# ARIC Manuscript Proposal Form

**ARIC Publication Admin Use Only**

Publication Committee Review Date: **08/13/24**
ARIC Manuscript Proposal Number: **# 4440**

**1.a. Full Title**: Polygenic risk scores (PRS), self-reported sleep phenotypes and association analyses with hypertension, and global cognitive function

**b. Abbreviated Title (Length 26 characters)**: Sleep-related association analyses

**2. Writing Group [please provide a middle initial if available; EX: Adam L Williams]:**

Writing group members

Tamar Sofer, Anna Wyss, Brian Spitzer, Heming Wang, Richa Saxena, Susan Redline
And by cohort:
HCHS/SOL: Wassim Tarraf, Robert Kaplan, Carmen Isasi, Melissa Lamar, Hector González, Charles DeCarli, Myriam Fornage
MESA: Jerome Rotter, Kent Taylor, Xiuqing Guo, Lekki Wood, Peter Liu, Susan Heckbert
ARIC: Alanna Morrison, Jan Bressler, Chloé Sarnowski, Bing Yu, Pamela Lutsey

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. ____ **[please confirm with your initials electronically or in writing]**

**First author [please provide a middle initial; EX: Adam L Williams]:**


| Tamar | | Sofer |
|---|---|---|
| First name | Middle initial | Last name |

Address:
Beth Israel Deaconess Medical Center
Harvard Medical School
330 Brookline Boston Ave
Boston, MA, 02215

Phone: (617) 667-7000
E-mail: tsofer@bidmc.harvard.edu


**ARIC author** to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (The ARIC author should be involved enough in ARIC to

be able to point the lead author to appropriate ancillary study PIs and to be able to search ARIC manuscript proposals if the lead author doesn't have the access needed to do such a search).

Name: Jan Bressler
Address: 1200 Pressler St, Houston TX 77030

Phone: 713-500-9919
E-mail: Jan.Bressler@uth.tmc.edu

**3. Timeline**: Work proposed as part of a funded R01: 09/2023 – 05/2028

**4. Rationale**:

There are multiple ways to measure phenotypes, whether self-reported, or objectively quantified. For example, in sleep research, insomnia typically is assessed using a questionnaire ascertaining symptoms of poor sleep quality which are used to classify insomnia status (yes/no) based on response patterns to these questionnaires. The questionnaire may differ across studies. For example, in the HCHS/SOL baseline exam the Women's Health Initiative Insomnia Rating Scale (WHIIRS) was used. The WHIIRS is composed of a set of 5 questions referring to the past 4 weeks, each with response scales from 1 ("No, not in the past 4 weeks") to 5 ("Yes, 5 or more times a week"). The questions are: Did you have trouble falling asleep? Did you wake up several times at night? Did you wake up earlier than you planned to? Did you have trouble getting back to sleep after you woke up too early? Overall, was your typical night's sleep during the past 4 weeks? (with here responses ranging from 0 "Very sound or restful", to 5 "Very restless"). In contrast, in MESA visit 4 there were three insomnia symptoms questions with a rating scale of 1 ("never") to 5 ("Almost always (16-30 times per month)"): Have trouble falling asleep, Wake up during the night and have difficulty going back to sleep, Wake up too early in the morning and be unable to get back to sleep. Another question in the same MESA questionnaire is: Have you ever been told by a doctor that you had insomnia? Similarly, there is a range of questions roughly along the same line but that are not exactly the same in other studies as well as in other exams from the same cohort studies.

Similarly, sleep duration may be self-reported as responses to various questions. In HCHS/SOL, participants were asked about their usual bed time and wake times during the week and during the weekend, allowing for specific hours and minutes format, according to which sleep duration is computed. In MESA visit 4, participants were asked: How much sleep do you usually get at night (or your main sleep period) on weekdays or workdays? with only whole numbers being used. In the WHI, the sleep duration question is phrased as: About how many hours of sleep did you get on a typical night during the past 4 weeks? And it allows for responses "5 hours or less", 6, 7, 8, 9, or "10 or more hours".

The problem of different questionnaires and instruments used to collect similar data means that sometimes we are unable to define phenotypes in the same way based on different questionnaires. Thus, if sleep duration may be categorized as "less than 5", 6, 7, 8, 9, or "more

than 10” in the Women Health Initiative (WHI), while in HCHS/SOL it is continuous, also “long sleep” (for example) may be defined differently in WHI compared to in HCHS/SOL. Further, even when two studies use the same specific instrument, like the WHIIRS, there is a disagreement about how to categorize it into “insomnia” or “no insomnia”, and researchers using data from the same study may categorize it in different ways in different analyses (e.g., all of WHIIRS ≥9,>9, ≥10 for defining insomnia have been used).

Here, we will assess the associations of polygenic risk scores (PRS) for sleep phenotypes across studies with sometimes different questions to measure sleep. PRSs are developed based on a genome-wide association study (GWAS) of a phenotype, potentially crudely measured. For example, a GWAS of excessive daytime sleepiness in the UK Biobank relied on a specific question, which may not be the optimal potential question to assess excessive daytime sleepiness. Our assumption is that PRSs, even if constructed based on a GWAS performed on crude measures, capture underlying genetically-determined phenotypes. We also note that these PRSs were already developed and constructed, and their associations with sleep phenotypes were validated in HCHS/SOL. In this manuscript, we will construct the sleep PRS across cohorts, and study their associations with sleep phenotypes harmonized across multiple cohorts: HCHS/SOL, MESA, and ARIC. In some cases, the phenotypes are measured slightly differently between cohorts.

## 5. Main Hypothesis/Study Aims:

Aims:

1. To construct PRS measures for self-reported sleep phenotypes
2. To study PRS associations with several definitions of sleep phenotypes.
3. To evaluate associations of selected sleep phenotypes with hypertension, the Digit Symbol Substitution test, and global cognitive function.
4. To evaluate the associations of sleep phenotypes and sleep PRSs with hypertension, scores of the Digit Symbol Substitution test, and global cognitive function.

This is a collaboration between three studies: HCHS/SOL, MESA, and ARIC. In the ARIC study, sleep measurements were carried out for 1,920 ARIC study participants from the Minnesota and Maryland field centers who were enrolled in the Sleep Heart Health Study (SHHS) in 1996-1998 (at about the same time as visit 4). Unattended polysomnography (PSG) was performed at home and the apnea-hypopnea index (AHI) was calculated as the number of obstructive apneas plus hypopneas per hour of sleep. The SHHS Sleep Habits Questionnaire was used to elicit information about habitual sleep duration, insomnia, and excessive daytime sleepiness (SHHS Files - Sleep Data - National Sleep Research Resource - NSRR). Other self-reported measures are available at visits 2 and 5 based on responses to questions relating to insomnia and daytime sleepiness. These is also daytime sleepiness questionnaire measured in visit 6, but we will not use it.

For each study, we will identify relevant sleep questionnaire items. We will study the association of each PRS with multiple potential definitions of the corresponding sleep phenotype. We will report the results across studies. The specific phenotype definitions may differ across studies because the specific questions may differ across studies.

Next, we will perform association analysis of genetically-harmonized binary sleep phenotypes (short and long sleep, insomnia, and excessive daytime sleepiness) with hypertension, with the Digit Symbol Substitution (DSS) test, and with global cognitive function, cross sectionally. We will also estimate the associations of sleep PRS with these phenotypes.
We may potentially perform incident association analysis with hypertension because appropriate data are available.

**6. Design and analysis - please address the following aspects:**

a) **inclusion/exclusion**
b) **study design**
c) **outcome and other variables of interest with specific reference to the time of their collection**
d) **summary of data analysis**
e) **Any anticipated methodologic limitations or challenges if present**

We already developed genome-wide sleep PRSs using PRS-CS based on multiple binary self-reported sleep phenotypes (short sleep, long sleep, insomnia, and daytime sleepiness) GWAS in European-ancestry populations (UK Biobank and FinnGen). We already showed in HCHS/SOL that the PRS based on GWAS for these phenotypes are associated with their respective phenotypes (unpublished). We will estimate the associations of these PRSs with binary sleep phenotypes in each study and visit where relevant data are available. Each study will use PLINK or PRSice to construct the PRS without further clumping. If some of the variants are missing, they will just be excluded from the PRS.

Next, each study/cohort (separately) will estimate the association of the sleep PRSs with the relevant sleep phenotypes while adjusting for age, sex, study-specific covariates such as study center, and principal components of genetic data as is standard for each study. In addition, each study/cohort will perform stratified analysis by sex, as many sleep phenotypes have different distribution in males and females.

Specific ways to define phenotypes may be discussed with each cohort according to data availability. Here is an example for each of the phenotypes:

Long sleep: we will use >8, ≥9, and >9; as well as based on "usual sleep", weekday, and average of weekday and weekend (depending on data availability).

Short sleep: we will use <6, ≤6, <7; as well as based on "usual sleep", weekday, and average of weekday and weekend (depending on data availability).

Excessive daytime sleepiness: we will use the Epworth Sleepiness Scale (ESS) when available with other questionnaire items, including "Feel excessively (overly) sleepy during the day" (MESA visit 4) with dichotomization of either (for instance, in MESA) "often" and

"always" or only "always" being defined as excessive daytime sleepiness. For the ESS, ESS>10 will be used to define excessive daytime sleepiness and we will consider alternative definitions based on specific items or definitions (e.g., having a score of 3 "high chance of nodding off" in at least 3 items) as excessive daytime sleepiness.

Insomnia: we will consider a few possible dichotomizations of the WHIIRS (WHIIRS>=9, >9, >=10). Notably, in ARIC, WHIIRS is not available, but three of the 5 WHIIRS questions are available. We plan to "scale" the sum of the relevant ARIC questions to get an equivalent scale to the full WHIIRS.

For sleep PRS-sleep trait association analysis, we will report results from each of the cohorts, and we will meta-analyze the results across cohorts.

In MESA, there are three exams with relevant questions. We will use the three exams (exam 2, 4, and 5).

We will then compare the association of the harmonized binary self-reported phenotypes (short sleep, long sleep, insomnia, and excessive daytime sleepiness) with measures that have known associations with suboptimal sleep: hypertension, and performance on the DSS cognitive test (available in HCHS/SOL, MESA, and ARIC), and meta-analyze the results across the studies. This final analysis would address the hypothesis that sleep phenotypes harmonized based on strong genetic associations in each study are more strongly associated, cross sectionally, with hypertension (defined by SBP≥130, DBP≥80, or current use of hypertensive medications), and with the DSS test, as well as with global cognitive function. We may also perform association analysis with incident hypertension. We will also perform association analysis of the sleep PRSs with these phenotypes. PRS analyses will be further adjusted for 5 genetic principal components.

Global cognitive function measure will be derived as the principal components of the three cognitive test: DSS, word frequency test, and the delayed word recall test, where we will first standardize these three tests in the sample so that none of them dominates the global cognitive function measures. We will orient the global cognitive function measure so that higher values will indicate better cognitive function.

**f) Will the author need Limited data to complete the proposed manuscript? ☒ Yes, Limited data is needed (Provide a brief (2-3 sentences) justification for requesting PHI data)** Access to genetic data will be necessary to derive the sleep PRS and evaluate association with sleep phenotypes, hypertension, DSST, and general cognitive function.

**☐ No, De-identified data will be sufficient.**

*Please note, Limited dataset access is strict and rarely provided. Limited data includes identifiable information such as dates (birthdays, visit dates, etc.). CMS, Genomic, Geocoded, Proteomics/Somalogic, and other -omic data all fall under the limited data category. De-identified data does not include dates. All dates are date adjusted to "Days since Visit 1".

**7.a. Will the data be used for non-ARIC analysis or by a for-profit organization in this manuscript? (Non-ARIC analysis means that the authors are not regarded as ARIC investigators and the "ARIC author" is essentially just a facilitator rather than an integral part of the writing group.)** ☒ **Yes** ☐ **No**

**b. If Yes, is the author aware that the current derived consent file ICTDER05 must be used to exclude persons with a value RES_OTH and/or RES_DNA = "ARIC only" and/or "Not for Profit" ?** ☒ **Yes** ☐ **No**
(The file ICTDER is distributed to ARIC PIs annually, and contains the responses to consent updates related to stored sample use for research.)

**8.a. Will the DNA data be used in this manuscript?** ☒ **Yes** ☐ **No**

**8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the current derived consent file ICTDER05 must be used to exclude those with value RES_DNA = "No use/storage DNA"?** ☒ **Yes** ☐ **No**

**9. The lead author or the "sponsoring" ARIC author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status.** ARIC Investigators have access to the publications lists under the Study Members Area of the website at: https://aric.cscc.unc.edu/aric9/proposalsearch [ARIC Website□Publications□Proposal Search]

☒ Yes ☐ No

**10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?**

03/06/2024 MS#4336 Rabinowitz JA, Associations of Accelerometry-Derived Sleep, Sedentary Behavior, and Physical Activity with Cognitive Outcomes: A Compositional Data Analysis Approach

03/31/2021 MS#3797 Pase MP, Sleep Architecture, Obstructive Sleep Apnea, and Cognitive Function in Adults.

01/13/2019 MS#3300 West NA, Cognition and 20-year subsequent sleep disturbances.

03/11/2013 MS#2096 Punjabi NM, Genetic analyses of obstructive sleep apnea and related traits in The Atherosclerosis Risk in Communities Study (ARIC)

05/02/2013 MS#2135 Lutsey PL, Obstructive Sleep Apnea and 15-Year Cognitive Decline: The Atherosclerosis Risk in Communities (ARIC) Study.

11/29/2011 MS#1877CARe Patel SR, Association of genomic loci with sleep apnea in European Americans and African-Americans: The Candidate Gene Association Resource (CARe)

04/26/2005 MS#884 Boland LL, Measures of cognitive function in persons with varying degrees of sleep-disordered breathing: the Sleep Heart Health Study.

**11.a. Is this manuscript proposal associated with any ARIC ancillary studies or does it use current [or ongoing] ancillary study data (this includes ACHIEVE)?** ☒ **Yes** ☐ **No → Skip to question 12**

**11.b. If yes to 11.a., is the proposal**

☒ **A. primarily the result of an ancillary study**

☐ **B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables)**

**11.c. If yes to 11.a., list number*** 2022.04__

*ancillary studies are listed by number
https://aric.cscc.unc.edu/aric9/researchers/ancillary_studies/approved_ancillary_studies [ARIC Website Ancillary Studies Approved Ancillary Studies]

**12a. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.**

**12b. The NIH instituted a Public Access Policy in April, 2008** which ensures that the public has access to the published results of NIH funded research. It is **your responsibility to upload manuscripts to PubMed Central** whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in https://aric.cscc.unc.edu/aric9/publications/policies_forms_and_guidelines [ARIC Website□Publications□Publication Policies, Forms, and Guidelines]. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.

References:


Mazzotti DR, Haendel MA, McMurry JA, Smith CJ, Buysse DJ, Roenneberg T, Penzel T, Purcell S, Redline S, Zhang Y, Merikangas KR, Menetski JP, Mullington J, Boudreau E. Sleep and circadian informatics data harmonization: a workshop report from the Sleep Research Society and Sleep Research Network. Sleep. 2022 Jun 13;45(6):zsac002. doi: 10.1093/sleep/zsac002. PMID: 35030631; PMCID: PMC9189941.

Zhang Y, Elgart M, Granot-Hershkovitz E, Wang H, Tarraf W, Ramos AR, Stickel AM, Zeng D, Garcia TP, Testai FD, Wassertheil-Smoller S, Isasi CR, Daviglus ML, Kaplan R, Fornage M, DeCarli C, Redline S, González HM, Sofer T. Genetic associations between sleep traits and cognitive ageing outcomes in the Hispanic Community Health Study/Study of Latinos. EBioMedicine. 2023 Jan;87:104393. doi: 10.1016/j.ebiom.2022.104393. Epub 2022 Dec 6. PMID: 36493726; PMCID: PMC9732133.

Li X, Sotres-Alvarez D, Gallo LC, Ramos AR, Aviles-Santa L, Perreira KM, Isasi CR, Zee PC, Savin KL, Schneiderman N, Wassertheil-Smoller S, Sofer T, Daviglus M, Redline S. Associations of Sleep-disordered Breathing and Insomnia with Incident Hypertension and Diabetes. The Hispanic Community Health Study/Study of Latinos. Am J Respir Crit Care Med. 2021 Feb 1;203(3):356-365. doi: 10.1164/rccm.201912-2330OC. PMID: 32758008; PMCID: PMC7874314.

Quan SF, Howard BV, Iber C, Kiley JP, Nieto FJ, O'Connor GT, Rapoport DM, Redline S, Robbins J, Samet JM, Wahl PW. The Sleep Heart Health Study: design, rationale, and methods. Sleep. 1997 Dec; 20 (12):1077-1085. PMID: 9493915

T Ge, CY Chen, Y Ni, YCA Feng, JW Smoller. Polygenic Prediction via Bayesian Regression and Continuous Shrinkage Priors. Nature Communications, 10:1776, 2019. PMID: 30992449, PMCID: PMC6467998.

Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic risk score software. Bioinformatics. 2015;31:1466–8. PMID: 25550326, PMCID: PMC4410663.

Ramos AR, Tarraf W, Daviglus M, Davis S, Gallo LC, Mossavar-Rahmani Y, Penedo FJ, Redline S, Rundek T, Sacco RL, Sotres-Alvarez D, Wright CB, Zee PC, González HM. Sleep Duration and Neurocognitive Function in the Hispanic Community Health Study/Study of

Latinos. Sleep. 2016 Oct 1;39(10):1843-1851. doi: 10.5665/sleep.6166. PMID: 27450689; PMCID: PMC5020366.

Scammell BH, Tchio C, Song Y, Nishiyama T, Louie TL, Dashti HS, Nakatochi M, Zee PC, Daghlas I, Momozawa Y, Cai J, Ollila HM, Redline S, Wakai K, Sofer T, Suzuki S, Lane JM, Saxena R. Multi-ancestry genome-wide analysis identifies shared genetic effects and common genetic variants for self-reported sleep duration. Hum Mol Genet. 2023 Sep 5;32(18):2797-2807. doi: 10.1093/hmg/ddad101. PMID: 37384397; PMCID: PMC10656946.

**To view publications materials, click "Log in" at the top right of the ARIC website. Click "Forgot Password" if you are experiencing issues with logging in.**