

Socioeconomic and lifestyle factors predict the association between sleep health and depression

Wen Liu^{1, 2}, Vincent Küppers^{1, 3}, Hanwen Bi^{1, 2}, Mostafa Mahdipour^{1, 2}, Jianxiao Wu^{1, 2}, Fateme Samea^{1, 2, 4}, Felix Hoffstaedter^{1, 2}, Kathrin Wolf⁵, Charlotte von Gall⁶, Agustin Ibanez^{7, 8, 9, 10, 11}, Simon B. Eickhoff^{1, 2}, Sarah Genon^{1, 2, 12}, Somayeh Maleki Balajoo^{1, 2#*}, Masoud Tahmasian^{1, 2, 3#*}

1. Institute of Neuroscience and Medicine, Brain and Behaviour (INM 7), Research Center Jülich, Jülich, Germany
2. Institute of Systems Neuroscience, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany
3. Department of Nuclear Medicine, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany
4. Clinic and Polyclinic for Child and Adolescent Psychiatry, Psychotherapy, and Psychosomatics, Medical Faculty, Leipzig University, Leipzig, Germany
5. Institute of Epidemiology, Helmholtz Zentrum München GmbH, German Research Center for Environmental Health, Neuherberg, Germany
6. Institute of Anatomy II, Medical Faculty, Heinrich Heine University, Düsseldorf, Germany
7. Latin America Brain Health Institute (BrainLat), Universidad Adolfo Ibanez, Santiago, Chile
8. Global Brain Health Institute (GBHI), Trinity College Dublin, Dublin, Ireland
9. Department of Biophysics, School of Medicine, Istanbul Medipol University, Istanbul, Türkiye
10. Barcelonaβeta Brain Research Center (BBRC), Pasqual Maragall Foundation, Barcelona, Spain
11. Cognitive Neuroscience Center (CNC), Universidad de San Andrés, Buenos Aires, Argentina
12. GIGA-CRC-Human Imaging, University of Liege, Liège, Belgium

These authors contributed equally to this work.

*Corresponding Author: Institute of Neuroscience and Medicine, Brain & Behaviour (INM-7), Research Center Jülich, Wilhelm-Johnen-Straße, 52428 Jülich, Germany. Phone: +49 2461 61-8785, Fax: +49 2461 61-1880. Email: m.tahmasian@fz-juelich.de

Abstract

Objective: Sleep health and depression are interconnected multidimensional constructs, yet their shared determinants remain obscure. Understanding the role of socioeconomic/lifestyle factors in predicting sleep-related depression (SRD) is critical for preventive strategies. This study aimed to identify the key socioeconomic/lifestyle predictors of SRD in the general population and patients with clinical depression.

Methods: To characterize SRD, we performed regularized canonical correlation analysis between sleep and depression to identify latent phenotypes of SRD in a general population subsample (GP1; $n = 87,405$) from the UK Biobank. Subsequently, machine-learning predictive models were developed in GP1 to predict SRD using socioeconomic/lifestyle factors. The best-performing predictive model was subsequently validated in GP2 at both baseline and follow-up (GP2; $n = 5,187$), and in clinical depression ($n = 7,454$) to assess its generalizability. Complementary analyses were conducted to assess other latent phenotypes (i.e., depression-related sleep, non-SRD, non-depression-related sleep, overall sleep health, and overall depression).

Results: A robust multivariate association was identified between sleep and depression in GP1 (canonical $r = 0.42$, $P_{FDR} < 0.001$). Socioeconomic/lifestyle factors moderately predicted SRD ($r = 0.25$; 95% CI: [0.24, 0.25]; $R^2 = 0.06$; 95% CI: [0.06, 0.06]; $rMSE = 1.08$; 95% CI: [1.08, 1.09]). The top predictors were less frequency of confiding in others, more sedentary television viewing, less vigorous physical activity, and passive smoking exposure. Out-of-sample validation of the predictive model showed similar patterns in GP2 at baseline, at follow-up, and in clinical depression subsamples. Similarly, less frequency of confiding in others and greater sedentary television viewing were the main predictors of other depression-related profiles, whereas more alcohol consumption frequency, less walking frequency, and less time spent outdoors in winter predicted poor sleep-related profiles.

Conclusions: Our generalizable predictive model identifies critical modifiable predictors of the association between sleep health and depression that could serve as potential targets for personalized interventions.

Introduction

Depression is the leading cause of disability worldwide, ranging from subclinical depressive symptoms in the general population to patients with major depressive disorder (MDD). Depression is inherently multidimensional, characterized by substantial heterogeneity across emotional, cognitive, social, biological, somatic, and sleep domains.^{1,2} Poor Sleep health (SH) represents pervasive features and a core domain of depression.^{1,3,4} Approximately 78% of patients with MDD report insomnia symptoms.⁵ SH is a multifaceted construct encompassing sleep duration, continuity, and quality, which support optimal daytime functioning, mental health, and well-being.^{6,7} Individuals with sleep disturbances and insomnia have a greater risk for the onset, recurrence, and exacerbation of MDD.^{8–13} Meanwhile, sleep disturbances are also considered a diagnostic criterion for MDD², reflecting their bidirectional relationship. Beyond symptom-level associations, converging genetic and neuroimaging evidence further underscores a tight coupling between poor SH and MDD,^{14–17} supporting a robust, multidimensional relationship and a shared mechanism underlying the sleep-depression nexus. However, prior studies predominantly relied on group-level comparisons or univariate correlations examining individual sleep characteristics (e.g., insomnia) and overall depression (total scores) in isolation, either in the general population or in patients with MDD.^{18–20} These approaches overlook inter-individual variability and fail to capture the complex, multifaceted nature of their associations, or their generalizability across populations. A multivariate framework is therefore needed to disentangle these overlapping constructs and identify their shared multidimensional structure. Thus, we introduce *sleep-related depression (SRD)* as a novel construct capturing the specific dimension of depressive symptoms that strongly covaries with SH characteristics at the individual level.

The modifiable factors contributing to their shared variation remain poorly understood, particularly at the individual level across both the general population and MDD samples. Socioeconomic and lifestyle factors are modifiable exposures that influence the severity of both depression^{21,22} and sleep problems^{23,24}. Poverty, social isolation, and unhealthy lifestyle, including physical inactivity, malnourishing diet, smoking, and high alcohol consumption, are associated with an increased risk of depressive symptoms and onset of MDD,^{25–27} whereas healthy lifestyle and social support are protective factors.^{27,28} Despite well-documented group-level evidence, existing works have examined socioeconomic and lifestyle factors in relation to either SH or depression. Consequently, it remains unclear whether and how these factors shape the shared variance between SH and depression, e.g., by predicting SRD.

This study leveraged large-scale data and multivariate association analyses to derive individual-level SRD scores, representing canonical variates of depression that maximized their correlation with corresponding SH canonical variates — thereby translating shared SH–

depression variance into an individual-level metric beyond conventional group-level measures. Machine learning (ML) models were then employed to predict SRD scores from socioeconomic/lifestyle factors and quantify their relative contributions to the shared SH-depression phenotype in the general population. Model generalizability was evaluated in independent general population subsamples (GP), both cross-sectionally and longitudinally, and in a clinical depression subsample (CD). Using complementary analyses, we examined the differential contributions of these factors across additional latent phenotype profiles, including depression-related sleep (DRS), non-sleep-related depression (NSRD), non-depression-related sleep (NDRS), overall SH, and overall depression. This systematic approach was adopted to disentangle the predictive role of socioeconomic/lifestyle factors on the shared and distinct latent dimensions underlying the SH and depression.

Methods

Participants and data handling

Data were obtained from the UK Biobank (UKB, ethical approval in eMethods 1). To calculate SRD scores, we analyzed four domains in both GP and CD: *i*) seven SH components; *ii*) depressive symptoms; *iii*) socioeconomic variables; and *iv*) lifestyle factors (eMethods 1, eTable 1). Variable selection was based on the availability of UKB data.

Participants were included if they had complete data across all domains and met our selection criteria (eTable 2-3, eMethods 2). Eligible participants were divided into four independent subsamples (Figure 1A). The primary subsample was GP1 ($n = 87,405$) with cross-sectional data, used to derive the SRD scores, ML predictive models, and complementary analyses. Generalizability of predictive models was assessed in independent subsamples using both baseline (GP2 baseline; $n = 5,187$) and follow-up (GP2 follow-up; mean follow-up 9.2 years; $n = 5,187$) assessments to evaluate the long-term roles of baseline socioeconomic/lifestyle factors in predicting follow-up SRD scores. We also assessed multivariate associations and the generalizability of predictive models in CD ($n = 7,454$). Participants in CD either had a documented ICD-10 diagnosis of either current depressive episode (F32) or recurrent depressive disorder (F33) or were taking antidepressant medication (eMethods 2) upon assessment.

Statistical Analyses

Analysis framework

To maximize data utilization, obtain unbiased individual-level SRD estimates, and prevent data leakage, we novelly combined regularized canonical correlation analysis (rCCA) and predictive modeling within a nested cross-validation framework²⁹ (Figure 1C). The GP1 was

randomly split into five outer folds. In each iteration, one outer fold (CCA fold, 20%) was used to build the rCCA model, while the remaining four folds (ML fold, 80%) were used for predictive models (Figure 1C). Within each CCA and ML fold, a nested cross-validation scheme was conducted.³⁰ To prevent data leakage, age and sex were regressed out using only training data, and the resulting regression parameters were applied to the validation and test sets for both rCCA and predictive models.³¹ (eMethods 3)

Regularized Canonical Correlation Analysis

RCCA is a multivariate approach that quantifies the maximum shared variance between at least two sets of variables while reducing overfitting through regularization, as previously applied.^{1,32–34} We applied rCCA in GP1 and CD to calculate SRD scores, reflecting the multivariate association between SH and depression characteristics after controlling for age and sex as covariates. The first canonical variate of depression, capturing the dimension most strongly associated with SH, was used to represent individual-level SRD scores. Canonical loadings were calculated as the associations between the original variable and its corresponding canonical variate. Redundancy was calculated as the mean variance of one set explained by the canonical variates of the other set.³⁵ To minimize overfitting and enhance robustness and generalizability, a regularized version (rCCA) was implemented within a ML framework,^{1,32,36} with a hyperparameter optimization procedure described in eMethods 4. Statistical significance for rCCA was assessed using a false discovery rate (FDR) correction with a threshold of $P_{FDR} < 0.05$ (eMethods 5).

Predictive analyses

To evaluate whether socioeconomic/lifestyle features can predict SRD scores, several linear and non-linear ML models, including ridge regression,³⁷ LASSO regression,³⁸ elastic net,³⁹ random forest,⁴⁰ extreme gradient boosting (XGBoost),⁴¹ linear support vector regression (SVR),⁴² and SVR with radial basis function kernel⁴³ were trained in GP1 (Figure 1B, eMethods). Multicollinearity among features was controlled using the variance inflation factor (VIF), with values > 5 indicating potentially problematic multicollinearity.⁴⁴ The best-performing model was calibrated (eMethods 6), and compared with a demographic model including age, sex, and years and education as predictors to assess whether the predictive model captured target-relevant information beyond demographic characteristics alone. The best-performing model with and without controlling confounds were compared to evaluate the potential risk from linear confound adjustment into non-linear models.⁴⁵

Feature contributions were quantified using SHapley Additive exPlanations (SHAP) values and examined across age and sex subgroups to assess potential heterogeneity in feature contributions. Age- and sex-stratified analyses were conducted within the best-

performing predictive model using a threshold of $P_{FDR} < 0.05$ for statistical significance (eMethods 7). The best-performing predictive model trained in GP1 was then validated in three independent subsamples (GP2 baseline, GP2 follow-up, and CD) to evaluate generalizability across populations using out-of-sample validation (eMethods 8).

Complementary analyses

We examined whether and how socioeconomic/lifestyle factors contributed differently to predicting other profiles of distinct latent phenotypes, including DRS, NSRD, NDRS, overall depression, and overall SH (eMethods 9). The same socioeconomic/lifestyle features were applied to GP1 using XGBoost to predict each latent profile. To derive latent dimensions comparable to SRD, principal component analysis was performed on each profile, and the first component, representing the largest proportion of variance, was selected as the representative dimension.

We also conducted sensitivity predictive analyses in CD by excluding participants with 1) recurrent depression ($n = 91$) and 2) both recurrent depression and antidepressant use ($n = 5,301$, eTable 4) to control for their impact.

Results

Demographics and phenotypic characteristics

The primary rCCA, ML analyses, and complementary analyses were conducted in GP1. Out-of-sample validation was performed in baseline and follow-up GP2 and CD (demographic and feature characteristics in eTable 5).

Multivariate associations between SH and depression

The first canonical dimension captured the strongest shared variance (17.6%) between SH and depression. Our rCCA results in GP1 showed a positive association between the corresponding canonical variates (canonical correlation $r_{\text{mean}} = 0.42$ [SD: 0.002], $P_{FDR} < 0.001$; Figure 2A). The canonical variates of depression from this dimension were therefore used to measure SRD scores. These individual-level SRD scores were used as the target for predictive models.

All SH components were positively associated with depression, except for sleep duration, for which shorter duration was associated with more severe depression. Among all components in SH, insomnia and difficulty getting up in the morning contributed most strongly to the SH variate, whereas frequency of tiredness or lethargy over the past two weeks showed the highest loading among depression measures. Collectively, an individual

with a higher SRD score had more sleep disturbances and greater depression severity (Figure 2B). The proportion of SH variance explained by SRD in GP1 was in eTable 6.

Predictability of SRD from lifestyle/socioeconomic factors in GP1

We assessed multicollinearity among input features prior to ML model training, and all VIF values were < 2 , indicating no strong multicollinearity (eTable 7). The distributions of SRD derived from the five ML folds are shown in eFigure 1. Across seven ML algorithms to evaluate whether the SRD scores could be predicted from socioeconomic/lifestyle factors in GP1 after controlling age and sex, eXtreme Gradient Boosting (XGBoost) demonstrated the best predictive performance ($r: 0.25 \pm 0.007$; 95% CI: [0.24, 0.25]; $R^2: 0.06 \pm 0.003$; 95% CI: [0.06, 0.06]; $rMSE: 1.08 \pm 0.024$; 95% CI: [1.08, 1.09]; Figure 3A, eFigure 2). XGBoost was calibrated with a slope of 1.03 and an intercept close to 0 (< 0.001 , eFigure 3). Feature importance from SHAP analyses revealed that lower social connection, indexed by less frequent confiding in others, was the main predictor of higher SRD scores. This was followed by physical inactivity factors, including longer hours of sedentary television viewing and fewer days of vigorous physical activity (lasting ≥ 10 minutes/week), and by smoking-related exposure, particularly greater passive smoking outside the home (Figure 3B).

Stratified analyses by age and sex demonstrated that passive smoking exposure was more strongly related to SRD in younger adults (< 65 years) and males, whereas vigorous physical activity was critical among older adults (≥ 65 years) and females (eFigure 4 and 5). Feature contribution rankings were largely stable across age and sex subgroups, though FDR-corrected analyses revealed subgroup-specific differences in contribution weights (eTable 8 and 9). The predictive model outperformed the demographic model (eTable 10), suggesting that the socioeconomic/lifestyle factors provided predictive information beyond demographic characteristics alone. Model performances were comparable with and without linear confound adjustment, indicating minimal risk of data leakage (eTable 11).

Out-of-sample validation in other independent subsamples

Generalizability was evaluated by applying the GP1-trained models to independent datasets. In both GP2 subsamples, the rCCA model trained in GP1 was applied directly to derive the SRD scores, which were subsequently used as the prediction targets. However, in CD, stronger multivariate associations between SH and depression were observed (canonical correlation $r_{mean} = 0.46$ [SD, 0.004]; eFigure 6), indicating a different structure of the SRD phenotype in clinical patients. To avoid underestimating SRD severity in CD and to ensure a representative prediction target, rCCA was retrained within CD using the same framework as GP1, prior to applying the GP1-trained XGBoost model.

Predictive performance was comparable to GP1 in GP2 at baseline for predicting SRD scores, but lower in GP2 for predicting follow-up SRD scores based on baseline features, with better predictive performance observed in CD (eTable 12, Figure 4A). Across all validation subsamples, the top predictors identified in GP1 remained largely stable. (Figure 4B). The proportion of SH explained by SRD in CD was reported in eTable 6.

Complementary analyses

To further understand different latent profiles of SH and depression, lower predictive performance was observed in DRS, NDRS, and NSRD, whereas higher performance was identified in overall SH and overall depression (eTable 13). Similar to SRD, less frequency of confiding and greater sedentary television viewing merged as the primary predictors of *depression-related profiles* (i.e., NSRD and overall depression), whereas more alcohol consumption frequency, less walking frequency, and less time spent outdoors in winter were the main predictors for *poor sleep-related profiles* (i.e., NDRS and overall sleep disturbances). In contrast, more frequent use of artificial light (sunlamps) was a stronger predictor of sleep disturbances (Figure 5A). These results indicate shared and distinct contributions of socioeconomic/lifestyle features in predicting SH and depression profiles (Figure 5B).

An identical canonical association between SH and depression was observed in non-recurrent depression (eTable 14, eFigure 7), whereas a weaker association was identified in untreated depression (eTable 14, eFigure 8). The predictive performance was comparable, and the top predictors remained stable when excluding recurrent depression, and excluding recurrent depression and antidepressants (eTable 15, eFigures 9-10).

Discussion

In this large-scale study, we employed a novel combined rCCA and ML approach within a nested cross-validation framework to avoid data leakage while maximizing data utilization. Our aim was to examine whether socioeconomic/lifestyle factors can predict the latent phenotype underlying the interplay between SH and depression. Our predictive models demonstrated moderate but generalizable patterns across cross-sectional and longitudinal GP and CD subsamples. Our rCCA findings identified insomnia and difficulty getting up in the morning as the strongest components linked to depression, such as tiredness and lethargy, in both GP and CD. Despite their robust associations, routine psychotherapies often target either insomnia^{46,47} or depression¹ in isolation and fail to alleviate the other condition to achieve full remission, suggesting the importance of simultaneously addressing

both conditions. Insomnia is a common residual symptom of depression, which may lead to relapses of depression, suicide, and poor prognosis,^{48,49} underscoring the importance of identifying shared influencing factors that may concurrently affect both SH and depression and have investable impacts on treatment outcomes.

Although the protective roles of healthy socioeconomic/lifestyle factors such as physical activity, Mediterranean diet, and social support, and the adverse effects of smoking and alcohol on SH and depression are well established,^{21,50–53} our results further highlighted their role in predicting SRD. Lower frequency of confiding, greater hours of sedentary television viewing, passive smoking exposure outside the home, and fewer days of vigorous physical activity emerged as the top predictors in the present study. Consistently, previous studies found that less confiding frequency, less vigorous physical activity (e.g., swimming or cycling), slower walking pace, and longer sedentary television viewing were associated with more depression,^{54,55} and confiding frequency together with sedentary television viewing were even suggested as causal factors to depression.⁵⁵ Notably, baseline measures of socioeconomic/lifestyle factors moderately predicted subsequent SRD scores assessed approximately 9 years later, underscoring their long-term impacts. The predictive model showed stronger performance in CD, regardless of the effects of recurrent episodes or medication use, suggesting a tighter association between socioeconomic/lifestyle factors and SRD in MDD. This finding was consistent with a meta-analysis reporting larger effect sizes for lifestyle interventions, including physical activity and dietary changes, in MDD patients compared to GP.⁵⁶

Although the top predictors identified in the pooled sample remained stable in age and sex stratified analyses, smoking was more prominent in males and younger adults, while vigorous physical activities were more important in females and older adults, suggesting a non-negligible role of age and sex in modulating the associations between socioeconomic/lifestyle factors and SRD. Prior studies show that SH and depression profiles drastically vary by age and sex,^{57,58} such as physical activities and social support.^{59–61}

Socioeconomic and lifestyle factors were more strongly predictive of depression-related profiles than of SH-related profiles. Less frequency of confiding and more sedentary television viewing contributed more strongly to different *depression-related profiles*, whereas less walking frequency and time outdoors in winter, and more alcohol consumption frequency showed greater links with *poor SH-related profiles*. Smoking history was more strongly related to overall sleep disturbances, secondhand smoke exposure outside the home predicted the shared variance of SH and depression (i.e., SRD and DRS), and current smoking status predicted NSRD. Higher frequency of alcohol consumption was predominantly associated with poor *sleep-related profiles*, whereas frequency of artificial light uses predicted SH. It has been shown that alcohol consumption was significantly

related to poor SH, including late chronotype, abnormal duration, snoring, insomnia, and daytime napping.⁶² More walking and sunlight exposure have been suggested to be beneficial for reducing sleep disturbances,^{60–64} but less evidence supports an association between exercise intensity and sleep quality improvement.⁶⁵ The frequency of confiding in others is internally associated with social isolation and may drive the model in predicting depression-related profiles.⁶⁶

Limitations

These findings should be cautiously interpreted, as the observed effect sizes were modest, accounting for only 6–11% of the variance. The generalizability of our findings based on middle-to-old-aged adults to younger adults and adolescents warrants further investigation. SH was not measured by objective assessment and was not stratified by workdays vs non-workdays, given the different sleep patterns across these periods.^{24,67,68} Moreover, because of the different SH-depression associations in GP and CD, the rCCA model was retrained in CD to derive SRD scores as the target for out-of-sample validation of the predictive model; thus, this is not a fully independent out-of-sample validation in CD. Finally, the current observational design precludes any causal inference.

Conclusions

Robust multivariate associations between SH and depression were consistently observed across four subsamples, and their shared covariance was moderately predictable from socioeconomic/lifestyle factors. Maladaptive lifestyle factors, including lower frequency of confiding, greater sedentary television viewing, secondhand smoke exposure, and less vigorous physical activity, emerged as consistent predictors of higher SRD scores across populations and time points. Our systematic approach disentangled the divergent and convergent roles of socioeconomic/lifestyle factors on the shared and distinct latent dimensions underlying the SH–depression relationships. The generalizability of predictive findings supports the utility of such modifiable factors as potential targets for preventive strategies aimed at reducing the interrelated burden of poor SH and depression.

Data availability

The data used in this study are derived from the UK Biobank resource were approved by the ethical committee of Heinrich Heine University Düsseldorf (2018-317-RetroDEuA). These data are available under restricted access due to ethical approval requirements, participant consent, and data protection regulations. Access can be obtained by researchers through application to the UK Biobank Access Management System

(<https://www.ukbiobank.ac.uk/>), subject to UK Biobank approval. Raw participant-level data are protected and cannot be publicly shared due to data privacy laws. Processed data that do not contain participant-level information are provided with this paper as Source Data files where applicable. Source data supporting this study are provided with this paper.

Code availability

The analysis code used in this study is publicly available on GitHub at (https://github.com/wenliu1/Socioeconomic-lifestyle_predicting_sleep_depression_profiles).

Acknowledgements

This research was conducted using data from the UK Biobank (<https://www.ukbiobank.ac.uk/>), a major biomedical database. We are grateful to UK Biobank for providing access to the data and to all study participants for generously contributing their time and information to this resource.

Author contributions

W.L., S.M.B., and M.T. had full access to all data in the study and took responsibility for the integrity and accuracy of the data and analysis. Study concept and design were by W.L., S.B.E., S.G., A.I., S.M.B., and M.T. Data preparation and preprocessing were carried out by W.L., V.K., F.H. Analyses and model development were conducted by W.L., H.B., J.W., F.S., and S.M.B. Data interpretation was by W.L., S.B.E., S.G., A.I., S.M.B., and M.T. Drafting of the manuscript was by W.L., S.M.B., and M.T. Critical revision of the manuscript for important intellectual content was done by V.K., H.B., J.W., F.S., F.H., C.V.G., S.G., S.B.E., K.W., M.M., and S.G. All authors approved the final version of the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. Wen Liu is supported by the Deutscher Akademischer Austauschdienst (DAAD) - Project-ID 57645448. Sarah Genon is supported by the Deutsche Forschungsgemeinschaft (DFG, GE 2835/2–1, GE 2835/9-1). Agustin Ibanez is supported by grants from the Multi-partner consortium to expand dementia research in Latin America [ReDLat2, supported by Fogarty International Center (FIC), National Institutes of Health, National Institutes of Aging (R01 AG057234, R01 AG075775, R01 AG21051, R01 AG083799, CARDS-NIH, R01 AG057234), Alzheimer's Association (SG-20-725707), Rainwater Charitable Foundation – The Bluefield project to cure FTD, and Global Brain Health Institute)], ANID/FONDECYT Regular (1250091 and 1210176 and 1220995); ANID/PIA/ANILLOS ACT210096; JPI JPND-Care, DISCeRN 2025 - Health and Social Care Research with a Focus on the Moderate and Late Stages of Neurodegenerative Diseases; FONDEF ID20I10152, and ANID/FONDAP 15150012; Wellcome Trust award for

BRAIN-CLIMA: Investigating the Combined Impact of Heat and Air Pollution on Blood-Brain Barrier Integrity and Brain Aging in Latin America, (335293/Z/25/Z), Wellcome Leap CARE Program (Grant Number: CARE-2025-0883490149) for the project “Advancing Female-Specific Predictive Models and Risk Assessment Tools for Alzheimer’s Disease in the US and Latin America, and the CliCBrain (Horizon ID: 101236426; DOI 10.3030/101236426, Marie Skłodowska-Curie Actions - MSCA). The contents of this publication are solely the responsibility of the authors and do not represent the official views of these institutions. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The contents of this publication are solely the author's responsibility and do not represent the official views of these institutions.

Competing interests

The authors declare no competing interests.

Reference:

1. Haritos R, Küppers V, Samea F, et al. The effect of psychotherapy on the multivariate association between insomnia and depressive symptoms in late-life depression. *Eur Psychiatry*. 2025;68(1):e120. doi:10.1192/j.eurpsy.2025.10088
2. Malhi GS, Mann JJ. Depression. *The Lancet*. 2018;392(10161):2299-2312. doi:10.1016/S0140-6736(18)31948-2
3. Freeman D, Sheaves B, Waite F, Harvey AG, Harrison PJ. Sleep disturbance and psychiatric disorders. *The Lancet Psychiatry*. 2020;7(7):628-637. doi:10.1016/S2215-0366(20)30136-X
4. Mendlewicz J. Sleep disturbances: Core symptoms of major depressive disorder rather than associated or comorbid disorders. *The World Journal of Biological Psychiatry*. 2009;10(4):269-275. doi:10.3109/15622970802503086
5. Toynbee M, Zygmunt D, Levett R, Moghaddacy S, Pappa S. The prevalence of insomnia in depressed adults: a systematic review. *Psychiatry Res*. 2026;356:116888. doi:10.1016/j.psychres.2025.116888
6. Hale L, Troxel W, Buysse DJ. Sleep Health: An Opportunity for Public Health to Address Health Equity. *Annu Rev Public Health*. 2020;41:81-99. doi:10.1146/annurev-publhealth-040119-094412
7. Tahmasian M, Küppers V, Genon S, Eickhoff SB, Golombek DA, Ibanez A. Elevating sleep to a global health priority: The One Sleep Health framework. *CR Med*. 2026;0(0). doi:10.1016/j.xcrm.2026.102828
8. Baglioni C, Battagliese G, Feige B, et al. Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. *J Affect Disord*. 2011;135(1-3):10-19. doi:10.1016/j.jad.2011.01.011
9. Hertenstein E, Benz F, Schneider CL, Baglioni C. Insomnia—A risk factor for mental disorders. *Journal of Sleep Research*. 2023;32(6):e13930. doi:10.1111/jsr.13930
10. Olfati M, Samea F, Faghihroohi S, et al. Prediction of depressive symptoms severity based on sleep quality, anxiety, and gray matter volume: a generalizable machine learning approach across three datasets. *EBioMedicine*. 2024;108:105313. doi:10.1016/j.ebiom.2024.105313
11. Tonon AC, Nexha A, Cunningham JEA, et al. One-Year Actigraphy Study of Sleep and Rest-Activity Rhythms as Markers of Relapse in Depression. *JAMA Psychiatry*. Published online February 11, 2026. doi:10.1001/jamapsychiatry.2025.4453
12. Marino C, Andrade B, Montplaisir J, et al. Testing Bidirectional, Longitudinal Associations Between Disturbed Sleep and Depressive Symptoms in Children and Adolescents Using Cross-Lagged Models. *JAMA Netw Open*. 2022;5(8):e2227119. doi:10.1001/jamanetworkopen.2022.27119
13. Bao YP, Han Y, Ma J, et al. Cooccurrence and bidirectional prediction of sleep disturbances and depression in older adults: Meta-analysis and systematic review. *Neurosci Biobehav Rev*. 2017;75:257-273. doi:10.1016/j.neubiorev.2017.01.032

14. Sun X, Liu B, Liu S, et al. Sleep disturbance and psychiatric disorders: a bidirectional Mendelian randomisation study. *Epidemiology and Psychiatric Sciences*. 2022;31:e26. doi:10.1017/S2045796021000810
15. Peng C, Wang K, Wang J, et al. Neural correlates of insomnia with depression and anxiety from a neuroimaging perspective: A systematic review. *Sleep Med Rev*. 2025;81:102093. doi:10.1016/j.smrv.2025.102093
16. Gibson MJ, Lawlor DA, Millard LAC. Identifying the potential causal role of insomnia symptoms on 11,409 health-related outcomes: a phenome-wide Mendelian randomisation analysis in UK Biobank. *BMC Med*. 2023;21(1):128. doi:10.1186/s12916-023-02832-8
17. Wang Y, Tang S, Zhang L, et al. Data-driven clustering differentiates subtypes of major depressive disorder with distinct brain connectivity and symptom features. *The British Journal of Psychiatry*. 2021;219(5):606-613. doi:10.1192/bjp.2021.103
18. Zhu Y, Chen F, Fang X, Liang N, Na G, Liu Z. U-shaped association between sleep duration and depression among postmenopausal women: Evidence from a population-based study. *Journal of Affective Disorders*. 2026;397:120934. doi:10.1016/j.jad.2025.120934
19. Zhou L, Saltoun K, Carrier J, Storch KF, Dunbar RIM, Bzdok D. Multimodal population study reveals the neurobiological underpinnings of chronotype. *Nat Hum Behav*. 2025;9(7):1442-1456. doi:10.1038/s41562-025-02182-w
20. Lyall LM, Wyse CA, Graham N, et al. Association of disrupted circadian rhythmicity with mood disorders, subjective wellbeing, and cognitive function: a cross-sectional study of 91 105 participants from the UK Biobank. *Lancet Psychiatry*. 2018;5(6):507-514. doi:10.1016/S2215-0366(18)30139-1
21. Firth J, Solmi M, Wootton RE, et al. A meta-review of “lifestyle psychiatry”: the role of exercise, smoking, diet and sleep in the prevention and treatment of mental disorders. *World Psychiatry*. 2020;19(3):360-380. doi:10.1002/wps.20773
22. Agerbo E, Trabjerg BB, Børghlum AD, et al. Risk of Early-Onset Depression Associated With Polygenic Liability, Parental Psychiatric History, and Socioeconomic Status. *JAMA Psychiatry*. 2021;78(4):387-397. doi:10.1001/jamapsychiatry.2020.4172
23. Gueye-Ndiaye S, Redline S. Sleep Health Disparities. *Annu Rev Med*. 2025;76(1):403-415. doi:10.1146/annurev-med-070323-103130
24. Aslamiyar D, Pilz LK, Gall C von. Relationships Between Self-Reported Sleep Quality, Quantity and Timing on Workdays vs Work-Free Days and Lifestyle Factors in Healthy Adults. *Nat Sci Sleep*. 2025;17:1641-1654. doi:10.2147/NSS.S537593
25. Sato A, Eguchi E, Hayashi F, et al. A prospective study of the association between lifestyle and the risk of depressive symptoms. *Journal of Affective Disorders*. 2025;376:269-279. doi:10.1016/j.jad.2025.01.040
26. Ridley M, Rao G, Schilbach F, Patel V. Poverty, depression, and anxiety: Causal evidence and mechanisms. *Science*. 2020;370(6522):eaay0214. doi:10.1126/science.aay0214

27. Zhao Y, Yang L, Sahakian BJ, et al. The brain structure, immunometabolic and genetic mechanisms underlying the association between lifestyle and depression. *Nat Mental Health*. 2023;1(10):736-750. doi:10.1038/s44220-023-00120-1
28. Zheng YB, Huang YT, Gong YM, et al. Association of lifestyle with sleep health in general population in China: a cross-sectional study. *Transl Psychiatry*. 2024;14(1):320. doi:10.1038/s41398-024-03002-x
29. Domingos P. A few useful things to know about machine learning. *Commun ACM*. 2012;55(10):78-87. doi:10.1145/2347736.2347755
30. Berrar D. Cross-Validation. Published online January 1, 2025:638-644. doi:10.1016/B978-0-323-95502-7.00032-4
31. Komeyer V, Nieto N, Eickhoff SB, Raimondo F, Patil KR. Overview of Challenges in Brain-Based Predictive Modeling: Toward Meaningful Predictive Insights. *Biological Psychiatry*. Published online September 12, 2025. doi:10.1016/j.biopsych.2025.09.003
32. Küppers V, Bi H, Nicolaisen-Sobesky E, et al. Multivariate associations of motor performance, sleep quality, depressive symptoms, and grey matter volume in younger and mid-to-older adults. *Sci Rep*. 2026;16(1):1318. doi:10.1038/s41598-025-34951-y
33. Balajoo SM, Plachti A, Nicolaisen-Sobesky E, et al. Discovery, Replicability, and Generalizability of a Left Anterior Hippocampus' Morphological Network Linked to Self-Regulation. *Hum Brain Mapp*. 2024;45(18):e70099. doi:10.1002/hbm.70099
34. Nicolaisen-Sobesky E, Maleki Balajoo S, Mahdipour M, et al. Cardiometabolic health and physical robustness map onto distinct patterns of brain structure and neurotransmitter systems. *PLoS Biol*. 2025;23(11):e3003498. doi:10.1371/journal.pbio.3003498
35. Stewart D, Love W. A general canonical correlation index. *Psychological Bulletin*. 1968;70(3, Pt.1):160-163. doi:10.1037/h0026143
36. Mihalik A, Ferreira FS, Moutoussis M, et al. Multiple Holdouts With Stability: Improving the Generalizability of Machine Learning Analyses of Brain–Behavior Relationships. *Biological Psychiatry*. 2020;87(4):368-376. doi:10.1016/j.biopsych.2019.12.001
37. McDonald GC. Ridge regression. *WIREs Computational Statistics*. 2009;1(1):93-100. doi:10.1002/wics.14
38. Ranstam J, Cook JA. LASSO regression. *Br J Surg*. 2018;105(10):1348. doi:10.1002/bjs.10895
39. Zou H, Hastie T. Regularization and Variable Selection Via the Elastic Net. *J R Stat Soc Ser B Stat Methodol*. 2005;67(2):301-320. doi:10.1111/j.1467-9868.2005.00503.x
40. Breiman L. Random Forests. *Machine Learning*. 2001;45(1):5-32. doi:10.1023/A:1010933404324
41. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. Association for Computing Machinery; 2016:785-794. doi:10.1145/2939672.2939785

42. Awad M, Khanna R. Support Vector Regression. In: Awad M, Khanna R, eds. *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*. Apress; 2015:67-80. doi:10.1007/978-1-4302-5990-9_4
43. Liu Q, Chen C, Zhang Y, Hu Z. Feature selection for support vector machines with RBF kernel. *Artif Intell Rev*. 2011;36(2):99-115. doi:10.1007/s10462-011-9205-2
44. Shrestha N. Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics*. 2020;8(2):39-42. doi:10.12691/ajams-8-2-1
45. Hamdan S, Love BC, von Polier GG, et al. Confound-leakage: confound removal in machine learning leads to leakage. *Gigascience*. 2022;12:giad071. doi:10.1093/gigascience/giad071
46. Blanken TF, Van Der Zweerde T, Van Straten A, Van Someren EJW, Borsboom D, Lancee J. Introducing Network Intervention Analysis to Investigate Sequential, Symptom-Specific Treatment Effects: A Demonstration in Co-Occurring Insomnia and Depression. *Psychother Psychosom*. 2019;88(1):52-54. doi:10.1159/000495045
47. Blom K, Forsell E, Hellberg M, Svanborg C, Jernelöv S, Kaldo V. Psychological Treatment of Comorbid Insomnia and Depression: A Double-Blind Randomized Placebo-Controlled Trial. *Psychother Psychosom*. 2024;93(2):100-113. doi:10.1159/000536063
48. Bei B, Asarnow LD, Krystal A, Edinger JD, Buysse DJ, Manber R. Treating insomnia in depression: Insomnia related factors predict long-term depression trajectories. *J Consult Clin Psychol*. 2018;86(3):282-293. doi:10.1037/ccp0000282
49. Inada K, Enomoto M, Yamato K, Marumoto T, Takeshima M, Mishima K. Effect of residual insomnia and use of hypnotics on relapse of depression: a retrospective cohort study using a health insurance claims database. *Journal of Affective Disorders*. 2021;281:539-546. doi:10.1016/j.jad.2020.12.040
50. Grigoriou I, Kotoulas SC, Porpodis K, et al. The Interactions between Smoking and Sleep. *Biomedicines*. 2024;12(8). doi:10.3390/biomedicines12081765
51. Cruz J, Llodio I, Iturricastillo A, Yanci J, Sánchez-Díaz S, Romarateabala E. Association of Physical Activity and/or Diet with Sleep Quality and Duration in Adolescents: A Scoping Review. *Nutrients*. 2024;16(19). doi:10.3390/nu16193345
52. Huang HH, Stubbs B, Chen LJ, et al. The effect of physical activity on sleep disturbance in various populations: a scoping review of randomized clinical trials. *Int J Behav Nutr Phys Act*. 2023;20(1):44. doi:10.1186/s12966-023-01449-7
53. Cleary JL, Fang Y, Zahodne LB, Bohnert ASB, Burmeister M, Sen S. Polygenic Risk and Social Support in Predicting Depression Under Stress. *Am J Psychiatry*. 2023;180(2):139-145. doi:10.1176/appi.ajp.21111100
54. Santos J, Ihle A, Peralta M, et al. Associations of Physical Activity and Television Viewing With Depressive Symptoms of the European Adults. *Front Public Health*. 2021;9:799870. doi:10.3389/fpubh.2021.799870
55. Choi KW, Stein MB, Nishimi KM, et al. An Exposure-Wide and Mendelian Randomization Approach to Identifying Modifiable Factors for the Prevention of Depression. *Am J Psychiatry*. 2020;177(10):944-954. doi:10.1176/appi.ajp.2020.19111158

56. Amiri S, Mahmood N, Javaid SF, Khan MA. The Effect of Lifestyle Interventions on Anxiety, Depression and Stress: A Systematic Review and Meta-Analysis of Randomized Clinical Trials. *Healthcare*. 2024;12(22):2263. doi:10.3390/healthcare12222263
57. Grandner M. Epidemiology of sleep health. In: *Sleep and Health*. Academic Press; 2026:13-22. doi:10.1016/B978-0-443-13954-3.00019-X
58. Negeri ZF, Levis B, Sun Y, et al. Accuracy of the Patient Health Questionnaire-9 for screening to detect major depression: updated systematic review and individual participant data meta-analysis. *BMJ*. 2021;375:n2183. doi:10.1136/bmj.n2183
59. Zhang J, Yen ST. Physical Activity, Gender Difference, and Depressive Symptoms. *Health Serv Res*. 2015;50(5):1550-1573. doi:10.1111/1475-6773.12285
60. Fukukawa Y, Nakashima C, Tsuboi S, et al. Age differences in the effect of physical activity on depressive symptoms. *Psychol Aging*. 2004;19(2):346-351. doi:10.1037/0882-7974.19.2.346
61. Amaike M, Yokoyama A, Tanaka Y, et al. Incidence of depressive symptoms and their associations with lifestyle and social support networks among community-dwelling older adults: a sex-stratified longitudinal study using the JAGES study. *BioPsychoSocial Med*. 2025;19(1):20. doi:10.1186/s13030-025-00342-y
62. Zheng JW, Ai SZ, Chang SH, et al. Association between alcohol consumption and sleep traits: observational and mendelian randomization studies in the UK biobank. *Mol Psychiatry*. 2024;29(3):838-846. doi:10.1038/s41380-023-02375-7
63. Siraji MA, Spitschan M, Kalavally V, Haque S. Light exposure behaviors predict mood, memory and sleep quality. *Sci Rep*. 2023;13(1):12425. doi:10.1038/s41598-023-39636-y
64. Wang Y, Yao Z, Wang M, et al. Daily steps and sleep in adults: A systematic review and meta-analysis. *Sleep Medicine*. 2026;138:108697. doi:10.1016/j.sleep.2025.108697
65. Liang Z, Zhao S, Tian S, et al. Determinants and interrelations of exercise effects on sleep quality: A multimethod meta-analysis. *Sleep Medicine Reviews*. 2026;86:102239. doi:10.1016/j.smr.2026.102239
66. Liu L, Gou Z, Zuo J. Social support mediates loneliness and depression in elderly people. *J Health Psychol*. 2016;21(5):750-758. doi:10.1177/1359105314536941
67. Gall C von, Holub L, Pfeffer M, Eickhoff S. Chronotype-Dependent Sleep Loss Is Associated with a Lower Amplitude in Circadian Rhythm and a Higher Fragmentation of REM Sleep in Young Healthy Adults. *Brain Sci*. 2023;13(10):1482. doi:10.3390/brainsci13101482
68. Aslmyar D, Gall C von. Sleep Compensation on Work-Free days Is Associated with Better Sleep Quality. *Nat Sci Sleep*. 2025;17:3137-3148. doi:10.2147/NSS.S562192

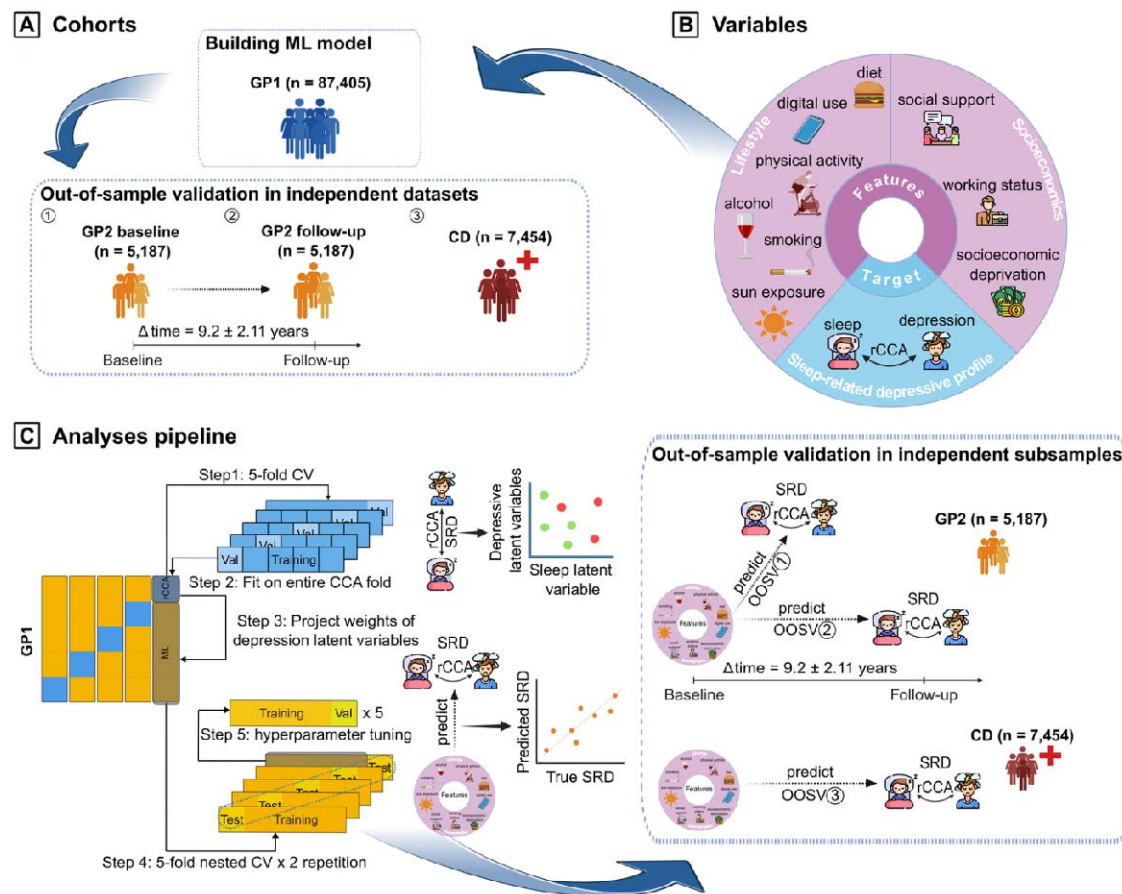


Figure 1: Overview of design and analyses. **A.** The UK Biobank cohort was divided into four independent subsamples. The largest one is for model training and testing (GP1), and three independent subsamples are used for out-of-sample validation (GP2 at baseline and at follow-up, as well as CD). **B.** Overview of variables as socioeconomic and lifestyle (input features) and sleep-related depression (SRD) profile (target) in the predictive model. **C.** rCCA identified the strongest multivariate association between SH and depression to obtain the individual-level SRD scores. Subsequently, linear and non-linear ML models were trained to predict SRD scores based on lifestyle and socioeconomic factors. The rCCA and predictive procedures were implemented in GP1 within an integrated nested cross-validation framework, with one fold for rCCA and the other four folds for prediction to avoid data leakage. Out-of-sample validation was performed on three independent samples to assess the generalizability of the predictive model trained and tested in GP1. GP: general population; CD: clinical depression; OOSV: out-of-sample validation; CV: cross-validation; rCCA: regularized canonical correlation analysis; ML: machine learning.

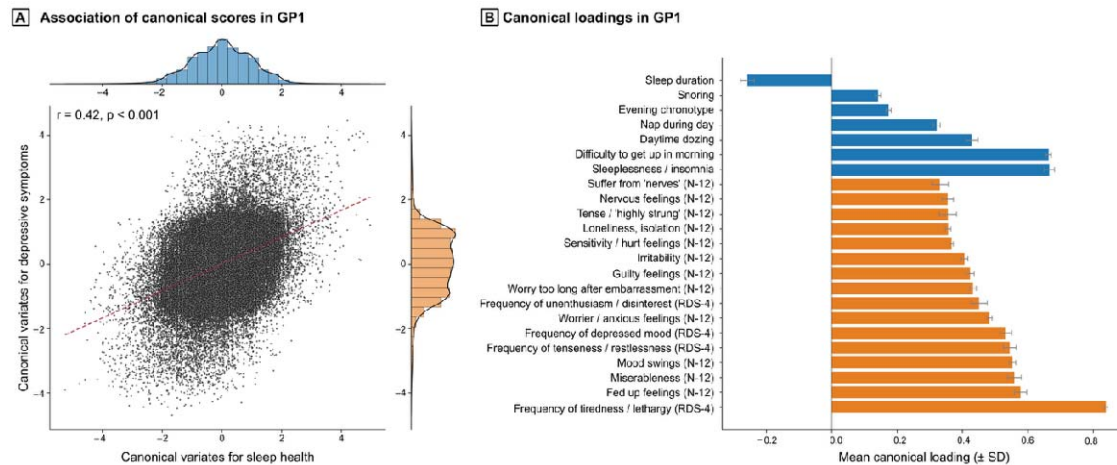


Figure 2: First significant latent dimension and loadings for sleep health (SH) and depression in the general population 1 (GP1) to calculate the sleep-related depression (SRD) profile. A. Multivariate associations between canonical variates for SH and canonical variates of depression in GP1. Each point represents an individual-level SRD score, with scores averaged across the test sets. Model performance is illustrated using the correlation coefficient (r). The blue and orange bar plots show the distributions of latent scores for SH and depression, respectively. **B.** Mean loadings across five outer CCA folds. Error bars indicate one standard deviation. Greater disturbances in SH loadings were associated with higher depression loadings. The blue bars represent SH, and the orange ones reflect depression. *N-12: neuroticism-12; RDS-4: recent depressive symptoms, indicating depressive symptoms in the past two weeks.*

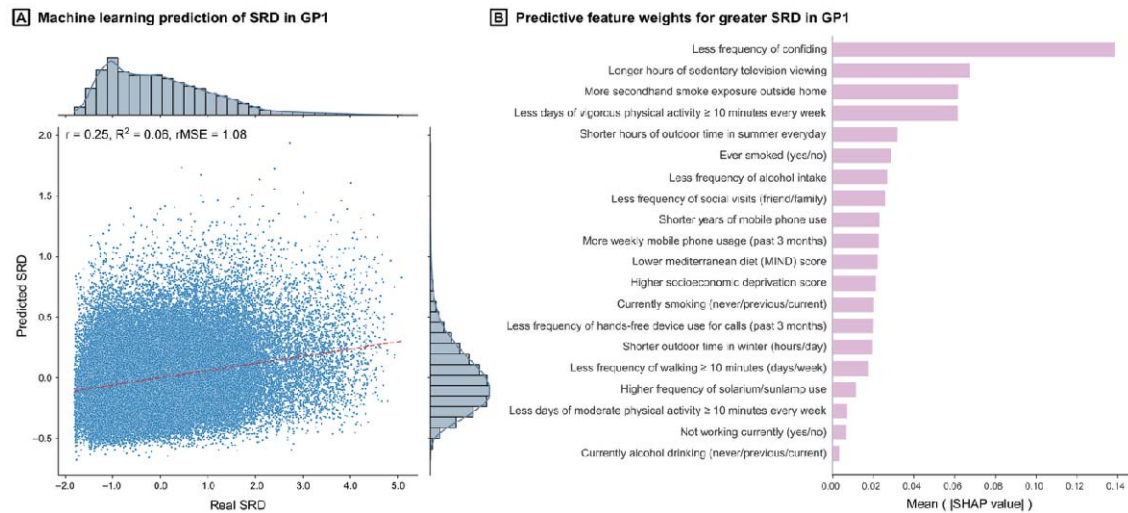


Figure 3: Predictive model performance in general population 1 (GP1). **A.** Each point represents an individual; bar plots represent the distribution of real and predicted SRD scores. Model performance is illustrated using the correlation coefficient (r), coefficient of determination (R^2), and root mean square error (rMSE). **B.** Weights for each sociodemographic and lifestyle feature in predicting higher SRD in GP1. The length of each bar represents the mean absolute SHAP values across the split test sets, reflecting the importance of each lifestyle and sociodemographic factor in predicting higher SRD.

GP: general population subsample; SRD: sleep-related depression; SHAP: SHapley Additive exPlanations.

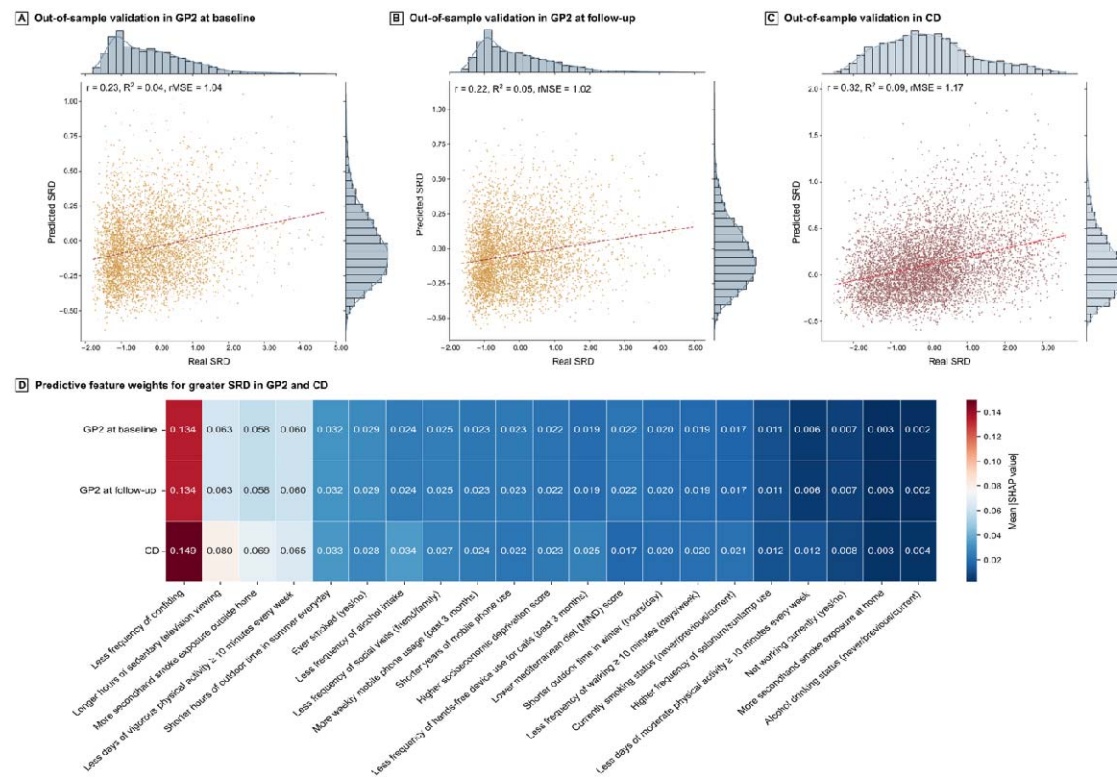


Figure 4: Out-of-sample validation of the prediction model. **A** and **B**. Out-of-sample validation performances in the independent subsamples (GP2) at baseline and follow-up. **C**. Shows out-of-sample validation performance in the CD. Each point represents an individual participant. The bar plots reflect the distribution of real and predicted SRD scores. Model performance is illustrated using the correlation coefficient (r), coefficient of determination (R^2), and root mean square error (rMSE). **D**. Weights for each feature in predicting greater SRD in the GP2 at baseline, follow-up, and CD. The colors represent the mean SHAP values for each predictor in the corresponding subsamples. GP: general population subsamples; CD: clinical depression subsamples; SRD: sleep-related depression.

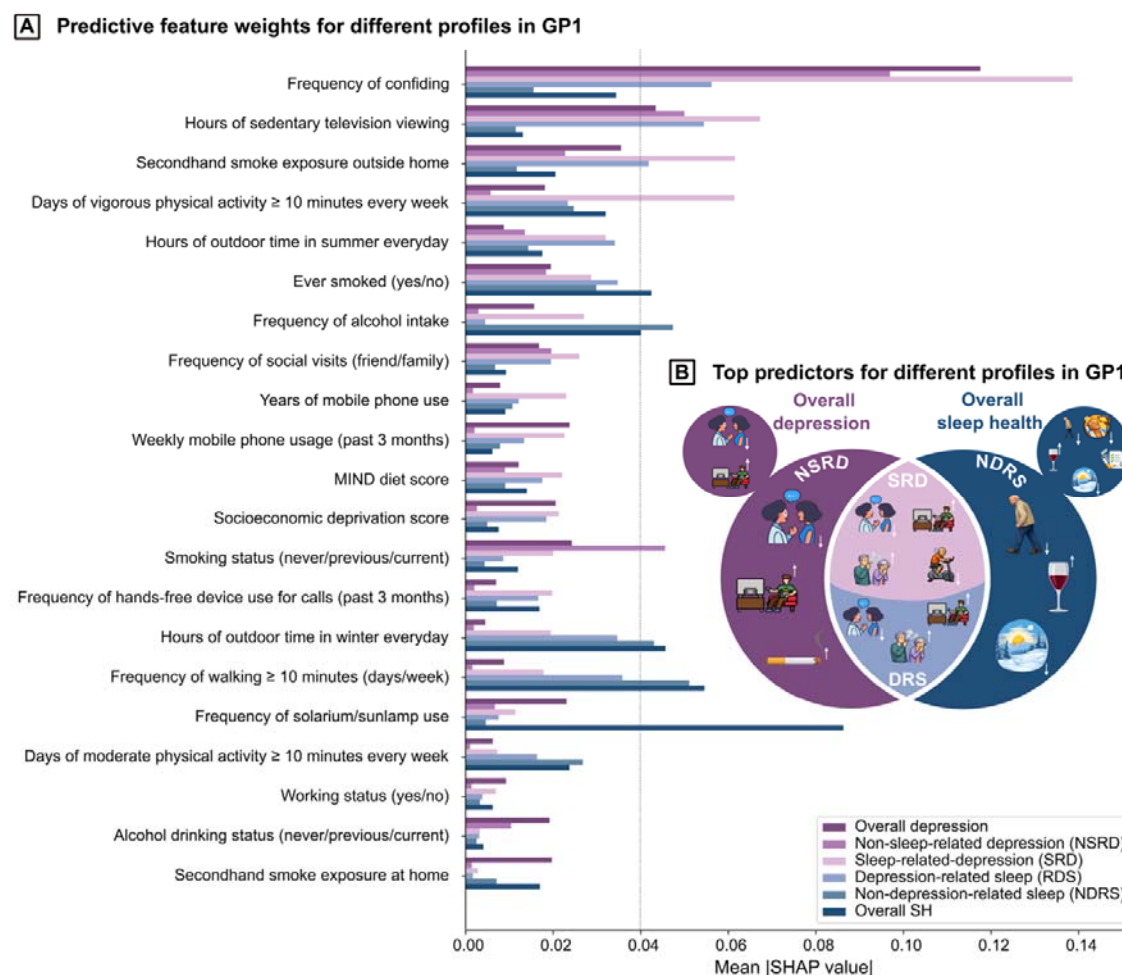


Figure 5. Predictive feature weights for six profiles in GP1. **A.** Feature weights for each sociodemographic /lifestyle factor in predicting different profiles of SH and depression in GP1. Bar length represents the mean absolute SHAP values across the split test sets, reflecting the relative importance of each sociodemographic /lifestyle factor in profile prediction. Colors indicate the different sleep and depression profiles. **B.** This schematic figure illustrates the most important predictors associated with greater symptom severity across six profiles: sleep-related depression (SRD), depression-related sleep (DRS), non-sleep-related depression (NSRD), non-depression-related sleep (NDRS), overall depression (the main dimension of variation in depression), and overall SH (the main dimension of variation in SH). Purple circles represent depression-related profiles, whereas dark blue circles denote sleep-related profiles. *SH: sleep health*.