0.89–0.96]), where sleep disorders are increasingly recognized as potential early indicators[37], and developmental delays and disorders (C-Index: 0.80 [95% CI: 0.77–0.84], AUROC: 0.84 [95% CI: 0.79–0.87]). Among circulatory conditions, the model effectively predicted hypertensive heart disease (C-Index: 0.84 [95% CI: 0.82–0.86], AUROC: 0.88 [95% CI: 0.85–0.91]) and intracranial hemorrhage (C-Index: 0.80 [95% CI: 0.74–0.86], AUROC: 0.82 [95% CI: 0.73–0.90]), consistent with established links between sleep disorders and cardiovascular risk[38]. Particularly noteworthy are our findings in the Neoplasm category, where the model showed strong predictive performance for several cancers: prostate cancer (C-Index: 0.89 [95% CI: 0.86–0.91], AUROC: 0.90 [95% CI: 0.87–0.93]), breast cancer (C-Index: 0.87 [95% CI: 0.83–0.90], AUROC: 0.90 [95% CI: 0.86–0.93]), and melanomas of skin (C-Index: 0.83 [95% CI: 0.77–0.87], AUROC: 0.83 [95% CI: 0.76–0.90]). These findings align with existing literature linking sleep patterns to cancer risk[39], including specific associations with breast cancer[40], cutaneous melanoma[41], and prostate cancer[42]. This concordance with previous literature further validates our modeling approach, demonstrating that SleepFM independently identifies conditions with established sleep-related associations.

Based on sleep expertise and previous literature, we identified 14 conditions of particular interest that are likely to have meaningful associations with sleep patterns. For instance, one study links sleep regularity with mortality risk[43], while another identifies prolonged sleep duration as a potential marker of early neurodegeneration[44]. Numerous other studies highlight

associations between sleep and dementia[45], type 2 diabetes[46], stroke[47], and various cardiovascular outcomes[14,48]. For several of these conditions, multiple related phecodes from the PheWAS catalog were consolidated into a single disease group in consultation with a medical doctor (mapping shown in Supp. Table 4). Results for these selected conditions, including death, stroke, heart failure, and dementia, are presented in Fig. 3. SleepFM achieves robust performance across these conditions, with particularly strong predictive accuracy for death (C-Index: 0.84 [95% CI: 0.81–0.87], AUROC: 0.84 [95% CI: 0.80–0.88]), heart failure (C-Index: 0.80 [95% CI: 0.77–0.83], AUROC: 0.83 [95% CI: 0.79–0.86]), chronic kidney disease (C-Index: 0.79 [95% CI: 0.77–0.81], AUROC: 0.82 [95% CI: 0.79–0.85]), dementia (C-Index: 0.85 [95% CI: 0.82–0.87], AUROC: 0.87 [95% CI: 0.84–0.91]), and stroke (C-Index: 0.78 [95% CI: 0.76–0.81], AUROC: 0.81 [95% CI: 0.78–0.85]). All selected conditions have statistically significant p-values (<0.01) after Bonferroni correction.

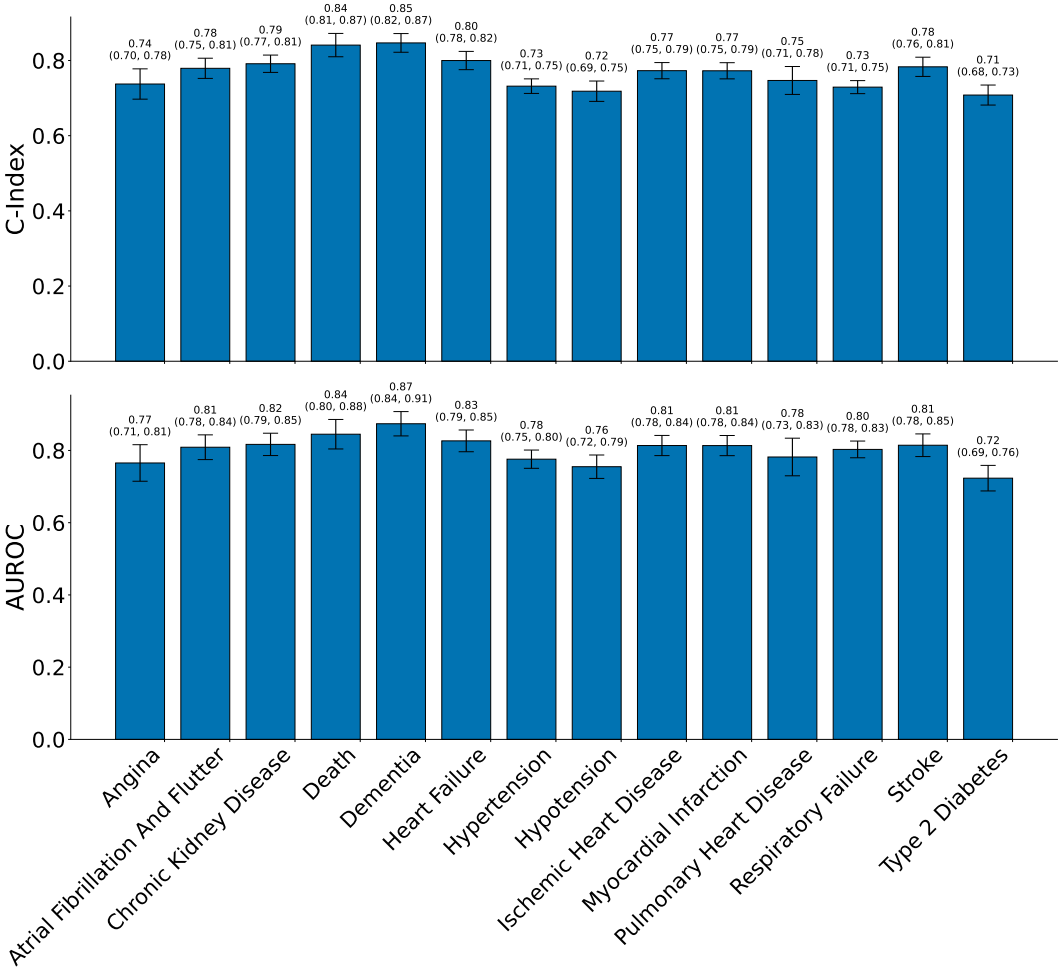


Figure 3. Prediction performance of SleepFM for key clinical outcomes on the test set, including C-Index and 6-year AUROC metrics for critical conditions such as death, heart failure, chronic kidney disease, dementia, and stroke. Error bars represent 95% confidence intervals (CI). All conditions are statistically significant with a p-value < 0.01 after Bonferroni correction. Both 95% CI and p-values were calculated using bootstrap resampling.

2.3 SleepFM Demonstrates Robust Generalization Across Time and cohorts

We evaluate SleepFM's generalization capabilities across two dimensions: temporal distribution shifts and external site validation. For temporal generalization, we test the model on a separate cohort comprising Stanford patients from 2020 onwards. All model pre-training and training was done on data prior to 2020. Despite the limited follow-up period (maximum four years) in this temporal test set, SleepFM maintains strong predictive performance. Supp. Fig. 5 shows results for our 14 selected conditions, with particularly robust and statistically significant performance (Bonferroni-corrected $p < 0.01$) for death (C-Index: 0.82 [95% CI: 0.78–0.87], AUROC: 0.83 [95% CI: 0.73–0.91]), heart failure (C-Index: 0.80 [95% CI: 0.76–0.83], AUROC: 0.80 [95% CI: 0.75–0.85]), and dementia (C-Index: 0.80 [95% CI: 0.76–0.84], AUROC: 0.83 [95% CI: 0.76–0.89]).

To assess cross-site generalization, we evaluate SleepFM's transfer learning capabilities on the SHHS, a dataset not used during pre-training. We extract embeddings using our pre-trained foundation model and fine-tune to predict specific conditions in SHHS. Due to differences in outcome data collection between SSC and SHHS cohorts, only a subset of conditions were available for evaluation in SHHS. This limited but important subset includes several key cardiovascular outcomes. This approach mimics real-world scenarios where foundation models must adapt to new clinical settings with minimal additional fine-tuning. As shown in Fig. 4, SleepFM demonstrates strong transfer learning performance across these available clinical outcomes. Notably, the model achieves statistically significant predictive accuracy (Bonferroni-corrected $p < 0.01$) for stroke (C-Index: 0.81 [95% CI: 0.76–0.85], AUROC: 0.82 [95% CI: 0.76–0.87]), congestive heart failure (C-Index: 0.83 [95% CI: 0.81–0.86], AUROC: 0.85 [95% CI: 0.82–0.88]), and cardiovascular disease-related mortality (C-Index: 0.86 [95% CI: 0.83–0.89], AUROC: 0.88 [95% CI: 0.83–0.91]).

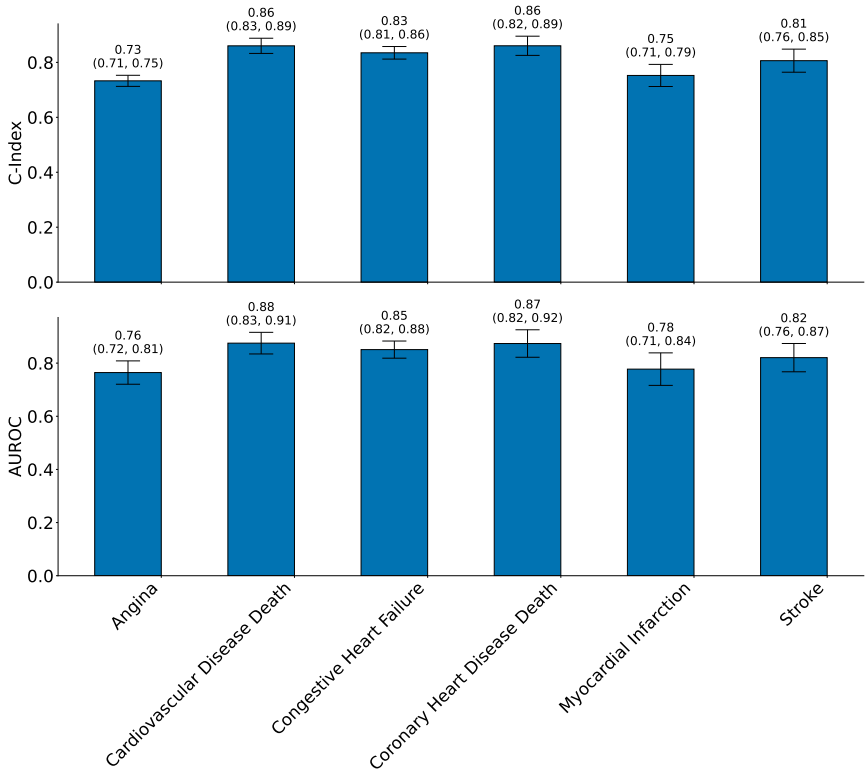


Figure 4. SleepFM prediction performance on the Sleep Heart Health Study (SHHS), an external dataset not used during pre-training. Due to differences in available outcome data between SHHS and Stanford datasets, evaluation was limited to a subset of conditions. Results demonstrate transfer learning capabilities across these key clinical outcomes, including stroke, congestive heart failure, and cardiovascular disease-related mortality. Performance is measured using C-Index and 6-year AUROC metrics, with error bars indicating 95% confidence intervals.

2.4 Differential Predictive Power Across Sleep Stages and Modalities

To evaluate the contribution of different components of sleep recordings to disease prediction, we analyzed the predictive power of individual sleep stages and signal modalities. For sleep stages, we fine-tuned separate models using recordings from individual stages (wake, stage 1/2, stage 3, and REM). Since stage 1 is a brief stage and is often combined with stage 2, we trained a model combining data from these two stages.

First, we evaluated the models on expert-selected conditions, as discussed in the previous section. The results are presented in Supp. Fig. 3. As shown in the heatmap, stage 1/2 and REM generally perform the best across multiple conditions. Stage 1/2 excel particularly in conditions such as angina, heart failure, and stroke, while REM performs exceptionally well for conditions like dementia and chronic kidney disease. Wake, stage 1/2, and REM show comparable performance in predicting mortality. In contrast, stage 3 is the worst-performing stage across multiple conditions.

To gain a broader understanding of the predictive power of different sleep stages, we analyzed the complete set of 1,041 phecodes from the PheWAS study and identified 62 conditions where at least one stage-specific model achieved a C-Index above

0.75. A comparison across these conditions is presented in Supp. Table 6, highlighting the strengths of individual sleep stages. Stage 1/2 demonstrates the highest predictive performance, excelling in 39 conditions, followed by REM sleep in 15 conditions and Wake in 6 conditions, while Stage 3 ranks the lowest with only 2 conditions. Among the strongest predictive associations, Stage 1/2 is most informative for conditions such as senile dementia, obstructive chronic bronchitis, type 2 diabetes with renal manifestations, and congestive heart failure, whereas REM sleep is particularly predictive for respiratory failure, hypertensive heart disease, and diabetic retinopathy.

Similar patterns of specialized predictive power are observed when analyzing individual signal modalities, as shown Supp. Fig. 4. For example, dementia is better predicted using BAS and RESP, while atrial fibrillation and flutter are more predictive using EKG and RESP. Mortality, heart failure, and stroke are associated with EKG and RESP, and BAS and RESP, respectively. Neurological conditions such as dementia and stroke are better predicted using BAS, while circulatory conditions like atrial fibrillation and pulmonary heart disease are better predicted using EKG. RESP is predictive across both neurological and circulatory conditions, underscoring its pivotal role across multiple physiological systems. EMG, however, is generally the least predictive among all modalities.

Expanding the analysis to all 1,041 phecodes, we identified 112 conditions where at least one modality-specific model achieved a C-Index above 0.75. EKG was the best-performing modality, excelling in 47 conditions, followed by BAS and RESP, each excelling in 32 conditions. EMG performed the worst overall. This analysis is summarized in Supp. Table 7. Like discussed before, BAS excels in predicting neurological and cognitive conditions, including autism, dementia, Parkinson's disease, Alzheimer's disease, and mild cognitive impairment. It also demonstrates strength in predicting developmental delays and disorders. Respiratory signals show high accuracy for pulmonary conditions and, intriguingly, certain skin conditions like melanomas, suggesting potential systemic manifestations in breathing patterns. EKG signals effectively predict cardiovascular conditions, including atrial flutter, atherosclerosis, and various forms of heart failure, aligning with their direct measurement of cardiac function.

Many conditions show complementary predictions across modalities. Respiratory failure and heart conditions are effectively predicted by both EKG and respiratory signals, while conditions like Parkinson's disease and dementia show strong signals in both BAS and respiratory modalities. Some conditions, such as developmental delays and disorders, demonstrate strong predictions across BAS, respiratory, and EKG signals. Notably, while EMG signals contribute to overall model performance, they do not show singular strong predictive power for any specific conditions, suggesting a more supportive role in conjunction with other modalities.

2.5 SleepFM Surpasses Supervised Baselines in Disease Prediction

We compare SleepFM against two baselines. The first is a demographics baseline, which includes features such as age, gender, body mass index (BMI), and race/ethnicity, implemented as a one-layer neural network. The second is an end-to-end PSG model that shares the same architecture as SleepFM but is trained directly on raw PSG signals without the foundation model pre-training stage. Both baselines are trained and evaluated on the same dataset as the fine-tuned SleepFM, using the same loss function and hyperparameters.

The overall performance comparison is shown in Supp. Fig. 6. The phecodes, represented by dots, are predominantly skewed below the diagonal line (towards the SleepFM axis), indicating that SleepFM generally outperforms both baselines. Compared to the demographics baseline, the average percentage improvement across all phecodes is 4.80% for the C-Index and 9.42% for the 6-year AUROC. For the end-to-end PSG model, the improvements are 4.48% for the C-Index and 8.88% for the 6-year AUROC.

Using the standard phecode-to-disease category mapping[35], we analyzed performance differences across various disease categories (Supp. Fig. 7). Compared to the demographics baseline, SleepFM demonstrates substantial improvements across neurological and mental disorders, circulatory conditions, endocrine/metabolic conditions, and respiratory conditions. In neurological and mental disorders, SleepFM achieves marked gains in conditions such as senile dementia (C-Index: 0.99 [95% CI: 0.98–1.00] vs. 0.87 [95% CI: 0.75–0.96]), myoneural disorders (0.81 [95% CI: 0.73–0.88] vs. 0.42 [95% CI: 0.28–0.55]), developmental delays and disorders (0.80 [95% CI: 0.77–0.84] vs. 0.58 [95% CI: 0.51–0.64]), and hallucinations (0.82 [95% CI: 0.71–0.92] vs. 0.58 [95% CI: 0.43–0.72]). For circulatory conditions, SleepFM notably outperforms in conditions such as atherosclerosis of native arteries (0.92 [95% CI: 0.88–0.95] vs. 0.74 [95% CI: 0.64–0.89]) and acute pulmonary heart disease (0.80 [95% CI: 0.75–0.85] vs. 0.74 [95% CI: 0.68–0.80]). Endocrine/metabolic conditions also show significant improvements, with disorders such as diabetes type 2 with peripheral circulatory disorders (0.87 [95% CI: 0.83–0.91] vs. 0.79 [95% CI: 0.74–0.85]) and diabetic retinopathy (0.81 [95% CI: 0.77–0.85] vs. 0.75 [95% CI: 0.69–0.80]). Additionally, SleepFM achieves superior performance in respiratory conditions, including respiratory insufficiency (0.79 [95% CI: 0.72–0.85] vs. 0.59 [95% CI: 0.51–0.67]) and respiratory failure (0.77 [95% CI: 0.73–0.80] vs. 0.70 [95% CI: 0.65–0.74]). A comprehensive comparison of performance between SleepFM and the demographics baseline is provided in Supp. Table 8 and Supp. Table 9.

The comparison with the end-to-end PSG model demonstrates the value of foundation model pre-training. Despite

both models having access to the same raw PSG signals and sharing the same architecture, SleepFM's self-supervised pre-training enables it to learn more robust and informative representations. This is evidenced by consistent improvements across neurological and mental disorders, circulatory conditions, endocrine/metabolic conditions, and respiratory conditions. In neurological and mental disorders, SleepFM outperforms the end-to-end model in conditions such as myoneural disorders (C-Index: 0.81 [95% CI: 0.73–0.88] vs. 0.60 [95% CI: 0.48–0.71]), developmental delays and disorders (0.80 [95% CI: 0.77–0.84] vs. 0.61 [95% CI: 0.54–0.67]), and speech and language disorders (0.81 [95% CI: 0.74–0.87] vs. 0.73 [95% CI: 0.64–0.82]). Circulatory conditions such as atherosclerosis of the extremities (0.81 [95% CI: 0.74–0.86] vs. 0.75 [95% CI: 0.69–0.81]) and acute pulmonary heart disease (0.80 [95% CI: 0.75–0.85] vs. 0.75 [95% CI: 0.69–0.81]) also show substantial improvements. Endocrine/metabolic conditions, including diabetes type 2 with peripheral circulatory disorders (0.87 [95% CI: 0.83–0.91] vs. 0.78 [95% CI: 0.72–0.84]), type 2 diabetes with neurological manifestations (0.81 [95% CI: 0.77–0.85] vs. 0.71 [95% CI: 0.68–0.75]), and diabetic retinopathy (0.81 [95% CI: 0.77–0.85] vs. 0.74 [95% CI: 0.69–0.79]) further highlight the advantages of pre-training. Improvements are also seen in respiratory conditions, such as respiratory insufficiency (C-Index: 0.79 [95% CI: 0.72–0.85] vs. 0.62 [95% CI: 0.53–0.70]) and respiratory failure (0.77 [95% CI: 0.73–0.80] vs. 0.72 [95% CI: 0.66–0.75]). A detailed performance comparison between SleepFM and the end-to-end PSG model is presented in Supp. Table 10 and Supp. Table 11.

In predicting all-cause mortality, SleepFM achieves a C-Index of 0.84 [95% CI: 0.81–0.87], significantly outperforming both the demographic baseline and end-to-end PSG model, which achieve C-Indices of 0.79 [95% CI: 0.75–0.82]. This improvement demonstrates the value of foundation model pre-training in capturing subtle mortality-related signals from PSG recordings beyond what is possible with demographic information alone.

3 Discussion

This study demonstrates the successful development and application of a large-scale foundation model for sleep analysis, developed on over 585,000 hours of PSG data from approximately 65,000 subjects. Our work makes several key contributions to the field of sleep medicine and biomedical research. First, we developed a foundation model that addresses fundamental challenges in sleep analysis through self-supervised learning - leveraging vast amounts of unlabeled data while handling variability in recording channels across different clinical sites. By designing a model that learns representations from unlabeled data and remains agnostic to differences in channel types and numbers, we enable broad exploration of sleep data across diverse clinical settings with minimal labeled data requirements for adaptation. Second, through extensive evaluation across 1,041 disease phenotypes, we demonstrate sleep's broad predictive power for diverse health outcomes. The model shows particularly strong performance in predicting death (C-Index: 0.84 [95% CI: 0.81–0.87]), heart failure (C-Index: 0.80 [95% CI: 0.77–0.83]), chronic kidney disease (C-Index: 0.79 [95% CI: 0.77–0.81]), and dementia (C-Index: 0.85 [95% CI: 0.82–0.87]). Third, we demonstrated remarkable transfer learning capabilities, a hallmark of foundation models, through strong performance on the SHHS dataset. Despite SHHS being entirely excluded from pre-training, our model maintains robust predictive power for key outcomes such as stroke (C-Index: 0.81 [95% CI: 0.76–0.85]), congestive heart failure (C-Index: 0.83 [95% CI: 0.81–0.86]), and death related to cardiovascular disease (C-Index: 0.86 [95% CI: 0.83–0.89]). Finally, SleepFM achieves competitive performance on standard sleep analysis tasks, including sleep staging and apnea detection. The model achieves mean F1 scores ranging from 0.69 to 0.78 for sleep staging across cohorts and an AUROC of 0.94 for apnea detection in the SHHS dataset, on par with state-of-the-art models such as U-Sleep and YASA[15,49], which are specifically optimized for these tasks.

SleepFM performs age estimation as a regression task, capturing aging-related information through pre-training embeddings, with a mean absolute error (MAE) of 7.3 years and a Pearson correlation coefficient of 0.88 across a wide age range (0–100 years). While previous studies using supervised learning have achieved an MAE as low as 4.19 years with PSG data[4], these results are not directly comparable due to differences in age distributions. For apnea diagnosis, SleepFM achieves an AUROC of 0.94 on the unseen SHHS dataset, comparable to models specifically trained for this task[13,50,51].

SleepFM predicts all-cause mortality more accurately than both the demographics-based model and the end-to-end PSG model, achieving a higher C-Index of 0.84 [95% CI: 0.81–0.87], compared to 0.79 [95% CI: 0.75–0.82] for both baselines. This improvement suggests that pre-training captures better subtle signals in the PSG data that are linked to mortality risk. Research shows that all-cause mortality is strongly associated with various sleep-related factors, including high arousal burden[52], low REM sleep proportion[53], untreated sleep-disordered breathing[54], hypoxemia, and low sleep efficiency[55,56]. Increased "brain age" derived from EEG has also been identified as a significant predictor of mortality[4,57]. SleepFM likely integrates many of these multifactorial contributors, capturing respiratory events, sleep fragmentation, arousal burden, and sleep efficiency, along with markers of cardiovascular, metabolic, and other diseases. These findings highlight the richness of sleep data in reflecting an individual's overall health trajectory, offering a valuable tool for early risk stratification and intervention.

Predictive and prognostic models for neurological and mental disorders are rapidly advancing, offering the potential for earlier and more individualized treatment when interventions are most effective. Among the top conditions predicted by SleepFM were Alzheimer's disease and Parkinson's disease, with C-indices of 0.91 [95% CI: 0.87–0.98] and 0.89 [95% CI:

0.85–0.92], respectively. Sleep disorders are strongly associated with preclinical Alzheimer's disease[58], including abnormalities in Non-REM sleep, such as reduced slow-wave activity[59], REM sleep disturbances[60,61], and decreased spindle activity[62]. In early Alzheimer's disease, REM sleep abnormalities have been linked to basal forebrain cholinergic lesions, which likely contribute to cognitive decline[63]. Similarly, Parkinson's disease is frequently preceded by REM sleep behavior disorder (RBD), characterized by REM sleep without atonia and abnormalities in BAS and ECG patterns[64–66]. Recent studies have also shown that respiratory signals can capture phenotypes specific to Parkinson's disease[67].

Consistent with these findings, SleepFM identified BAS as the strongest predictor of neurological and mental disorders, while respiratory signals were particularly effective in predicting senile dementia, suggesting that they contain information beyond sleep depth and apnea. Most studies in this domain rely on imaging modalities such as MRI and fMRI to predict dementia. For example, one study using hippocampal MRI achieved a C-index of 0.864[68], while another using fMRI reported an AUROC of 0.824 for predicting dementia up to nine years in advance[69]. Although direct performance comparisons are challenging due to differences in sample distributions, the ability of SleepFM to leverage PSG data to predict neurological and mental disorders underscores its potential as an alternative to imaging-based approaches.

Other established biomarkers for Alzheimer's disease—such as amyloid PET, decreased CSF β-amyloid$_{42}$, and increased CSF phosphorylated tau (e.g., p-tau$_{129}$)[70–72]—have been widely used for diagnosis and prognosis. More recently, plasma p-tau$_{217}$ has emerged as a promising less invasive marker[73]. Sleep biomarkers from PSG data offer a complementary, non-invasive tool for the prognosis of dementia and mild cognitive impairment. Future research could explore an integrated approach combining PSG with established imaging, CSF, and plasma markers to potentially enhance predictive accuracy.

SleepFM accurately modeled cardiovascular disease in both the SSC and SHHS datasets, leveraging data-driven methods commonly used in prognostic modeling of cardiovascular disease, particularly with ECG data[74] and lead II ECG from PSG studies[14]. Foundation models have demonstrated state-of-the-art performance with ECG data in various cross-sectional tasks[23]. For predicting cardiovascular mortality over 10 years, a previous study reported an AUROC of 0.84 [95% CI: 0.78–0.89] in a subset of SHHS participants with sleep apnea, whereas SleepFM achieved a slightly higher AUROC of 0.88 [95% CI: 0.83–0.91]. Similarly, for atrial fibrillation, earlier work reported an AUROC of 0.82[14], which aligns with SleepFM's performance of 0.81 [95% CI: 0.78–0.84]. Our ablation study further demonstrated that both ECG and respiratory signals contribute to the prediction of circulatory system phenotypes, suggesting that SleepFM integrates information on sleep apnea and heart activity in ways that are consistent with known disease mechanisms[75].

Most disease categories, including respiratory, endocrine/metabolic, and neoplasms, were predicted with significantly improved performance by SleepFM compared to the demographics-based and end-to-end PSG baseline models. Many of these diseases are either associated with sleep (e.g., type 2 diabetes[76]) or directly influenced by the signal modalities (e.g., heart arrhythmia), providing predictive power for a wide range of phenotypes. Disrupted and unhealthy sleep contributes to dysfunction across multiple physiological systems, increasing the risk of diseases such as obesity, type 2 diabetes, hypertension, stroke, and cardiovascular disease[77,78]. Sleep-specific conditions, including sleep apnea[75] and less conclusively periodic leg movements[79], are also linked to cardiovascular outcomes. Additionally, specific EEG waveforms, such as coupled slow-wave and spindle activity, have been identified as markers of next-day blood glucose regulation[80]. SleepFM excels in integrating information across all modalities, making it particularly effective for diseases with complex interactions between physiological systems.

Despite these promising results, several limitations should be acknowledged. While our dataset is large, it primarily consists of patients referred for sleep studies due to suspected sleep disorders or other medical conditions requiring overnight monitoring. This selection bias means our cohort is not representative of the general population, as individuals without sleep complaints or those with limited access to specialized sleep clinics are underrepresented. Future work should validate these findings in more diverse populations. The model's performance shows some degradation in temporal test sets, highlighting the challenge of maintaining predictive accuracy over time as clinical practices and patient populations evolve. Additionally, interpreting the predictions made by SleepFM is inherently challenging due to the complexity of the learned features during training by a deep CNN and transformer based models. To mitigate this, we stratified the model's performance across sleep stages and data modalities, and conducted evaluations on temporal test sets and unseen datasets to gain insights into its behavior. However, further work is needed to enhance case-level interpretability and understand the specific sleep patterns and features driving these predictions. Additional studies are needed to evaluate the model's impact on clinical decision-making and patient outcomes in real-world settings.

Methods

Dataset and Preprocessing

Our dataset includes PSG recordings from four different sites: SSC, BioSerenity, MESA[30,31], and MROS[30,32], with SHHS[30,33] serving as an external validation dataset. Among these, MESA, MROS, and SHHS are publicly available datasets, while SSC is

Table 1. Distribution of polysomnography recordings across different sleep centers and dataset splits. The model is first pre-trained on SSC, BioSerenity, MESA, and MROS data, then these same recordings are used for task-specific fine-tuning. SHHS dataset is reserved exclusively for evaluating transfer learning capabilities and is only used during fine-tuning, not during pre-training. The temporal test set consists of SSC recordings from 2020 onwards, used to evaluate model robustness to temporal distribution shifts. Dashes (-) indicate no data available for that split.

Split	**SSC**	**BioSerenity**	**MESA**	**MROS**	**SHHS**[†]	**Total**
Train	24,137	18,869	1,747	3,340	3,291	**51,384**
Validation	764	100	10	18	500	**1,392**
Test	5,019	—	150	572	2,000	**7,741**
Temporal Test	5,132	—	—	—	—	**5,132**
Total	**35,052**	**18,969**	**1,907**	**3,930**	**5,791**	**65,649**

[†] SHHS data used only for fine-tuning, not included in pre-training

our proprietary dataset. The BioSerenity dataset, provided by the BioSerenity company, contains 18,869 overnight recordings lasting 7-11 hours each. This dataset is a subset of a larger collection from SleepMed and BioSerenity sleep laboratories, gathered between 2004 and 2019 across 240 US facilities[29]. At the time of this study, approximately 20K de-identified PSGs were available for analysis. The dataset distribution across different splits is shown in Table 1, with SSC constituting the largest cohort. To prevent data leakage, participants with multiple PSG recordings were assigned to a single split. For MESA, MROS, and SHHS details, we refer readers to their original publications. Below, we describe our internal SSC dataset in more detail.

The SSC dataset comprises 35,052 recordings, each lasting approximately 8 hours overnight. It includes diverse waveforms such as BAS, ECG, EMG, and respiratory channels, making it a high-quality resource for sleep-related research. The dataset spans recordings from 1999 to 2024 and includes participants aged 2 to 96. The patient demographic statistics for SSC and BioSerenity are summarized in Supp. Table 1 and Supp. Table 2 respectively.

Our preprocessing strategy minimizes alterations to preserve raw signal characteristics crucial for identifying nuanced patterns. Each recording contains up to four modalities: BAS, ECG, EMG, and respiratory, with variable numbers of channels. For BAS, we allowed up to 10 channels, for ECG 2 channels, for EMG 4 channels, and for respiratory 7 channels. The number and type of channels vary across sites and even between patients within the same site, depending on the study type. The types of channels available across sites are described in Supp. Table 12 to Supp. Table 16. BAS includes channels that measure brain activity from different regions (frontal, central, occipital) as well as EOG for eye movements. EMG records electrical activity in muscles, while ECG captures cardiac electrical function. Respiratory channels measure chest and abdominal movements, pulse readings, and nasal/oral airflow. These channels were selected based on their relevance to sleep studies, guided by sleep experts[1].

Each PSG recording is resampled to 128 Hz to standardize sampling rates across participants and sites. Signals are then segmented into 5-second patches, with each segment embedded into a vector representation for transformer model processing. To prevent data leakage, PSGs were split into pre-train, train, validation, test, and temporal test sets early in the preprocessing pipeline. While there is overlap between the pre-training and training sets, no overlap exists with the validation, test, or temporal test sets. The SHHS serves as an independent dataset not used during pre-training, instead being used to evaluate the model's ability to adapt to a new site through lightweight fine-tuning.

During pre-training, the only required labels are the modality types of the signals. A self-supervised contrastive learning (CL) objective is employed for pre-training. For downstream evaluations, we consider canonical tasks such as age/gender prediction, sleep stage classification, sleep apnea diagnosis, and various patient conditions extracted from EHR data. Sleep staging and apnea labels for SSC, MESA, MROS, and SHHS were annotated by sleep experts. SHHS also includes diagnostic information for conditions such as myocardial infarction, stroke, angina, congestive heart failure, and death. For SSC, we paired PSG data with Stanford EHR data using de-identified patient IDs to extract demographic and diagnostic information. As BioSerenity lacks associated labels, it was used exclusively for pre-training.

SleepFM Model Architecture

Our model architecture is illustrated in Fig. 1. The architecture includes several key components that differ slightly between the pre-training and fine-tuning stages. During pre-training, we employ contrastive learning (CL) as the objective function for representation learning, which is explained in detail in Section 3. A single model processes all four modalities.

The first component of the architecture is the *Encoder*, a 1D convolutional neural network (CNN) that processes raw signal data for each modality separately. The encoder takes raw input vectors, where the length of each vector corresponds to a 5-second segment of the signal, referred to as a token. The input dimensions are (B,T,C), where B is the batch size, T

is the raw temporal length of the input, and C is the number of channels for each modality. These inputs are reshaped into (B,C,S,L), where S is the sequence length representing the number of tokens ($S = T/L$), and L corresponds to the raw vector length for a single token (e.g., 640 samples). Each token is then processed individually through a stack of six convolutional layers, each followed by normalization and ELU activation layers. These layers progressively reduce the temporal resolution while increasing the number of feature channels, converting the input from one channel to 128 channels. After this, adaptive average pooling further reduces the temporal dimensions, and a fully connected layer compresses the representation into a 128-dimensional embedding for each token. The final output of the encoder has dimensions (B,C,S,D), where $D = 128$.

Following the encoder, a sequence of transformer-based operations is applied to extract and aggregate modality-specific and temporal features. The first step is *channel pooling*, which aggregates token embeddings from all channels within a given modality. This operation uses an attention pooling mechanism based on a transformer layer to compute attention scores for each channel and produces a single aggregated embedding per time segment by averaging over the channel dimension. The resulting embeddings, with dimensions (B,S,D), are then passed through a *temporal transformer*, which operates along the temporal dimension to capture dependencies between tokens. The *temporal transformer* applies sinusoidal positional encoding to the token embeddings, followed by 2 transformer blocks consisting of self-attention and feedforward layers, enabling the model to learn contextual relationships across the sequence. After temporal modeling, the embeddings are processed through *temporal pooling*, which aggregates token embeddings over the sequence length (S) for each modality. Similar to *channel pooling*, *temporal pooling* uses an attention mechanism to compute weighted averages, generating a compact representation of size $(B,128)$ per modality. These steps collectively ensure that the model captures both spatial and temporal dependencies while reducing dimensionality for computational efficiency.

The final output is a single 128-dimensional embedding for each modality, used for contrastive learning during pre-training (detailed in Section 3). While the 5-minute recordings are used exclusively for pre-training, we retain the 5-second-level embeddings for each modality for downstream tasks such as sleep staging and disease classification.

Baseline Models

We evaluate SleepFM against two carefully chosen baseline approaches to demonstrate the value of our foundation model methodology. The first baseline is a simple demographic model that processes only patient characteristics, including age, gender, body-mass index (BMI), and race/ethnicity information. This demographic baseline is implemented as a one-layer neural network to establish a minimum performance threshold using only basic patient data available in most clinical settings.

The second baseline is a more sophisticated end-to-end PSG model that directly processes raw sleep recordings. This model uses an identical architecture to SleepFM, including the same CNN encoder, transformer layers, and fully-connected layer, but crucially, it is trained directly on downstream tasks without the benefit of foundation model pre-training.

Both baseline models are trained and evaluated using the same dataset, optimization objectives, and hyperparameters as the fine-tuned SleepFM to ensure a fair comparison. This experimental design allows us to quantify both the advantage of incorporating PSG data over demographics alone and the additional benefit gained through our foundation model approach compared to direct task-specific training.

Model Training

Model training can be categorized into two segments: pre-training and fine-tuning. Pre-training stage involves self-supervised representation learning with a CL objective and fine-tuning involves training the model with supervised learning objective for specific tasks such as sleep stage classification, sleep apnea diagnosis, and disease prediction. We describe these in more details below.

Pre-training

Model pre-training is performed using a self-supervised learning objective called CL. Specifically, we employ a CL objective for multiple modalities, referred to as leave-one-out contrastive learning (LOO-CL), which we proposed in our previous work[16]. The key idea behind CL is to bring positive pairs of embeddings from different modalities closer in the latent space while pushing apart negative pairs. Positive pairs are derived from temporally aligned 5-minute aggregated embeddings, obtained after temporal pooling as described in Section 3, across four different modalities. All other non-matching instances within a training batch are treated as negative pairs.

In LOO-CL, we define a predictive task where an embedding from one modality attempts to identify the corresponding embeddings from the remaining modalities. For each modality i, we construct an embedding $\bar{x}_k^{-i}$ by averaging over embeddings from all other modalities, excluding modality i. We then apply a contrastive loss between the embedding of modality i and this leave-one-out representation:

$$\mathscr{L}_{i,k} = -\log \frac{\exp\left(\mathrm{sim}(x_k^i, \bar{x}_k^{-i})/\tau\right)}{\sum_{m=1}^{N} \exp\left(\mathrm{sim}(x_k^i, \bar{x}_m^{-i})/\tau\right)},$$

where $\mathscr{L}_{i,k}$ is the loss for a sample k from modality i in a given batch, $\text{sim}(\cdot)$ represents a similarity function (e.g., cosine similarity), and τ is a temperature scaling parameter. The numerator computes the similarity between the embedding of modality i and the leave-one-out representation of the corresponding sample, while the denominator sums the similarities across all samples within the batch. The motivation behind the leave-one-out method is to encourage each embedding to align semantically with all other modalities.

Fine-tuning

After pre-training with the CL objective, we extract 5-second embeddings for all patient PSG data across modalities. We standardize the temporal context to 9 hours for all patients - longer recordings are cropped and shorter ones are zero-padded to ensure consistent input dimensions. For example, for a patient's standardized 9-hour sleep data, the resulting patient matrix has dimensions $(4 \times 6480 \times 128)$, where 4 represents the number of modalities, 6480 is the number of 5-second embeddings for 9 hours of sleep, and 128 is the embedding vector dimension.

During fine-tuning, we first apply a channel pooling operation across different modalities, reducing the dimensions to (6480×128) for our example patient matrix. The pooled embeddings are then processed through a 2-layer LSTM block, which is designed to handle temporal sequences. For sleep staging tasks, these 5-second embeddings are directly passed through a classification layer. For all other tasks, the embeddings are first pooled along the temporal dimension before being passed through an output layer.

For disease classification, we append age and gender features to the mean-pooled embedding vector after the LSTM block, before passing it to the final output layer. This addition empirically improves performance and surpasses the demographic baseline, as demonstrated in Supp. Fig. 6.

The fine-tuning objective for disease prediction uses the CoxPH loss function, a standard approach in survival analysis for modeling time-to-event data. The CoxPH loss maximizes the partial likelihood and is defined for a single label as:

$$\mathscr{L}_{\text{CoxPH}} = -\frac{1}{N_e} \sum_{i=1}^{n} \delta_i \left(h_i - \log \sum_{j \in R(t_i)} \exp(h_j) \right),$$

where h_i is the predicted hazard for the i-th patient, δ_i is the event indicator (1 for event occurrence, 0 otherwise), t_i is the event or censoring time, $R(t_i)$ represents the risk set of all patients with event times greater than or equal to t_i, n is the total number of patients, and $N_e = \sum_{i=1}^{n} \delta_i$ is the number of events.

For our multi-label setup with 1,041 labels, we extend the CoxPH loss by computing it independently for each label and summing the results:

$$\mathscr{L}_{\text{total}} = \sum_{k=1}^{L} \mathscr{L}_{\text{CoxPH}}^{(k)},$$

where L is the total number of labels.

Given the large dataset size, computing the loss for all patients in a single batch is computationally infeasible. Therefore, we calculate the loss in smaller batches of 32 samples, with patients sorted by event time in descending order to ensure correct computation of the partial likelihood. This batching strategy, combined with the summation of per-label losses, provides an efficient and scalable approach for multi-label time-to-event modeling.

Implementation Details

All implementations were carried out using PyTorch, a widely-used library for deep learning[81]. For pre-training, the model was trained with a batch size of 32, a learning rate of 0.001, 8 pooling heads, 3 transformer layers, and a dropout rate of 0.3. As previously described, each patch size corresponds to a 5-second segment, and the total sequence length is 5 minutes for the transformer model. The total parameter count for the model was approximately 4.44 million. Pre-training was performed on 432,000 hours of sleep data collected from 48,000 participants for one epoch, using an NVIDIA A100 GPU. The entire pre-training process took approximately 15 hours.

For fine-tuning, the batch size was also set to 32, with a learning rate of 0.001, 4 pooling heads, 2 LSTM layers, and a dropout rate of 0.3. The fine-tuned model had approximately 0.88 million learnable parameters. Training was conducted on patient data, with each token embedding represented as a 128-dimensional vector, over 10 epochs. The fine-tuning process was performed on an NVIDIA A100 GPU, with the total training time per epoch ranging from 2 to 5 minutes, depending on the task.

Data Availability

Of the five data sources used in this study, three datasets are publicly available and can be accessed at the following links: SHHS (https://archive.physionet.org/physiobank/database/shhpsgdb/), MrOS (https://sleepdata.

org/datasets/mros), and MESA (https://sleepdata.org/datasets/mesa). These datasets are also cited in Section 3. The Bioserenity dataset is proprietary and has been shared with Stanford for research and development purposes. We will also release the SSC PSG data upon paper publication.

Code Availability

All of the SleepFM code is open source and available at https://github.com/zou-group/sleepfm-clinical

Ethics & Inclusion Statement

This study was conducted in collaboration with researchers across multiple institutions, ensuring equitable contributions to study design, data collection, analysis, and manuscript preparation. All co-authors meet the authorship criteria outlined by Nature Portfolio. Ethical approval was obtained from the Stanford University Institutional Review Board (protocol number: 69873), and all procedures adhered to applicable ethical guidelines. Informed consent was obtained from all human research participants. The study prioritized participant safety, confidentiality, and regulatory compliance. No risks of stigmatization, discrimination, or harm were introduced, and all necessary measures were taken to ensure data privacy. No biological materials, cultural artifacts, or traditional knowledge were transferred out of the country in a manner requiring benefit-sharing agreements. Relevant local and regional research has been appropriately cited to ensure inclusivity and acknowledgment of prior work.

References

1. Berry, R. B. *et al.* The AASM manual for the scoring of sleep and associated events. *Rules, Terminol. Tech. Specif. Darien, Illinois, Am. Acad. Sleep Medicine* **176**, 2012 (2012).
2. Kryger, M. H., Roth, T. & Dement, W. C. Principles and practice of sleep medicine fifth edition (2010).
3. Lim, D. C. *et al.* The need to promote sleep health in public health agendas across the globe. *The Lancet Public Heal.* **8**, e820–e826 (2023).
4. Brink-Kjaer, A. *et al.* Age estimation from sleep studies using deep learning predicts life expectancy. *NPJ Digit.* **5**, 103 (2022).
5. Leary, E. B., Stone, K. L. & Mignot, E. Living to dream—reply. *JAMA Neurol.* **78**, 495–496 (2021).
6. Riemann, D. Insomnia and comorbid psychiatric disorders. *Sleep medicine* **8**, S15–S20 (2007).
7. André, C. *et al.* Association of sleep-disordered breathing with alzheimer disease biomarkers in community-dwelling older adults: a secondary analysis of a randomized clinical trial. *JAMA Neurol.* **77**, 716–724 (2020).
8. Addo, P. N. O. *et al.* Associations between sleep duration, sleep disturbance and cardiovascular disease biomarkers among adults in the united states. *BMC Public Heal.* **24**, 947 (2024).
9. Uwishema, O. *et al.* The pulse of sleep: novel interventions in understanding the sleep-cardiovascular connection: a literature review. *Annals Medicine Surg.* **86**, 5283–5291 (2024).
10. Boashash, B. & Ouelha, S. Automatic signal abnormality detection using time-frequency features and machine learning: A newborn EEG seizure case study. *Knowledge-Based Syst.* **106**, 38–50 (2016).
11. Hassan, A. R. & Bhuiyan, M. I. H. Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting. *Comput. methods programs biomedicine* **140**, 201–210 (2017).
12. Perslev, M. *et al.* U-sleep: resilient high-frequency sleep staging. *NPJ Digit.* **4**, 72 (2021).
13. Nassi, T. E. *et al.* Automated scoring of respiratory events in sleep with a single effort belt and deep neural networks. *IEEE transactions on biomedical engineering* **69**, 2094–2104 (2021).
14. Koscova, Z. *et al.* From sleep patterns to heart rhythms: Predicting atrial fibrillation from overnight polysomnograms. *medRxiv* 2024–06 (2024).
15. Stephansen, J. B. *et al.* Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy. *Nat. Commun.* **9**, 5229 (2018).
16. Thapa, R. *et al.* SleepFM: Multi-modal representation learning for sleep across brain activity, ECG and respiratory signals. *arXiv preprint arXiv:2405.17766* (2024).
17. Liu, G.-R. *et al.* Large-scale assessment of consistency in sleep stage scoring rules among multiple sleep centers using an interpretable machine learning algorithm. *J. Clin. Sleep Medicine* **17**, 159–166 (2021).

18. Bommasani, R. *et al.* On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258* (2021).

19. Saab, K. *et al.* Capabilities of Gemini models in medicine. *arXiv preprint arXiv:2404.18416* (2024).

20. Xu, H. *et al.* A whole-slide foundation model for digital pathology from real-world data. *Nature* 1–8 (2024).

21. Zhao, T. *et al.* BiomedParse: a biomedical foundation model for image parsing of everything everywhere all at once. *arXiv preprint arXiv:2405.12971* (2024).

22. Nguyen, E. *et al.* Sequence modeling and design from molecular to genome scale with Evo. *Science* **386**, eado9336 (2024).

23. McKeen, K. *et al.* Ecg-fm: An open electrocardiogram foundation model. *arXiv preprint arXiv:2408.05178* (2024).

24. Cui, W. *et al.* Neuro-gpt: Towards a foundation model for eeg. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, 1–5 (IEEE, 2024).

25. Abbaspourazad, S. *et al.* Large-scale training of foundation models for wearable biosignals. *arXiv preprint arXiv:2312.05409* (2023).

26. Pendergrass, S. *et al.* The use of phenome-wide association studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery. *Genet. epidemiology* **35**, 410–422 (2011).

27. Wei, W.-Q. *et al.* Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. *PloS one* **12**, e0175508 (2017).

28. Hanif, U. *et al.* Automatic detection of chronic insomnia from polysomnographic and clinical variables using machine learning. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 1–5 (IEEE, 2023).

29. Hanif, U. *et al.* Associations between self-reported parasomnias and psychiatric illness in 370,000 patients with sleep disorders. *Psychiatry Clin. Neurosci.* **78**, 667–677 (2024).

30. Zhang, G.-Q. *et al.* The national sleep research resource: towards a sleep data commons. *J. Am. Med. Informatics Assoc.* **25**, 1351–1358 (2018).

31. Chen, X. *et al.* Racial/ethnic differences in sleep disturbances: the multi-ethnic study of atherosclerosis (MESA). *Sleep* **38**, 877–888 (2015).

32. Blackwell, T. *et al.* Associations between sleep architecture and sleep-disordered breathing and cognition in older community-dwelling men: the osteoporotic fractures in men sleep study. *J. Am. Geriatr. Soc.* **59**, 2217–2225 (2011).

33. Quan, S. F. *et al.* The sleep heart health study: design, rationale, and methods. *Sleep* **20**, 1077–1085 (1997).

34. Younes, M., Raneri, J. & Hanly, P. Staging sleep in polysomnograms: analysis of inter-scorer variability. *J. Clin. Sleep Medicine* **12**, 885–894 (2016).

35. Wu, P. *et al.* Mapping ICD-10 and ICD-10-CM codes to phecodes: workflow development and initial evaluation. *JMIR medical informatics* **7**, e14325 (2019).

36. Wennberg, A. M., Wu, M. N., Rosenberg, P. B. & Spira, A. P. Sleep disturbance, cognitive decline, and dementia: a review. In *Seminars in neurology*, vol. 37, 395–406 (Thieme Medical Publishers, 2017).

37. Stefani, A. & Högl, B. Sleep in parkinson's disease. *Neuropsychopharmacology* **45**, 121–128 (2020).

38. Ravichandran, R. *et al.* The interplay between sleep disorders and cardiovascular diseases: A systematic review. *Cureus* **15** (2023).

39. Song, C. *et al.* Sleep quality and risk of cancer: findings from the english longitudinal study of aging. *Sleep* **44**, zsaa192 (2021).

40. Shigesato, M. *et al.* Association between sleep duration and breast cancer incidence: the multiethnic cohort. *Int. journal cancer* **146**, 664–670 (2020).

41. Martinez-Garcia, M. A. *et al.* Cancer and sleep apnea: cutaneous melanoma as a case study. *Am. journal respiratory critical care medicine* **200**, 1345–1353 (2019).

42. Freeman, J. R. *et al.* Actigraphy-derived measures of sleep and risk of prostate cancer in the uk biobank. *JNCI: J. Natl. Cancer Inst.* **116**, 434–444 (2024).

43. Windred, D. P. *et al.* Sleep regularity is a stronger predictor of mortality risk than sleep duration: A prospective cohort study. *Sleep* **47**, zsad253 (2024).

44. Westwood, A. J. *et al.* Prolonged sleep duration as a marker of early neurodegeneration predicting incident dementia. *Neurology* **88**, 1172–1179 (2017).

45. Shi, L. *et al.* Sleep disturbances increase the risk of dementia: a systematic review and meta-analysis. *Sleep medicine reviews* **40**, 4–16 (2018).

46. Darraj, A. The link between sleeping and type 2 diabetes: A systematic review. *Cureus* **15** (2023).

47. Mc Carthy, C. E. *et al.* Sleep patterns and the risk of acute stroke: results from the interstroke international case-control study. *Neurology* **100**, e2191–e2203 (2023).

48. Cappuccio, F. P., Cooper, D., D'Elia, L., Strazzullo, P. & Miller, M. A. Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. *Eur. heart journal* **32**, 1484–1492 (2011).

49. Vallat, R. & Walker, M. P. An open-source, high-performance tool for automated sleep staging. *Elife* **10**, e70092 (2021).

50. Van Steenkiste, T., Groenendaal, W., Deschrijver, D. & Dhaene, T. Automated sleep apnea detection in raw respiratory signals using long short-term memory neural networks. *IEEE journal biomedical health informatics* **23**, 2354–2364 (2018).

51. Lakhan, P., Ditthapron, A., Banluesombatkul, N. & Wilaiprasitporn, T. Deep neural networks with weighted averaged overnight airflow features for sleep apnea-hypopnea severity classification. In *TENCON 2018-2018 IEEE region 10 conference*, 0441–0445 (Ieee, 2018).

52. Shahrbabaki, S. S., Linz, D., Hartmann, S., Redline, S. & Baumert, M. Sleep arousal burden is associated with long-term all-cause and cardiovascular mortality in 8001 community-dwelling older men and women. *Eur. heart journal* **42**, 2088–2099 (2021).

53. Leary, E. B. *et al.* Association of rapid eye movement sleep with mortality in middle-aged and older adults. *JAMA neurology* **77**, 1241–1251 (2020).

54. Young, T. *et al.* Sleep disordered breathing and mortality: eighteen-year follow-up of the wisconsin sleep cohort. *Sleep* **31**, 1071–1078 (2008).

55. Wallace, M. L. *et al.* Physiological sleep measures predict time to 15-year mortality in community adults: application of a novel machine learning framework. *J. sleep research* **30**, e13386 (2021).

56. Yan, B. *et al.* Objective sleep efficiency predicts cardiovascular disease in a community population: the sleep heart health study. *J. Am. Hear. Assoc.* **10**, e016201 (2021).

57. Paixao, L. *et al.* Excess brain age in the sleep electroencephalogram predicts reduced life expectancy. *Neurobiol. aging* **88**, 150–155 (2020).

58. Bubu, O. M. *et al.* Sleep, cognitive impairment, and alzheimer's disease: a systematic review and meta-analysis. *Sleep* **40**, zsw032 (2017).

59. Ju, Y.-E. S. *et al.* Slow wave sleep disruption increases cerebrospinal fluid amyloid-β levels. *Brain* **140**, 2104–2111 (2017).

60. Falgàs, N. & Walsh, C. M. The importance of rem sleep and its implications for alzheimer's disease (2024).

61. Montplaisir, J., Petit, D., Lorrain, D. & Gauthier, S. Sleep in alzheimer's disease: further considerations on the role of brainstem and forebrain cholinergic populations in sleep-wake mechanisms. *Sleep: J. Sleep Res. & Sleep Medicine* (1995).

62. Weng, Y.-Y., Lei, X. & Yu, J. Sleep spindle abnormalities related to alzheimer's disease: a systematic mini-review. *Sleep medicine* **75**, 37–44 (2020).

63. André, C. *et al.* Rem sleep is associated with the volume of the cholinergic basal forebrain in amci individuals. *Alzheimer's Res. & Ther.* **15**, 151 (2023).

64. Brink-Kjaer, A. *et al.* End-to-end deep learning of polysomnograms for classification of rem sleep behavior disorder. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2941–2944 (IEEE, 2022).

65. Brink-Kjær, A. *et al.* Arousal characteristics in patients with parkinson's disease and isolated rapid eye movement sleep behavior disorder. *Sleep* **44**, zsab167 (2021).

66. Sorensen, G. L., Kempfner, J., Zoetmulder, M., Sorensen, H. B. & Jennum, P. Attenuated heart rate response in rem sleep behavior disorder and parkinson's disease. *Mov. disorders* **27**, 888–894 (2012).

67. Yang, Y. *et al.* Artificial intelligence-enabled detection and assessment of parkinson's disease using nocturnal breathing signals. *Nat. medicine* **28**, 2207–2215 (2022).

68. Li, H. *et al.* A deep learning model for early prediction of alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. *Alzheimer's & Dementia* **15**, 1059–1070 (2019).

69. Ereira, S., Waters, S., Razi, A. & Marshall, C. R. Early detection of dementia with default-mode network effective connectivity. *Nat. Mental Heal.* 1–14 (2024).

70. Blennow, K. & Hampel, H. Csf markers for incipient alzheimer's disease. *The Lancet Neurol.* **2**, 605–613 (2003).

71. Hansson, O. *et al.* Association between csf biomarkers and incipient alzheimer's disease in patients with mild cognitive impairment: a follow-up study. *The Lancet Neurol.* **5**, 228–234 (2006).

72. Klunk, W. E. *et al.* Imaging brain amyloid in alzheimer's disease with pittsburgh compound-b. *Annals Neurol. Off. J. Am. Neurol. Assoc. Child Neurol. Soc.* **55**, 306–319 (2004).

73. Janelidze, S. *et al.* Plasma p-tau181 in alzheimer's disease: relationship to other biomarkers, differential diagnosis, neuropathology and longitudinal progression to alzheimer's dementia. *Nat. medicine* **26**, 379–386 (2020).

74. Moreno-Sánchez, P. A. *et al.* Ecg-based data-driven solutions for diagnosis and prognosis of cardiovascular diseases: A systematic review. *Comput. Biol. Medicine* 108235 (2024).

75. Mitra, A. K., Bhuiyan, A. R. & Jones, E. A. Association and risk factors for obstructive sleep apnea and cardiovascular diseases: a systematic review. *Diseases* **9**, 88 (2021).

76. Barone, M. T. & Menna-Barreto, L. Diabetes and sleep: a complex cause-and-effect relationship. *Diabetes research clinical practice* **91**, 129–137 (2011).

77. Yang, C. *et al.* Associations of sleep with cardiometabolic risk factors and cardiovascular diseases: an umbrella review of observational and mendelian randomization studies. *Sleep Medicine Rev.* 101965 (2024).

78. Zhang, C. & Qin, G. Irregular sleep and cardiometabolic risk: Clinical evidence and mechanisms. *Front. Cardiovasc. Medicine* **10**, 1059257 (2023).

79. Figorilli, M., Puligheddu, M., Congiu, P. & Ferri, R. The clinical importance of periodic leg movements in sleep. *Curr. treatment options neurology* **19**, 1–15 (2017).

80. Vallat, R., Shah, V. D. & Walker, M. P. Coordinated human sleeping brainwaves map peripheral body glucose homeostasis. *Cell Reports Medicine* **4** (2023).

81. Paszke, A. *et al.* Pytorch: An imperative style, high-performance deep learning library. *Adv. neural information processing systems* **32** (2019).

Acknowledgements

The Multi-Ethnic Study of Atherosclerosis (MESA) Sleep Ancillary study was funded by NIH-NHLBI Association of Sleep Disorders with Cardiovascular Health Across Ethnic Groups (RO1 HL098433). MESA is supported by NHLBI funded contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168 and N01-HC-95169 from the National Heart, Lung, and Blood Institute, and by cooperative agreements UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420 funded by NCATS. The National Sleep Research Resource was supported by the National Heart, Lung, and Blood Institute (R24 HL114473, 75N92019R002).

The National Heart, Lung, and Blood Institute provided funding for the ancillary MrOS Sleep Study, "Outcomes of Sleep Disorders in Older Men," under the following grant numbers: R01 HL071194, R01 HL070848, R01 HL070847, R01 HL070842, R01 HL070841, R01 HL070837, R01 HL070838, and R01 HL070839. The National Sleep Research Resource was supported by the National Heart, Lung, and Blood Institute (R24 HL114473, 75N92019R002).

The Sleep Heart Health Study (SHHS) was supported by National Heart, Lung, and Blood Institute cooperative agreements U01HL53916 (University of California, Davis), U01HL53931 (New York University), U01HL53934 (University of Minnesota), U01HL53937 and U01HL64360 (Johns Hopkins University), U01HL53938 (University of Arizona), U01HL53940 (University of Washington), U01HL53941 (Boston University), and U01HL63463 (Case Western Reserve University). The National Sleep Research Resource was supported by the National Heart, Lung, and Blood Institute (R24 HL114473, 75N92019R002).

R.T. is supported by the Knight-Hennessy Scholars funding. E.M. and and B.W. are supported by a grant from the National Heart, Lung, and Blood Institute of the NIH (R01HL161253). J.Z. is supported by funding from the Chan-Zuckerberg Biohub.

Disclosures

B.W. is a co-founder, scientific advisor, consultant to, and has personal equity interest in Beacon Biosignals.

Author Contributions Statement

R.T. and M.R.K. contributed equally to brainstorming the project, running experiments, and writing the manuscript. B.H., I.C., P.J., A.B.K., and B.W. provided high-level brainstorming and contributed to writing and editing the paper. U.H., G.G. and H.M. assisted with data access. E.M. and J.Z., as senior co-authors, conceived the project and provided overall guidance. All authors reviewed and approved the final manuscript.

A Additional Results

Supplementary Table 1. Demographic characteristics of the Stanford Sleep Clinic (SSC) cohort by dataset split. Values are presented as mean $\pm$ 2 standard deviations for age, and as counts (percentages) for categorical variables.

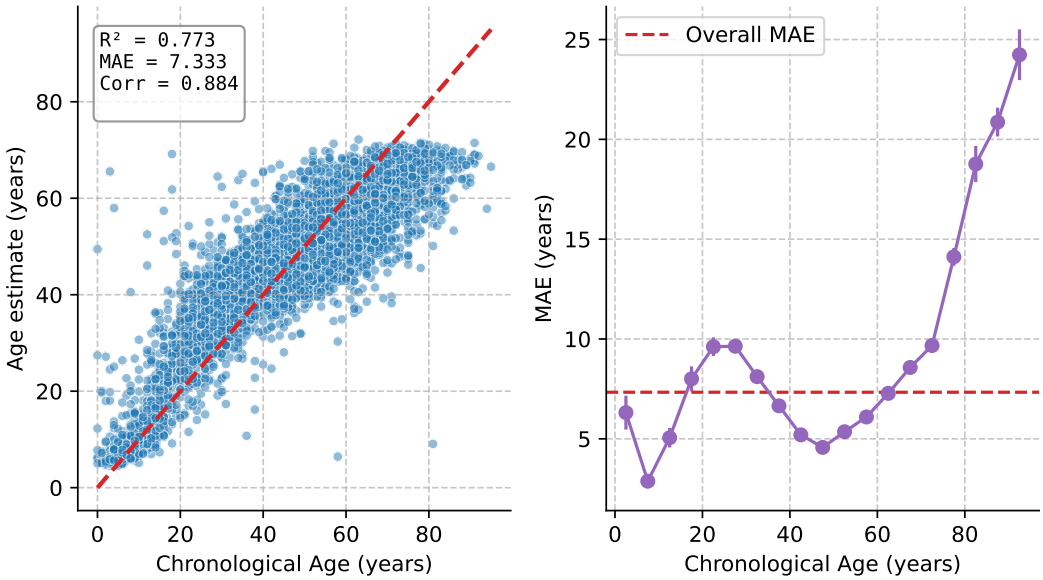
Demographics	**Train**	**Validation**	**Test**	**Temporal Test**
Age (mean ± 2SD)	45.00 ± 40.59	44.65 ± 42.89	45.71 ± 40.21	49.25 ± 39.36
Male	13230 (58.95%)	418 (59.80%)	2819 (60.27%)	2693 (54.61%)
Female	9205 (41.02%)	281 (40.20%)	1856 (39.68%)	2237 (45.37%)
Unknown	8 (0.04%)	0 (0.00%)	2 (0.04%)	1 (0.02%)
White	12807 (57.06%)	381 (54.51%)	2748 (58.76%)	2476 (50.21%)
Asian	2864 (12.76%)	95 (13.59%)	571 (12.21%)	914 (18.54%)
Black	628 (2.80%)	23 (3.29%)	144 (3.08%)	205 (4.16%)
Pacific Islander	225 (1.00%)	9 (1.29%)	34 (0.73%)	52 (1.05%)
Native American	76 (0.34%)	1 (0.14%)	16 (0.34%)	31 (0.63%)
Other	2290 (10.20%)	65 (9.30%)	476 (10.18%)	896 (18.17%)
Unknown	3553 (15.83%)	125 (17.88%)	688 (14.71%)	357 (7.24%)
Non-Hispanic	16914 (75.36%)	513 (73.39%)	3600 (76.97%)	3828 (77.63%)
Hispanic/Latino	1978 (8.81%)	57 (8.15%)	392 (8.38%)	736 (14.93%)
Other	2290 (10.20%)	65 (9.30%)	476 (10.18%)	896 (18.17%)
Unknown	3551 (15.82%)	129 (18.45%)	685 (14.65%)	367 (7.44%)

Supplementary Table 2. Demographic characteristics of the Bioserenity cohort. Values are presented as mean $\pm$ 2 standard deviations for age, and as counts (percentages) for categorical variables.

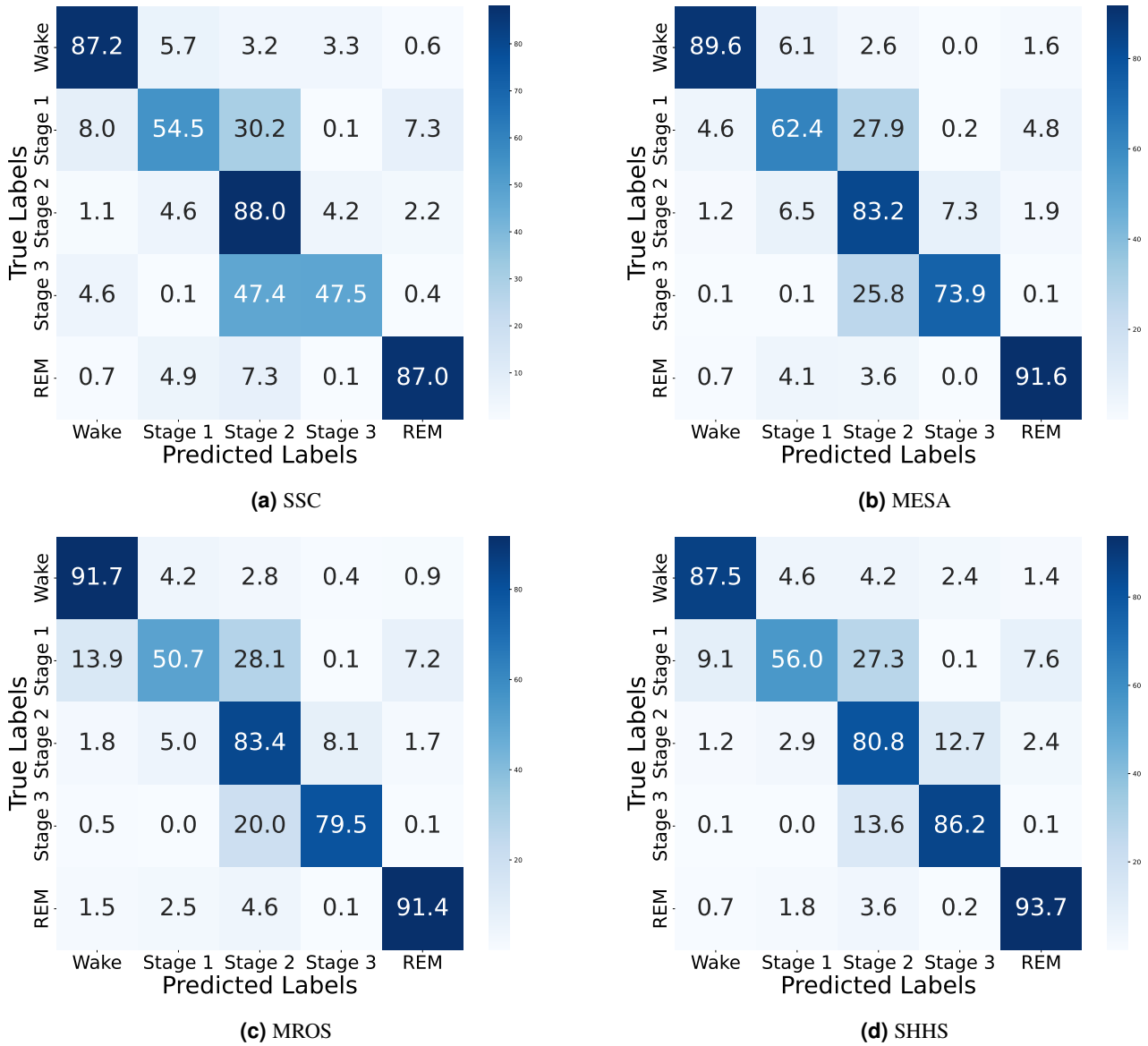
Demographics	**Train**	**Validation**
Age (mean ± 2SD)	48.71 ± 38.59	49.11 ± 41.44
Female	9684 (52.06%)	47 (47.00%)
Male	8916 (47.94%)	53 (53.00%)
African American	4636 (24.92%)	29 (29.00%)
Alaska Native	0 (0.00%)	0 (0.00%)
American Indian	124 (0.67%)	1 (1.00%)
Asian	105 (0.56%)	0 (0.00%)
Hispanic	538 (2.89%)	4 (4.00%)
Native Hawaiian	3 (0.02%)	0 (0.00%)
Pacific Islander	22 (0.12%)	0 (0.00%)
White (Caucasian)	12755 (68.58%)	64 (64.00%)
Unknown	417 (2.24%)	2 (2.00%)

Supplementary Table 3. Sleep staging results for SleepFM across the SSC, MESA, MROS, and SHHS cohorts. For comparison, F1 scores for U-Sleep on the MESA, MROS, and SHHS cohorts are also provided. Overall, SleepFM achieves performance comparable to the state-of-the-art U-Sleep, despite the latter being trained exclusively for sleep staging.

	SleepFM				U-Sleep			
Stage	**SSC**	**MESA**	**MROS**	**SHHS**	**SSC**	**MESA**	**MROS**	**SHHS**
Wake	$0.92_{(0.92, 0.93)}$	$0.94_{(0.94, 0.94)}$	$0.94_{(0.94, 0.94)}$	$0.92_{(0.92, 0.92)}$	0.61	0.92	0.93	0.93
Stage 1	$0.48_{(0.48, 0.48)}$	$0.56_{(0.56, 0.56)}$	$0.41_{(0.40, 0.41)}$	$0.49_{(0.48, 0.49)}$	0.35	0.59	0.46	0.51
Stage 2	$0.87_{(0.87, 0.87)}$	$0.83_{(0.83, 0.83)}$	$0.86_{(0.86, 0.86)}$	$0.84_{(0.84, 0.84)}$	0.75	0.87	0.87	0.87
Stage 3	$0.39_{(0.38, 0.39)}$	$0.68_{(0.68, 0.69)}$	$0.65_{(0.64, 0.65)}$	$0.72_{(0.72, 0.72)}$	0.51	0.65	0.68	0.76
REM	$0.86_{(0.86, 0.86)}$	$0.88_{(0.88, 0.88)}$	$0.90_{(0.90, 0.90)}$	$0.91_{(0.91, 0.91)}$	0.83	0.90	0.88	0.92
Mean	**0.70**	**0.78**	**0.75**	**0.78**	**0.61**	**0.79**	**0.77**	**0.80**



Supplementary Figure 1. Age estimation performance. Left: Scatter plot showing predicted versus chronological age across all patients, with the diagonal line representing perfect prediction. Right: Mean Absolute Error (MAE) across different age groups, with error bars indicating the standard error of the mean. The horizontal dashed line represents the overall MAE. Our model achieves an MAE comparable to state-of-the-art models and demonstrates better age prediction performance for younger age groups compared to older ones.



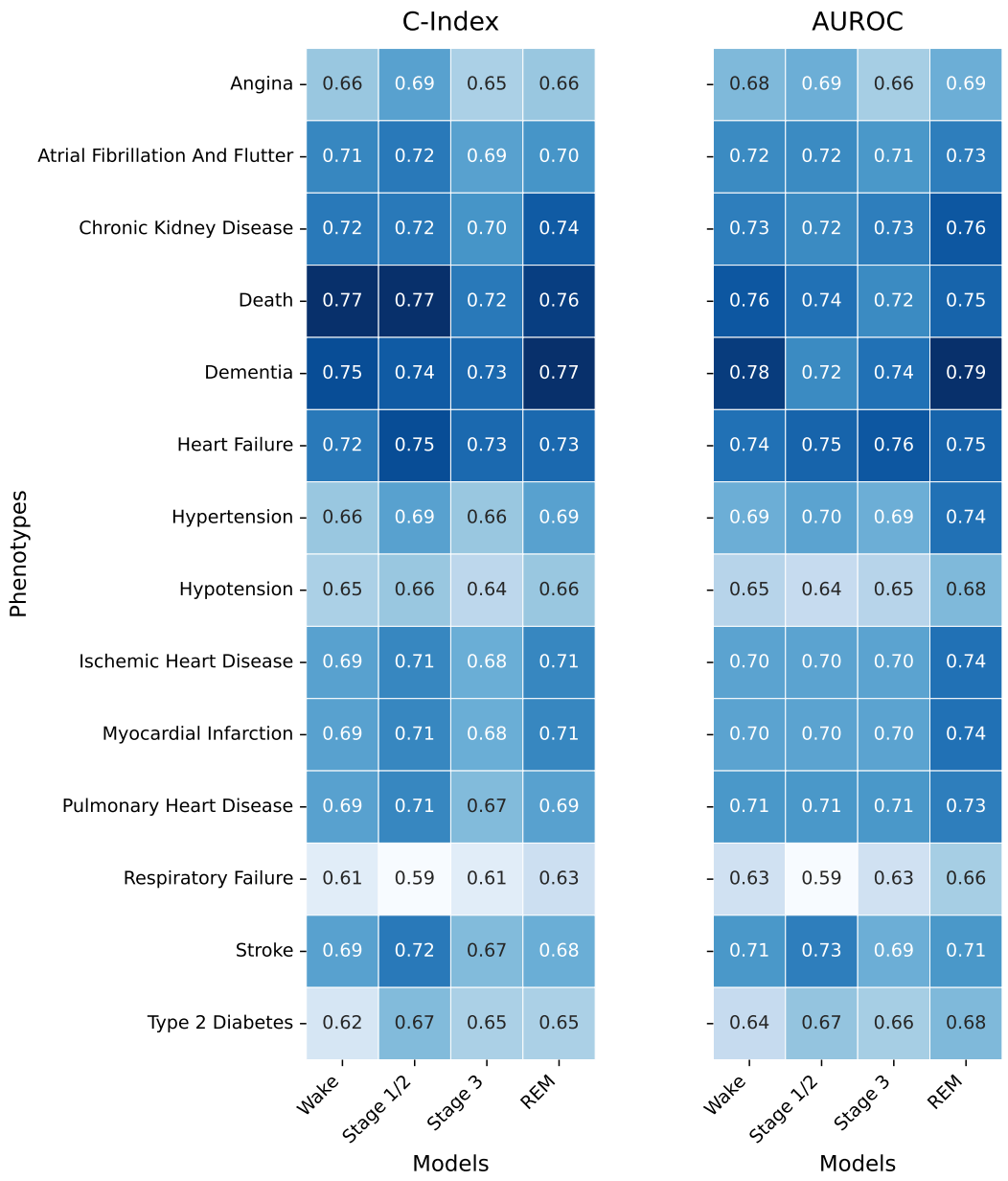
Supplementary Figure 2. SleepFM sleep staging confusion matrices for the Stanford, MESA, MROS, and SHHS cohorts. SleepFM generally performs well in differentiating between sleep stages, with mild confusion observed between Stage 1 and Stage 2, as well as between Stage 2 and Stage 3.

Disease	Phecode	Phenotype
Dementia	290.0	Delirium dementia and amnestic and other cognitive disorders
Dementia	290.1	Dementias
Dementia	290.16	Vascular dementia
Dementia	290.12	Dementia with cerebral degenerations
Dementia	290.11	Alzheimer's disease
Dementia	290.13	Senile dementia
Myocardial Infarction	411.0	Ischemic Heart Disease
Myocardial Infarction	411.2	Myocardial infarction
Ischemic Heart Disease	411.0	Ischemic Heart Disease
Ischemic Heart Disease	411.8	Other chronic ischemic heart disease, unspecified
Ischemic Heart Disease	411.9	Other acute and subacute forms of ischemic heart disease
Angina	411.1	Unstable angina (intermediate coronary syndrome)
Angina	411.3	Angina pectoris
Hypertension	401.0	Hypertension
Hypertension	401.1	Essential hypertension
Hypertension	401.2	Hypertensive heart and/or renal disease
Hypertension	401.21	Hypertensive heart disease
Hypertension	401.3	Other hypertensive complications
Hypertension	401.22	Hypertensive chronic kidney disease
Hypotension	458.0	Hypotension
Hypotension	458.1	Orthostatic hypotension
Hypotension	458.2	Iatrogenic hypotension
Hypotension	458.9	Hypotension NOS
Pulmonary Heart Disease	415.0	Pulmonary heart disease
Pulmonary Heart Disease	415.2	Chronic pulmonary heart disease
Pulmonary Heart Disease	415.1	Acute pulmonary heart disease
Pulmonary Heart Disease	415.21	Primary pulmonary hypertension
Pulmonary Heart Disease	415.11	Pulmonary embolism and infarction, acute
Atrial Fibrillation and Flutter	427.0	Cardiac dysrhythmias
Atrial Fibrillation and Flutter	427.2	Atrial fibrillation and flutter
Atrial Fibrillation and Flutter	427.21	Atrial fibrillation
Atrial Fibrillation and Flutter	427.22	Atrial flutter
Respiratory Failure	509.0	Respiratory failure, insufficiency, arrest
Respiratory Failure	509.1	Respiratory failure
Respiratory Failure	509.2	Respiratory insufficiency
Respiratory Failure	509.3	Pulmonary insufficiency or respiratory failure following trauma and surgery
Respiratory Failure	509.5	Respiratory arrest
Respiratory Failure	509.8	Dependence on respirator [Ventilator] or supplemental oxygen

Supplementary Table 4. Hand-selected diseases and their corresponding phecode groupings. These diseases were chosen for their relevance to sleep, in consultation with a sleep expert. All phecode groupings were reviewed and curated by a medical professional.

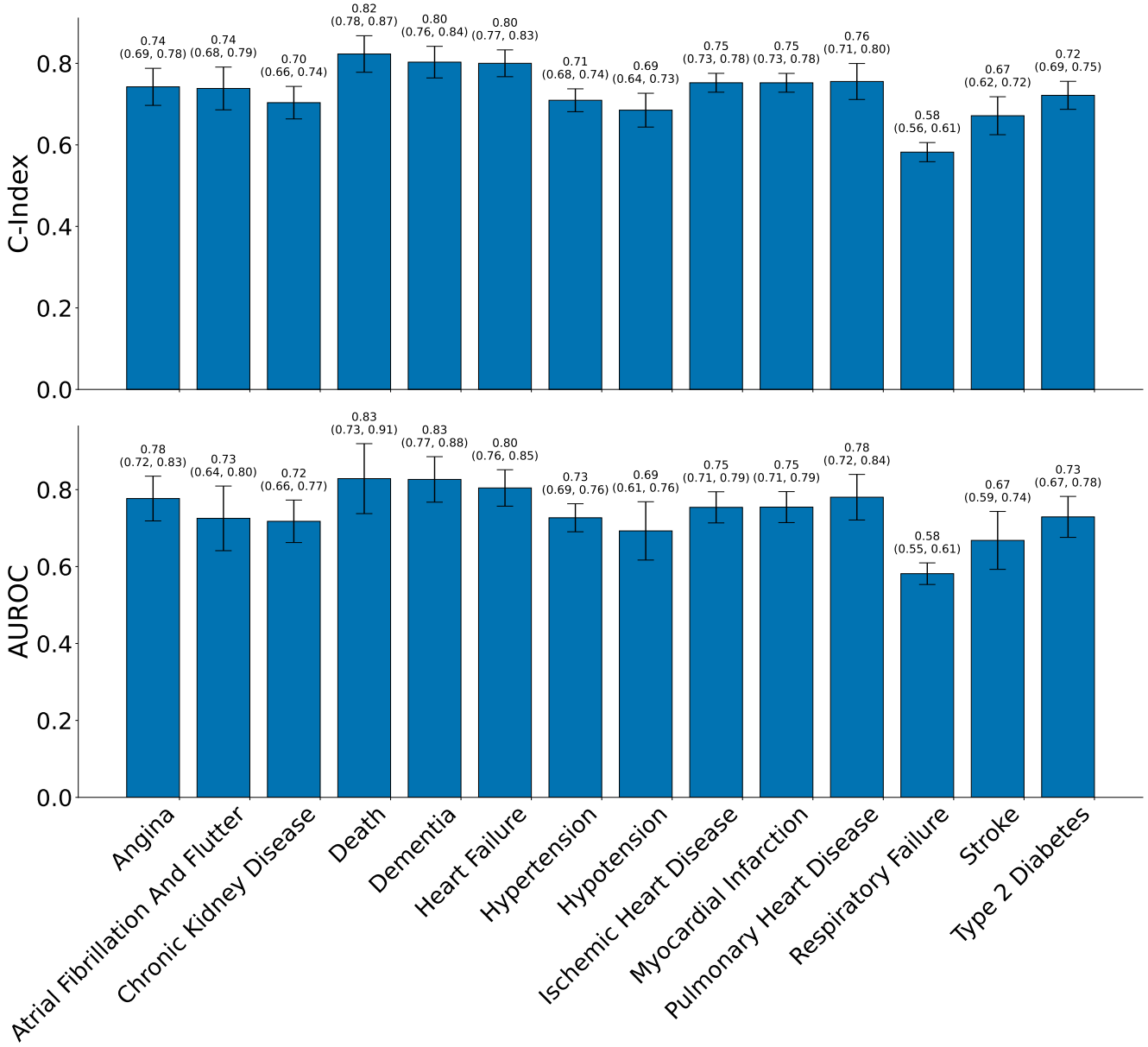
Phecode	Phenotype	Category	C-Index	AUROC (6 Years)	Prevalence (%)
290.11	Alzheimer's disease	mental disorders	0.91 (0.87, 0.94)	0.96 (0.94, 0.98)	1.18
332.0	Parkinson's disease	neurological	0.89 (0.85, 0.92)	0.93 (0.89, 0.96)	1.33
185.0	Cancer of prostate	neoplasms	0.89 (0.86, 0.91)	0.90 (0.87, 0.93)	1.97
290.1	Dementias	mental disorders	0.88 (0.85, 0.90)	0.91 (0.87, 0.94)	2.99
218.1	Uterine leiomyoma	neoplasms	0.87 (0.85, 0.89)	0.88 (0.85, 0.91)	1.68
428.4	Heart failure with preserved EF [Diastolic heart failure]	circulatory system	0.87 (0.85, 0.90)	0.90 (0.86, 0.93)	2.98
218.0	Benign neoplasm of uterus	neoplasms	0.87 (0.85, 0.89)	0.88 (0.84, 0.91)	1.71
250.25	Diabetes type 2 with peripheral circulatory disorders	endocrine/metabolic	0.87 (0.83, 0.91)	0.89 (0.85, 0.93)	1.07
174.0	Breast cancer	neoplasms	0.87 (0.83, 0.90)	0.90 (0.86, 0.93)	1.47
174.11	Malignant neoplasm of female breast	neoplasms	0.87 (0.83, 0.90)	0.89 (0.85, 0.93)	1.52
504.0	Other alveolar and parietoalveolar pneumonopathy	respiratory	0.85 (0.80, 0.89)	0.90 (0.84, 0.95)	1.1
290.0	Delirium dementia and amnestic and other cognitive disorders	mental disorders	0.85 (0.82, 0.87)	0.87 (0.84, 0.91)	4.78
401.21	Hypertensive heart disease	circulatory system	0.84 (0.82, 0.86)	0.88 (0.85, 0.91)	4.01
496.21	Obstructive chronic bronchitis	respiratory	0.83 (0.79, 0.87)	0.90 (0.86, 0.93)	1.38
276.42	Alkalosis	endocrine/metabolic	0.83 (0.77, 0.88)	0.87 (0.77, 0.93)	1.07
447.7	Aortic ectasia	circulatory system	0.83 (0.80, 0.86)	0.90 (0.87, 0.93)	2.36
172.22	Squamous cell carcinoma	neoplasms	0.83 (0.78, 0.87)	0.83 (0.78, 0.88)	1.52
474.2	Chronic tonsillitis and adenoiditis	respiratory	0.83 (0.81, 0.85)	0.82 (0.79, 0.84)	7.59
250.22	Type 2 diabetes with renal manifestations	endocrine/metabolic	0.83 (0.80, 0.85)	0.87 (0.84, 0.90)	3.16
172.11	Melanomas of skin	neoplasms	0.83 (0.77, 0.87)	0.83 (0.76, 0.90)	1.2
395.2	Nonrheumatic aortic valve disorders	circulatory system	0.83 (0.79, 0.86)	0.86 (0.82, 0.90)	2.95
741.3	Difficulty in walking	musculoskeletal	0.83 (0.79, 0.86)	0.89 (0.84, 0.93)	2.51
474.0	Acute and chronic tonsillitis	respiratory	0.82 (0.80, 0.84)	0.81 (0.79, 0.84)	7.86
440.9	Atherosclerosis of aorta	circulatory system	0.82 (0.79, 0.86)	0.86 (0.82, 0.89)	1.83
427.22	Atrial flutter	circulatory system	0.82 (0.78, 0.85)	0.88 (0.84, 0.92)	3.04
443.9	Peripheral vascular disease, unspecified	circulatory system	0.82 (0.79, 0.85)	0.85 (0.81, 0.89)	3.3
433.1	Occlusion and stenosis of precerebral arteries	circulatory system	0.82 (0.79, 0.84)	0.85 (0.81, 0.88)	3.51
426.23	Second degree AV block	circulatory system	0.82 (0.77, 0.86)	0.89 (0.85, 0.92)	1.14
290.2	Delirium due to conditions classified elsewhere	mental disorders	0.82 (0.76, 0.87)	0.81 (0.74, 0.88)	1.98
250.24	Type 2 diabetes with neurological manifestations	endocrine/metabolic	0.81 (0.77, 0.85)	0.86 (0.81, 0.90)	2.42
250.7	Diabetic retinopathy	endocrine/metabolic	0.81 (0.77, 0.85)	0.84 (0.79, 0.89)	1.79
274.2	Crystal arthropathies	endocrine/metabolic	0.81 (0.76, 0.85)	0.83 (0.75, 0.90)	1.16
433.21	Cerebral artery occlusion, with cerebral infarction	circulatory system	0.81 (0.78, 0.84)	0.85 (0.80, 0.89)	2.7
411.2	Myocardial infarction	circulatory system	0.81 (0.78, 0.84)	0.85 (0.82, 0.88)	5.53
362.27	Drusen (degenerative) of retina	sense organs	0.81 (0.77, 0.85)	0.83 (0.78, 0.88)	1.37
426.92	Cardiac defibrillator in situ	circulatory system	0.81 (0.75, 0.87)	0.85 (0.79, 0.89)	1.08
292.2	Mild cognitive impairment	mental disorders	0.81 (0.78, 0.84)	0.84 (0.80, 0.88)	3.29
440.2	Atherosclerosis of the extremities	circulatory system	0.81 (0.74, 0.86)	0.84 (0.75, 0.90)	1.21
526.3	Anomalies of jaw size/symmetry	digestive	0.81 (0.77, 0.84)	0.84 (0.80, 0.88)	2.66
433.2	Occlusion of cerebral arteries	circulatory system	0.81 (0.77, 0.84)	0.85 (0.80, 0.89)	2.81
362.2	Degeneration of macula and posterior pole of retina	sense organs	0.81 (0.78, 0.83)	0.84 (0.81, 0.87)	4.34
411.8	Other chronic ischemic heart disease, unspecified	circulatory system	0.81 (0.76, 0.85)	0.81 (0.76, 0.86)	1.64
707.2	Chronic ulcer of leg or foot	dermatologic	0.80 (0.76, 0.85)	0.84 (0.79, 0.88)	1.12
428.3	Heart failure with reduced EF [Systolic or combined heart failure]	circulatory system	0.80 (0.77, 0.84)	0.83 (0.78, 0.87)	3.7
430.0	Intracranial hemorrhage	circulatory system	0.80 (0.74, 0.86)	0.82 (0.73, 0.90)	1.49
315.0	Develomental delays and disorders	mental disorders	0.80 (0.77, 0.84)	0.84 (0.79, 0.87)	2.65
415.1	Acute pulmonary heart disease	circulatory system	0.80 (0.75, 0.85)	0.84 (0.77, 0.90)	1.73
172.1	Melanomas of skin, dx or hx	neoplasms	0.80 (0.74, 0.86)	0.79 (0.69, 0.88)	1.34
415.11	Pulmonary embolism and infarction, acute	circulatory system	0.80 (0.75, 0.85)	0.85 (0.78, 0.91)	1.71
440.0	Atherosclerosis	circulatory system	0.80 (0.77, 0.83)	0.84 (0.80, 0.87)	3.66
401.2	Hypertensive heart and/or renal disease	circulatory system	0.80 (0.77, 0.82)	0.83 (0.80, 0.86)	7.58
428.0	Congestive heart failure; nonhypertensive	circulatory system	0.80 (0.77, 0.82)	0.83 (0.79, 0.85)	6.32
740.2	Osteoarthrosis, generalized	musculoskeletal	0.80 (0.76, 0.83)	0.85 (0.80, 0.89)	2.09
426.2	Atrioventricular [AV] block	circulatory system	0.79 (0.76, 0.83)	0.84 (0.80, 0.88)	4.26
374.6	Dermatochalasis	sense organs	0.79 (0.75, 0.83)	0.85 (0.80, 0.90)	2.21
249.0	Secondary diabetes mellitus	endocrine/metabolic	0.79 (0.74, 0.84)	0.82 (0.76, 0.88)	2.63
509.2	Respiratory insufficiency	respiratory	0.79 (0.72, 0.85)	0.82 (0.72, 0.91)	1.23
743.1	Osteoporosis	musculoskeletal	0.79 (0.77, 0.82)	0.81 (0.78, 0.85)	6
411.4	Coronary atherosclerosis	circulatory system	0.79 (0.77, 0.81)	0.83 (0.81, 0.86)	10.3
743.11	Osteoporosis NOS	musculoskeletal	0.79 (0.77, 0.82)	0.82 (0.78, 0.85)	6.24
426.32	Left bundle branch block	circulatory system	0.79 (0.76, 0.83)	0.82 (0.77, 0.87)	3.91
172.3	Carcinoma in situ of skin	neoplasms	0.79 (0.73, 0.84)	0.83 (0.76, 0.89)	1.01
285.21	Anemia in chronic kidney disease	hematopoietic	0.79 (0.74, 0.83)	0.78 (0.70, 0.85)	1.89
427.11	Paroxysmal supraventricular tachycardia	circulatory system	0.79 (0.76, 0.81)	0.87 (0.84, 0.90)	7.57
362.26	Macular puckering of retina	sense organs	0.79 (0.76, 0.82)	0.83 (0.79, 0.87)	2.4
275.6	Hypercalcemia	endocrine/metabolic	0.79 (0.73, 0.84)	0.83 (0.74, 0.91)	1.59
550.4	Umbilical hernia	digestive	0.79 (0.74, 0.83)	0.84 (0.78, 0.90)	1.25
365.1	Open-angle glaucoma	sense organs	0.78 (0.73, 0.83)	0.83 (0.77, 0.88)	1.16
427.1	Paroxysmal tachycardia, unspecified	circulatory system	0.78 (0.76, 0.81)	0.87 (0.84, 0.89)	8.33
350.2	Abnormality of gait	neurological	0.78 (0.76, 0.81)	0.83 (0.80, 0.86)	7.69
427.21	Atrial fibrillation	circulatory system	0.78 (0.75, 0.81)	0.81 (0.77, 0.84)	6.29
426.21	First degree AV block	circulatory system	0.78 (0.74, 0.82)	0.83 (0.78, 0.87)	3.45
427.8	Sinoatrial node dysfunction (Bradycardia)	circulatory system	0.78 (0.73, 0.83)	0.77 (0.70, 0.84)	2.23
411.1	Unstable angina (intermediate coronary syndrome)	circulatory system	0.78 (0.74, 0.82)	0.80 (0.75, 0.85)	1.66
285.1	Acute posthemorrhagic anemia	hematopoietic	0.78 (0.75, 0.81)	0.83 (0.79, 0.87)	4.57
427.2	Atrial fibrillation and flutter	circulatory system	0.78 (0.75, 0.81)	0.81 (0.78, 0.84)	6.38
427.12	Paroxysmal ventricular tachycardia	circulatory system	0.78 (0.74, 0.82)	0.84 (0.79, 0.88)	3.54
428.1	Congestive heart failure (CHF) NOS	circulatory system	0.78 (0.74, 0.81)	0.80 (0.76, 0.84)	5.06
442.1	Aortic aneurysm	circulatory system	0.78 (0.72, 0.83)	0.82 (0.74, 0.89)	1.55
366.2	Senile cataract	sense organs	0.78 (0.75, 0.80)	0.82 (0.79, 0.85)	9.07
427.61	Supraventricular premature beats	circulatory system	0.77 (0.75, 0.80)	0.83 (0.80, 0.86)	6.95
394.1	Mitral valve stenosis and aortic valve stenosis	circulatory system	0.77 (0.72, 0.83)	0.78 (0.71, 0.85)	1.35
411.0	Ischemic Heart Disease	circulatory system	0.77 (0.75, 0.79)	0.81 (0.78, 0.84)	12.22

Supplementary Table 5. Summary of SleepFM performance metrics, including C-Index and AUROC (6 years) on the Stanford dataset. Results are sorted by C-Index, ensuring that both AUROC and C-Index exceed 0.80, with 95% confidence intervals that do not overlap with 0.5 (random chance). All conditions have statistically significant p-values (<0.01) after Bonferroni correction. Notable top conditions include Alzheimer's disease, Parkinson's disease, and dementias.



Supplementary Figure 3. Performance comparison of models trained on different sleep stages across selected conditions. Stage 1/2 demonstrates the highest predictive performance, followed by REM.

Supplementary Figure 4. Performance comparison of models trained on different sleep modalities across selected conditions. BAS, EKG, and RESP exhibit similar predictive performance across various conditions, while EMG is the least predictive.



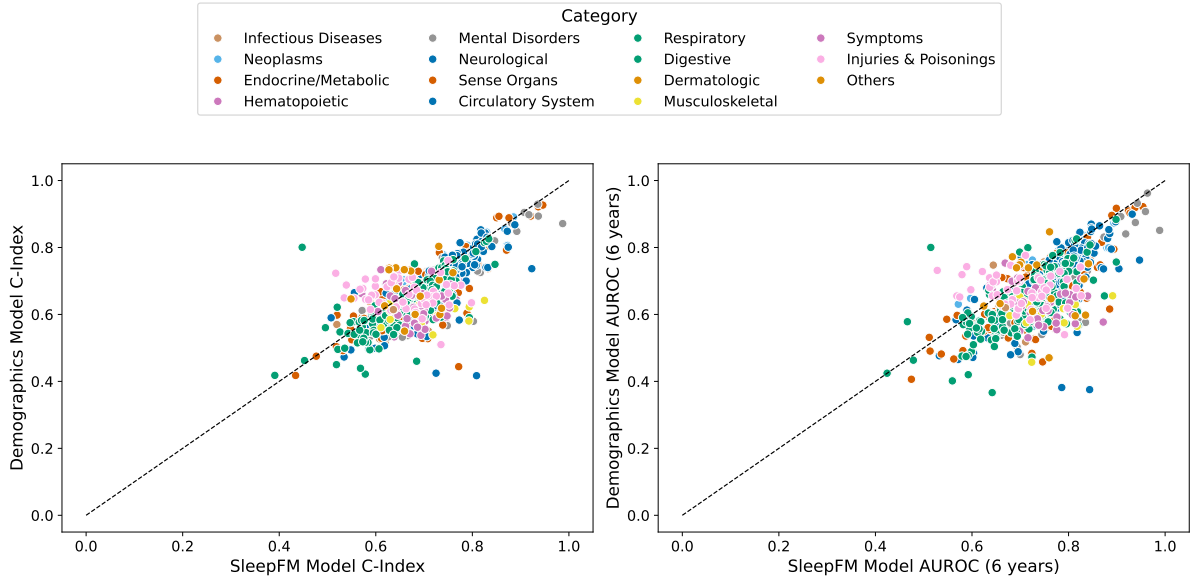
Supplementary Figure 5. Performance across clinically relevant diseases evaluated on the temporal test set (Stanford data from 2020 onwards). Performance is evaluated using multiple metrics: C-Index and 6-year AUROC. The selected conditions include critical health outcomes such as death, heart failure, stroke, and dementia. Despite the temporal distribution shift, SleepFM maintains robust predictive performance across these important clinical endpoints, demonstrating effective generalization to more recent patients.

Phecode	Phenotype	Category	Wake	Stage1/2	Stage3	REM	Best
290.12	Dementia With Cerebral Degenerations	Mental Disorders	0.88 (0.82, 0.93)	0.87 (0.78, 0.93)	0.80 (0.68, 0.90)	0.82 (0.69, 0.91)	Wake
440.1	Atherosclerosis Of Renal Artery	Circulatory System	0.78 (0.65, 0.88)	0.76 (0.53, 0.94)	0.73 (0.50, 0.89)	0.74 (0.55, 0.90)	Wake
427.22	Atrial Flutter	Circulatory System	0.77 (0.73, 0.81)	0.77 (0.72, 0.81)	0.75 (0.70, 0.79)	0.75 (0.72, 0.79)	Wake
290.1	Dementias	Mental Disorders	0.77 (0.73, 0.80)	0.75 (0.71, 0.79)	0.75 (0.71, 0.79)	0.75 (0.71, 0.80)	Wake
306.1	Mental Disorders Durring/After Pregnancy	Mental Disorders	0.77 (0.68, 0.86)	0.72 (0.53, 0.86)	0.70 (0.51, 0.87)	0.67 (0.48, 0.83)	Wake
474.0	Acute And Chronic Tonsillitis	Respiratory	0.76 (0.74, 0.79)	0.76 (0.73, 0.79)	0.75 (0.72, 0.78)	0.76 (0.74, 0.79)	Wake
440.21	Atherosclerosis Of Native Arteries Of The Extremities With Ulceration Or Gangrene	Circulatory System	0.82 (0.63, 0.95)	0.95 (0.89, 1.00)	0.83 (0.74, 0.93)	0.83 (0.71, 0.95)	Stage 1/2
290.13	Senile Dementia	Mental Disorders	0.90 (0.85, 0.97)	0.94 (0.89, 0.99)	0.66 (0.39, 1.00)	0.89 (0.85, 0.94)	Stage 1/2
264.9	Lack Of Normal Physiological Development, Unspecified	Endocrine/Metabolic	0.87 (0.78, 0.95)	0.91 (0.84, 0.96)	0.91 (0.84, 0.96)	0.86 (0.77, 0.94)	Stage 1/2
313.3	Autism	Mental Disorders	0.88 (0.82, 0.93)	0.88 (0.83, 0.93)	0.88 (0.81, 0.93)	0.87 (0.79, 0.92)	Stage 1/2
411.8	Other Chronic Ischemic Heart Disease, Unspecified	Circulatory System	0.77 (0.73, 0.82)	0.83 (0.79, 0.86)	0.73 (0.69, 0.77)	0.78 (0.72, 0.82)	Stage 1/2
426.24	Atrioventricular Block, Complete	Circulatory System	0.82 (0.76, 0.87)	0.82 (0.75, 0.89)	0.80 (0.74, 0.87)	0.82 (0.76, 0.87)	Stage 1/2
250.25	Diabetes Type 2 With Peripheral Circulatory Disorders	Endocrine/Metabolic	0.81 (0.76, 0.87)	0.82 (0.76, 0.88)	0.75 (0.69, 0.80)	0.81 (0.76, 0.86)	Stage 1/2
496.21	Obstructive Chronic Bronchitis	Respiratory	0.74 (0.68, 0.79)	0.82 (0.78, 0.86)	0.71 (0.65, 0.76)	0.78 (0.73, 0.83)	Stage 1/2
264.0	Lack Of Normal Physiological Development	Endocrine/Metabolic	0.78 (0.68, 0.87)	0.81 (0.72, 0.88)	0.78 (0.67, 0.87)	0.79 (0.70, 0.87)	Stage 1/2
426.92	Cardiac Defibrillator In Situ	Circulatory System	0.76 (0.68, 0.82)	0.79 (0.72, 0.87)	0.79 (0.74, 0.85)	0.74 (0.67, 0.81)	Stage 1/2
276.42	Alkalosis	Endocrine/Metabolic	0.74 (0.65, 0.82)	0.79 (0.71, 0.85)	0.72 (0.63, 0.80)	0.77 (0.70, 0.83)	Stage 1/2
250.22	Type 2 Diabetes With Renal Manifestations	Endocrine/Metabolic	0.72 (0.68, 0.76)	0.79 (0.75, 0.82)	0.75 (0.70, 0.79)	0.77 (0.73, 0.81)	Stage 1/2
331.0	Other Cerebral Degenerations	Neurological	0.74 (0.65, 0.81)	0.79 (0.72, 0.85)	0.75 (0.67, 0.81)	0.74 (0.67, 0.80)	Stage 1/2
427.4	Cardiac Arrest And Ventricular Fibrillation	Circulatory System	0.69 (0.60, 0.77)	0.78 (0.70, 0.85)	0.68 (0.58, 0.78)	0.66 (0.58, 0.75)	Stage 1/2
256.4	Polycystic Ovaries	Endocrine/Metabolic	0.76 (0.66, 0.84)	0.78 (0.70, 0.85)	0.73 (0.66, 0.79)	0.63 (0.54, 0.73)	Stage 1/2
250.1	Type 1 Diabetes	Endocrine/Metabolic	0.69 (0.61, 0.76)	0.78 (0.72, 0.83)	0.64 (0.56, 0.71)	0.71 (0.65, 0.77)	Stage 1/2
440.9	Atherosclerosis Of Aorta	Circulatory System	0.73 (0.68, 0.79)	0.77 (0.72, 0.81)	0.69 (0.63, 0.75)	0.75 (0.70, 0.80)	Stage 1/2
707.2	Chronic Ulcer Of Leg Or Foot	Dermatologic	0.72 (0.64, 0.79)	0.77 (0.71, 0.82)	0.69 (0.63, 0.75)	0.71 (0.63, 0.77)	Stage 1/2
707.1	Decubitus Ulcer	Dermatologic	0.74 (0.67, 0.81)	0.77 (0.72, 0.82)	0.69 (0.61, 0.76)	0.75 (0.68, 0.81)	Stage 1/2
256.0	Ovarian Dysfunction	Endocrine/Metabolic	0.74 (0.65, 0.81)	0.77 (0.69, 0.84)	0.71 (0.64, 0.77)	0.63 (0.54, 0.71)	Stage 1/2
287.32	Secondary Thrombocytopenia	Hematopoietic	0.75 (0.67, 0.82)	0.77 (0.68, 0.84)	0.68 (0.57, 0.77)	0.72 (0.62, 0.80)	Stage 1/2
426.9	Cardiac Pacemaker/Device In Situ	Circulatory System	0.72 (0.68, 0.76)	0.76 (0.72, 0.81)	0.75 (0.71, 0.80)	0.71 (0.67, 0.75)	Stage 1/2
426.21	First Degree Av Block	Circulatory System	0.75 (0.71, 0.79)	0.76 (0.72, 0.80)	0.69 (0.64, 0.74)	0.74 (0.69, 0.79)	Stage 1/2
428.3	Heart Failure With Reduced Ef [Systolic Or Combined Heart Failure]	Circulatory System	0.74 (0.71, 0.78)	0.76 (0.73, 0.80)	0.71 (0.67, 0.75)	0.73 (0.70, 0.77)	Stage 1/2
430.0	Intracranial Hemorrhage	Circulatory System	0.73 (0.65, 0.80)	0.76 (0.70, 0.81)	0.70 (0.63, 0.77)	0.74 (0.67, 0.81)	Stage 1/2
433.8	Late Effects Of Cerebrovascular Disease	Circulatory System	0.74 (0.66, 0.80)	0.76 (0.69, 0.81)	0.70 (0.62, 0.76)	0.69 (0.61, 0.76)	Stage 1/2
443.9	Peripheral Vascular Disease, Unspecified	Circulatory System	0.75 (0.71, 0.78)	0.76 (0.72, 0.80)	0.69 (0.65, 0.73)	0.76 (0.72, 0.79)	Stage 1/2
415.11	Pulmonary Embolism And Infarction, Acute	Circulatory System	0.72 (0.66, 0.77)	0.76 (0.70, 0.81)	0.74 (0.68, 0.79)	0.70 (0.63, 0.77)	Stage 1/2
250.24	Type 2 Diabetes With Neurological Manifestations	Endocrine/Metabolic	0.72 (0.66, 0.77)	0.76 (0.71, 0.81)	0.68 (0.63, 0.72)	0.75 (0.71, 0.79)	Stage 1/2
172.11	Melanomas Of Skin	Neoplasms	0.68 (0.60, 0.75)	0.76 (0.70, 0.82)	0.75 (0.67, 0.82)	0.70 (0.63, 0.77)	Stage 1/2
332.0	Parkinson'S Disease	Neurological	0.76 (0.69, 0.81)	0.76 (0.70, 0.81)	0.73 (0.67, 0.78)	0.75 (0.68, 0.82)	Stage 1/2
377.1	Optic Atrophy	Sense Organs	0.70 (0.64, 0.77)	0.76 (0.70, 0.82)	0.66 (0.59, 0.73)	0.69 (0.63, 0.77)	Stage 1/2
415.1	Acute Pulmonary Heart Disease	Circulatory System	0.72 (0.66, 0.77)	0.75 (0.69, 0.81)	0.74 (0.68, 0.80)	0.70 (0.63, 0.76)	Stage 1/2
440.0	Atherosclerosis	Circulatory System	0.72 (0.68, 0.75)	0.75 (0.71, 0.79)	0.69 (0.65, 0.73)	0.71 (0.67, 0.75)	Stage 1/2
428.0	Congestive Heart Failure; Nonhypertensive	Circulatory System	0.72 (0.69, 0.75)	0.75 (0.72, 0.78)	0.72 (0.69, 0.75)	0.73 (0.70, 0.76)	Stage 1/2
457.2	Encounter For Long-Term (Current) Use Of Antiplatelets/Antithrombotics	Circulatory System	0.71 (0.66, 0.76)	0.75 (0.72, 0.79)	0.71 (0.67, 0.76)	0.72 (0.67, 0.77)	Stage 1/2
426.32	Left Bundle Branch Block	Circulatory System	0.74 (0.71, 0.78)	0.75 (0.71, 0.79)	0.72 (0.68, 0.76)	0.74 (0.70, 0.78)	Stage 1/2
509.1	Respiratory Failure	Respiratory	0.70 (0.66, 0.74)	0.75 (0.71, 0.79)	0.68 (0.64, 0.73)	0.72 (0.67, 0.76)	Stage 1/2
509.2	Respiratory Insufficiency	Respiratory	0.71 (0.64, 0.77)	0.75 (0.70, 0.81)	0.71 (0.63, 0.78)	0.69 (0.61, 0.76)	Stage 1/2
426.91	Cardiac Pacemaker In Situ	Circulatory System	0.74 (0.69, 0.78)	0.78 (0.73, 0.82)	0.79 (0.74, 0.83)	0.73 (0.68, 0.77)	Stage 3
425.2	Secondary/Extrinsic Cardiomyopathies	Circulatory System	0.59 (0.49, 0.69)	0.60 (0.51, 0.68)	0.77 (0.66, 0.86)	0.58 (0.48, 0.69)	Stage 3
428.4	Heart Failure With Preserved Ef [Diastolic Heart Failure]	Circulatory System	0.80 (0.76, 0.83)	0.81 (0.78, 0.85)	0.76 (0.73, 0.80)	0.82 (0.78, 0.85)	REM
290.16	Vascular Dementia	Mental Disorders	0.80 (0.72, 0.87)	0.76 (0.67, 0.84)	0.77 (0.67, 0.86)	0.80 (0.71, 0.87)	REM
362.29	Macular Degeneration (Senile) Of Retina Nos	Sense Organs	0.77 (0.72, 0.83)	0.78 (0.72, 0.83)	0.70 (0.62, 0.77)	0.79 (0.74, 0.84)	REM
315.2	Speech And Language Disorder	Mental Disorders	0.77 (0.67, 0.85)	0.74 (0.65, 0.82)	0.77 (0.68, 0.84)	0.78 (0.68, 0.85)	REM
401.21	Hypertensive Heart Disease	Circulatory System	0.72 (0.68, 0.76)	0.74 (0.71, 0.78)	0.74 (0.70, 0.77)	0.77 (0.73, 0.80)	REM
411.9	Other Acute And Subacute Forms Of Ischemic Heart Disease	Circulatory System	0.71 (0.62, 0.79)	0.69 (0.59, 0.79)	0.67 (0.58, 0.77)	0.77 (0.69, 0.85)	REM
426.8	Other Cardiac Conduction Disorders	Circulatory System	0.76 (0.72, 0.79)	0.75 (0.71, 0.79)	0.69 (0.65, 0.73)	0.77 (0.74, 0.81)	REM
290.0	Delirium Dementia And Amnestic And Other Cognitive Disorders	Mental Disorders	0.75 (0.72, 0.79)	0.74 (0.70, 0.77)	0.73 (0.70, 0.76)	0.77 (0.73, 0.80)	REM
290.2	Delirium Due To Conditions Classified Elsewhere	Mental Disorders	0.75 (0.70, 0.80)	0.73 (0.67, 0.79)	0.69 (0.63, 0.75)	0.77 (0.72, 0.82)	REM
474.2	Chronic Tonsillitis And Adenoiditis	Respiratory	0.77 (0.74, 0.79)	0.77 (0.74, 0.79)	0.75 (0.72, 0.78)	0.77 (0.74, 0.79)	REM
250.7	Diabetic Retinopathy	Endocrine/Metabolic	0.74 (0.67, 0.79)	0.73 (0.68, 0.78)	0.67 (0.61, 0.73)	0.76 (0.71, 0.81)	REM
264.2	Failure To Thrive (Childhood)	Endocrine/Metabolic	0.74 (0.62, 0.86)	0.73 (0.56, 0.86)	0.66 (0.51, 0.83)	0.75 (0.66, 0.86)	REM
260.2	Severe Protein-Calorie Malnutrition	Endocrine/Metabolic	0.73 (0.67, 0.78)	0.70 (0.64, 0.75)	0.62 (0.55, 0.68)	0.75 (0.70, 0.80)	REM
333.1	Essential Tremor	Neurological	0.69 (0.63, 0.76)	0.70 (0.62, 0.77)	0.65 (0.58, 0.71)	0.75 (0.68, 0.82)	REM
504.0	Other Alveolar And Parietoalveolar Pneumonopathy	Respiratory	0.72 (0.65, 0.79)	0.73 (0.65, 0.80)	0.75 (0.67, 0.83)	0.75 (0.67, 0.83)	REM

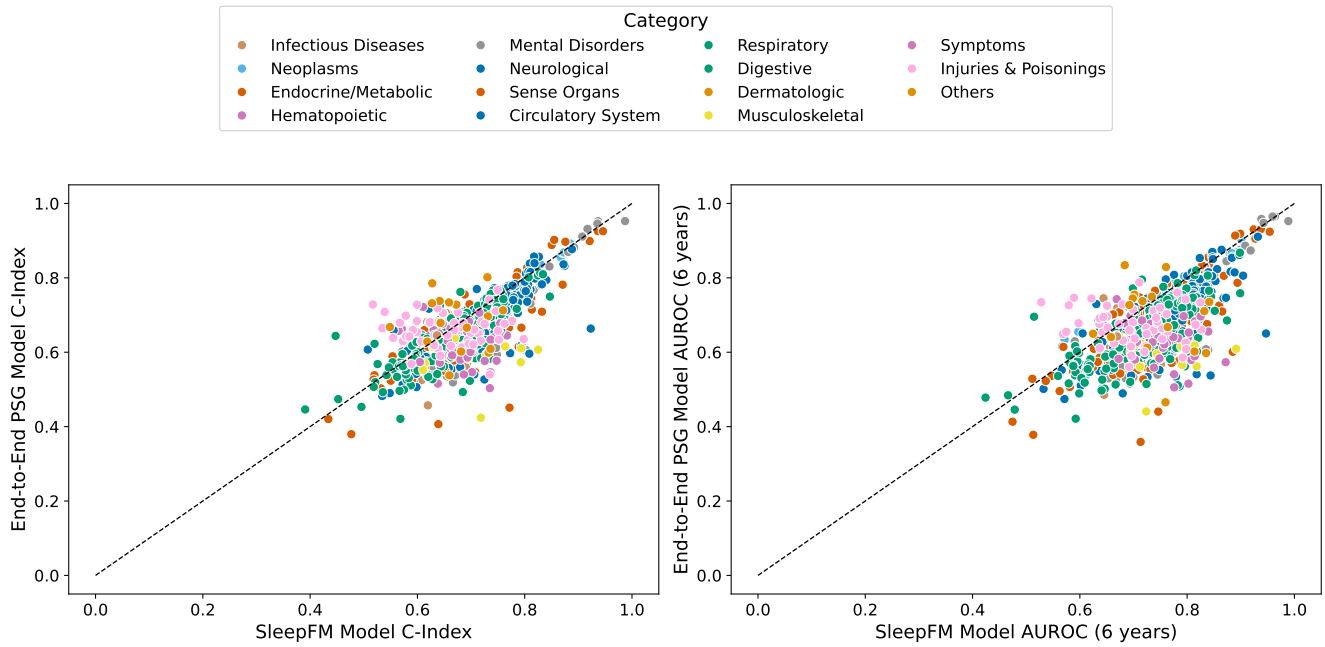
Supplementary Table 6. Performance comparison of models trained on different sleep stages across conditions. For each sleep stage, C-Index values are reported with 95% confidence intervals in parentheses. Only conditions where at least one sleep stage achieves a C-Index above 0.75 are included. Results are grouped by disease category and sorted alphabetically within each category. Stage 1/2 demonstrates the highest predictive performance across 39 conditions, followed by REM sleep with 15 conditions.

Phecode	Phenotype	Category	BAS	Respiratory	EKG	EMG	Best
313.3	Autism	Mental Disorders	0.89 (0.85, 0.93)	0.84 (0.77, 0.90)	0.86 (0.80, 0.91)	0.85 (0.80, 0.91)	BAS
290.12	Dementia With Cerebral Degenerations	Mental Disorders	0.88 (0.78, 0.95)	0.82 (0.75, 0.89)	0.83 (0.73, 0.91)	0.76 (0.67, 0.84)	BAS
290.16	Vascular Dementia	Mental Disorders	0.87 (0.82, 0.92)	0.78 (0.68, 0.87)	0.83 (0.75, 0.89)	0.78 (0.70, 0.85)	BAS
331.0	Other Cerebral Degenerations	Neurological	0.85 (0.80, 0.90)	0.74 (0.64, 0.82)	0.71 (0.62, 0.77)	0.74 (0.65, 0.82)	BAS
447.7	Aortic Ectasia	Circulatory System	0.82 (0.79, 0.85)	0.81 (0.78, 0.85)	0.80 (0.76, 0.84)	0.82 (0.78, 0.85)	BAS
250.7	Diabetic Retinopathy	Endocrine/Metabolic	0.82 (0.77, 0.86)	0.78 (0.74, 0.83)	0.80 (0.76, 0.84)	0.75 (0.69, 0.80)	BAS
264.0	Lack Of Normal Physiological Development	Endocrine/Metabolic	0.82 (0.74, 0.89)	0.77 (0.68, 0.85)	0.75 (0.65, 0.85)	0.80 (0.72, 0.87)	BAS
332.0	Parkinson'S Disease	Neurological	0.82 (0.76, 0.87)	0.82 (0.77, 0.86)	0.79 (0.73, 0.84)	0.78 (0.72, 0.83)	BAS
250.22	Type 2 Diabetes With Renal Manifestations	Endocrine/Metabolic	0.81 (0.78, 0.84)	0.78 (0.75, 0.81)	0.81 (0.78, 0.84)	0.77 (0.73, 0.80)	BAS
290.1	Dementias	Mental Disorders	0.81 (0.77, 0.84)	0.80 (0.76, 0.83)	0.78 (0.74, 0.81)	0.78 (0.74, 0.81)	BAS
362.29	Macular Degeneration (Senile) Of Retina Nos	Sense Organs	0.81 (0.76, 0.86)	0.74 (0.68, 0.80)	0.69 (0.61, 0.77)	0.75 (0.69, 0.81)	BAS
707.1	Decubitus Ulcer	Dermatologic	0.80 (0.74, 0.86)	0.71 (0.63, 0.79)	0.74 (0.67, 0.81)	0.78 (0.72, 0.83)	BAS
290.0	Delirium Dementia And Amnestic And Other Cognitive Disorders	Mental Disorders	0.80 (0.76, 0.83)	0.79 (0.75, 0.81)	0.78 (0.75, 0.81)	0.76 (0.72, 0.78)	BAS
504.0	Other Alveolar And Parietoalveolar Pneumonopathy	Respiratory	0.80 (0.73, 0.87)	0.80 (0.72, 0.87)	0.77 (0.70, 0.84)	0.79 (0.73, 0.86)	BAS
290.11	Alzheimer'S Disease	Mental Disorders	0.79 (0.74, 0.84)	0.76 (0.69, 0.82)	0.72 (0.66, 0.79)	0.74 (0.68, 0.79)	BAS
433.21	Cerebral Artery Occlusion, With Cerebral Infarction	Circulatory System	0.78 (0.74, 0.82)	0.75 (0.70, 0.79)	0.75 (0.71, 0.79)	0.77 (0.73, 0.81)	BAS
433.2	Occlusion Of Cerebral Arteries	Circulatory System	0.78 (0.73, 0.82)	0.74 (0.70, 0.79)	0.75 (0.70, 0.79)	0.77 (0.73, 0.81)	BAS
443.9	Peripheral Vascular Disease, Unspecified	Circulatory System	0.78 (0.74, 0.82)	0.75 (0.71, 0.79)	0.77 (0.73, 0.80)	0.75 (0.71, 0.79)	BAS
250.23	Type 2 Diabetes With Ophthalmic Manifestations	Endocrine/Metabolic	0.78 (0.71, 0.83)	0.74 (0.68, 0.79)	0.76 (0.71, 0.82)	0.71 (0.65, 0.77)	BAS
315.0	Develomental Delays And Disorders	Mental Disorders	0.78 (0.73, 0.82)	0.77 (0.73, 0.81)	0.77 (0.73, 0.81)	0.72 (0.67, 0.77)	BAS
333.2	Myoclonus	Neurological	0.78 (0.70, 0.85)	0.62 (0.54, 0.70)	0.68 (0.58, 0.77)	0.69 (0.60, 0.78)	BAS
433.1	Occlusion And Stenosis Of Precerebral Arteries	Circulatory System	0.77 (0.73, 0.80)	0.74 (0.70, 0.78)	0.73 (0.69, 0.77)	0.75 (0.71, 0.79)	BAS
264.2	Failure To Thrive (Childhood)	Endocrine/Metabolic	0.77 (0.67, 0.86)	0.67 (0.51, 0.83)	0.62 (0.46, 0.80)	0.76 (0.63, 0.86)	BAS
292.2	Mild Cognitive Impairment	Mental Disorders	0.77 (0.73, 0.81)	0.73 (0.68, 0.77)	0.73 (0.68, 0.77)	0.73 (0.69, 0.76)	BAS
202.0	Cancer Of Other Lymphoid, Histiocytic Tissue	Neoplasms	0.77 (0.70, 0.84)	0.75 (0.65, 0.83)	0.75 (0.66, 0.83)	0.73 (0.65, 0.81)	BAS
430.0	Intracranial Hemorrhage	Circulatory System	0.76 (0.70, 0.82)	0.69 (0.63, 0.76)	0.71 (0.64, 0.78)	0.70 (0.64, 0.76)	BAS
433.8	Late Effects Of Cerebrovascular Disease	Circulatory System	0.76 (0.70, 0.81)	0.69 (0.63, 0.76)	0.72 (0.65, 0.79)	0.70 (0.64, 0.75)	BAS
260.6	Anorexia	Endocrine/Metabolic	0.76 (0.72, 0.80)	0.72 (0.67, 0.77)	0.70 (0.64, 0.75)	0.73 (0.68, 0.77)	BAS
202.2	Non-Hodgkins Lymphoma	Neoplasms	0.76 (0.67, 0.84)	0.74 (0.62, 0.84)	0.72 (0.62, 0.82)	0.71 (0.62, 0.79)	BAS
350.2	Abnormality Of Gait	Neurological	0.76 (0.73, 0.78)	0.73 (0.70, 0.76)	0.73 (0.70, 0.76)	0.73 (0.70, 0.76)	BAS
362.26	Macular Puckering Of Retina	Sense Organs	0.76 (0.71, 0.79)	0.73 (0.68, 0.78)	0.71 (0.67, 0.76)	0.73 (0.69, 0.78)	BAS
362.2	Degeneration Of Macula And Posterior Pole Of Retina	Sense Organs	0.75 (0.72, 0.78)	0.73 (0.69, 0.76)	0.70 (0.67, 0.73)	0.72 (0.69, 0.75)	BAS
290.13	Senile Dementia	Mental Disorders	0.88 (0.78, 1.00)	0.92 (0.86, 0.99)	0.89 (0.87, 0.91)	0.67 (0.54, 0.83)	RESP
428.4	Heart Failure With Preserved Ef [Diastolic Heart Failure]	Circulatory System	0.83 (0.79, 0.86)	0.85 (0.82, 0.87)	0.84 (0.80, 0.87)	0.81 (0.78, 0.84)	RESP
401.21	Hypertensive Heart Disease	Circulatory System	0.81 (0.78, 0.84)	0.83 (0.80, 0.86)	0.82 (0.79, 0.85)	0.79 (0.76, 0.82)	RESP
440.1	Atherosclerosis Of Renal Artery	Circulatory System	0.73 (0.52, 0.94)	0.82 (0.69, 0.94)	0.81 (0.65, 0.93)	0.77 (0.62, 0.96)	RESP
496.21	Obstructive Chronic Bronchitis	Respiratory	0.76 (0.71, 0.81)	0.82 (0.77, 0.86)	0.80 (0.76, 0.85)	0.77 (0.71, 0.83)	RESP
426.23	Second Degree Av Block	Circulatory System	0.78 (0.73, 0.83)	0.81 (0.76, 0.84)	0.76 (0.70, 0.81)	0.79 (0.73, 0.83)	RESP
250.24	Type 2 Diabetes With Neurological Manifestations	Endocrine/Metabolic	0.78 (0.74, 0.83)	0.81 (0.76, 0.85)	0.80 (0.76, 0.83)	0.76 (0.71, 0.80)	RESP
741.3	Difficulty In Walking	Musculoskeletal	0.77 (0.72, 0.81)	0.81 (0.77, 0.84)	0.78 (0.73, 0.82)	0.76 (0.72, 0.81)	RESP
277.4	Disorders Of Bilirubin Excretion	Endocrine/Metabolic	0.79 (0.71, 0.85)	0.80 (0.72, 0.86)	0.70 (0.62, 0.77)	0.74 (0.65, 0.81)	RESP
276.11	Hyperosmolality And/Or Hypernatremia	Endocrine/Metabolic	0.74 (0.63, 0.83)	0.80 (0.72, 0.88)	0.78 (0.71, 0.85)	0.76 (0.66, 0.83)	RESP
315.2	Speech And Language Disorder	Mental Disorders	0.78 (0.69, 0.86)	0.80 (0.73, 0.86)	0.77 (0.69, 0.84)	0.74 (0.65, 0.82)	RESP
411.2	Myocardial Infarction	Circulatory System	0.78 (0.75, 0.81)	0.79 (0.76, 0.81)	0.78 (0.76, 0.81)	0.76 (0.73, 0.79)	RESP
250.1	Type 1 Diabetes	Endocrine/Metabolic	0.74 (0.67, 0.80)	0.78 (0.73, 0.83)	0.75 (0.69, 0.80)	0.71 (0.63, 0.77)	RESP
172.11	Melanomas Of Skin	Neoplasms	0.72 (0.65, 0.78)	0.78 (0.72, 0.83)	0.76 (0.69, 0.83)	0.71 (0.65, 0.77)	RESP
172.22	Squamous Cell Carcinoma	Neoplasms	0.71 (0.65, 0.77)	0.78 (0.73, 0.83)	0.75 (0.70, 0.80)	0.74 (0.68, 0.80)	RESP
427.11	Paroxysmal Supraventricular Tachycardia	Circulatory System	0.76 (0.73, 0.78)	0.77 (0.75, 0.80)	0.77 (0.74, 0.79)	0.76 (0.73, 0.79)	RESP
427.1	Paroxysmal Tachycardia, Unspecified	Circulatory System	0.76 (0.73, 0.78)	0.77 (0.75, 0.79)	0.77 (0.75, 0.79)	0.76 (0.73, 0.78)	RESP
505.0	Other Pulmonary Inflamation Or Edema	Respiratory	0.73 (0.65, 0.81)	0.77 (0.71, 0.83)	0.72 (0.65, 0.79)	0.70 (0.63, 0.77)	RESP
509.1	Respiratory Failure	Respiratory	0.72 (0.68, 0.76)	0.77 (0.72, 0.81)	0.76 (0.72, 0.80)	0.75 (0.71, 0.79)	RESP
509.2	Respiratory Insufficiency	Respiratory	0.69 (0.62, 0.77)	0.77 (0.69, 0.83)	0.76 (0.70, 0.82)	0.72 (0.64, 0.80)	RESP
415.1	Acute Pulmonary Heart Disease	Circulatory System	0.72 (0.67, 0.78)	0.76 (0.71, 0.82)	0.76 (0.71, 0.81)	0.73 (0.67, 0.78)	RESP
428.0	Congestive Heart Failure; Nonhypertensive	Circulatory System	0.74 (0.71, 0.77)	0.76 (0.73, 0.79)	0.76 (0.72, 0.79)	0.72 (0.69, 0.75)	RESP
415.11	Pulmonary Embolism And Infarction, Acute	Circulatory System	0.73 (0.67, 0.79)	0.76 (0.71, 0.81)	0.76 (0.71, 0.81)	0.73 (0.67, 0.78)	RESP
440.21	Atherosclerosis Of Native Arteries Of The Extremities With Ulceration Or Gangrene	Circulatory System	0.90 (0.82, 0.95)	0.86 (0.75, 0.94)	0.93 (0.87, 0.98)	0.86 (0.79, 0.92)	EKG
264.9	Lack Of Normal Physiological Development, Unspecified	Endocrine/Metabolic	0.86 (0.77, 0.93)	0.88 (0.82, 0.93)	0.90 (0.84, 0.95)	0.88 (0.81, 0.94)	EKG
426.24	Atrioventricular Block, Complete	Circulatory System	0.87 (0.80, 0.92)	0.85 (0.77, 0.91)	0.87 (0.82, 0.91)	0.81 (0.75, 0.88)	EKG
426.92	Cardiac Defibrillator In Situ	Circulatory System	0.79 (0.73, 0.84)	0.79 (0.73, 0.85)	0.87 (0.82, 0.91)	0.75 (0.69, 0.82)	EKG
250.25	Diabetes Type 2 With Peripheral Circulatory Disorders	Endocrine/Metabolic	0.83 (0.78, 0.88)	0.83 (0.80, 0.87)	0.86 (0.82, 0.89)	0.81 (0.75, 0.86)	EKG
411.8	Other Chronic Ischemic Heart Disease, Unspecified	Circulatory System	0.76 (0.70, 0.81)	0.77 (0.72, 0.82)	0.84 (0.80, 0.88)	0.75 (0.69, 0.80)	EKG
426.91	Cardiac Pacemaker In Situ	Circulatory System	0.77 (0.73, 0.81)	0.74 (0.69, 0.79)	0.82 (0.78, 0.86)	0.74 (0.69, 0.79)	EKG
276.42	Alkalosis	Endocrine/Metabolic	0.81 (0.74, 0.87)	0.81 (0.74, 0.87)	0.81 (0.74, 0.87)	0.77 (0.70, 0.84)	EKG
287.32	Secondary Thrombocytopenia	Hematopoietic	0.62 (0.52, 0.73)	0.77 (0.69, 0.84)	0.81 (0.73, 0.87)	0.69 (0.59, 0.77)	EKG
427.22	Atrial Flutter	Circulatory System	0.77 (0.73, 0.80)	0.80 (0.76, 0.83)	0.80 (0.77, 0.84)	0.74 (0.70, 0.78)	EKG
428.3	Heart Failure With Reduced Ef [Systolic Or Combined Heart Failure]	Circulatory System	0.75 (0.71, 0.79)	0.78 (0.75, 0.82)	0.80 (0.75, 0.84)	0.74 (0.70, 0.77)	EKG
426.8	Other Cardiac Conduction Disorders	Circulatory System	0.74 (0.70, 0.78)	0.80 (0.76, 0.83)	0.80 (0.77, 0.83)	0.72 (0.68, 0.76)	EKG
474.0	Acute And Chronic Tonsillitis	Respiratory	0.78 (0.75, 0.80)	0.77 (0.74, 0.79)	0.80 (0.78, 0.82)	0.74 (0.71, 0.77)	EKG
474.2	Chronic Tonsillitis And Adenoiditis	Respiratory	0.78 (0.76, 0.81)	0.78 (0.75, 0.80)	0.80 (0.78, 0.83)	0.75 (0.72, 0.77)	EKG
427.4	Cardiac Arrest And Ventricular Fibrillation	Circulatory System	0.66 (0.57, 0.74)	0.72 (0.63, 0.80)	0.79 (0.72, 0.85)	0.71 (0.64, 0.78)	EKG
426.9	Cardiac Pacemaker/Device In Situ	Circulatory System	0.74 (0.70, 0.78)	0.72 (0.68, 0.77)	0.79 (0.75, 0.83)	0.71 (0.66, 0.75)	EKG
249.0	Secondary Diabetes Mellitus	Endocrine/Metabolic	0.78 (0.73, 0.83)	0.78 (0.74, 0.82)	0.79 (0.75, 0.83)	0.75 (0.70, 0.79)	EKG
440.9	Atherosclerosis Of Aorta	Circulatory System	0.74 (0.70, 0.78)	0.76 (0.71, 0.81)	0.78 (0.73, 0.82)	0.70 (0.65, 0.74)	EKG
426.2	Atrioventricular [Av] Block	Circulatory System	0.72 (0.69, 0.76)	0.76 (0.72, 0.79)	0.78 (0.74, 0.81)	0.72 (0.67, 0.75)	EKG
428.2	Heart Failure Nos	Circulatory System	0.73 (0.64, 0.81)	0.74 (0.65, 0.82)	0.78 (0.70, 0.86)	0.69 (0.59, 0.77)	EKG
395.2	Nonrheumatic Aortic Valve Disorders	Circulatory System	0.77 (0.73, 0.81)	0.74 (0.70, 0.78)	0.78 (0.74, 0.81)	0.71 (0.66, 0.76)	EKG
427.8	Sinoatrial Node Dysfunction (Bradycardia)	Circulatory System	0.73 (0.67, 0.78)	0.78 (0.74, 0.82)	0.78 (0.74, 0.83)	0.70 (0.65, 0.75)	EKG
285.21	Anemia In Chronic Kidney Disease	Hematopoietic	0.76 (0.70, 0.81)	0.77 (0.72, 0.82)	0.78 (0.74, 0.82)	0.73 (0.67, 0.78)	EKG
426.21	First Degree Av Block	Circulatory System	0.70 (0.66, 0.74)	0.74 (0.70, 0.78)	0.77 (0.73, 0.81)	0.70 (0.66, 0.74)	EKG
401.2	Hypertensive Heart And/Or Renal Disease	Circulatory System	0.75 (0.72, 0.78)	0.77 (0.74, 0.79)	0.77 (0.74, 0.79)	0.73 (0.70, 0.75)	EKG
426.32	Left Bundle Branch Block	Circulatory System	0.72 (0.68, 0.75)	0.74 (0.70, 0.77)	0.77 (0.73, 0.81)	0.72 (0.68, 0.75)	EKG
395.3	Nonrheumatic Tricuspid Valve Disorders	Circulatory System	0.63 (0.56, 0.70)	0.75 (0.69, 0.80)	0.77 (0.72, 0.82)	0.65 (0.59, 0.71)	EKG
411.9	Other Acute And Subacute Forms Of Ischemic Heart Disease	Circulatory System	0.73 (0.64, 0.81)	0.76 (0.69, 0.83)	0.77 (0.70, 0.82)	0.70 (0.61, 0.78)	EKG
275.6	Hypercalcemia	Endocrine/Metabolic	0.75 (0.69, 0.80)	0.75 (0.69, 0.81)	0.77 (0.70, 0.82)	0.74 (0.66, 0.80)	EKG
427.12	Paroxysmal Ventricular Tachycardia	Circulatory System	0.75 (0.71, 0.79)	0.76 (0.71, 0.79)	0.76 (0.72, 0.80)	0.76 (0.71, 0.80)	EKG
571.51	Cirrhosis Of Liver Without Mention Of Alcohol	Digestive	0.67 (0.58, 0.76)	0.73 (0.65, 0.80)	0.76 (0.67, 0.82)	0.71 (0.62, 0.79)	EKG
440.2	Atherosclerosis Of The Extremities	Circulatory System	0.74 (0.68, 0.80)	0.72 (0.65, 0.79)	0.75 (0.68, 0.81)	0.68 (0.62, 0.74)	EKG
426.3	Bundle Branch Block	Circulatory System	0.71 (0.68, 0.74)	0.73 (0.69, 0.76)	0.75 (0.72, 0.78)	0.70 (0.67, 0.73)	EKG
420.3	Endocarditis	Circulatory System	0.66 (0.54, 0.77)	0.73 (0.63, 0.82)	0.75 (0.68, 0.82)	0.61 (0.50, 0.71)	EKG
427.6	Premature Beats	Circulatory System	0.73 (0.71, 0.75)	0.74 (0.71, 0.76)	0.75 (0.73, 0.77)	0.72 (0.70, 0.74)	EKG

Supplementary Table 7. Performance comparison of different modalities across conditions. For each modality, C-Index values are reported with 95% confidence intervals in parentheses. Only conditions where at least one modality achieves a C-Index above 0.75 are included. Results are grouped by disease category and sorted alphabetically within each category. BAS and EKG demonstrate the highest predictive performance, followed closely by RESP. EMG is the least predictive, with no top conditions where it performs best.

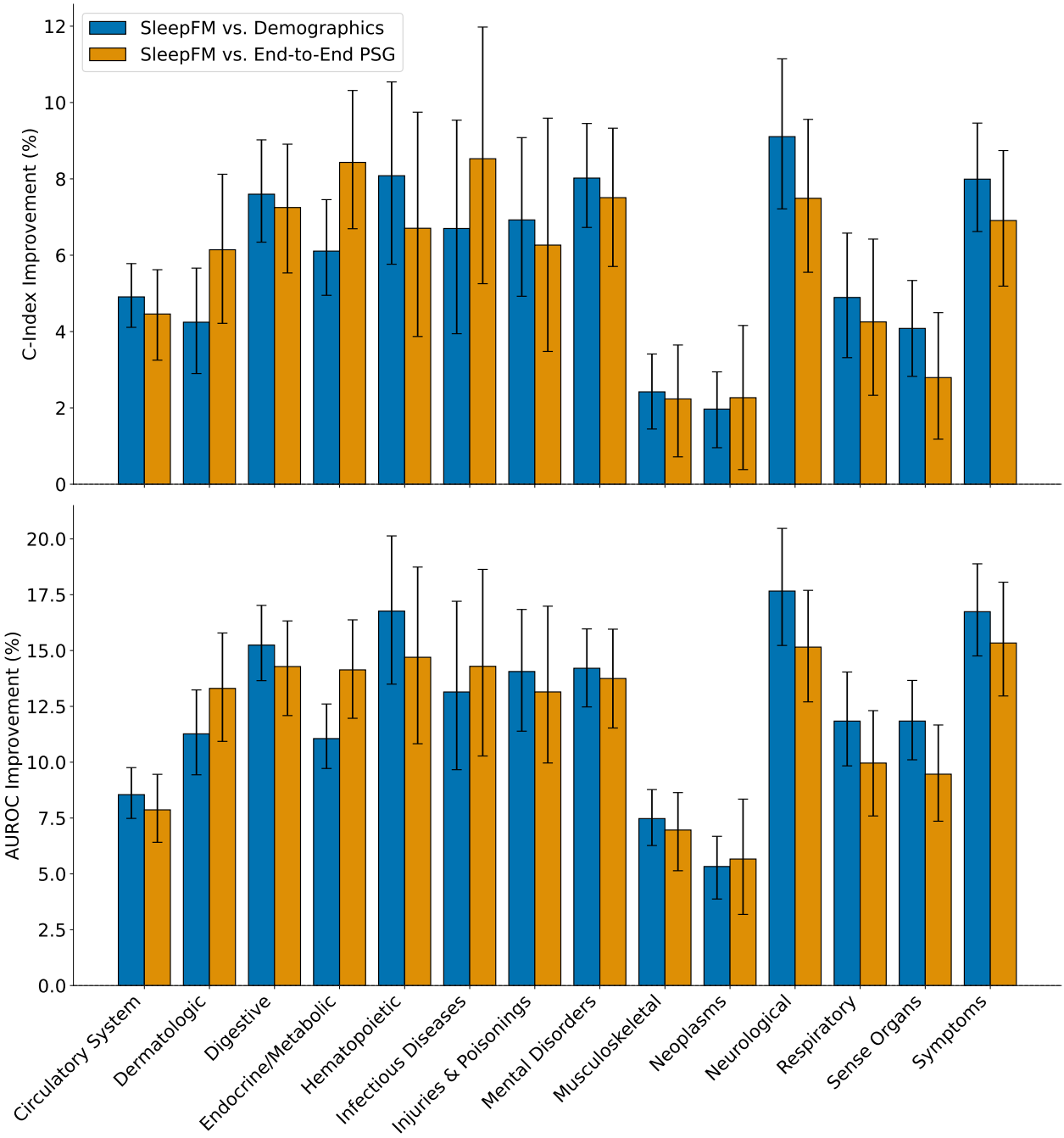


(a) Comparison with demographics-only baseline

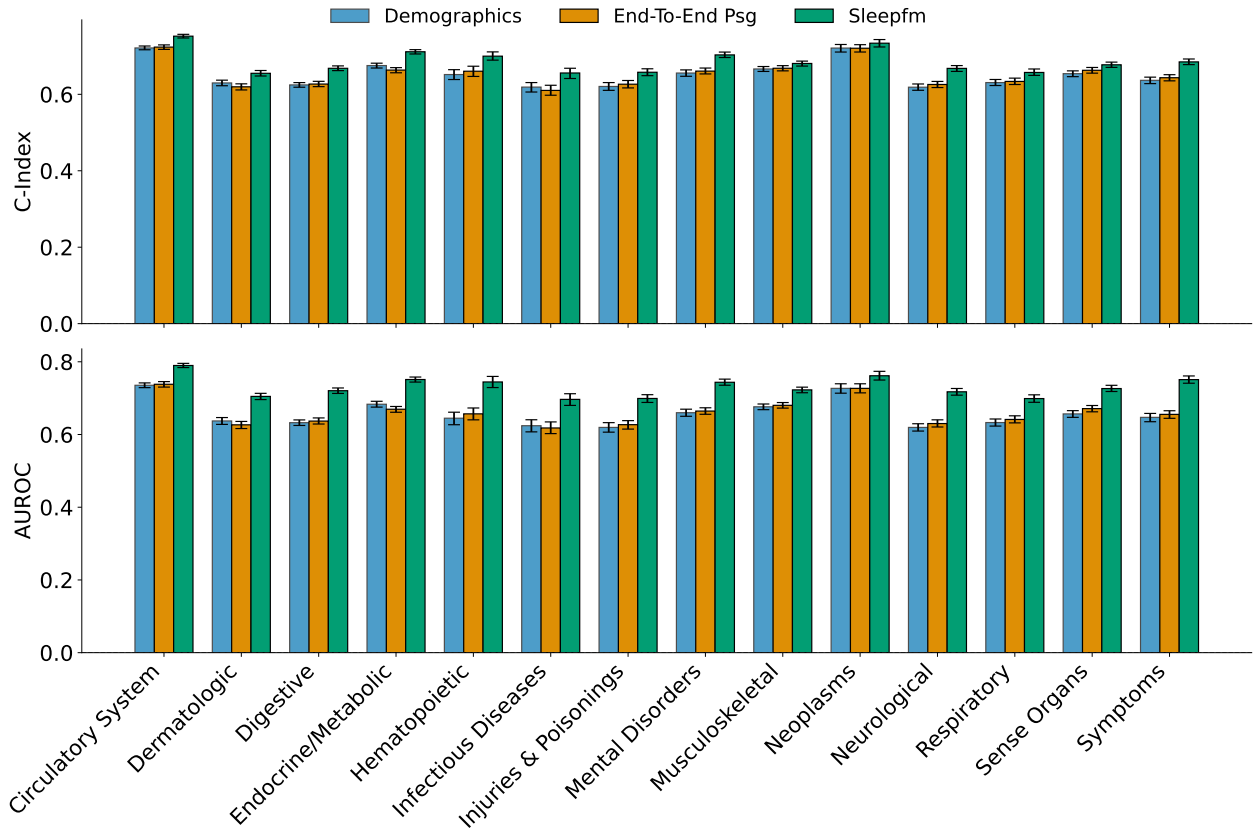


(b) Comparison with End-to-End PSG model

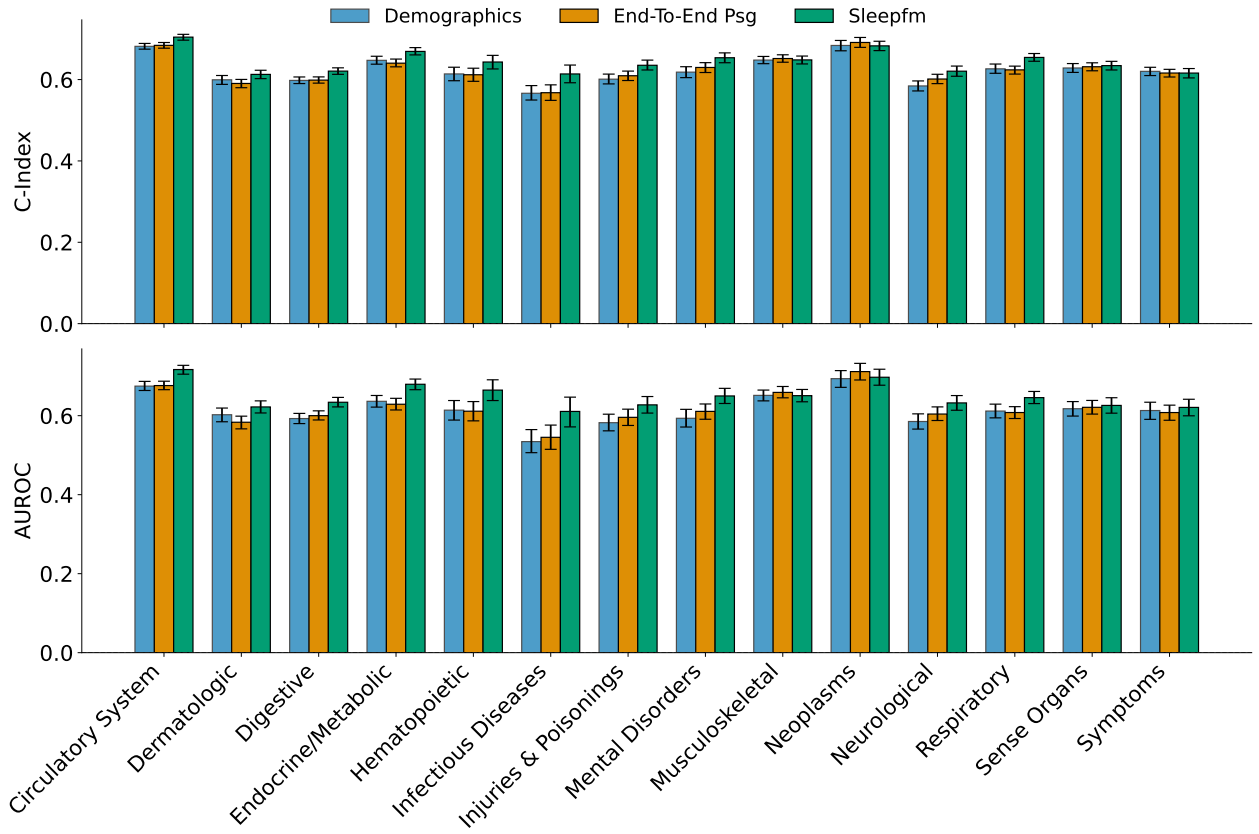
Supplementary Figure 6. Performance comparison of SleepFM with baseline models on the test set using C-Index and 6-year AUROC metrics. Each point represents a disease phenotype. (a) Comparison against the demographics-only baseline shows points consistently below the diagonal, indicating superior performance by SleepFM. (b) Comparison against the End-to-End PSG model highlights the advantages of foundation model pre-training over direct supervised learning from raw PSG signals. Overall, most points fall below the diagonal, demonstrating that SleepFM outperforms the baseline models for the majority of conditions.



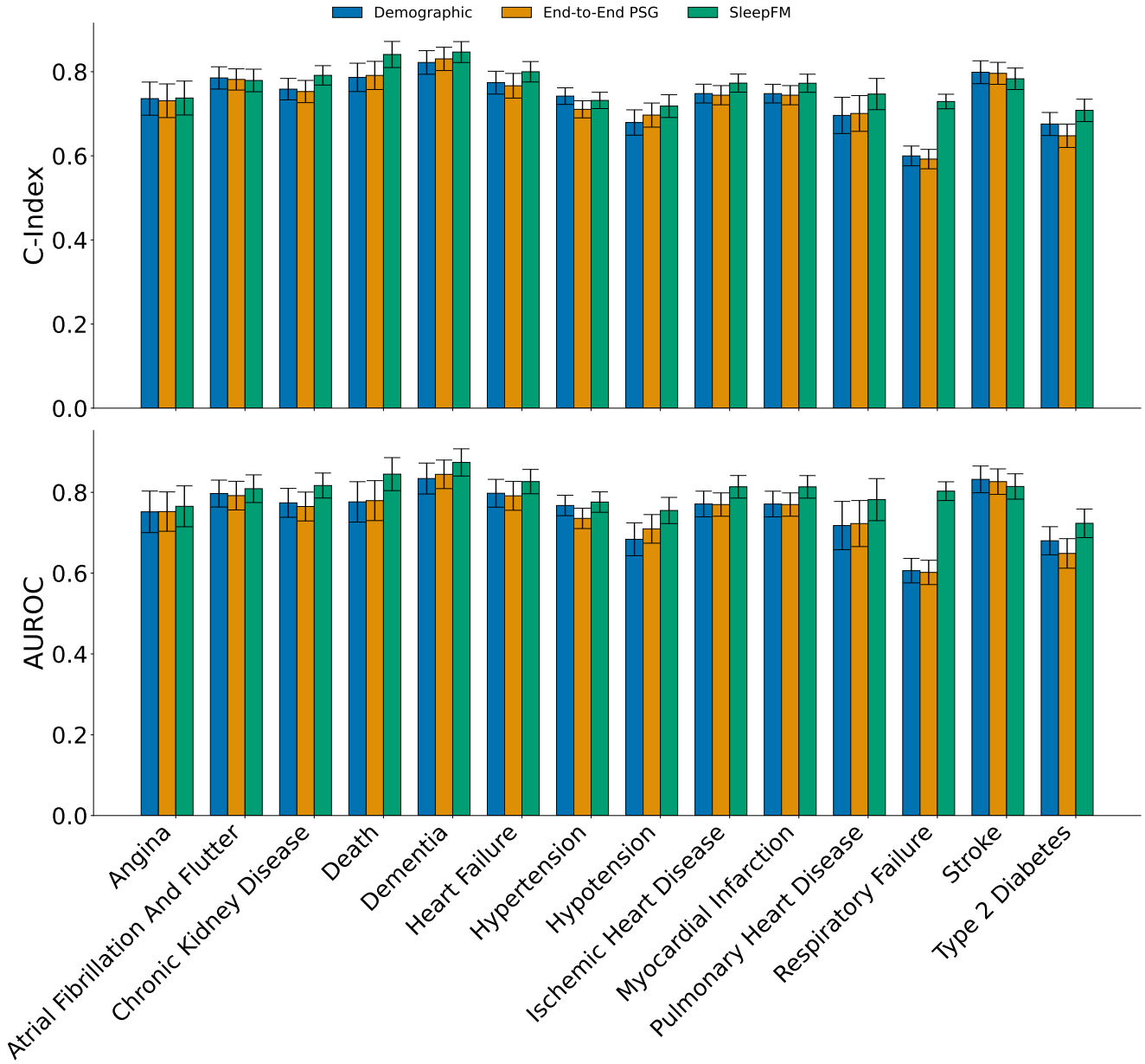
Supplementary Figure 7. Performance improvements of SleepFM over baseline models across disease categories. Results show percentage improvements in both C-Index and 6-year AUROC metrics on the test set. The comparison against the demographics-only model (using age, gender, BMI, and race/ethnicity features) shows substantial improvements across all disease categories, with particularly strong gains in neurological conditions and congenital anomalies. The comparison with the End-to-End PSG model, which is trained directly on raw PSG signals without pre-training, demonstrates the benefits of foundation model pre-training for learning robust sleep representations.



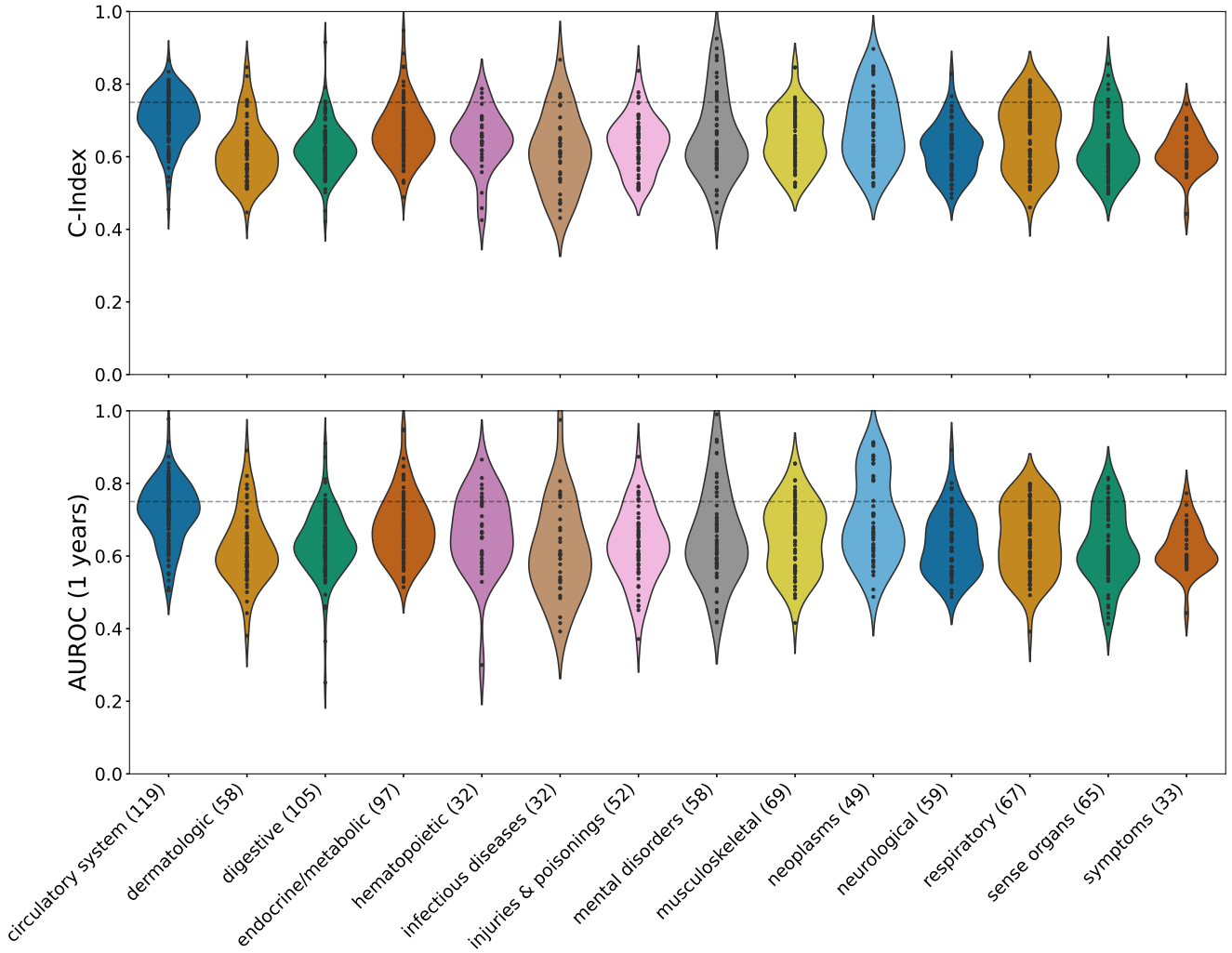
Supplementary Figure 8. Performance comparison of SleepFM with two baseline models across disease categories on the test set, showing average C-Index and 6-year AUROC metrics. The demographics baseline uses only clinical features (age, gender, BMI, and race/ethnicity), while the End-to-End PSG baseline is trained directly on raw PSG signals without pre-training. SleepFM demonstrates consistent performance advantages over both baselines across most disease categories, highlighting the benefits of foundation model pre-training on PSG data.



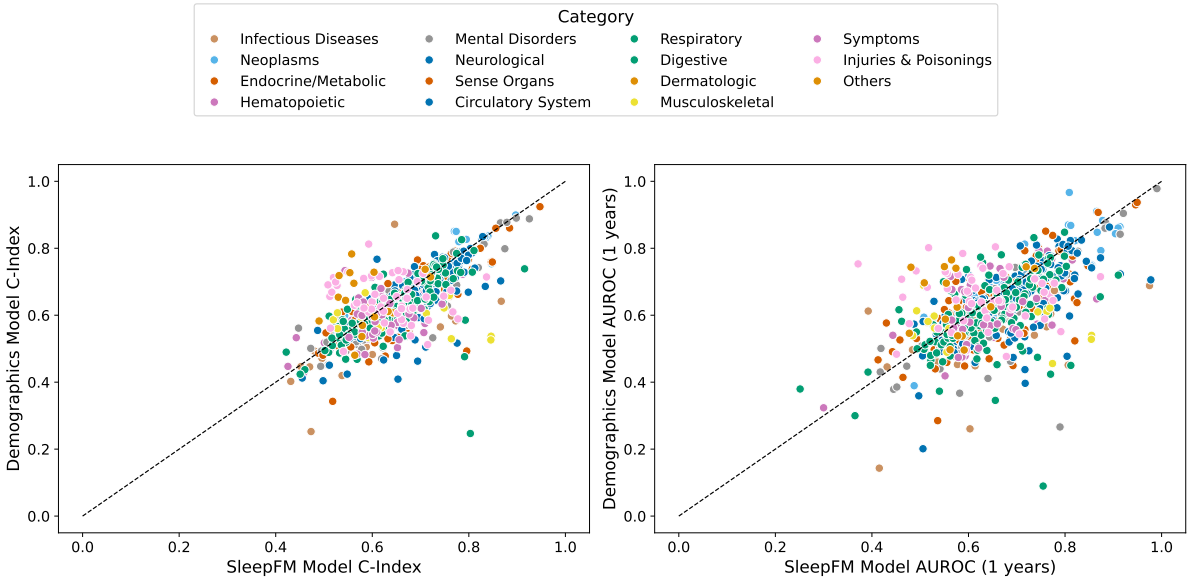
Supplementary Figure 9. Performance comparison of SleepFM with baseline models on the temporal test set (2020 onwards), showing average C-Index and 6-year AUROC metrics across disease categories. The demographics baseline uses only clinical features (age, gender, BMI, and race/ethnicity), while the End-to-End PSG baseline is trained directly on raw PSG signals without pre-training. Despite the temporal distribution shift, SleepFM maintains its performance advantages over both baselines across most disease categories, demonstrating robust generalization to more recent patients.



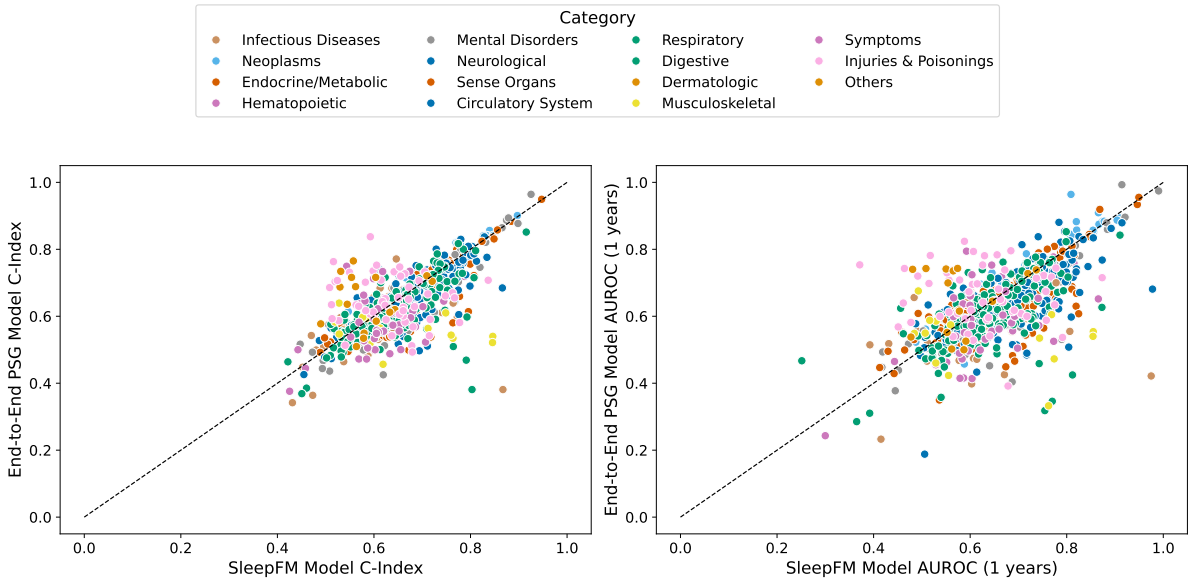
Supplementary Figure 10. Performance comparison across clinically relevant diseases selected in consultation with sleep experts. The plot compares SleepFM against two baselines: a demographics-only model using clinical features (age, gender, BMI, and race/ethnicity) and an End-to-End PSG model trained directly on raw PSG signals. Performance is evaluated on the test set using multiple metrics: C-Index and 6-year AUROC. The selected conditions include critical health outcomes such as death, heart failure, stroke, and dementia. SleepFM demonstrates superior predictive performance across these important clinical endpoints compared to both baseline approaches.



Supplementary Figure 11. SleepFM performance across disease categories on the temporal test set (Stanford data from 2020 onwards). Performance is measured using C-Index and 6-year AUROC metrics to assess the model's robustness to temporal distribution shifts. Despite these shifts, SleepFM maintains strong performance across a broad spectrum of diseases.

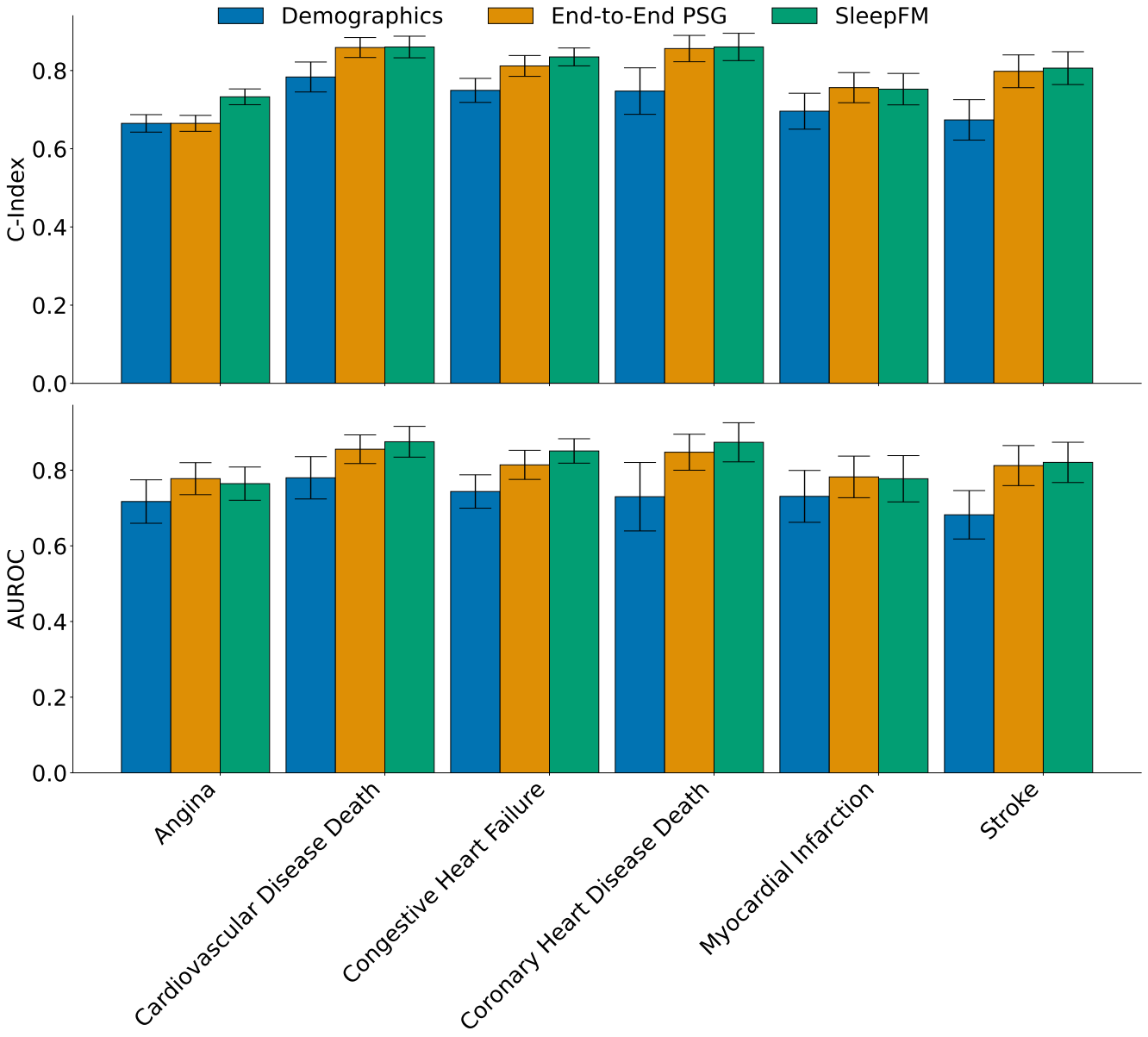


(a) Comparison with demographics-only baseline



(b) Comparison with End-to-End PSG model

Supplementary Figure 12. Performance comparison on the temporal test set (Stanford data from 2020 onwards) using C-Index and 6-year AUROC metrics. Each point represents a disease phenotype. (a) Comparison against demographics baseline shows points below the diagonal, indicating SleepFM's superior performance. (b) Comparison against End-to-End PSG model demonstrates the benefits of foundation model pre-training even under temporal distribution shift. These results highlight SleepFM's robust generalization to more recent patients. Overall, most points fall below the diagonal, demonstrating that SleepFM outperforms the baseline models for the majority of conditions.



Supplementary Figure 13. Performance comparison on the Sleep Heart Health Study (SHHS) dataset, which was not used during model pre-training. The plot compares three approaches: SleepFM, which extracts embeddings using the pre-trained foundation model and then fine-tunes on SHHS data; a demographics-only model trained directly on SHHS using clinical features (age, gender, BMI, and race/ethnicity); and an End-to-End PSG model trained from scratch on SHHS data. Performance is evaluated using C-Index and AUROC metrics. Despite SHHS being an entirely unseen dataset during pre-training, SleepFM's transfer learning approach demonstrates strong generalization performance compared to both baseline models trained specifically on SHHS data.

Phecode	Phenotype	Category	SleepFM	Demographics	End-to-End PSG
290.13	Senile dementia	mental disorders	0.99 (0.98, 1.00)	0.87 (0.75, 0.96)	0.95 (0.91, 0.98)
440.21	Atherosclerosis of native arteries of the extremities with ulceration or gangrene	circulatory system	0.92 (0.88, 0.95)	0.74 (0.64, 0.89)	0.66 (0.50, 0.89)
358.0	Myoneural disorders	neurological	0.81 (0.73, 0.88)	0.42 (0.28, 0.55)	0.60 (0.48, 0.71)
264.2	Failure to thrive (childhood)	endocrine/metabolic	0.77 (0.68, 0.88)	0.44 (0.26, 0.67)	0.45 (0.26, 0.67)
315.0	Develomental delays and disorders	mental disorders	0.80 (0.77, 0.84)	0.58 (0.51, 0.64)	0.61 (0.54, 0.67)
509.2	Respiratory insufficiency	respiratory	0.79 (0.72, 0.85)	0.59 (0.51, 0.67)	0.62 (0.53, 0.69)
277.4	Disorders of bilirubin excretion	endocrine/metabolic	0.79 (0.70, 0.85)	0.60 (0.46, 0.75)	0.61 (0.47, 0.75)
344.0	Other paralytic syndromes	neurological	0.77 (0.72, 0.83)	0.58 (0.48, 0.69)	0.60 (0.48, 0.69)
626.8	Infertility, female	genitourinary	0.89 (0.84, 0.93)	0.80 (0.73, 0.87)	0.88 (0.83, 0.92)
614.51	Cervicitis and endocervicitis	genitourinary	0.88 (0.85, 0.91)	0.79 (0.74, 0.83)	0.86 (0.83, 0.89)
619.2	Disorders of uterus, NEC	genitourinary	0.87 (0.84, 0.90)	0.79 (0.76, 0.82)	0.84 (0.81, 0.88)
609.0	Male infertility and abnormal spermatozoa	genitourinary	0.82 (0.78, 0.86)	0.71 (0.66, 0.76)	0.82 (0.78, 0.86)
504.0	Other alveolar and parietoalveolar pneumonopathy	respiratory	0.85 (0.80, 0.89)	0.75 (0.68, 0.81)	0.75 (0.68, 0.81)
619.3	Noninflammatory disorders of cervix	genitourinary	0.86 (0.82, 0.89)	0.77 (0.72, 0.81)	0.82 (0.79, 0.85)
250.25	Diabetes type 2 with peripheral circulatory disorders	endocrine/metabolic	0.87 (0.83, 0.91)	0.79 (0.74, 0.85)	0.78 (0.72, 0.84)
426.24	Atrioventricular block, complete	circulatory system	0.87 (0.83, 0.92)	0.80 (0.73, 0.87)	0.83 (0.76, 0.89)
348.8	Encephalopathy, not elsewhere classified	neurological	0.77 (0.70, 0.83)	0.63 (0.56, 0.71)	0.71 (0.64, 0.77)
276.42	Alkalosis	endocrine/metabolic	0.83 (0.77, 0.88)	0.74 (0.66, 0.80)	0.71 (0.63, 0.78)
249.0	Secondary diabetes mellitus	endocrine/metabolic	0.79 (0.74, 0.84)	0.68 (0.63, 0.72)	0.67 (0.62, 0.71)
447.7	Aortic ectasia	circulatory system	0.83 (0.80, 0.86)	0.74 (0.69, 0.79)	0.78 (0.74, 0.82)
280.2	Iron deficiency anemia secondary to blood loss (chronic)	hematopoietic	0.77 (0.73, 0.81)	0.64 (0.59, 0.69)	0.64 (0.60, 0.69)
218.1	Uterine leiomyoma	neoplasms	0.87 (0.85, 0.89)	0.81 (0.79, 0.83)	0.87 (0.84, 0.89)
300.9	Posttraumatic stress disorder	mental disorders	0.75 (0.70, 0.79)	0.62 (0.56, 0.68)	0.64 (0.59, 0.69)
331.0	Other cerebral degenerations	neurological	0.83 (0.76, 0.88)	0.74 (0.65, 0.82)	0.79 (0.70, 0.87)
290.2	Delirium due to conditions classified elsewhere	mental disorders	0.82 (0.76, 0.87)	0.73 (0.67, 0.78)	0.77 (0.71, 0.82)
218.0	Benign neoplasm of uterus	neoplasms	0.87 (0.85, 0.89)	0.81 (0.79, 0.83)	0.87 (0.85, 0.89)
276.4	Acid-base balance disorder	endocrine/metabolic	0.77 (0.72, 0.82)	0.66 (0.61, 0.72)	0.67 (0.61, 0.72)
250.24	Type 2 diabetes with neurological manifestations	endocrine/metabolic	0.81 (0.77, 0.85)	0.73 (0.68, 0.77)	0.71 (0.68, 0.75)
315.2	Speech and language disorder	mental disorders	0.81 (0.74, 0.87)	0.73 (0.63, 0.82)	0.73 (0.64, 0.82)
530.2	Esophageal bleeding (varices/hemorrhage)	digestive	0.76 (0.72, 0.79)	0.65 (0.60, 0.71)	0.69 (0.64, 0.74)
260.6	Anorexia	endocrine/metabolic	0.76 (0.71, 0.80)	0.65 (0.59, 0.72)	0.68 (0.62, 0.74)
458.2	Iatrogenic hypotension	circulatory system	0.76 (0.68, 0.83)	0.66 (0.58, 0.74)	0.67 (0.60, 0.74)
585.31	Renal dialysis	genitourinary	0.78 (0.73, 0.82)	0.68 (0.63, 0.74)	0.66 (0.60, 0.71)
426.8	Other cardiac conduction disorders	circulatory system	0.79 (0.75, 0.82)	0.70 (0.65, 0.74)	0.71 (0.66, 0.75)
276.11	Hyperosmolality and/or hypernatremia	endocrine/metabolic	0.76 (0.66, 0.86)	0.67 (0.58, 0.75)	0.68 (0.59, 0.76)
264.0	Lack of normal physiological development	endocrine/metabolic	0.77 (0.67, 0.86)	0.69 (0.56, 0.81)	0.69 (0.55, 0.81)
707.2	Chronic ulcer of leg or foot	dermatologic	0.80 (0.76, 0.85)	0.73 (0.68, 0.79)	0.70 (0.65, 0.75)
622.1	Polyp of corpus uteri	genitourinary	0.85 (0.81, 0.89)	0.79 (0.75, 0.83)	0.85 (0.81, 0.88)
594.3	Calculus of ureter	genitourinary	0.75 (0.69, 0.81)	0.67 (0.60, 0.73)	0.67 (0.62, 0.73)
250.7	Diabetic retinopathy	endocrine/metabolic	0.81 (0.77, 0.85)	0.75 (0.69, 0.80)	0.74 (0.69, 0.79)
513.8	Disorders of diaphragm	respiratory	0.75 (0.65, 0.84)	0.67 (0.56, 0.76)	0.68 (0.57, 0.77)
592.11	Acute cystitis	genitourinary	0.81 (0.78, 0.83)	0.74 (0.72, 0.77)	0.74 (0.71, 0.77)
285.1	Acute posthemorrhagic anemia	hematopoietic	0.78 (0.75, 0.81)	0.71 (0.67, 0.74)	0.72 (0.68, 0.75)
440.2	Atherosclerosis of the extremities	circulatory system	0.81 (0.74, 0.86)	0.75 (0.68, 0.80)	0.75 (0.69, 0.81)
362.27	Drusen (degenerative) of retina	sense organs	0.81 (0.77, 0.85)	0.75 (0.69, 0.80)	0.83 (0.79, 0.87)
501.0	Pneumonitis due to inhalation of food or vomitus	respiratory	0.75 (0.70, 0.80)	0.68 (0.62, 0.73)	0.69 (0.62, 0.76)
415.11	Pulmonary embolism and infarction, acute	circulatory system	0.80 (0.75, 0.85)	0.74 (0.68, 0.79)	0.75 (0.69, 0.81)
427.11	Paroxysmal supraventricular tachycardia	circulatory system	0.79 (0.76, 0.81)	0.72 (0.69, 0.76)	0.73 (0.70, 0.76)
415.1	Acute pulmonary heart disease	circulatory system	0.80 (0.75, 0.85)	0.74 (0.68, 0.80)	0.75 (0.69, 0.81)
411.8	Other chronic ischemic heart disease, unspecified	circulatory system	0.81 (0.76, 0.85)	0.75 (0.69, 0.79)	0.77 (0.72, 0.81)
440.22	Atherosclerosis of native arteries of the extremities with intermittent claudication	circulatory system	0.80 (0.73, 0.86)	0.75 (0.68, 0.83)	0.74 (0.65, 0.81)
411.9	Other acute and subacute forms of ischemic heart disease	circulatory system	0.78 (0.72, 0.85)	0.72 (0.64, 0.79)	0.72 (0.66, 0.79)
509.1	Respiratory failure	respiratory	0.77 (0.73, 0.80)	0.70 (0.65, 0.74)	0.69 (0.64, 0.73)
740.2	Osteoarthrosis, generalized	musculoskeletal	0.80 (0.76, 0.83)	0.74 (0.69, 0.79)	0.79 (0.74, 0.82)

Supplementary Table 8. Top conditions where sleep recordings provide substantial predictive value beyond demographic information alone. For each condition, we compare the predictive performance (C-Index) of SleepFM against demographic-only and end-to-end PSG models. The table highlights conditions where SleepFM achieved both strong absolute performance (C-Index > 0.75) and demonstrated meaningful improvement over demographic features (difference in C-Index > 0.5). Conditions are ranked by their relative improvement over demographics, emphasizing areas where sleep patterns contribute most significantly to prediction accuracy. Notable top conditions where SleepFM performs much better than baseline, include dementia, respiratory insufficiency, and developmental delays and disorders.

Phecode	Phenotype	Category	SleepFM	Demographics	End-to-End PSG
290.13	Senile dementia	mental disorders	0.99 (0.98, 1.00)	0.87 (0.75, 0.96)	0.95 (0.91, 0.98)
440.21	Atherosclerosis of native arteries of the extremities with ulceration or gangrene	circulatory system	0.92 (0.88, 0.95)	0.74 (0.64, 0.89)	0.66 (0.50, 0.89)
358.0	Myoneural disorders	neurological	0.81 (0.73, 0.88)	0.42 (0.28, 0.55)	0.60 (0.48, 0.71)
264.2	Failure to thrive (childhood)	endocrine/metabolic	0.77 (0.68, 0.88)	0.44 (0.26, 0.67)	0.45 (0.26, 0.67)
315.0	Develomental delays and disorders	mental disorders	0.80 (0.77, 0.84)	0.58 (0.51, 0.64)	0.61 (0.54, 0.67)
509.2	Respiratory insufficiency	respiratory	0.79 (0.72, 0.85)	0.59 (0.51, 0.67)	0.62 (0.53, 0.69)
277.4	Disorders of bilirubin excretion	endocrine/metabolic	0.79 (0.70, 0.85)	0.60 (0.46, 0.75)	0.61 (0.47, 0.75)
344.0	Other paralytic syndromes	neurological	0.77 (0.72, 0.83)	0.58 (0.48, 0.69)	0.60 (0.48, 0.69)
504.0	Other alveolar and parietoalveolar pneumonopathy	respiratory	0.85 (0.80, 0.89)	0.75 (0.68, 0.81)	0.75 (0.68, 0.81)
250.25	Diabetes type 2 with peripheral circulatory disorders	endocrine/metabolic	0.87 (0.83, 0.91)	0.79 (0.74, 0.85)	0.78 (0.72, 0.84)
426.24	Atrioventricular block, complete	circulatory system	0.87 (0.83, 0.92)	0.80 (0.73, 0.87)	0.83 (0.76, 0.89)
348.8	Encephalopathy, not elsewhere classified	neurological	0.77 (0.70, 0.83)	0.63 (0.56, 0.71)	0.71 (0.64, 0.77)
276.42	Alkalosis	endocrine/metabolic	0.83 (0.77, 0.88)	0.74 (0.66, 0.80)	0.71 (0.63, 0.78)
249.0	Secondary diabetes mellitus	endocrine/metabolic	0.79 (0.74, 0.84)	0.68 (0.63, 0.72)	0.67 (0.62, 0.71)
447.7	Aortic ectasia	circulatory system	0.83 (0.80, 0.86)	0.74 (0.69, 0.79)	0.78 (0.74, 0.82)
280.2	Iron deficiency anemia secondary to blood loss (chronic)	hematopoietic	0.77 (0.73, 0.81)	0.64 (0.59, 0.69)	0.64 (0.60, 0.69)
218.1	Uterine leiomyoma	neoplasms	0.87 (0.85, 0.89)	0.81 (0.79, 0.83)	0.87 (0.84, 0.89)
300.9	Posttraumatic stress disorder	mental disorders	0.75 (0.70, 0.79)	0.62 (0.56, 0.68)	0.64 (0.59, 0.69)
331.0	Other cerebral degenerations	neurological	0.83 (0.76, 0.88)	0.74 (0.65, 0.82)	0.79 (0.70, 0.87)
290.2	Delirium due to conditions classified elsewhere	mental disorders	0.82 (0.76, 0.87)	0.73 (0.67, 0.78)	0.77 (0.71, 0.82)
218.0	Benign neoplasm of uterus	neoplasms	0.87 (0.85, 0.89)	0.81 (0.79, 0.83)	0.87 (0.85, 0.89)
276.4	Acid-base balance disorder	endocrine/metabolic	0.77 (0.72, 0.82)	0.66 (0.61, 0.72)	0.67 (0.61, 0.72)
250.24	Type 2 diabetes with neurological manifestations	endocrine/metabolic	0.81 (0.77, 0.85)	0.73 (0.68, 0.77)	0.71 (0.68, 0.75)
315.2	Speech and language disorder	mental disorders	0.81 (0.74, 0.87)	0.73 (0.63, 0.82)	0.73 (0.64, 0.82)
530.2	Esophageal bleeding (varices/hemorrhage)	digestive	0.76 (0.72, 0.79)	0.65 (0.60, 0.71)	0.69 (0.64, 0.74)
260.6	Anorexia	endocrine/metabolic	0.76 (0.71, 0.80)	0.65 (0.59, 0.72)	0.68 (0.62, 0.74)
458.2	Iatrogenic hypotension	circulatory system	0.76 (0.68, 0.83)	0.66 (0.58, 0.74)	0.67 (0.60, 0.74)
426.8	Other cardiac conduction disorders	circulatory system	0.79 (0.75, 0.82)	0.70 (0.65, 0.74)	0.71 (0.66, 0.75)
276.11	Hyperosmolality and/or hypernatremia	endocrine/metabolic	0.76 (0.66, 0.86)	0.67 (0.58, 0.75)	0.68 (0.59, 0.76)
264.0	Lack of normal physiological development	endocrine/metabolic	0.77 (0.67, 0.86)	0.69 (0.56, 0.81)	0.69 (0.55, 0.81)
707.2	Chronic ulcer of leg or foot	dermatologic	0.80 (0.76, 0.85)	0.73 (0.68, 0.79)	0.70 (0.65, 0.75)
250.7	Diabetic retinopathy	endocrine/metabolic	0.81 (0.77, 0.85)	0.75 (0.69, 0.80)	0.74 (0.69, 0.79)
513.8	Disorders of diaphragm	respiratory	0.75 (0.65, 0.84)	0.67 (0.56, 0.76)	0.68 (0.57, 0.77)
285.1	Acute posthemorrhagic anemia	hematopoietic	0.78 (0.75, 0.81)	0.71 (0.67, 0.74)	0.72 (0.68, 0.75)
440.2	Atherosclerosis of the extremities	circulatory system	0.81 (0.74, 0.86)	0.75 (0.68, 0.80)	0.75 (0.69, 0.81)
362.27	Drusen (degenerative) of retina	sense organs	0.81 (0.77, 0.85)	0.75 (0.69, 0.80)	0.83 (0.79, 0.87)
501.0	Pneumonitis due to inhalation of food or vomitus	respiratory	0.75 (0.70, 0.80)	0.68 (0.62, 0.73)	0.69 (0.62, 0.76)
415.11	Pulmonary embolism and infarction, acute	circulatory system	0.80 (0.75, 0.85)	0.74 (0.68, 0.79)	0.75 (0.69, 0.81)
427.11	Paroxysmal supraventricular tachycardia	circulatory system	0.79 (0.76, 0.81)	0.72 (0.69, 0.76)	0.73 (0.70, 0.76)
415.1	Acute pulmonary heart disease	circulatory system	0.80 (0.75, 0.85)	0.74 (0.68, 0.80)	0.75 (0.69, 0.81)
411.8	Other chronic ischemic heart disease, unspecified	circulatory system	0.81 (0.76, 0.85)	0.75 (0.69, 0.79)	0.77 (0.72, 0.81)
440.22	Atherosclerosis of native arteries of the extremities with intermittent claudication	circulatory system	0.80 (0.73, 0.86)	0.75 (0.68, 0.83)	0.74 (0.65, 0.81)
411.9	Other acute and subacute forms of ischemic heart disease	circulatory system	0.78 (0.72, 0.85)	0.72 (0.64, 0.79)	0.72 (0.66, 0.79)
509.1	Respiratory failure	respiratory	0.77 (0.73, 0.80)	0.70 (0.65, 0.74)	0.69 (0.64, 0.73)
740.2	Osteoarthrosis, generalized	musculoskeletal	0.80 (0.76, 0.83)	0.74 (0.69, 0.79)	0.79 (0.74, 0.82)
427.1	Paroxysmal tachycardia, unspecified	circulatory system	0.78 (0.76, 0.81)	0.73 (0.70, 0.75)	0.73 (0.71, 0.76)
427.4	Cardiac arrest and ventricular fibrillation	circulatory system	0.77 (0.68, 0.84)	0.71 (0.63, 0.78)	0.73 (0.65, 0.81)
569.2	Gastrointestinal complications	digestive	0.76 (0.70, 0.81)	0.69 (0.62, 0.77)	0.66 (0.59, 0.72)
427.12	Paroxysmal ventricular tachycardia	circulatory system	0.78 (0.74, 0.82)	0.72 (0.68, 0.77)	0.72 (0.67, 0.77)
260.2	severe protein-calorie malnutrition	endocrine/metabolic	0.79 (0.74, 0.83)	0.73 (0.68, 0.78)	0.71 (0.67, 0.76)
285.2	Anemia of chronic disease	hematopoietic	0.76 (0.72, 0.79)	0.70 (0.66, 0.75)	0.71 (0.67, 0.75)
459.9	Circulatory disease NEC	circulatory system	0.77 (0.74, 0.79)	0.71 (0.68, 0.74)	0.71 (0.68, 0.75)
276.13	Hyperpotassemia	endocrine/metabolic	0.75 (0.71, 0.80)	0.70 (0.65, 0.75)	0.70 (0.65, 0.74)
574.2	Calculus of bile duct	digestive	0.76 (0.70, 0.80)	0.70 (0.63, 0.78)	0.70 (0.63, 0.78)
250.1	Type 1 diabetes	endocrine/metabolic	0.75 (0.69, 0.81)	0.70 (0.64, 0.75)	0.68 (0.62, 0.75)

Supplementary Table 9. Top conditions where sleep recordings demonstrate strong predictive performance for 6-year risk assessment. For each condition, we compare the 6-year AUROC of SleepFM against demographic-only and end-to-end PSG models. The table includes conditions where SleepFM achieved both strong absolute performance (AUROC > 0.75) and substantial improvement over the demographic baseline. Conditions are ranked by their relative improvement over demographics, highlighting diseases where sleep patterns provide the most significant additional predictive value. Notable top conditions where SleepFM performs much better than baseline, include dementia, respiratory insufficiency, and developmental delays and disorders.

Phecode	Phenotype	Category	C-Index		AUROC (6 Years)	
			SleepFM	Demographics	SleepFM	Demographics
290.13	Senile dementia	mental disorders	0.99 (0.98, 1.00)	0.87 (0.75, 0.96)	0.99 (0.98, 1.00)	0.85 (0.75, 0.97)
440.21	Atherosclerosis of native arteries of the extremities with ulceration or gangrene	circulatory system	0.92 (0.88, 0.95)	0.74 (0.64, 0.89)	0.95 (0.91, 0.98)	0.76 (0.72, 0.80)
358.0	Myoneural disorders	neurological	0.81 (0.73, 0.88)	0.42 (0.28, 0.55)	0.84 (0.75, 0.91)	0.38 (0.25, 0.53)
264.2	Failure to thrive (childhood)	endocrine/metabolic	0.77 (0.68, 0.88)	0.44 (0.26, 0.67)	0.75 (0.65, 0.87)	0.46 (0.29, 0.66)
315.0	Develomental delays and disorders	mental disorders	0.80 (0.77, 0.84)	0.58 (0.51, 0.64)	0.84 (0.79, 0.87)	0.58 (0.49, 0.66)
509.2	Respiratory insufficiency	respiratory	0.79 (0.72, 0.85)	0.59 (0.51, 0.67)	0.82 (0.72, 0.91)	0.57 (0.46, 0.67)
277.4	Disorders of bilirubin excretion	endocrine/metabolic	0.79 (0.70, 0.85)	0.60 (0.46, 0.75)	0.89 (0.84, 0.93)	0.62 (0.41, 0.80)
344.0	Other paralytic syndromes	neurological	0.77 (0.72, 0.83)	0.58 (0.48, 0.69)	0.82 (0.75, 0.88)	0.55 (0.41, 0.68)
504.0	Other alveolar and parietoalveolar pneumonopathy	respiratory	0.85 (0.80, 0.89)	0.75 (0.68, 0.81)	0.90 (0.84, 0.95)	0.76 (0.66, 0.84)
250.25	Diabetes type 2 with peripheral circulatory disorders	endocrine/metabolic	0.87 (0.83, 0.91)	0.79 (0.74, 0.85)	0.89 (0.85, 0.93)	0.79 (0.72, 0.87)
426.24	Atrioventricular block, complete	circulatory system	0.87 (0.83, 0.92)	0.80 (0.73, 0.87)	0.89 (0.84, 0.94)	0.80 (0.70, 0.88)
348.8	Encephalopathy, not elsewhere classified	neurological	0.77 (0.70, 0.83)	0.63 (0.56, 0.71)	0.78 (0.70, 0.86)	0.63 (0.54, 0.73)
276.42	Alkalosis	endocrine/metabolic	0.83 (0.77, 0.88)	0.74 (0.66, 0.80)	0.87 (0.77, 0.93)	0.75 (0.64, 0.84)
249.0	Secondary diabetes mellitus	endocrine/metabolic	0.79 (0.74, 0.84)	0.68 (0.63, 0.72)	0.82 (0.76, 0.88)	0.67 (0.59, 0.74)
447.7	Aortic ectasia	circulatory system	0.83 (0.80, 0.86)	0.74 (0.69, 0.79)	0.90 (0.87, 0.93)	0.74 (0.67, 0.80)
280.2	Iron deficiency anemia secondary to blood loss (chronic)	hematopoietic	0.77 (0.73, 0.81)	0.64 (0.59, 0.69)	0.84 (0.79, 0.89)	0.66 (0.59, 0.72)
218.1	Uterine leiomyoma	neoplasms	0.87 (0.85, 0.89)	0.81 (0.79, 0.83)	0.88 (0.85, 0.91)	0.80 (0.77, 0.83)
300.9	Posttraumatic stress disorder	mental disorders	0.75 (0.70, 0.79)	0.62 (0.56, 0.68)	0.82 (0.76, 0.87)	0.61 (0.54, 0.68)
331.0	Other cerebral degenerations	neurological	0.83 (0.76, 0.88)	0.74 (0.65, 0.82)	0.85 (0.77, 0.91)	0.75 (0.65, 0.84)
290.2	Delirium due to conditions classified elsewhere	mental disorders	0.82 (0.76, 0.87)	0.73 (0.67, 0.78)	0.81 (0.74, 0.88)	0.71 (0.64, 0.78)
218.0	Benign neoplasm of uterus	neoplasms	0.87 (0.85, 0.89)	0.81 (0.79, 0.83)	0.88 (0.84, 0.91)	0.80 (0.77, 0.83)
276.4	Acid-base balance disorder	endocrine/metabolic	0.77 (0.72, 0.82)	0.66 (0.61, 0.72)	0.79 (0.73, 0.85)	0.67 (0.59, 0.74)
250.24	Type 2 diabetes with neurological manifestations	endocrine/metabolic	0.81 (0.77, 0.85)	0.73 (0.68, 0.77)	0.86 (0.81, 0.90)	0.74 (0.68, 0.80)
315.2	Speech and language disorder	mental disorders	0.81 (0.74, 0.87)	0.73 (0.63, 0.82)	0.83 (0.74, 0.90)	0.72 (0.60, 0.83)
530.2	Esophageal bleeding (varices/hemorrhage)	digestive	0.76 (0.72, 0.79)	0.65 (0.60, 0.71)	0.87 (0.84, 0.90)	0.66 (0.57, 0.73)
260.6	Anorexia	endocrine/metabolic	0.76 (0.71, 0.80)	0.65 (0.59, 0.72)	0.84 (0.79, 0.88)	0.65 (0.55, 0.74)
458.2	Iatrogenic hypotension	circulatory system	0.76 (0.68, 0.83)	0.66 (0.58, 0.74)	0.77 (0.65, 0.87)	0.66 (0.53, 0.77)
426.8	Other cardiac conduction disorders	circulatory system	0.79 (0.75, 0.82)	0.70 (0.65, 0.74)	0.82 (0.77, 0.87)	0.69 (0.62, 0.75)
276.11	Hyperosmolality and/or hypernatremia	endocrine/metabolic	0.76 (0.66, 0.86)	0.67 (0.58, 0.75)	0.77 (0.59, 0.94)	0.78 (0.68, 0.88)
264.0	Lack of normal physiological development	endocrine/metabolic	0.77 (0.67, 0.86)	0.69 (0.56, 0.81)	0.72 (0.60, 0.84)	0.64 (0.49, 0.78)
707.2	Chronic ulcer of leg or foot	dermatologic	0.80 (0.76, 0.85)	0.73 (0.68, 0.79)	0.84 (0.79, 0.88)	0.75 (0.67, 0.81)
250.7	Diabetic retinopathy	endocrine/metabolic	0.81 (0.77, 0.85)	0.75 (0.69, 0.80)	0.84 (0.79, 0.89)	0.77 (0.70, 0.82)
513.8	Disorders of diaphragm	respiratory	0.75 (0.65, 0.84)	0.67 (0.56, 0.76)	0.73 (0.58, 0.86)	0.63 (0.48, 0.77)
285.1	Acute posthemorrhagic anemia	hematopoietic	0.78 (0.75, 0.81)	0.71 (0.67, 0.74)	0.83 (0.79, 0.87)	0.71 (0.66, 0.77)
440.2	Atherosclerosis of the extremities	circulatory system	0.81 (0.74, 0.86)	0.75 (0.68, 0.80)	0.84 (0.75, 0.90)	0.77 (0.68, 0.84)
362.27	Drusen (degenerative) of retina	sense organs	0.81 (0.77, 0.85)	0.75 (0.69, 0.80)	0.83 (0.78, 0.88)	0.72 (0.65, 0.78)
501.0	Pneumonitis due to inhalation of food or vomitus	respiratory	0.75 (0.70, 0.80)	0.68 (0.62, 0.73)	0.75 (0.69, 0.81)	0.65 (0.56, 0.72)
415.11	Pulmonary embolism and infarction, acute	circulatory system	0.80 (0.75, 0.85)	0.74 (0.68, 0.79)	0.85 (0.78, 0.91)	0.76 (0.68, 0.83)
427.11	Paroxysmal supraventricular tachycardia	circulatory system	0.79 (0.76, 0.81)	0.72 (0.69, 0.76)	0.87 (0.84, 0.90)	0.74 (0.69, 0.78)
415.1	Acute pulmonary heart disease	circulatory system	0.80 (0.75, 0.85)	0.74 (0.68, 0.80)	0.84 (0.77, 0.90)	0.76 (0.68, 0.82)
411.8	Other chronic ischemic heart disease, unspecified	circulatory system	0.81 (0.76, 0.85)	0.75 (0.69, 0.79)	0.81 (0.76, 0.86)	0.73 (0.66, 0.79)
440.22	Atherosclerosis of native arteries of the extremities with intermittent claudication	circulatory system	0.80 (0.73, 0.86)	0.75 (0.68, 0.83)	0.82 (0.74, 0.89)	0.75 (0.67, 0.84)
411.9	Other acute and subacute forms of ischemic heart disease	circulatory system	0.78 (0.72, 0.85)	0.72 (0.64, 0.79)	0.79 (0.71, 0.88)	0.74 (0.65, 0.82)
509.1	Respiratory failure	respiratory	0.77 (0.73, 0.80)	0.70 (0.65, 0.74)	0.76 (0.71, 0.82)	0.69 (0.63, 0.75)
740.2	Osteoarthrosis, generalized	musculoskeletal	0.80 (0.76, 0.83)	0.74 (0.69, 0.79)	0.85 (0.80, 0.89)	0.77 (0.71, 0.83)
427.1	Paroxysmal tachycardia, unspecified	circulatory system	0.78 (0.76, 0.81)	0.73 (0.70, 0.75)	0.87 (0.84, 0.89)	0.74 (0.70, 0.77)
427.4	Cardiac arrest and ventricular fibrillation	circulatory system	0.77 (0.68, 0.84)	0.71 (0.63, 0.78)	0.76 (0.66, 0.86)	0.73 (0.65, 0.81)
569.2	Gastrointestinal complications	digestive	0.76 (0.70, 0.81)	0.69 (0.62, 0.77)	0.79 (0.71, 0.87)	0.68 (0.59, 0.77)
427.12	Paroxysmal ventricular tachycardia	circulatory system	0.78 (0.74, 0.82)	0.72 (0.68, 0.77)	0.84 (0.79, 0.88)	0.71 (0.65, 0.77)
260.2	severe protein-calorie malnutrition	endocrine/metabolic	0.79 (0.74, 0.83)	0.73 (0.68, 0.78)	0.83 (0.75, 0.90)	0.77 (0.70, 0.84)
285.2	Anemia of chronic disease	hematopoietic	0.76 (0.72, 0.79)	0.70 (0.66, 0.75)	0.77 (0.72, 0.82)	0.69 (0.62, 0.75)
459.9	Circulatory disease NEC	circulatory system	0.77 (0.74, 0.79)	0.71 (0.68, 0.74)	0.84 (0.80, 0.87)	0.74 (0.69, 0.78)
276.13	Hyperpotassemia	endocrine/metabolic	0.75 (0.71, 0.80)	0.70 (0.65, 0.75)	0.79 (0.73, 0.85)	0.69 (0.64, 0.75)
574.2	Calculus of bile duct	digestive	0.76 (0.70, 0.80)	0.70 (0.63, 0.78)	0.79 (0.72, 0.85)	0.72 (0.65, 0.79)
250.1	Type 1 diabetes	endocrine/metabolic	0.75 (0.69, 0.81)	0.70 (0.64, 0.75)	0.77 (0.70, 0.82)	0.72 (0.65, 0.78)

Supplementary Table 10. Comparison of performance metrics, including C-Index and AUROC (6 Years) for SleepFM and demographics model. Conditions are ranked by their relative improvement over demographics model, retaining only conditions where the C-Index for SleepFM exceeds 0.75. All metrics include 95% confidence intervals calculated using bootstrapping.

Phecode	Phenotype	Category	C-Index		AUROC (6 Years)	
			SleepFM	End-to-End PSG	SleepFM	End-to-End PSG
440.21	Atherosclerosis of native arteries of the extremities with ulceration or gangrene	circulatory system	0.92 (0.88, 0.95)	0.66 (0.50, 0.89)	0.95 (0.92, 0.98)	0.65 (0.61, 0.69)
264.2	Failure to thrive (childhood)	endocrine/metabolic	0.77 (0.68, 0.88)	0.45 (0.26, 0.67)	0.75 (0.65, 0.87)	0.44 (0.26, 0.66)
358.0	Myoneural disorders	neurological	0.81 (0.73, 0.88)	0.60 (0.48, 0.71)	0.84 (0.75, 0.91)	0.54 (0.40, 0.69)
315.0	Develomental delays and disorders	mental disorders	0.80 (0.77, 0.84)	0.61 (0.54, 0.67)	0.84 (0.79, 0.87)	0.61 (0.52, 0.69)
509.2	Respiratory insufficiency	respiratory	0.79 (0.72, 0.85)	0.62 (0.53, 0.69)	0.82 (0.72, 0.91)	0.64 (0.54, 0.73)
277.4	Disorders of bilirubin excretion	endocrine/metabolic	0.79 (0.70, 0.85)	0.61 (0.47, 0.75)	0.89 (0.84, 0.93)	0.60 (0.41, 0.78)
344.0	Other paralytic syndromes	neurological	0.77 (0.72, 0.83)	0.60 (0.48, 0.69)	0.82 (0.75, 0.88)	0.54 (0.40, 0.68)
276.42	Alkalosis	endocrine/metabolic	0.83 (0.77, 0.88)	0.71 (0.63, 0.78)	0.87 (0.77, 0.93)	0.71 (0.59, 0.81)
250.25	Diabetes type 2 with peripheral circulatory disorders	endocrine/metabolic	0.87 (0.83, 0.91)	0.78 (0.72, 0.84)	0.89 (0.85, 0.93)	0.79 (0.70, 0.87)
504.0	Other alveolar and parietoalveolar pneumonopathy	respiratory	0.85 (0.80, 0.89)	0.75 (0.68, 0.81)	0.90 (0.84, 0.95)	0.76 (0.66, 0.84)
249.0	Secondary diabetes mellitus	endocrine/metabolic	0.79 (0.74, 0.84)	0.67 (0.62, 0.71)	0.82 (0.76, 0.88)	0.66 (0.59, 0.72)
250.24	Type 2 diabetes with neurological manifestations	endocrine/metabolic	0.81 (0.77, 0.85)	0.71 (0.68, 0.75)	0.86 (0.81, 0.90)	0.73 (0.67, 0.78)
280.2	Iron deficiency anemia secondary to blood loss (chronic)	hematopoietic	0.77 (0.73, 0.81)	0.64 (0.60, 0.69)	0.84 (0.79, 0.89)	0.66 (0.60, 0.71)
707.2	Chronic ulcer of leg or foot	dermatologic	0.80 (0.76, 0.85)	0.70 (0.65, 0.75)	0.84 (0.79, 0.88)	0.73 (0.67, 0.79)
276.4	Acid-base balance disorder	endocrine/metabolic	0.77 (0.72, 0.82)	0.67 (0.61, 0.72)	0.79 (0.73, 0.85)	0.67 (0.60, 0.74)
300.9	Posttraumatic stress disorder	mental disorders	0.75 (0.70, 0.79)	0.64 (0.59, 0.69)	0.82 (0.76, 0.87)	0.63 (0.56, 0.69)
315.2	Speech and language disorder	mental disorders	0.81 (0.74, 0.87)	0.73 (0.64, 0.82)	0.83 (0.74, 0.90)	0.71 (0.60, 0.83)
569.2	Gastrointestinal complications	digestive	0.76 (0.70, 0.81)	0.66 (0.59, 0.72)	0.79 (0.71, 0.87)	0.65 (0.57, 0.73)
426.8	Other cardiac conduction disorders	circulatory system	0.79 (0.75, 0.82)	0.71 (0.66, 0.75)	0.82 (0.77, 0.87)	0.69 (0.62, 0.76)
458.2	Iatrogenic hypotension	circulatory system	0.76 (0.68, 0.83)	0.67 (0.60, 0.74)	0.77 (0.65, 0.87)	0.67 (0.56, 0.77)
250.7	Diabetic retinopathy	endocrine/metabolic	0.81 (0.77, 0.85)	0.74 (0.69, 0.79)	0.84 (0.79, 0.89)	0.76 (0.69, 0.82)
276.11	Hyperosmolality and/or hypernatremia	endocrine/metabolic	0.76 (0.66, 0.86)	0.68 (0.59, 0.76)	0.77 (0.59, 0.94)	0.74 (0.62, 0.85)
264.0	Lack of normal physiological development	endocrine/metabolic	0.77 (0.67, 0.86)	0.69 (0.55, 0.81)	0.72 (0.60, 0.84)	0.63 (0.48, 0.77)
440.22	Atherosclerosis of native arteries of the extremities with intermittent claudication	circulatory system	0.80 (0.73, 0.86)	0.74 (0.65, 0.81)	0.82 (0.74, 0.89)	0.74 (0.65, 0.83)
274.11	Gouty arthropathy	endocrine/metabolic	0.76 (0.71, 0.81)	0.68 (0.63, 0.73)	0.77 (0.70, 0.84)	0.67 (0.60, 0.74)
509.1	Respiratory failure	respiratory	0.77 (0.73, 0.80)	0.69 (0.64, 0.73)	0.76 (0.71, 0.82)	0.68 (0.62, 0.74)
260.2	severe protein-calorie malnutrition	endocrine/metabolic	0.79 (0.74, 0.83)	0.71 (0.67, 0.76)	0.83 (0.75, 0.90)	0.72 (0.64, 0.80)
550.4	Umbilical hernia	digestive	0.79 (0.74, 0.83)	0.72 (0.64, 0.79)	0.84 (0.78, 0.90)	0.75 (0.66, 0.83)
513.8	Disorders of diaphragm	respiratory	0.75 (0.65, 0.84)	0.68 (0.57, 0.77)	0.73 (0.58, 0.86)	0.63 (0.48, 0.79)
260.6	Anorexia	endocrine/metabolic	0.76 (0.71, 0.80)	0.68 (0.62, 0.74)	0.84 (0.79, 0.88)	0.68 (0.58, 0.77)
530.2	Esophageal bleeding (varices/hemorrhage)	digestive	0.76 (0.72, 0.79)	0.69 (0.64, 0.74)	0.87 (0.84, 0.90)	0.69 (0.61, 0.75)
440.2	Atherosclerosis of the extremities	circulatory system	0.81 (0.74, 0.86)	0.75 (0.69, 0.81)	0.84 (0.75, 0.90)	0.78 (0.71, 0.85)
411.9	Other acute and subacute forms of ischemic heart disease	circulatory system	0.78 (0.72, 0.85)	0.72 (0.66, 0.79)	0.79 (0.71, 0.88)	0.75 (0.68, 0.82)
227.0	Benign neoplasm of other endocrine glands and related structures	neoplasms	0.75 (0.66, 0.84)	0.69 (0.59, 0.77)	0.82 (0.69, 0.93)	0.72 (0.59, 0.83)
250.1	Type 1 diabetes	endocrine/metabolic	0.75 (0.69, 0.81)	0.68 (0.62, 0.75)	0.77 (0.70, 0.82)	0.70 (0.64, 0.78)
285.1	Acute posthemorrhagic anemia	hematopoietic	0.78 (0.75, 0.81)	0.72 (0.68, 0.75)	0.83 (0.79, 0.87)	0.73 (0.68, 0.78)
415.1	Acute pulmonary heart disease	circulatory system	0.80 (0.75, 0.85)	0.75 (0.69, 0.81)	0.84 (0.77, 0.90)	0.76 (0.69, 0.83)
348.8	Encephalopathy, not elsewhere classified	neurological	0.77 (0.70, 0.83)	0.71 (0.64, 0.77)	0.78 (0.70, 0.86)	0.70 (0.62, 0.78)
427.11	Paroxysmal supraventricular tachycardia	circulatory system	0.79 (0.76, 0.81)	0.73 (0.70, 0.76)	0.87 (0.84, 0.90)	0.75 (0.70, 0.79)
427.12	Paroxysmal ventricular tachycardia	circulatory system	0.78 (0.74, 0.82)	0.72 (0.67, 0.77)	0.84 (0.79, 0.88)	0.72 (0.65, 0.78)
428.2	Heart failure NOS	circulatory system	0.77 (0.70, 0.84)	0.71 (0.64, 0.78)	0.81 (0.74, 0.88)	0.72 (0.64, 0.80)
427.1	Paroxysmal tachycardia, unspecified	circulatory system	0.78 (0.76, 0.81)	0.73 (0.71, 0.76)	0.87 (0.84, 0.89)	0.75 (0.71, 0.79)
501.0	Pneumonitis due to inhalation of food or vomitus	respiratory	0.75 (0.70, 0.80)	0.69 (0.62, 0.76)	0.75 (0.69, 0.81)	0.66 (0.57, 0.74)
276.13	Hyperpotassemia	endocrine/metabolic	0.75 (0.71, 0.80)	0.70 (0.65, 0.74)	0.79 (0.73, 0.85)	0.69 (0.63, 0.74)
459.9	Circulatory disease NEC	circulatory system	0.77 (0.74, 0.79)	0.71 (0.68, 0.75)	0.84 (0.80, 0.87)	0.74 (0.69, 0.78)
574.2	Calculus of bile duct	digestive	0.76 (0.70, 0.80)	0.70 (0.63, 0.78)	0.79 (0.72, 0.85)	0.72 (0.64, 0.80)

Supplementary Table 11. Comparison of performance metrics, including C-Index and AUROC (6 Years) for SleepFM and End-to-End PSG model. Conditions are ranked by their relative improvement over End-to-End PSG model, retaining only conditions where the C-Index for SleepFM exceeds 0.75. All metrics include 95% confidence intervals calculated using bootstrapping.

Modality	Channel	Prevalence (%)
RESP	Airflow	100.00
RESP	HR	99.92
RESP	Abdominal	73.92
RESP	SaO2	73.92
RESP	Thoracic	73.92
RESP	ABD	26.08
RESP	Chest	26.08
RESP	SpO2	26.08
BAS	C3	99.85
BAS	C4	99.85
BAS	LOC	73.92
BAS	ROC	73.92
BAS	A1	73.74
BAS	A2	73.74
BAS	M1	26.11
BAS	M2	26.11
BAS	E1	26.08
BAS	E2	26.08
EKG	ECG L	73.77
EKG	ECG R	73.77
EKG	ECGL	26.08
EKG	ECGR	26.08
EKG	ECG L-ECG R	0.15
EMG	Leg L	73.92
EMG	Leg R	73.92
EMG	L Chin	73.77
EMG	R Chin	73.77
EMG	LChin	26.08
EMG	LegL	26.08
EMG	LegR	26.08
EMG	RChin	26.08
EMG	L Chin-R Chin	0.15

Supplementary Table 12. Channels and prevalence percentages by modality for MROS

modality	channel	prevalence
RESP	Abdo	100.00
RESP	HR	100.00
RESP	Snore	100.00
RESP	SpO2	100.00
RESP	Thor	100.00
RESP	Therm	99.37
BAS	EEG1	100.00
BAS	EEG2	100.00
BAS	EEG3	100.00
BAS	EOG-L	100.00
BAS	EOG-R	100.00
EKG	EKG	100.00
EMG	EMG	100.00
EMG	Leg	100.00
EMG	Pleth	100.00

Supplementary Table 13. Channels and prevalence percentages by modality for MESA

modality	channel	prevalence
RESP	ABDO RES	100.00
RESP	SaO2	100.00
RESP	THOR RES	100.00
RESP	AIRFLOW	75.94
RESP	H.R.	68.69
RESP	NEW AIR	46.64
RESP	New Air	9.43
RESP	NEWAIR	4.14
RESP	AIRFLOW-0	1.08
RESP	AIRFLOW-1	1.08
EKG	ECG	100.00
BAS	EEG	100.00
BAS	EOG(L)	100.00
BAS	EOG(R)	100.00
BAS	EEG(sec)	97.95
BAS	EEG 2	1.15
BAS	EEG2	0.65
BAS	EEG sec	0.15
BAS	EEG(SEC)	0.09
EMG	EMG	100.00
EMG	LEG(L)	0.02
EMG	LEG(R)	0.02

Supplementary Table 14. Channels and prevalence percentages by modality for SHHS

modality	channel	prevalence
BAS	LOC	70.10
BAS	ROC	70.10
BAS	A1A2	68.81
BAS	C3M2	60.02
BAS	C4M1	60.02
BAS	O1M2	60.02
BAS	O2M1	60.02
BAS	F3M2	58.78
BAS	F4M1	58.78
BAS	C3A2	36.61
RESP	Abd	83.75
RESP	SpO2	72.65
RESP	Snore	70.40
RESP	Pulse	27.17
RESP	SAO2	26.81
RESP	SNOR	16.34
RESP	PPG	13.07
RESP	ABD	12.29
RESP	Abdomen	2.03
RESP	Thorax	2.03
EMG	LLeg	75.01
EMG	RLeg	75.01
EMG	Chin	67.73
EMG	CHIN	26.86
EMG	LLEG	18.28
EMG	RLEG	18.28
EMG	InterEMG	2.91
EMG	Pleth	2.42
EMG	CHINEMG	1.96
EMG	LLEGEMG	1.96
EKG	ECG	92.22
EKG	EKG	2.35
EKG	ECGI-0	1.83
EKG	ECGI-1	1.83
EKG	ECGII-0	1.83
EKG	ECGII-1	1.83
EKG	ECG1	1.27
EKG	ECGI	1.23
EKG	ECG2	0.41
EKG	ECG II	0.23

Supplementary Table 15. Channels and prevalence percentages by modality for BioSerenity

modality	channel	prevalence
RESP	Snore	88.04
RESP	Chest	83.96
RESP	SpO2	66.98
RESP	Abd	65.96
RESP	Pulse Rate	45.14
RESP	Nasal	43.81
RESP	Nasal Pressure	43.49
RESP	Oral Therm	42.88
RESP	Pulse	39.97
RESP	Oral-CO2	37.94
BAS	E1-Cz	43.53
BAS	C3-Cz	43.31
BAS	C4-Cz	43.31
BAS	M1-Cz	43.31
BAS	M2-Cz	43.31
BAS	O1-Cz	43.31
BAS	C3	43.11
BAS	C4	43.11
BAS	O1	43.11
BAS	O2	43.11
EMG	Chin	31.40
EMG	Chin2	28.35
EMG	Chin EMG	27.09
EMG	Chin_Ctr-Cz	26.01
EMG	Chin_L-Cz	26.01
EMG	Chin_R-Cz	26.01
EMG	Feet-R	19.18
EMG	Pleth	17.60
EMG	Arms-1	17.52
EMG	Arms-2	17.52
EKG	EKG_L-Cz	43.31
EKG	EKG_R-Cz	43.31
EKG	ECG	17.78
EKG	EKG-L	12.01
EKG	EKG-R	12.01
EKG	ECG II	11.87
EKG	EKG	9.07
EKG	ECG 2	1.48
EKG	EKG2	0.02
EKG	ECG1-EC	0.00

Supplementary Table 16. Channels and prevalence percentages by modality for Stanford