Progress in Cardiovascular Diseases xxx (xxxx) xxx

Contents lists available at ScienceDirect

Progress in Cardiovascular Diseases

journal homepage: www.onlinepcd.com

Research Paper

Wearable-derived REM sleep as a modifiable risk factor for cardiovascular disease: A multimodal prediction model in the All of Us Research Program

Jiseung Kang [a,b,k], Hyeon Jin Kim [c,d,k], Yesol Yim [c,d,k], Min Seo Kim [e,f], Jaeyu Park [c,d], Yejun Son [c,d], Guillaume Fond [g], Laurent Boyer [g], Dong Keon Yon [c,d,h,*], Christa J. Nehs [i,j,**]

[a] School of Health and Environmental Science, Korea University College of Health Science, Seoul, South Korea
[b] Department of Health and Safety Convergence Science, Korea University Graduate School, Seoul, South Korea
[c] Center for Digital Health, Medical Science Research Institute, Kyung Hee University College of Medicine, Seoul, South Korea
[d] Department of Precision Medicine, Kyung Hee University College of Medicine, Seoul, South Korea
[e] Cardiovascular Disease Initiative, Broad Institute of MIT and Harvard, Cambridge, MA, USA
[f] Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA
[g] Research Centre on Health Services and Quality of Life, Assistance Publique-Hopitaux de Marseille, Aix Marseille University, Marseille, France
[h] Department of Pediatrics, Kyung Hee University Medical Center, Kyung Hee University College of Medicine, Seoul, South Korea
[i] Division of Sleep Medicine, Harvard Medical School, Boston, MA, USA
[j] Mass General Brigham Department of Anesthesiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

ARTICLE INFO

Keywords:
All of Us Research Program
Cardiovascular diseases
Machine learning
Multimodal data
Prediction models
REM sleep
United States

ABSTRACT

Sleep is increasingly recognized as a modifiable factor for cardiovascular diseases (CVDs), yet the roles of specific sleep stages, particularly rapid eye movement (REM) sleep, and the utility of wearable-derived sleep parameters for CVD prediction remain unknown. Therefore, we aimed to investigate the associations of sleep with CVD risk and to develop multimodal prediction models for incident CVD. Using data from the All of Us Research Program, 23,413 adults who consented to share both Fitbit and electronic health record data between May 2018 and October 2023 were included in the analysis. Sleep data were obtained from Fitbit wearable devices, including sleep duration and stage-specific information such as REM sleep. Incident CVD was defined as occurring at least six months after the initiation of Fitbit monitoring. Cox proportional hazards models were applied to estimate adjusted hazard ratios (aHR) with 95% confidence intervals (CIs) for sleep duration and REM sleep in association with CVD. For the prediction of CVD within six years, two modeling strategies were implemented, including a non-invasive multimodal model using demographic, self-reported, and wearable-derived data and an invasive multimodal model that additionally incorporated laboratory measures. To compare real-world model performance, we applied and compared several conventional risk prediction models, including SCORE2, the Framingham Risk Score, AutoPrognosis 2.0, and the Pooled Cohort Equations. Among 23,413 participants (mean [standard deviation] age, 56.7 [15.5] years; 70.9% female), both short (<7 h) and long (≥9 h) sleep durations were associated with higher risk of CVDs compared with 7 to 9 h of sleep (short sleep: aHR, 1.16 [95% CI, 1.07 to 1.27]; long sleep: aHR, 1.32 [95% CI, 1.10 to 1.58]), with U-shaped relationship. Individuals with less than 20% REM sleep had a greater risk of CVDs (aHR, 1.31 [95% CI, 1.18 to 1.45]) compared with those with 20% to 25% REM sleep. These associations were consistent across CVD subtypes. In prediction analyses, soft-voting ensemble models with LightGBM and XGBoost integrating wearable, clinical, and laboratory data achieved high performance (AUROC of non-invasive models, 0.748; AUROC of invasive models, 0.782), outperforming conventional

Abbreviations: AI, artificial intelligence; AHA, American Heart Association; AoURP, All of Us Research Program; AUROC, area under the receiver operating characteristic curve; BMI, body mass index; CI, confidence interval; CVD, cardiovascular disease; DBP, diastolic blood pressure; EHR, electronic health record; GBM, gradient boosting machine; HPA, hypothalamic-pituitary-adrenal; HR, hazard ratio; ICD-10-CM, international classification of diseases, 10th revision, clinical modification; LDL, low-density lipoprotein; LightGBM, light gradient boosting machine; ML, machine learning; MR, mendelian randomization; OSA, obstructive sleep apnea; PS, propensity score; REM, rapid eye movement; RF, random forest; ROC, receiver operating characteristic; SBP, systolic blood pressure; SD, standard deviation; SHAP, shapley additive explanation; TC, total cholesterol; XGBoost, extreme gradient boosting.

* Correspondence to: D.K. Yon, Center for Digital Health, Medical Science Research Institute, Kyung Hee University College of Medicine, Seoul, South Korea.
** Correspondence to: C.J. Nehs, Division of Sleep Medicine, Harvard Medical School, Boston, MA, USA.
E-mail addresses: yonkkang@gmail.com (D.K. Yon), CNEHS@mgh.harvard.edu (C.J. Nehs).
[k] These authors contributed equally to this work as the first authors.

https://doi.org/10.1016/j.pcad.2026.03.002

Available online 4 March 2026
0033-0620/© 2026 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

risk scores (SCORE2, the Framingham Risk Score, AutoPrognosis 2.0, and the Pooled Cohort Equations: AUROC range, 0.649 to 0.685). Short and long sleep durations, as well as reduced REM sleep, were associated with increased risk of CVD. We also derived and validated a multimodal model to predict the incidence of CVD within six years. These findings support the potential value of integrating objective sleep-stage measures and multimodal digital health data into future cardiovascular prevention strategies and guideline development.

Introduction

Sleep is a recognized modifiable risk factor for all-cause mortality and chronic diseases, including cardiovascular diseases (CVDs).[1] The American Heart Association (AHA) recently added healthy sleep to its cardiovascular health checklist, Life's Essential 8, emphasizing the growing evidence that sleep behaviors influence cardiovascular health and related risk factors.[2] Previous studies have shown that both short (<7 h) and long (>9 h) sleep durations are significantly associated with cardiometabolic diseases, including obesity, hypertension, coronary heart disease, type 2 diabetes, and stroke.[3]

Recently, research interest has shifted from sleep duration to sleep architecture and patterns. While previous studies have predominantly focused on total sleep time or overall sleep quality, much less is known about how specific sleep stages contribute to cardiovascular disease risk.[1,4,5] Rapid eye movement (REM) sleep, which is characterized by vivid dreaming and heightened autonomic activity, is of particular interest but remains uninvestigated in this context. In a multicenter, population-based study, each 5% decrement in REM sleep was associated with a 13% increase in all-cause mortality, with parallel increases observed in cardiovascular mortality among older adults.[6] These findings implied that insufficient REM sleep could be detrimental to long-term cardiovascular health. However, REM sleep has not yet been firmly suggested as a modifiable risk factor for incident CVD, in part due to limitations in measuring sleep stages at scale.

Most existing observational studies rely on self-reported sleep, which is prone to recall bias and misreporting, especially in populations with or at risk for insomnia and lower concentrations due to sleep deprivation.[4,7,8] Some studies have used actigraphy-based sleep measures, but these studies only involved short-term data collection periods from 7 to 14 days, which cannot discern REM sleep or capture long-term sleep patterns.[1,5,9] Recent advances in commercial wearable devices, including wrist-worn trackers, now allow long-term, objective, and continuous monitoring of sleep stages with validated accuracy in real-world settings. These devices enable large-scale investigations of sleep architecture in relation to clinical outcomes.

The National Institutes of Health All of Us Research Program (AoURP) provides an unprecedented multimodal data resource, integrating electronic health records (EHRs), laboratory results, and survey information from a diverse cohort, alongside long-term Fitbit wearable data on physical activity and sleep patterns.[10] In this study, we utilized the longitudinal multimodal data, including EHR and commercial wearable devices records, and regular health checkup data in AoURP to investigate the associations between sleep patterns, particularly REM sleep, and incident cardiovascular disease. Furthermore, we developed a multimodal prediction model that integrates wearable-derived REM sleep parameters with demographic, clinical, and lifestyle variables from multimodal datasets to predict incident CVD within six years. This study aims to clarify the role of REM sleep as a potentially modifiable risk factor for CVD and to demonstrate the utility of wearable-derived sleep data in multimodal cardiovascular risk prediction.

Methods

Overview

Using the AoURP dataset, which integrates diverse data sources, this study incorporated actigraphy-derived sleep measures to explore CVD risk using two complementary methodological frameworks. First, statistical analyses were conducted to estimate hazard ratios (HRs) that assess the role of sleep, particularly REM sleep, in the development of CVD. Second, multimodal models were developed to predict the incidence of CVD within six years using sleep-related features. To evaluate the incremental value of different data layers, two modeling strategies were constructed: a non-invasive model using demographic, self-reported, and wearable-derived data, and an invasive model that additionally incorporated laboratory results. Ensemble methods were applied within each modeling strategy to derive the final predictive model. The overall study flow was described in Fig. 1.

Data source and study population

The AoURP is a longitudinal cohort study that aims to advance precision medicine and improve population health through comprehensive data collection from one million or more diverse participants across the United States.[10] For this analysis, we used data from the AoURP Controlled Tier Dataset v8 (C2024Q3R8), which included participants aged 18 years and older who were enrolled between May 2018 and October 2023 across more than 340 clinics and medical centers.[10] The dataset included multiple data domains, such as EHRs, laboratory results, survey responses, and wearable data. Participants provided informed consent at enrollment, and those who owned a Fitbit device and agreed to share both Fitbit and EHR data were eligible for inclusion.[11] The protocol was reviewed by the Institutional Review Board of the AoURP (#2016–05-TN-Master). In accordance with regulations and guidance from the national institutes of health Office for Human Research Protections, research conducted through the AoURP maintained consistent standards for protecting participants' rights and welfare. All participants provided informed consent to share their electronic health records, surveys, and other research data with qualified researchers. This study used de-identified data from the AoURP (https://www.researchallofus.org/), which is publicly available to registered researchers.[12] Detailed descriptions of the AoURP cohort is provided in the **Supplementary Methods 1.1**. The Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis statement was used to guide both the design and reporting of this study.[13]

From the AoURP dataset, we identified 54,322 individuals with Fitbit sleep data (31,844,165 nights) and 393,596 individuals with EHR data. We applied sequential exclusion criteria for data quality. We included only sleep periods flagged as 'main sleep' (longest sleep per day) by the Fitbit algorithm, and we used only those with available sleep stage data for analysis.[1] We restricted analysis to participants with plausible sleep durations (>0 and ≤ 24 h/night) and excluded those with <4 h of sleep on ≥30% of nights to ensure consistent device wear and to reflect typical sleep patterns.[1] We further limited the cohort to participants with at least 6 months of continuous Fitbit monitoring. After linking Fitbit and EHR data, a final analytic cohort of 23,413 participants was included in the study (**Fig. S1**).

Fitbit sleep data

Sleep data were obtained from participants who owned a Fitbit device and voluntarily shared their information through the "Bring Your Own Device" program.[1] In this study, Fitbit-derived daily summary data, including sleep duration, time in bed, and sleep stage durations (light,

deep, and REM sleep), were used. The device identified sleep periods as episodes lasting at least 1 h, with wake intervals longer than 3 min treated as separate segments. For sleep periods exceeding 3 h, the device estimates sleep stages using a proprietary algorithm based on heart rate and movement data.[1] According to validation studies, the device categorizes 'light' sleep as N1 and N2 stages, 'deep' sleep as N3, and 'REM' as rapid eye movement sleep.

Individual sleep duration and sleep stage metrics were quantified as averages across each participant's monitoring period, with sleep stages expressed as the proportion of total sleep time spent in each stage rather than as absolute durations.[1,6]

Outcome

The primary outcome was the incidence of CVD occurring at least six months after initiation of Fitbit wearable device use. CVD outcomes were identified using EHR diagnosis records from diverse clinics and medical centers. Diagnosis codes were recorded in the International Classification of Diseases (ICD), 10th Revision, Clinical Modification (ICD-10-CM) format from the AoURP Curated Data Repository.[14] The ICD-based CVD definition was aligned with previous research and categorized as follows:[15] (1) cerebrovascular diseases including stroke and transient ischemic attack; (2) dysrhythmia including atrial flutter, atrial fibrillation, and ventricular arrhythmia; (3) ischemic heart diseases including acute coronary disease, myocardial infarction, and angina; (4) thrombotic disorders including arterial thromboembolism, pulmonary embolism, and deep vein thrombosis; and (5) other cardiac disorders including pericarditis, myocarditis, heart failure, cardiac arrest, and cardiogenic shock. Participants with CVD were defined as having experienced a cardiovascular outcome if diagnosed with any of these events. Detailed ICD codes for each disease category are provided in **Table S1**.

Study variables

To construct a harmonized analytic dataset, variables were extracted from multiple data domains available in the AoURP. Demographic characteristics, including age, sex at birth (male, female, and unknown), and race (Hispanic or Latino, White, Black or African American, other, and unknown), were obtained from core participant data. Survey-derived variables included education background (graduate or professional degree, college degree, and high school or below), smoking status (never and ever ≥100 cigarettes), alcohol consumption (yes and no), and family history of CVD. From the laboratory data domain, body mass index (BMI, kg/m^2), systolic blood pressure (SBP, mmHg), diastolic blood pressure (DBP, mmHg), total cholesterol (TC, mg/dL), low-density lipoprotein cholesterol (LDL, mg/dL), and high-density lipoprotein cholesterol (HDL, mg/dL) were obtained,[16] with values recorded nearest

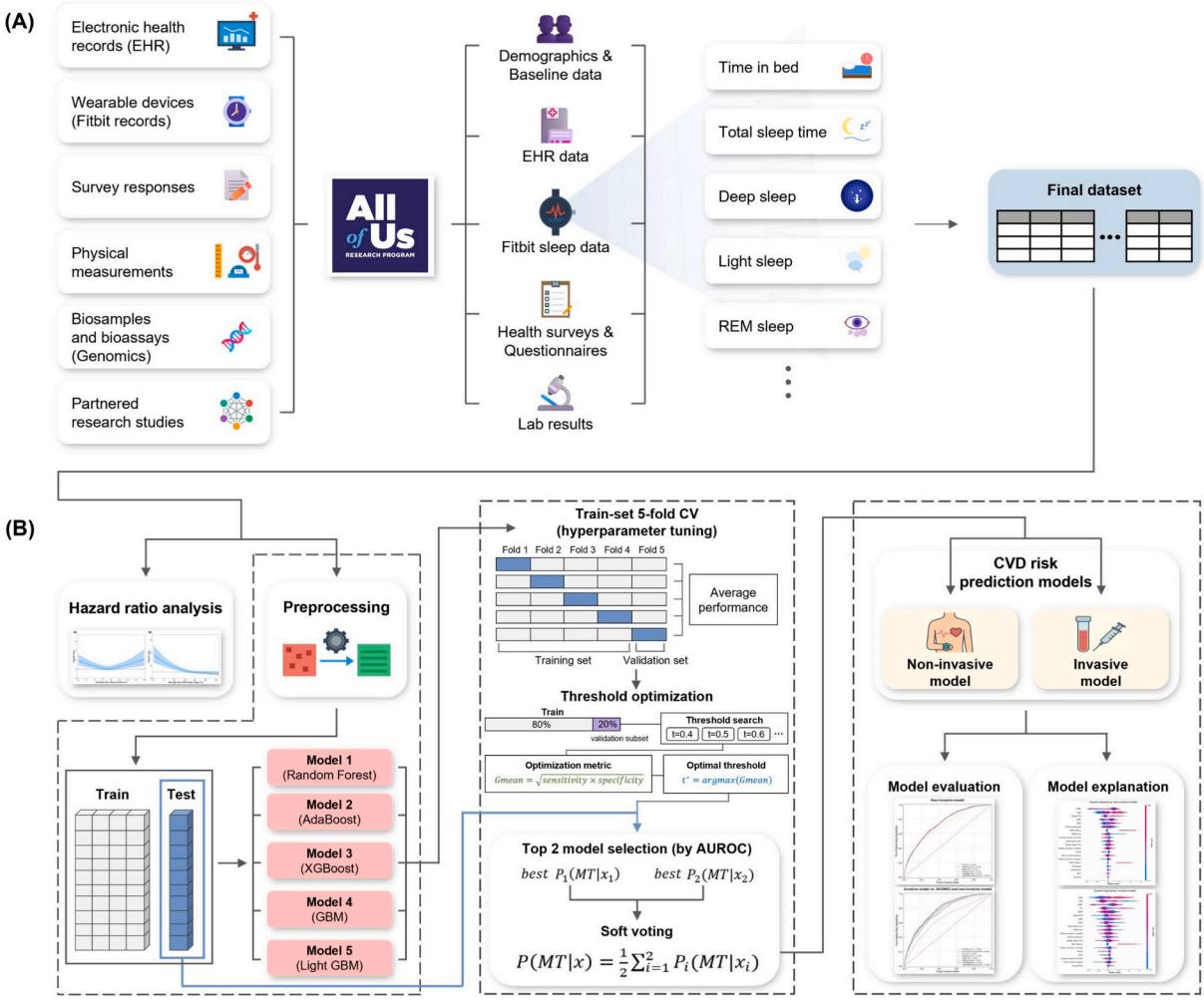


Fig. 1. Overview of this study. (A) Multiple data sources from the All of Us Research Program including EHRs, wearable Fitbit sleep data, surveys, physical measurements, and lab results were integrated to construct the final analytical dataset; (B) The final dataset was preprocessed and used to train five machine learning models, and a soft-voting ensemble combined their predictions to estimate cardiovascular risk in both non-invasive and invasive model settings.
Abbreviations: CVD, cardiovascular disease; EHR, electronic health record; REM, rapid eye movement.

to the initiation of Fitbit monitoring selected. In the EHR domain, disease-related information, including CVD and a history of diabetes and chronic kidney disease, was available. In addition, wearable-derived measures included daily sleep duration, time in bed, wake duration, and sleep stage metrics (light, deep, and REM sleep).

Statistical analysis

Cox proportional hazards models were used to examine associations between Fitbit-derived sleep duration and incident CVD, for which HRs and 95% confidence intervals (CIs) were estimated.[11] The follow-up began at the initiation of Fitbit monitoring and continued until the earliest occurrence of CVD diagnosis or October 1, 2023, with person-time accrued in days. Models were adjusted for seven baseline covariates, including age, sex at birth, race, BMI, education background, smoking status, and alcohol consumption. To examine potential threshold-based differences, we further classified average sleep duration into $<$7 h, 7–9 h, and $\geq$ 9 h per night,[17] and mean REM sleep proportion into $<$20%, 20–25%, and $>$ 25%[18] All statistical analyses were performed using SAS software (version 9.4; SAS Institute Inc., Cary, NC, USA) and R software (version 4.5.1; R Foundation for Statistical Computing, Vienna, Austria).

Sensitivity analyses

Two sensitivity analyses were conducted to assess the robustness of our findings. First, to evaluate whether the association between REM sleep percentage and CVD was sensitive to the choice of binary cutoff, we repeated the Cox regression analyses using alternative thresholds of 15%, 17%, and 18%, in addition to the primary 20% cutoff. These cutoffs were selected based on prior literature and to facilitate comparison with previous polysomnography-based studies.[6] Second, to address potential confounding by conditions and medications that may reduce REM sleep percentage,[19] we constructed an additional model that included previous diagnoses of insomnia and obstructive sleep apnea (OSA), as well as antidepressant use, in addition to the covariates in the primary model. Participants were considered to have a history of insomnia, OSA, or antidepressant use if they had two or more diagnoses or prescriptions recorded before wearable device initiation. Insomnia and OSA were identified using ICD-10-CM codes G47.0 and G47.33, respectively.[20,21] Antidepressant use was identified using the Anatomical Therapeutic Chemical classification system code N06A.

Data preprocessing and feature engineering for the model

The following processes were undertaken to support CVD prediction and improve model performance. Laboratory values that were mismapped or fell outside physiologically reasonable ranges were identified and excluded. In addition, multiple imputation using chained equations was employed to iteratively estimate and reconstruct missing data based on the relationships among variables.[16,22] Continuous features were standardized using the mean and standard deviation (SD). To capture physiologically and clinically relevant features not directly represented in the original variables, additional derived measures were generated. Pulse pressure was derived from SBP and DBP.[23] A sleep propensity score (PS) was estimated using logistic regression, with REM sleep proportion modeled using a 20% cutoff and seven baseline covariates, including age, sex at birth, race, BMI, educational background, smoking status, and alcohol consumption.[6] This PS was derived to account for baseline differences associated with REM sleep. The distribution of PSs for the REM sleep groups was presented in **Fig. S2**. All steps were performed separately for the training and testing sets using the same methodology to prevent data leakage.

Model development and training

To assess the contribution of different data domains, two modeling strategies were implemented. The non-invasive model included demographic, self-reported, and wearable-derived variables, whereas the invasive model further incorporated laboratory measures. The variables included in each model are presented in **Table S2**.

The dataset was split into training and testing sets at a 7:3 ratio, with model development performed exclusively on the training set using 5-fold cross-validation. Given the class imbalance between CVD and non-CVD cases, we employed class weighting to address disproportionate representation during model training.[24] Class weights were computed using the balanced weighting strategy, which assigns higher penalties to misclassifications of the minority class (CVD cases), inversely proportional to their frequency in the training data.[25] This approach improves model sensitivity to underrepresented CVD cases without altering the original data distribution.[24]

We employed five algorithms: random forest (RF), AdaBoost (AdaBoost), extreme gradient boosting (XGBoost), gradient boosting machine (GBM), and light gradient boosting machine (LightGBM). Hyperparameters were optimized using grid search with cross-validation,[26] where the training dataset was randomly shuffled while preserving the class distribution across all folds to ensure robust generalization.[27] Classification thresholds were optimized on a validation subset (20% of training data) to maximize the geometric mean of sensitivity and specificity, balancing case detection with positive predictive value.[26] Following individual model optimization, we selected the top two models with the highest area under the receiver operating characteristic curve (AUROC) for ensemble construction.[28] These models were combined via soft-voting, which aggregates probability predictions from individual classifiers.[29] Optimized hyperparameters are provided in **Table S3** and **Table S4**.

Model evaluation

The final ensemble model was evaluated on an independent test dataset that was reserved exclusively for performance assessment. The test dataset for evaluation was not involved in any stage of training or internal validation, ensuring an unbiased assessment of the model's generalizability.[16] Model performance was assessed using sensitivity, specificity, accuracy, balanced accuracy, and AUROC. Using ROC curve analysis, we compared the predictive performance of the non-invasive and invasive models with established cardiovascular risk prediction models, including SCORE2, the Framingham Risk Score, AutoPrognosis 2.0, and the Pooled Cohort Equations. Additional methodological details are provided in **Supplementary Methods 1.2**. To interpret model predictions and identify key predictors, we calculated Shapley Additive Explanation (SHAP) values for both non-invasive and invasive models, quantifying the contribution and directionality of each feature to CVD risk predictions by ranking feature importance.[30]

To enhance clinical interpretability, predicted probabilities were calibrated using Platt scaling as a post-hoc adjustment procedure.[31] Model calibration in the test set was assessed using two complementary metrics, mean calibration and calibration slope. Mean calibration reflected the difference between the average predicted probability and the observed outcome proportion. In contrast, the calibration slope evaluated the correspondence between predicted probabilities and observed outcomes across the range of model outputs.[32] A mean calibration value closer to 0 indicated minimal average bias, while a calibration slope closer to 1 indicated better proportional alignment between predicted and observed outcomes. In addition, decision curve analysis was conducted in the test set to quantify the potential clinical benefit of model use across a range of threshold probabilities and to compare performance against default strategies, including treating all individuals or treating none.[33] All predictive modeling procedures were conducted using Python (version 3.11.11; Python Software Foundation, Beaverton,

OR, USA) with the following libraries: NumPy (2.1.0), Pandas (2.2.3), Scikit-learn (1.6.1), and Matplotlib (3.10.1).

Results

Study population

We used multimodal medical big data from AoURP, including EHR, Fitbit data, demographics and baseline data, health surveys and questionnaires, and lab results (Fig. 1A). As shown in **Fig. S1**, we identified 54,322 individuals with Fitbit sleep data, including 31,844,165 night records. A total of 23,528 adult participants with linked EHR data, a Fitbit monitored a period of at least 6 months and had fewer than 30% of nights with <4 h sleep, were included for analysis. After removal of participants with prevalent CVD before Fitbit device wear, 23,413 eligible participants (age, 56.7 [SD, 15.5] years; 70.9% female) were included in the final analysis (Table 1). For sleep stages, the average percentage spent in REM, light, and deep sleep were 17.7% (SD, 3.6), 56.6% (5.1), and 13.2% (3.0), respectively. Of 23,413 adults included in the study, we identified 6170 adults with <7 h of sleep (age, 58.7 [SD, 14.2] years; 62.1% female), 16,192 adults with 7–9 h of sleep (age, 56.0 [SD, 15.9] years; 73.6% female), and 1051 adults with ≥9 h of sleep (age, 56.6 [SD, 16.4] years; 81.5% female).

Table 1
Baseline characteristics of study participants.

	Total (n = 23,413)	<7 sleep hours (n = 6170)	7–9 sleep hours (n = 16,192)	≥9 sleep hours (n = 1051)
Age, *years*, mean (SD)	56.7 (15.5)	58.7 (14.2)	56.0 (15.9)	56.6 (16.4)
Sex at birth, n (%)				
Male	6717 (28.7)	2310 (37.4)	4217 (26.0)	190 (18.1)
Female	16,602 (70.9)	3833 (62.1)	11,912 (73.6)	857 (81.5)
Unknown	94 (0.4)	27 (0.4)	63 (0.4)	4 (0.4)
Race, n (%)				
Hispanic or Latino	1884 (8.0)	554 (9.0)	1263 (7.8)	67 (6.4)
White	17,779 (75.9)	4138 (67.1)	12,763 (78.8)	878 (83.5)
Black or African American	1375 (5.9)	729 (11.8)	623 (3.8)	23 (2.2)
Other	847 (3.6)	289 (4.7)	538 (3.3)	20 (1.9)
Unknown	1528 (6.5)	460 (7.5)	1005 (6.2)	63 (6.0)
Education background, n (%)				
Graduate or professional degree	15,851 (67.7)	3840 (62.2)	11,332 (70.0)	679 (64.6)
College degree	5785 (24.7)	1751 (28.4)	3761 (23.2)	273 (26.0)
High school or below	1777 (7.6)	579 (9.4)	1099 (6.8)	99 (9.4)
BMI, kg/m^2, mean (SD)	28.9 (6.2)	30.5 (6.3)	28.4 (6.0)	28.2 (6.2)
Smoking status, n (%)				
Never	15,300 (65.3)	3786 (61.4)	10,833 (66.9)	681 (64.8)
Ever (≥100 cigarettes)	8113 (34.7)	2384 (38.6)	5359 (33.1)	370 (35.2)
Alcohol consumption, n (%)				
No	826 (3.5)	283 (4.6)	504 (3.1)	39 (3.7)
Yes	22,587 (96.5)	5887 (95.4)	15,688 (96.9)	1012 (96.3)
Average daily sleep stage, mean (SD)				
REM	17.7 (3.6)	16.9 (3.8)	18.0 (3.4)	17.7 (3.2)
Light	56.6 (5.1)	57.4 (5.6)	56.3 (4.9)	57.4 (4.5)
Deep	13.2 (3.0)	13.2 (3.3)	13.3 (3.0)	12.1 (2.7)

Abbreviations: BMI, body mass index; REM, rapid eye movement; SD, standard deviation.

Association between sleep and risk of CVDs

To evaluate the association between sleep and CVDs, we performed Cox proportional hazard analyses for CVDs (Fig. 1B). Both short and long sleep duration were associated with a higher risk of CVDs compared with 7 to 9 h of sleep. A U-shaped association was observed, with the lowest HR near 7 to 8 h of sleep (Fig. 2A). Participants sleeping fewer than 7 h had a higher risk of CVD (aHR, 1.16 [95% CI, 1.07 to 1.27]), as did those sleeping 9 h or more (aHR, 1.32 [95% CI, 1.10 to 1.58]), relative to the reference group (adult with 7–9 h of sleep) in Table 2. For REM sleep stages, a lower REM sleep percentage was notably associated with a greater risk of CVD (Fig. 2B). Individuals with less than 20% REM sleep had a greater risk of CVDs (aHR, 1.31 [95% CI, 1.18 to 1.45]) compared with those with 20% to 25% REM sleep. This association remained consistent when alternative binary cutoffs of 15%, 17%, and 18% were applied (aHR range, 1.30 to 1.44; **Table S5**) and after additional adjustment for insomnia, obstructive sleep apnea, and antidepressant use (aHR, 1.26 [95% CI, 1.14 to 1.40]; **Table S6**). Although REM sleep greater than 25% was not statistically associated with CVD risk, the point estimates suggested a potential decreasing trend. In analyses of additional sleep characteristics, a higher proportion of deep sleep was associated with a lower risk of CVD, whereas increasing light sleep percentage showed a steep, monotonic increase in risk of CVD (**Fig. S2**). Higher sleep efficiency was modestly associated with a lower risk of CVD (**Fig. S2**).

In subgroup analyses of CVD subtypes, a similar U-shaped association was observed, with the lowest HR near 7 to 8 h of sleep across CVD subtypes (**Fig. S3**). Short sleep duration (<7 h) was associated with higher risks of ischemic heart disease (aHR, 1.33 [95% CI, 1.14 to 1.55]) and other cardiac disorders (aHR, 1.64 [95% CI, 1.19 to 2.28]) compared with 7 to 9 h of sleep (**Table S3**). Long sleep duration (≥9 h) was also associated with increased risk of dysrhythmia (aHR, 1.35 [95% CI, 1.07 to 1.70]; **Table S3**). For REM sleep stages, REM sleep <20% was similarly associated with higher risks of dysrhythmia (aHR, 1.36 [95% CI, 1.19 to 1.55]) and other cardiac disorders (aHR, 1.82 [95% CI, 1.12 to 2.95]) compared with 20% to 25% REM sleep (**Table S4**). Additional stratified analyses by demographic characteristics are shown in **Fig. S4** and **Tables S7-S10**.

Five-fold cross-validation results

To develop prediction models for CVDs within six years, we trained five algorithms on the preprocessed dataset and evaluated their performance using five-fold cross-validation, with model predictions combined through a soft-voting ensemble (Fig. 1B). Generally, two types of prediction models were constructed: a non-invasive model and an invasive model, distinguished by whether laboratory results obtained through invasive procedures (eg, blood samples) were included. Table 3 presents the five-fold cross-validation of various models using sensitivity, specificity, accuracy, balanced accuracy, and AUROC as evaluation metrics.

Among the non-invasive models, we selected the ensemble model with the best prediction performance (**Fig. S5**). Regarding the balanced accuracy and AUROC, the XGBoost and LightGBM models provided the best performance when only a single model was considered. To further enhance the prediction performance, we explored the ensemble approach, which combines both stacking and voting, utilizing a blend of LightGBM and XGBoost. The soft-voting ensemble model with LightGBM and XGBoost showed the highest performance in the test set, with an AUROC of 0.748, sensitivity of 73.6%, specificity of 59.7%, accuracy of 61.2%, and balanced accuracy of 66.6% (Table 3).

For invasive multimodal models, which incorporated laboratory results, the ensemble model also yielded the highest performance

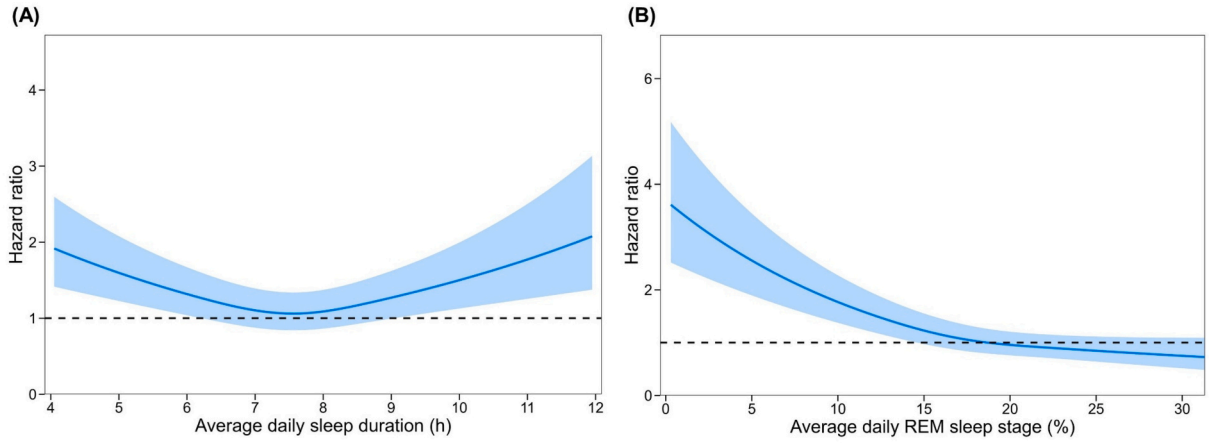


Fig. 2. Associations between sleep characteristics and cardiovascular disease risk. (A) Association between average daily sleep duration and CVD; (B) Association between average daily REM sleep stage percentage and CVD.
Abbreviations: BMI, body mass index; CI, confidence interval; CVD, cardiovascular disease; HR, hazard ratio; REM, rapid eye movement. Hazards models adjusted for age, sex at birth (male, female, and unknown), race (Hispanic or Latino, White, Black or African American, other, and unknown), education background (graduate or professional degree, college degree, and high school or below), BMI (kg/m^2), smoking status (never and ever $\geq$*100 cigarettes*), and alcohol consumption (yes and no). The dashed horizontal line represents HR = 1.0 (reference).

Table 2
Cox proportional hazards models for the association between Fitbit-derived sleep duration, REM sleep stage percentage, and CVD.

Sleep category	<7 sleep hours	7–9 sleep hours	≥9 sleep hours
Event (%)	793 (12.7)	1676 (10.4)	122 (11.9)
Adjusted HR (95% CI)*	**1.16 (1.07–1.27)**	1 (reference)	**1.32 (1.10–1.58)**
REM sleep stage category	<20% REM sleep	20–25% REM sleep	>25% REM sleep
Event (%)	2098 (12.3)	481 (7.8)	12 (5.8)
Adjusted HR (95% CI)*	**1.31 (1.18–1.45)**	1 (reference)	0.78 (0.44–1.39)

Abbreviations: BMI, body mass index; CI, confidence interval; CVD, cardiovascular disease; HR, hazard ratio; REM, rapid eye movement.
Bold indicates statistically significant associations ($p < 0.05$).
* Models were adjusted for age, sex at birth (male, female, and unknown), race (Hispanic or Latino, White, Black or African American, other, and unknown), education background (graduate or professional degree, college degree, and high school or below), BMI (kg/m^2), smoking status (never and ever $\geq$*100 cigarettes*), and alcohol consumption (yes and no).

(**Fig. S5**). Regarding balanced accuracy and AUROC, the XGBoost and LightGBM models provided the highest performance when only a single model was considered. Therefore, we made a soft-voting ensemble model with LightGBM and XGBoost. As a result, the ensemble model had the best performance, with AUROC of 0.782, sensitivity of 72.5%, specificity of 67.8%, accuracy of 68.4%, and balanced accuracy of 70.2%. Across models, invasive approaches consistently outperformed non-invasive models, and both exceeded the performance than the established reference models, including SCORE2, the Framingham Risk Score, AutoPrognosis 2.0, and the Pooled Cohort Equations (AUROC range, 0.649 to 0.685; Fig. 3).

SHAP value analysis

To analyze the effect of features in predicting incident CVD, we calculated the SHAP value for the soft-voting of the non-invasive model (ensemble model with LightGBM and XGBoost) and invasive model (ensemble model with LightGBM and XGBoost), presented in **Fig. S6**. SHAP value analysis indicated that systolic and diastolic blood pressure, age, sleep performance score, BMI, and pulse pressure were among the most influential predictors in the non-invasive model, whereas age, LDL cholesterol, diastolic and systolic blood pressure, total cholesterol, and BMI were the strongest contributors in the invasive model (**Fig. S6**). Across both models, REM sleep percentage, light sleep percentage, and mean sleep duration also contributed to prediction, although with comparatively lower importance than established cardiometabolic risk factors.

Calibration plots

The calibration plots show that both the non-invasive and invasive models showed improved agreement between predicted and observed risks after calibration, with calibration slopes near 1.0 and mean calibration errors close to zero (**Fig. S7** and **Fig. S8**). In decision curve analysis, the calibrated models provided higher net benefit than the uncalibrated models across clinically relevant threshold probabilities.

Discussion

Key findings

In this large, multimodal AoURP cohort integrating wearable sleep data with EHRs, we found several important associations between sleep characteristics and risk of CVD. First, both short and long sleep duration were associated with a higher risk of CVD, showing a clear U-shaped relationship. Second, reduced REM sleep, particularly <20% of total sleep, was consistently linked to increased risk of CVD. Third, these associations were observed not only for overall CVD risk but also for specific CVD subtypes, including dysrhythmia, ischemic heart disease, and other cardiac disorders. Fourth, this study presented two types of ML-based approaches for predicting the incidence of CVD, using data from multimodal medical data. In the invasive model, the ensemble model, which uses a soft voting approach with LightGBM and XGBoost, found superior predictive performance compared with single models, achieving an AUROC of 0.748. Invasive models incorporating laboratory results outperformed non-invasive models, with an AUROC of 0.782. Across conventional risk models, invasive approaches consistently outperformed non-invasive models, and both exceeded the performance than the established reference models, including SCORE2, the Framingham Risk Score, AutoPrognosis 2.0, and the Pooled Cohort Equations. Together, these findings suggest that clinically significant relationships between sleep and the onset of CVDs, and detailed sleep metrics from consumer wearable devices, may offer clinically

Table 3
Performance of non-invasive and invasive machine learning models for CVD prediction.

	Sensitivity	Specificity	Accuracy	Balanced accuracy	AUROC
Non-invasive model					
Train set, mean (SD)					
RF	71.9% (1.3)	64.3% (3.7)	65.2% (3.3)	68.1% (2.1)	0.742 (0.017)
AdaBoost	71.7% (1.6)	63.7% (3.0)	64.6% (2.8)	67.7% (2.0)	0.735 (0.023)
GBM	70.3% (0.1)	60.6% (5.2)	61.7% (4.6)	65.5% (2.6)	0.717 (0.022)
XGBoost	71.1% (1.1)	65.2% (4.3)	65.8% (3.9)	68.1% (2.4)	0.743 (0.019)
LightGBM	72.2% (1.3)	64.7% (4.5)	65.5% (3.9)	68.5% (2.0)	0.743 (0.019)
LightGBM & XGBoost*	**70.9% (0.9)**	**65.8% (4.2)**	**66.4% (3.8)**	**68.4% (2.4)**	**0.745 (0.020)**
Test set					
RF	75.9%	57.3%	59.4%	66.6%	0.741
AdaBoost	80.1%	53.5%	56.4%	66.8%	0.744
GBM	71.6%	56.9%	58.5%	64.2%	0.719
XGBoost	73.1%	60.4%	61.8%	66.7%	0.746
LightGBM	76.6%	57.8%	59.9%	67.2%	0.747
LightGBM & XGBoost*	**73.6%**	**59.7%**	**61.2%**	**66.6%**	**0.748**
Invasive model					
Train set, mean (SD)					
RF	74.0% (2.3)	68.1% (2.3)	68.8% (1.9)	71.1% (0.7)	0.781 (0.008)
AdaBoost	73.9% (2.6)	67.2% (3.8)	67.9% (3.1)	70.5% (0.6)	0.772 (0.010)
GBM	73.1% (2.8)	66.7% (4.8)	67.4% (4.0)	69.9% (1.4)	0.756 (0.015)
XGBoost	**71.6% (1.5)**	**71.0% (2.2)**	**71.0% (1.9)**	**71.3% (0.7)**	**0.784 (0.010)**
LightGBM	73.8% (3.9)	67.4% (4.2)	68.1% (3.4)	70.6% (0.7)	0.773 (0.013)
LightGBM & XGBoost*	74.8% (3.1)	67.0% (3.0)	67.9% (2.3)	70.9% (0.6)	0.782 (0.011)
Test set					
RF	77.8%	63.2%	64.8%	70.5%	0.773
AdaBoost	64.6%	72.5%	71.6%	68.5%	0.765
GBM	66.5%	69.8%	69.5%	68.2%	0.750
XGBoost	68.7%	69.6%	69.5%	69.2%	0.777
LightGBM	79.1%	63.0%	64.7%	71.0%	0.781
LightGBM & XGBoost*	**72.5%**	**67.8%**	**68.4%**	**70.2%**	**0.782**

Abbreviations: AdaBoost, adaptiveboosting; AUROC, area under the receiver operating characteristic curve; CVD, cardiovascular disease; GBM, gradient boosting machine; RF, random forest; SD, standard deviation; XGBoost, extreme gradient boosting.
Bold indicates the best performance.
* Ensemble model applied a soft voting method.

meaningful information for assessing future risk of CVD.

Comparisons with previous studies

In previous studies on CVD risk prediction, most artificial intelligence (AI) models have been developed using structured EHRs as the predominant data modality, and they typically applied classical ML algorithms.[34–36] In contrast, our study utilized a multimodal medical big-data framework, integrating EHRs, wearable device sleep data, baseline demographics, health surveys and questionnaires, and laboratory results from AoURP, and thereby extended previous work on sleep and risk of CVDs. Our findings of a U-shaped association between sleep duration and cardiovascular events aligned with earlier investigations based on self-reported sleep and risk of coronary heart disease and myocardial infarction, which reported the lowest risk among intermediate sleep durations.[7] Despite such findings, most of these studies have inherent limitations from their study design and cohort size, which limited their ability to fully address the long-term associations between sleep and risk of CVDs.[1,5,9] First, most previous studies relied on self-reported sleep over short periods (e.g., 7 days) using wearable devices provided to participants.[1,5,9] Second, the lack of detailed sleep-stage data, categorized by light, deep, and REM sleep, resulted in their limited ability to evaluate long-term associations and the differential impacts of specific sleep stages. In contrast, our use of continuous wearable-derived sleep metrics combined with longitudinal clinical data increases the granularity and scope of risk stratification. Third, most previous studies, including those using the AoURP, have primarily evaluated the association between sleep characteristics and risk of CVD.[1,37] However, AI-based prediction models for future CVD integrating wearable-derived sleep metrics with longitudinal EHR data have not been conducted. Our study addresses this gap by developing multimodal prediction models for incident CVD. Moreover, in the domain of CVD risk prediction, our multimodal prediction models build on earlier work by demonstrating improved discrimination over conventional scores (SCORE2, the Framingham Risk Score, AutoPrognosis 2.0, and the Pooled Cohort Equations) when rich digital and longitudinal data are incorporated.[34,38–40] These advances suggest the potential of combining digital-health and AI approaches to enhance cardiovascular risk stratification.

Plausible underlying mechanism

Consistent with our findings, previous studies also showed a U-shaped relationship between sleep duration and risk of cardiovascular events, with approximately 7 to 8 h of sleep associated with the lowest risk.[37,41,42] Both short and prolonged sleep duration have been linked to adverse cardiometabolic profiles through distinct biological pathways. Chronic sleep deprivation triggers overactivation in the sympathetic nervous system and stimulation in the hypothalamic-pituitary-adrenal (HPA) axis, leading to increased nocturnal catecholamines, cortisol, and proinflammatory cytokines, such as interleukin-1β, interleukin-6, and C-reactive protein.[41,43] These changes impaired endothelial function, promoted hypertension and insulin resistance, and accelerated atherosclerosis over time.[41,43] In contrast, abnormally prolonged sleep duration may be associated with underdiagnosed diseases or comorbidities, as a marker. Those underlying health issues or inefficient, fragmented sleep architecture.[44] For example, comorbidities such as depression or OSA can induce prolonged and/or fragmented sleep, which in turn can heighten systemic inflammation and risk of CVDs. In addition, there is evidence that those with a subjective prolonged sleep pattern may have poor sleep efficiency, which is also linked with adverse health outcomes.[45] Consistent with this interpretation, recent Mendelian randomization (MR) studies have shown that genetically predicted short sleep is associated with an increased risk of coronary artery disease, myocardial infarction, and heart failure, whereas no such causal association was observed for long sleep duration.[46,47] Taken together, these findings reinforce the notion that short sleep duration may be causally related to CVD and represents a modifiable intervention target, whereas long sleep duration may serve as a clinical indicator warranting evaluation for underlying conditions that predispose to both prolonged sleep and adverse cardiovascular outcomes.

Growing evidence indicates that decreased REM sleep is associated with increased objective sleep monitoring; a lower portion of REM sleep was associated with significantly higher incidence of all-cause and cardiovascular mortality, with roughly a 13% increase in mortality rates for every 5% decrement in REM sleep.[6] One potential mechanism is that reduced REM sleep reflects greater sleep fragmentation and chronic activation of stress pathways.[48] Individuals with abnormally low REM sleep showed signs of hyperarousal, including elevated nocturnal cortisol levels and impaired insulin sensitivity, and dysregulated HPA axis.[49] This neuroendocrine imbalance can contribute to hypertension,

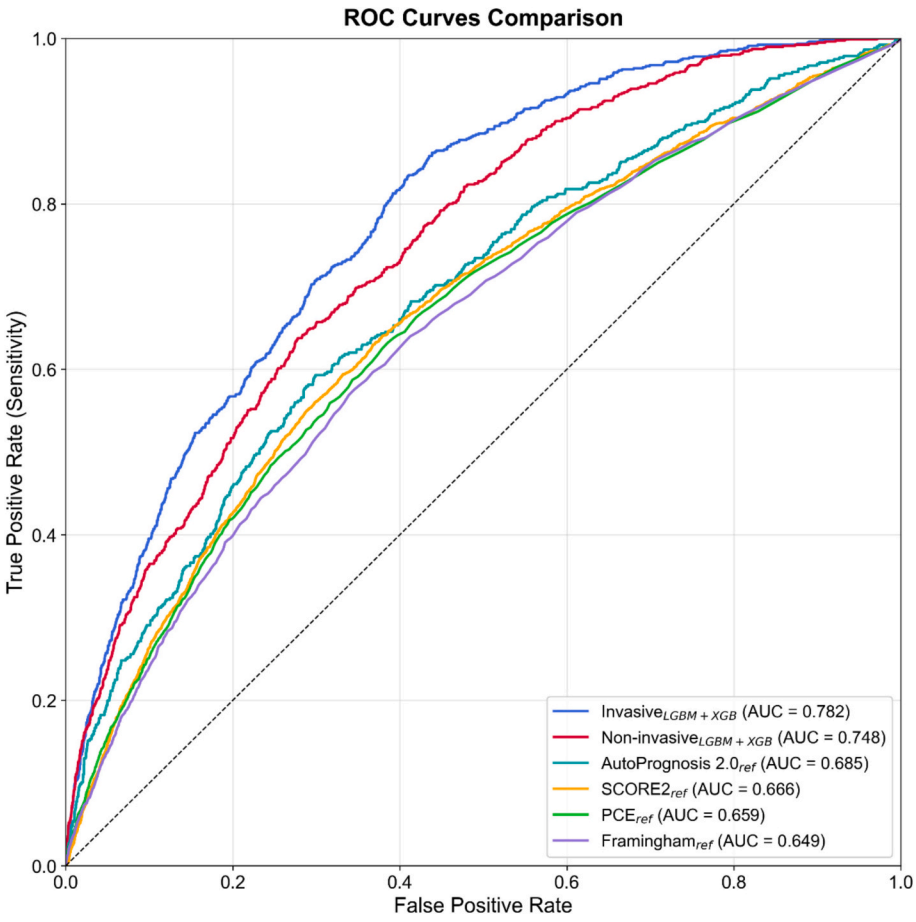


Fig. 3. ROC curves comparing non-invasive and invasive models with reference risk prediction models.
Abbreviations: AUC, area under the curve; LGBM, light gradient boosting machine; PCE, pooled cohort equations; ROC, receiver-operating characteristic curve; XGB, extreme gradient boosting. LGBM+XGB refers to an ensemble model generated by combining LightGBM and XGBoost using soft voting.

endothelial dysfunction, and arrhythmogenesis. Indeed, deficits in REM sleep have been linked to downstream cardiovascular events. For instance, shorter REM duration was associated with an increased risk of new-onset atrial fibrillation in older adults.[1]

Sleep-related risk factors extend beyond aggregate CVD to influence specific cardiovascular conditions, including arrhythmias, ischemic heart disease, and heart failure. Short sleep or insomnia can accelerate coronary artery disease via sustained blood pressure elevation, adverse lipid profiles, and inflammation.[50] For example, genetic predisposition to insomnia has been associated with significantly higher odds of myocardial infarction and congestive heart failure in MR analyses.[50] Moreover, sleep deficiency has been linked to prothrombotic and inflammatory states that can precipitate acute ischemic events. Such as stroke, acute coronary syndromes. Notably, recent analyses integrating wearable sleep metrics found that individuals with lower amounts of restorative deep and REM sleep had higher odds of developing atrial fibrillation over time.[1] Together, these findings underscore that poor sleep health, whether in the form of insufficient sleep or insomnia, can target multiple organ systems and specific cardiac substrates, thereby elevating the risk of dysrhythmias, coronary ischemia, and other adverse cardiac outcomes.

Integrating sleep metrics with other health data through AI approaches is emerging as a powerful strategy to improve CVD prediction. Advanced algorithms can synthesize multimodal inputs, including EHR-based clinical variables, wearable-derived sleep characteristics, and laboratory biomarkers, to detect complex patterns and early risk signals not captured by traditional risk scores. In preliminary studies, such ML models have outperformed conventional cardiovascular risk algorithms in predicting incident events.[51] Likewise, longitudinal analyses from AoURP that integrate high-resolution wearable sleep data with clinical variables enhance the prediction of future CVD. Together, these findings highlight the potential of multimodal models to improve cardiovascular risk stratification by capturing granular sleep behaviors alongside established clinical risk factors.

Clinical policy implications

Our findings have important implications for clinical practice and CVD prevention policy. Current guidelines from the AHA Life's Essential 8 recently elevated sleep to a core modifiable cardiovascular health metric, recommending 7 to 9 h of sleep per night for optimal cardiovascular health.[2] However, existing guidelines primarily focus on sleep duration and provide limited integration of sleep architecture or objective wearable-derived measurements. The observed U-shaped association between sleep duration and risk of CVDs, together with the elevated risk associated with reduced REM sleep, suggests that cardiovascular risk assessment could be strengthened by incorporating multidimensional sleep indicators rather than duration alone. Our findings support the potential value of embedding sleep-stage metrics, particularly REM sleep proportion, into preventive cardiology frameworks. Moreover, the superior performance of multimodal prediction models over conventional risk scores highlights opportunities to expand clinical guidelines toward precision risk stratification using integrated clinical and digital phenotypes.[51] Also, in settings where blood-based testing is not feasible or where individuals are reluctant to undergo blood collection, noninvasive wearable-derived models may provide a

practical alternative for early cardiovascular risk screening. Such approaches could enable population-level risk stratification and support preventive strategies in resource-limited or community-based care settings. Collectively, these findings align with and extend current policy directions, indicating that future cardiovascular health guidelines may benefit from formal incorporation of objective sleep-stage metrics, long-term wearable monitoring, and multimodal data-driven approaches to enhance early detection and prevention of CVDs.

Limitations

To the best of our knowledge, this is the first study to utilize linked wearable device data and EHR data to investigate the association between sleep metrics and incident CVD, and to develop multimodal prediction models incorporating actigraphy-derived sleep measures. However, several limitations should be acknowledged. First, although this study was conducted using a cohort design, the influence of potential confounding factors cannot be fully excluded in assessing the causal association between sleep variables and the development of CVD, and therefore, the findings should be interpreted with caution. Second, the study population was restricted to participants who owned a Fitbit device, and the cohort was predominantly White, female at birth, and relatively highly educated. As such, the findings may not be fully representative of the general population. However, extensive adjustment for relevant variables was undertaken to minimize potential bias related to this demographic imbalance. Third, sleep metrics used in the study were indirectly estimated using algorithms embedded in Fitbit wearable devices and may have limited accuracy compared to gold-standard sleep assessments such as polysomnography. Nonetheless, Fitbit-based sleep estimation algorithms have been validated in multiple prior studies,[52,53] and their ability to capture long-term, large-scale sleep data in real-world settings supports their practical relevance. Fourth, prior studies suggest that, in addition to wearable-derived sleep parameters included in our prediction models, including light sleep (%), deep sleep (%), REM sleep (%), total sleep duration, time in bed, and wake duration, electrophysiologic sleep markers (e.g., delta wave activity in deep sleep) may also be associated with risk of CVD.[54] However, such markers were not available in our dataset and therefore could not be evaluated. Fifth, the reference models used for comparison were developed from distinct populations and rely on differing variable sets and modeling assumptions. Sixth, for some laboratory variables, missing data required imputation, which may have introduced a degree of uncertainty, as the imputed values may not fully reflect actual measurements, despite the use of established methods. Lastly, external validation using an independent cohort was not performed, and therefore, the generalizability of the developed models to other populations or settings should be interpreted with caution.

Conclusion

In this large multimodal cohort, both short and long sleep duration and lower REM sleep were associated with higher risks of CVD, with consistent patterns across CVD subtypes. ML models incorporating clinical, laboratory, and wearable-derived sleep data demonstrated improved prediction of incident CVD within six years, particularly for invasive models. Against conventional risk models, our models consistently outperformed SCORE2, the Framingham Risk Score, AutoPrognosis 2.0, and the Pooled Cohort Equations. These findings highlight the potential utility of integrating multidimensional sleep metrics and multimodal data into cardiovascular risk assessment. Further validation in diverse populations and real-world clinical settings is needed, but the results support future consideration of objective sleep measures in CVD prevention and guideline development.

CRediT authorship contribution statement

Jiseung Kang: Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Hyeon Jin Kim:** Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Yesol Yim:** Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Min Seo Kim:** Writing – review & editing. **Jaeyu Park:** Writing – review & editing. **Yejun Son:** Writing – review & editing. **Guillaume Fond:** Writing – review & editing. **Laurent Boyer:** Writing – review & editing. **Dong Keon Yon:** Writing – original draft, Supervision, Software, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Christa J. Nehs:** Writing – review & editing, Writing – original draft, Supervision, Software, Methodology, Conceptualization.

Funding

This research was supported by the Ministry of Science and ICT (RS-2024-00509257 and IITP-2024-RS-2024-00438239) and the Ministry of Health & Welfare (RS-2025-02220492), South Korea. This work was supported by the United States of America National Institutes of Health National Institute on Aging (R01AG076704). The funders had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Declaration of competing interest

The authors declare that they have no financial interests related to the material in the manuscript. This research was conducted independently of any business or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pcad.2026.03.002.

Data availability

Dr. JK had full access to the All of Us research program dataset in the study and took responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. Zheng NS, Annis J, Master H, et al. Sleep patterns and risk of chronic disease as measured by long-term monitoring with commercial wearable devices in the All of Us Research Program. *Nat Med*. 2024;30:2648–2656.
2. Lloyd-Jones DM, Allen NB, Anderson CAM, et al. Life's essential 8: updating and enhancing the American Heart Association's construct of cardiovascular health: a presidential advisory from the American Heart Association. *Circulation*. 2022;146:e18–e43.
3. St-Onge MP, Grandner MA, Brown D, et al. Sleep duration and quality: impact on lifestyle behaviors and cardiometabolic health: a scientific statement from the American Heart Association. *Circulation*. 2016;134:e367–e386.
4. Diao T, Liu K, Lyu J, et al. Changes in sleep patterns, genetic susceptibility, and incident cardiovascular disease in China. *JAMA Netw Open*. 2024;7, e247974.
5. Windred DP, Burns AC, Rutter MK, et al. Light exposure at night and cardiovascular disease incidence. *JAMA Netw Open*. 2025;8, e2539031.
6. Leary EB, Watson KT, Ancoli-Israel S, et al. Association of rapid eye movement sleep with mortality in middle-aged and older adults. *JAMA Neurol*. 2020;77:1241–1251.
7. Ayas NT, White DP, Manson JE, et al. A prospective study of sleep duration and coronary heart disease in women. *Arch Intern Med*. 2003;163:205–209.
8. Moon C, Hagen EW, Johnson HM, Brown RL, Peppard PE. Longitudinal sleep characteristics and hypertension status: results from the Wisconsin sleep cohort study. *J Hypertens*. 2021;39:683–691.

9. Full KM, Huang T, Shah NA, et al. Sleep irregularity and subclinical markers of cardiovascular disease: the multi-ethnic study of atherosclerosis. *J Am Heart Assoc*. 2023;12, e027361.
10. All of Us Research Program I, Denny JC, Rutter JL, et al. The "All of Us" research program. *N Engl J Med*. 2019;381:668–676.
11. Master H, Annis J, Huang S, et al. Association of step counts over time with the risk of chronic disease in the All of Us Research Program. *Nat Med*. 2022;28:2301–2308.
12. Ramirez AH, Sulieman L, Schlueter DJ, et al. The All of Us Research Program: data quality, utility, and diversity. *Patterns (N Y)*. 2022;3, 100570.
13. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385, e078378.
14. Ronquillo JG, Lester WT. Precision medicine landscape of genomic testing for patients with cancer in the National Institutes of Health all of us database using informatics approaches. *JCO Clin Cancer Inform*. 2022;6, e2100152.
15. Lee S, Hwang SH, Park S, et al. Burden of cardiovascular outcomes after SARS-CoV-2 infection in South Korea and Japan: a binational population-based cohort study. *Circulation*. 2025;152(1):1–15.
16. Vu T, Kokubo Y, Inoue M, et al. Machine learning model for predicting coronary heart disease risk: development and validation using insights from a Japanese population-based study. *JMIR Cardio*. 2025;9, e68066.
17. Full KM, Shi H, Lipworth L, Dauer LT, Mumma MT, Xiao Q. Sleep trajectories and all-cause mortality among low-income adults. *JAMA Netw Open*. 2025;8, e2462117.
18. Floyd JA, Janisse JJ, Jenuwine ES, Ager JW. Changes in REM-sleep percentage over the adult lifespan. *Sleep*. 2007;30:829–836.
19. McCarthy A, Wafford K, Shanks E, Ligocki M, Edgar DM, Dijk D-J. REM sleep homeostasis in the absence of REM sleep: effects of antidepressants. *Neuropharmacology*. 2016;108:415–425.
20. Sofer T, Kurniansyah N, Murray M, et al. Genome-wide association study of obstructive sleep apnoea in the million veteran program uncovers genetic heterogeneity by sex. *EBioMedicine*. 2023;90, 104536.
21. Simmonds E, Levine KS, Han J, et al. Sleep disturbances as risk factors for neurodegeneration later in life. *NPJ Dement*. 2025;1, 6.
22. Finocchi A, Campagnini S, Mannini A, et al. Multiple imputation integrated to machine learning: predicting post-stroke recovery of ambulation after intensive inpatient rehabilitation. *Sci Rep*. 2024;14, 25188.
23. Li C, Zhu Y, Ma Y, Hua R, Zhong B, Xie W. Association of cumulative blood pressure with cognitive decline, dementia, and mortality. *J Am Coll Cardiol*. 2022;79:1321–1335.
24. De Angeli K, Gao S, Danciu I, et al. Class imbalance in out-of-distribution datasets: improving the robustness of the TextCNN for the classification of rare cancer types. *J Biomed Inform*. 2022;125, 103957.
25. Osawa I, Goto T, Yamamoto Y, Tsugawa Y. Machine-learning-based prediction models for high-need high-cost patients using nationwide clinical and claims data. *NPJ Digit Med*. 2020;3:148.
26. Lee H, Hwang SH, Park S, et al. Prediction model for type 2 diabetes mellitus and its association with mortality using machine learning in three independent cohorts from South Korea, Japan, and the UK: a model development and validation study. *EClinicalMedicine*. 2025;80, 103069.
27. Wang H, Huang H, Li F, et al. Risk-stratified classification of pulmonary nodule malignancy via a machine learning model integrating imaging and cell-free DNA: a model development and validation study (DECIPHER-NODL). *Lancet Reg Health West Pac*. 2025;64.
28. Hu C, Xia Y, Zheng Z, et al. AI-based large-scale screening of gastric cancer from noncontrast CT imaging. *Nat Med*. 2025;31:3011–3019.
29. Sherazi SWA, Bae JW, Lee JY. A soft voting ensemble classifier for early prediction and diagnosis of occurrences of major adverse cardiovascular events for STEMI and NSTEMI during 2-year follow-up in patients with acute coronary syndrome. *PloS One*. 2021;16, e0249338.
30. Hou F, Zhu Y, Zhao H, et al. Development and validation of an interpretable machine learning model for predicting the risk of distant metastasis in papillary thyroid cancer: a multicenter study. *EClinicalMedicine*. 2024;77, 102913.
31. Huang Y, Li W, Macheret F, Gabriel RA, Ohno-Machado L. A tutorial on calibration measurements and calibration models for clinical prediction models. *J Am Med Inform Assoc*. 2020;27:621–633.
32. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol*. 2016;74:167–176.
33. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res*. 2019;3:18.
34. group Sw, collaboration ESCCr. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *Eur Heart J*. 2021;42:2439–2454.
35. Forrest IS, Petrazzini BO, Duffy Á, et al. Machine learning-based marker for coronary artery disease: derivation and validation in two longitudinal cohorts. *Lancet*. 2023;401:215–225.
36. Singh M, Kumar A, Khanna NN, et al. Artificial intelligence for cardiovascular disease risk assessment in personalised framework: a scoping review. *EClinicalMedicine*. 2024;73, 102660.
37. Fan M, Sun D, Zhou T, et al. Sleep patterns, genetic susceptibility, and incident cardiovascular disease: a prospective study of 385 292 UK biobank participants. *Eur Heart J*. 2020;41:1182–1189.
38. D'Agostino Sr RB, Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham heart study. *Circulation*. 2008;117:743–753.
39. Alaa AM, Bolton T, Di Angelantonio E, Rudd JHF, van der Schaar M. Cardiovascular disease risk prediction using automated machine learning: a prospective study of 423,604 UK Biobank participants. *PloS One*. 2019;14, e0213653.
40. Goff Jr DC, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association task force on practice guidelines. *J Am Coll Cardiol*. 2014;63:2935–2959.
41. Wang YH, Wang J, Chen SH, et al. Association of Longitudinal Patterns of habitual sleep duration with risk of cardiovascular events and all-cause mortality. *JAMA Netw Open*. 2020;3, e205246.
42. Saad R, Lillaney P, Profant DA, et al. Cardiovascular burden of individuals diagnosed with idiopathic hypersomnia: real-world idiopathic hypersomnia total health model (CV-RHYTHM). *Sleep Med*. 2025;133, 106587.
43. Mullington JM, Haack M, Toth M, Serrador JM, Meier-Ewert HK. Cardiovascular, inflammatory, and metabolic consequences of sleep deprivation. *Prog Cardiovasc Dis*. 2009;51:294–302.
44. Stamatakis KA, Punjabi NM. Long sleep duration: a risk to health or a marker of risk? *Sleep Med Rev*. 2007;11:337–339.
45. Reinhard W, Plappert N, Zeman F, et al. Prognostic impact of sleep duration and sleep efficiency on mortality in patients with chronic heart failure. *Sleep Med*. 2013;14:502–509.
46. Ai S, Zhang J, Zhao G, et al. Causal associations of short and long sleep durations with 12 cardiovascular diseases: linear and nonlinear Mendelian randomization analyses in UK Biobank. *Eur Heart J*. 2021;42:3349–3357.
47. Daghlas I, Dashti HS, Lane J, et al. Sleep duration and myocardial infarction. *J Am Coll Cardiol*. 2019;74:1304–1314.
48. Parhizkar S, Holtzman DM. The night's watch: exploring how sleep protects against neurodegeneration. *Neuron*. 2025;113:817–837.
49. Zhao B, Jin X, Yang J, et al. Increased rapid eye movement sleep is associated with a reduced risk of heart failure in middle-aged and older adults. *Front Cardiovasc Med*. 2022;9, 771280.
50. Larsson SC, Markus HS. Genetic liability to insomnia and cardiovascular disease risk. *Circulation*. 2019;140:796–798.
51. Kakadiaris IA, Vrigkas M, Yen AA, Kuznetsova T, Budoff M, Naghavi M. Machine learning outperforms ACC / AHA CVD risk calculator in MESA. *J Am Heart Assoc*. 2018;7, e009476.
52. Stucky B, Clark I, Azza Y, et al. Validation of Fitbit charge 2 sleep and heart rate estimates against polysomnographic measures in shift workers: naturalistic study. *J Med Internet Res*. 2021;23, e26476.
53. Schyvens AM, Peters B, Van Oost NC, et al. A performance validation of six commercial wrist-worn wearable sleep-tracking devices for sleep stage scoring compared to polysomnography. *Sleep Adv*. 2025;6, zpaf021.
54. Ai S, Ye S, Li G, et al. Association of disrupted delta wave activity during sleep with long-term cardiovascular disease and mortality. *J Am Coll Cardiol*. 2024;83:1671–1684.