



ORIGINAL ARTICLE

Polygenic risk score identifies associations between sleep duration and diseases determined from an electronic medical record biobank

Hassan S. Dashti^{1,2}, Susan Redline³ and Richa Saxena^{1,2,4,*}

¹Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, ²Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, ³Departments of Medicine, Brigham and Women's Hospital and Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA and ⁴Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, MA

*Corresponding author. Richa Saxena, Center for Genomic Medicine, Massachusetts General Hospital, 185 Cambridge Street, CPZN 5.806, Boston, MA 02114. Email: rsaxena@broadinstitute.org.

Abstract

Study Objectives: We aimed to detect cross-sectional phenotype and polygenic risk score (PRS) associations between sleep duration and prevalent diseases using the Partners Biobank, a hospital-based cohort study linking electronic medical records (EMR) with genetic information.

Methods: Disease prevalence was determined from EMR, and sleep duration was self-reported. A PRS for sleep duration was derived using 78 previously associated SNPs from genome-wide association studies (GWAS) for self-reported sleep duration. We tested for associations between (1) self-reported sleep duration and 22 prevalent diseases ($n = 30251$), (2) the PRS and self-reported sleep duration ($n = 6903$), and (3) the PRS and the 22 prevalent diseases ($n = 16033$). For observed PRS-disease associations, we tested causality using two-sample Mendelian randomization (MR).

Results: In the age-, sex-, and race-adjusted model, U-shaped associations were observed for sleep duration and asthma, depression, hypertension, insomnia, obesity, obstructive sleep apnea, and type 2 diabetes, where both short and long sleepers had higher odds for these diseases than normal sleepers ($p < 2.27 \times 10^{-3}$). Next, we confirmed associations between the PRS and longer sleep duration (0.65 ± 0.19 SD minutes per effect allele; $p = 7.32 \times 10^{-6}$). The PRS collectively explained 1.4% of the phenotypic variance in sleep duration. After adjusting for age, sex, genotyping array, and principal components of ancestry, we observed that the PRS was also associated with congestive heart failure (CHF; $p = 0.015$), obesity ($p = 0.019$), hypertension ($p = 0.039$), restless legs syndrome (RLS; $p = 0.041$), and insomnia ($p = 0.049$). Associations were maintained following additional adjustment for obesity status, except for hypertension and insomnia. For all diseases, except RLS, carrying a higher genetic burden of the 78 sleep duration-increasing alleles (i.e. higher sleep duration PRS) associated with lower odds for prevalent disease. In MR, we estimated causal associations between genetically defined longer sleep duration with decreased risk of CHF (inverse variance weighted [IVW] OR per minute of sleep [95% CI] = 0.978 [0.961–0.996]; $p = 0.019$) and hypertension (IVW OR [95% CI] = 0.993 [0.986–1.000]; $p = 0.049$), and increased risk of RLS (IVW OR [95% CI] = 1.018 [1.000–1.036]; $p = 0.045$).

Conclusions: By validating the PRS for sleep duration and identifying cross-phenotype associations, we lay the groundwork for future investigations on the intersection between sleep, genetics, clinical measures, and diseases using large EMR datasets.

Statement of Significance

Recent implementations of electronic medical records (EMR) with genetic data across the healthcare sector have enabled advancement of clinical research. However, these approaches remain underutilized in sleep research likely due to the paucity of sleep information. We tested polygenic risk score (PRS) for sleep duration using 78 signals identified for self-reported sleep duration. We verified associations between the PRS and self-reported sleep duration in a hospital-based population and observed associations of higher PRS with lower odds of EMR-derived congestive heart failure, obesity, hypertension, and insomnia and higher odds of restless legs syndrome. This study demonstrates the feasibility and power of using genetic markers of sleep in clinical cohorts, laying the groundwork for future investigations on sleep, genetics, clinical measures, and diseases using large EMR datasets.

Key words: sleep duration; polygenic risk score; cardiometabolic diseases; psychiatric disorders; epidemiology

Submitted: 9 August, 2018; Revised: 7 November, 2018

© Sleep Research Society 2018. Published by Oxford University Press on behalf of the Sleep Research Society. All rights reserved. For permissions, please e-mail journals.permissions@oup.com.

Introduction

Sleep duration is a complex phenotype driven by genetic and lifestyle factors [1]. Prospective epidemiologic studies have indicated that deviating from the recommended sleep duration of 7–8 hours is associated with increased risk for various diseases [2]. Indeed, U-shaped relationships have been observed between habitual self-reported short (<6, 7 hours per night) and long (>8, 9 hours per night) sleep duration and cognitive and psychiatric, metabolic, cardiovascular, and immunological dysfunction as well as all-cause mortality, compared to sleeping 7–8 hours per night [3–9]. These observations have been made predominantly in population-based cohort studies relying on a single estimate of sleep duration at baseline via questionnaires [9–12].

Recent implementations of electronic medical/health records (EMR) across the healthcare sector have enabled significant advancement of clinical research through linkage of large and diverse medical data enriched for disease states with DNA biorepositories and lifestyle surveys, as has been implemented in the Partners Biobank [13] and eMERGE Network [14]. Robust algorithms using a combination of codified and narrative EMR data are applied to define disease states, enabling systematic comparisons with clinical measures and genetic variants [15, 16]. Although a promising resource for health services and cost-effectiveness research, EMR are currently underutilized in sleep research, likely as a result of a paucity of sleep assessment or polysomnography information associated with EMR data.

As EMR are increasingly linked to genetic data collected on patients within health care systems, the genetic data can be used to derive genetic instruments, such as polygenic risk scores (PRSs) [17], to approximate the genetic architecture of a trait and dissect lifestyle from genetic exposures. PRS provide a quantitative measure of an aggregated genetic burden of disease in each person and are powerful tools to validate genetic links to disease and to dissect pleiotropic or causal associations across traits [17–19]. For example, in coronary artery disease, a trait with a sizable genetic component, PRSs comprised of top genetic signals identified from genome-wide association studies (GWAS) or aggregating variants spanning the entire genome have been shown to be valuable in disease risk prediction [20, 21]. Sleep duration is also a heritable trait, and twin- and family-based studies have estimated that 9% to 45% of variability in self-reported sleep duration is influenced by genetic factors [22, 23]. Recent GWAS in up to 446 118 participants from the UK Biobank identified 78 signals for habitual self-reported sleep duration [23]. Here, we postulated that the PRS comprised of these signals could serve as a viable marker for sleep duration, providing estimates of the likelihood of a sleep phenotype that is not limited by sampling biases related to surveys and possibly also reflecting long-term sleep duration exposure.

Thus, in this study from the Partners Biobank, we used EMR and genetic data supplemented with a sleep habits questionnaire to (1) test cross-sectional associations between self-reported sleep duration and 22 prevalent diseases determined from EMR (19 from narrative data; three from codified data), (2) validate a PRS comprised of 78 signals associated with self-reported sleep duration from GWAS, and (3) systematically investigate whether the PRS associates with any of the 22 prevalent diseases. We further derive a genome-wide sleep duration PRS, and for observed PRS-disease associations, we test causality using two-sample Mendelian randomization (MR).

Methods

Partners healthcare biobank

The Partners Biobank is a hospital-based cohort study from the Partners HealthCare hospitals with EMR and genetic data supplemented with electronic health and lifestyle surveys [13]. Recruitment for the Partners Biobank launched in 2010 and is active at participating clinics at Brigham and Women's Hospital (BWH), Massachusetts General Hospital (MGH), Spaulding Rehabilitation Hospital (SRH), Faulkner Hospital (FH) and McLean Hospital (MCL), Newton-Wellesley Hospital (NWH), North Shore Medical Center (NSMC). All patients provided consent upon enrollment. To date (February 2018), a total of 78 726 participants were consented. The current analysis was restricted to 43 058 adults ≥ 18 years with either self-reported sleep and/or high-quality genotyping with EMR data.

Disease status ascertainment

Disease prevalence was determined from EMR using both structured and unstructured data [13]. Natural Language Processing (NLP) was used to extract data from narrative text including coded diagnoses, medications, procedures, and vital signs, as previously described [24], to screen enrichment/frequencies of predictive disease features, such as comorbidities or symptoms, identified from Wikipedia and Medscape articles using an automated feature extraction protocol (AFEP). The feature set is narrowed to those that are most relevant using the adaptive least absolute shrinkage and selection operator (LASSO) procedure, and a gold-standard set of patients is used to train the model to accurately predict the phenotype based on the refined definition. Thus, for each disease phenotype, algorithm development included the following steps: (1) filtering Biobank participants by presence of the ICD9-CM billing code for each disease; (2) randomly selecting 100 participants with each code; (3) chart reviews by board-certified clinician to define disease status in a Training Set ([Supplementary Methods](#)); (4) automated feature extraction and feature selection to EMR narrative text; (5) LASSO penalized regression with features predicting disease status in Training Set; and (6) applying the algorithm to remaining participants to define disease phenotype set. The process produced robust phenotype algorithms that were evaluated using metrics such as positive predictive value (PPV), the proportion of individuals classified as cases by the algorithm, and True Positive Rates (TPR), which reflects the sensitivity or the proportion of true positives correctly identified as such. A total of 19 disease phenotypes were determined from EMR using this approach: asthma (PPV = 90%; TPR = 0.761; $n = 1886$), bipolar disorder (PPV = 89%; TPR = 0.200; $n = 178$), breast cancer (PPV = 90%; TPR = 0.963; $n = 2302$), chronic obstructive pulmonary disorder (PPV = 90%; TPR = 0.434; $n = 400$), congestive heart failure (CHF; PPV = 90%; TPR = 0.880; $n = 594$), coronary artery disease (CAD; PPV = 99%; TPR = 0.970; $n = 3618$), Crohn's disease (PPV = 90%; TPR = 0.960; $n = 755$), depression (PPV = 90%; TPR = 0.805; $n = 4651$), epilepsy (PPV = 90%; TPR = 0.932; $n = 1167$), gout (PPV = 90%; TPR = 0.931; $n = 1886$), hypertension (PPV = 90%; TPR = 0.952; $n = 16569$), multiple sclerosis (MS; PPV = 90%; TPR = 0.810; $n = 368$), obesity (PPV = 90%; TPR = 0.870; $n = 13102$), rheumatoid arthritis (RA; PPV = 90%; TPR = 0.905; $n = 1161$), schizophrenia (PPV = 90%; TPR = 0.832; $n = 53$), stroke (PPV = 90%;

TPR = 0.623; $n = 511$), type 1 diabetes (T1D; PPV = 90%; TPR = 0.784; $n = 271$), type 2 diabetes (T2D; PPV = 99%; TPR = 0.880; $n = 3106$), and ulcerative colitis (UC; PPV = 90%; TPR = 0.660; $n = 561$). In addition, three sleep-related disease phenotypes were determined from codified data using physician diagnoses ICD-10 codes: insomnia (G47.0; $n = 5091$), restless legs syndrome (RLS; G25.81; $n = 1016$), and obstructive sleep apnea (G47.3; $n = 6247$); however, these diagnoses were not validated by chart review.

Sleep duration and covariate measures

Study participants were invited following enrollment to self-report information regarding their lifestyle, environment, and family history via the Partners Biobank Health Information Survey, an optional online survey. Sleep was assessed by the questions “In considering your longest sleep period, what time do you usually go to bed on WEEKDAYS or WORK or SCHOOL days?” and “In considering your longest sleep period, what time do you usually wake up on WEEKDAYS or WORK or SCHOOL days?” Similar questions were asked for “WEEKENDS or DAYS OFF.” Responses were in half-hour increments. To date, a total of 31221 participants have completed the sleep survey. Weighted average weekly sleep duration was computed using self-reported weekday and weekend bed and wake-up times with 5/7 weighting for weekdays and 2/7 for weekends. Sleep duration was also categorized as short (<7 hours per night), normal (7 to <9 hours per night), and long (≥ 9 hours per night).

Age, sex, and race data were obtained from EMR. Study participants further self-reported height, weight, alcohol intake, employment status, exercise, and smoking status. Smoking was assessed by the question, “Have you smoked at least 100 cigarettes in your lifetime?” and categorized into “Yes, currently smoke,” “Yes, smoked in the past, but quit,” or “No.” Alcohol was assessed by the question, “During the past year, how many alcoholic drinks (glass/bottle/can of beer; 4oz glass of wine; drink or shot of liquor) did you usually drink in a typical week?” and categorized as the following: None, or less than 1 per month; 1–3 per month; 1 per week; 2–4 per week; 5–6 per week; 1–2 per day; 3–4 per day; 5–6 per day; More than 6 per day. Exercise was assessed by the question, “During the past year, what was your average time spent per week at each of the following recreational activity:” for eight activities, and total moderate to high-intensity exercise (excluding walking/hiking) was estimated in hours per week. Employment was assessed by the question, “Which of the following best describes your usual work schedule?” with the following response options: day shift, afternoon shift, night shift, split shift, irregular shift/on-call, rotating shifts or do not work. Responses were collapsed to day shift, do not work, or non-day shift (all remaining options). Education was assessed by the question, “What is the highest grade in school that you finished?” with the following options: Grade school (1–4 years); Grade school (5–8 years); Some high school (9–11 years); High school diploma or GED (finished high school); Some college; 2-year college or vocational school; 4-year college; Masters, doctoral, or professional degree. Weight and height were self-reported and body mass index (BMI) was calculated as weight (kg) / height² (m²). Missing covariates (all <5% missing) were imputed by using sex-specific median values for continuous variables (i.e. BMI, alcohol intake, exercise, and Charlson Index), or using a missing indicator approach for categorical variables (i.e. smoking, employment).

Genetic data genotyping, imputation and quality control and generation of PRS

DNA from participants was genotyped using ~1.6 million SNPs on the Illumina Multi-Ethnic GWAS/Exome SNP Array. Imputation was performed using Minimac3 [25] using the HRC (Version r1.1 2016) reference panel [26] for imputation. This HRC panel consists of 64940 haplotypes of predominantly European ancestry. Haplotype phasing was performed using SHAPEIT [27]. To date, 20038 participants have been genotyped, of which 7982 also completed the sleep survey. Participant ancestry was determined using TRACE [28] and the Human Genome Diversity Project (HGDP) [29] as a reference panel. To correct for population stratification, we further computed principal components (PCs) using the same software in the subset with genetically European ancestry. Next, we derived a PRS for self-reported sleep duration using 78 SNPs previously reported to be associated at the genome-wide significance level in the UK Biobank ($p < 5 \times 10^{-8}$) [23]. All SNPs had a minor allele frequency (MAF) > 1% and an imputation quality (minimac rsq) ≥ 0.70 (Supplementary Table S1). Individual participant scores were created by summing the number of risk alleles (allele associated with longer sleep duration) at each genetic variant, which were weighted by the respective allelic effect sizes on longer sleep duration in the UK Biobank. Additionally, we generated a genome-wide PRS for each individual by summing sleep increasing risk alleles across the genome, each weighted by the beta estimate for that allele from the sleep duration GWAS. We included 1201079 SNPs after excluding X chromosome variants and, at each site, clumped SNPs based on association p value (the variant with the smallest p value within a 250kb range was retained and all those in linkage disequilibrium, $r^2 > 0.1$, were removed). LD clumping and genome-wide PRS generation were done using PRSice [30], and the best fit genome-wide sleep duration PRS encompassed 2096 SNPs at p value threshold = 0.00015 (Supplementary Figure S1).

Statistical analysis

The relationship between sleep duration and prevalent diseases were examined using categorical logistic regression with adjustments for age, sex, and race (model 1); then further adjusted for BMI, other than obesity outcome (model 2); and then further adjusted for additional established risk factors: alcohol intake, Charlson comorbidity index, education, employment status, exercise, and smoking status (model 3). Consistent with previous reports, improbable sleep durations <3 or >14 hours [31] ($n = 968$; resulting primarily from am/pm coding errors by participants) and improbable BMI (>100 kg/m²; $n = 1$) were excluded. Final total sample size for association analyses between estimated sleep duration and prevalent diseases was 30251. Phenotypic associations were considered significant at the Bonferroni correction accounting for 22 diseases ($p < 2.27 \times 10^{-3}$ [=0.05/22]).

For genetic analyses, samples with high-quality genotyping and EMR data were included ($n = 20038$). Participants of non-European ancestry based on genetic ancestry were excluded from all genetic analysis to minimize the influence of differences in population structure ($n = 4005$ removed). Independent SNP associations with self-reported sleep duration were conducted in the genetic subset with self-reported sleep duration and tested using linear trends and an additive genetic model adjusted for age, sex, genotyping array, and 5 PCs of ancestry ($n = 7155$).

Independent SNP associations were considered significant at the Bonferroni-adjusted threshold accounting for 78 tests ($p < 6.4 \times 10^{-04}$). Post hoc power calculations for the independent SNP associations were conducted using Quanto version 1.2.4 (<http://biostats.usc.edu/Quanto.html>). Validation of the PRS was conducted in the genetic subset with self-reported sleep duration ($n = 6903$) by estimating linear trends of the PRS adjusted for age, sex, genotyping array, and five PCs of ancestry, and separately for weekday ($n = 6805$) and weekend ($n = 6481$) self-reported sleep duration. Systematic PRS association analyses with all 22 prevalent diseases were conducted in the full genetic dataset ($n = 16033$), and further adjusted for obesity status determined from EMR. PRS associations were considered significant at $p < 0.05$. For all significant PRS associations ($p < 0.05$), we assessed for associations between quartiles of PRS and odds of disease prevalence. Similarly, validation of the genome-wide PRS was conducted with self-reported sleep duration and tested for associations with the five diseases significant in the 78 SNP PRS-disease association analyses.

Mendelian randomization

Two-sample MR [32] was carried out using MRCIEU/TwoSampleMR package in R, using the inverse variance weighted (IVW) approach as our main analysis method [33], and MR-Egger [34] and weighted median estimation [35] as sensitivity analyses, as previously described [23] (Supplementary Figure S2). For our two-sample MR analyses, for all 78 signals for habitual self-reported sleep duration [23], we looked for the per allele difference in disease outcomes (those showing associations with sleep duration PRS) from the Partners Biobank. Results are therefore a measure of genetically “longer sleep duration”. Sample 1 is UK Biobank and sample 2 the disease phenotype from Partners Biobank.

Results

In a sample of 43058 adult participants (56.95 [16.84] years; 57.6% female) from the Partners Biobank, prevalent cases of diseases determined from EMR ranged from 0.12% for schizophrenia to 38.5% for hypertension. Sleep data were available for a subset of 30251 (70.3%) participants. Self-reported sleep duration was normally distributed with a mean of 8.24 hours ($SD = 1.38$ hours) per night (Table 1), and 11.9% (<7 hours per night; $n = 3604$), 63.1% (7 to <9 hours per night; $n = 19084$), and 25.0% (≥ 9 hours per night; $n = 7563$) were short, normal and long sleepers, respectively. Weekday self-reported sleep duration was shorter than average weekend self-reported sleep duration (weekday = 7.93 [1.30] hours per night vs. weekend = 9.05 [2.80] hours per night). Overall, short sleepers were more likely to be male and non-White, had higher BMI, consumed less alcohol, more likely to be current smokers and night workers, and had less education (all $p < 0.001$). The genetic subset, which was limited to participants of European ancestry, included 16033 participants and was significantly older (60.35 [16.51] years; $p < 0.05$) and consisted of more males (53.8% female; $p < 0.05$), compared to the subset who had self-reported sleep data.

In the age-, sex-, and race-adjusted model (model 1), U-shaped associations were observed for sleep duration and asthma, depression, hypertension, insomnia, obesity,

obstructive sleep apnea, and type 2 diabetes, where both short and long sleepers had higher odds for these diseases than normal sleepers (Table 2). In addition, compared to normal sleepers, short sleepers had higher odds for chronic obstructive pulmonary disorder, gout, RLS, and type 1 diabetes, whereas long sleepers had higher odds for bipolar disorder and epilepsy. Overall, observed disease associations for short and long sleep were attenuated when further adjusted for BMI (model 2) and several additional established risk factors (model 3) (Table 2, Supplementary Table S2). In model 3, U-shaped associations remained significant for insomnia and obstructive sleep apnea, short sleep associations remained significant for obesity and type 1 diabetes, and long sleep associations remained significant for bipolar disorder and depression.

Among participants of European ancestry with genetic data ($n = 16033$), we generated a combined weighted PRS of the 78 self-reported sleep duration signals (Supplementary Table S1). In participants with self-reported sleep data ($n = 6903$), the PRS was associated with a 0.65 (0.19) minutes longer sleep per effect allele ($p = 7.32 \times 10^{-04}$) (Table 3). The PRS showed stronger associations with weekday ($p = 3.19 \times 10^{-04}$) than weekend ($p = 0.07$) self-reported sleep duration. The 5% of participants carrying most of the 78 sleep duration-increasing alleles had an estimated 8.4 minutes longer sleep duration compared to the 5% carrying the fewest. The 78 independent variants collectively explained ~1.4% of the phenotypic variance in estimated sleep duration. As expected, we did not observe association of the individual signals with sleep duration (all $p > 6.4 \times 10^{-04}$; Supplementary Table S1) likely due to low statistical power (all $p < 0.80$). Using genome-wide PRS, we also observed significant associations with sleep duration ($p = 3.34 \times 10^{-08}$), whereby the 5% of participants carrying most of the genome-wide sleep duration-increasing alleles had an estimated 28.2 minutes longer sleep duration compared to the 5% carrying the fewest.

Using the PRS as a marker for sleep duration, we investigated whether sleep duration associated with 22 prevalent diseases curated from EMR among all 16033 participants with genetic data. After adjusting for age, sex, genotyping array, and PCs of ancestry, we observed that the PRS was associated with CHF ($p = 0.015$), obesity ($p = 0.019$), hypertension ($p = 0.039$), RLS ($p = 0.041$), and insomnia ($p = 0.049$; Table 4). For all diseases, except RLS, carrying more of the 78 sleep duration-increasing alleles (i.e. higher PRS), associated with lower odds for prevalent disease, whereas for RLS, carrying more of the 78 sleep duration-increase alleles associated with higher odds for prevalence disease. Compared to the lowest quartile of PRS, the highest quartile of PRS was associated with a 29.7% lower odds for CHF ($p = 0.01$), 11.3% lower odds for hypertension ($p = 0.02$), 10.6% lower odds for insomnia ($p = 0.09$), 9.7% lower odds for obesity ($p = 0.03$), and 23.2% higher odds for RLS ($p = 0.13$; Table 4). When adjusted for obesity status, PRS associations remained significant and point estimates were consistent for CHF ($p = 0.017$) and RLS ($p = 0.031$; Table 3). In models unadjusted or adjusted for obesity, we observed that the genome-wide PRS associated with CHF ($p = 2.86 \times 10^{-04}$), but none of the other disease phenotypes (all $p < 0.05$). In MR analyses between sleep duration and PRS associated diseases using the 78 SNP PRS as the genetic instrument, we estimated causal associations between genetically defined longer sleep duration with decreased risk of CHF (IVW OR per minute of sleep [95% CI] = 0.978 [0.961–0.996]; $p = 0.019$) and hypertension (IVW OR [95% CI] = 0.993 [0.986–1.000]; $p = 0.049$),

Table 1. General characteristics of Partners Biobank participants with self-reported sleep duration estimates (n = 30251)

Characteristic	All 30251	Sleep duration (hours/day)			
		<7 hours 3604 (11.9%)	7 to <9 hours 19084 (63.1%)	≥9 hours 7563 (25.0%)	
Age, years	56.23 (16.48)	56.49 (14.70)	55.77 (16.16)	57.28 (17.97)	<0.001
Sex, female	18061 (59.7)	1869 (51.9)	11348 (59.5)	4844 (64.0)	<0.001
Race, n (%)					<0.001
Asian	713 (2.4)	106 (2.9)	470 (2.5)	137 (1.8)	
Black	780 (2.6)	197 (5.5)	397 (2.1)	186 (2.5)	
White	27010 (89.3)	3031 (84.1)	17190 (90.1)	6789 (89.8)	
Other	296 (1.0)	53 (1.5)	165 (0.9)	78 (1.0)	
Unknown	1452 (4.8)	217 (6.0)	862 (4.5)	373 (4.9)	
Sleep duration, hours					
Average	8.24 (1.38)	6.13 (0.80)	7.94 (0.52)	9.99 (1.08)	<0.001
Weekday	7.93 (1.30)	5.90 (0.93)	7.77 (0.62)	9.35 (1.19)	<0.001
Weekend	9.05 (2.80)	6.81 (1.49)	8.41 (0.98)	11.72 (4.17)	<0.001
BMI, kg/m ²	27.28 (6.01)	28.59 (6.36)	27.00 (5.80)	27.37 (6.25)	<0.001
Charlson Index	4.83 (2.56)	4.95 (2.51)	4.70 (2.54)	5.11 (2.61)	<0.001
Alcohol, g/day	3.21 (1.94)	2.95 (1.95)	3.31 (1.90)	3.07 (2.00)	<0.001
Exercise, hours/week	2.64 (3.73)	2.43 (3.93)	2.83 (3.78)	2.25 (3.45)	<0.001
Smoking status, n (%)					<0.001
Never	17600 (58.2)	1952 (54.2)	11512 (60.3)	4136 (54.7)	
Past	11144 (36.8)	1358 (37.7)	6782 (35.5)	3004 (39.7)	
Current	1410 (4.7)	285 (7.9)	736 (3.9)	389 (5.1)	
Other	97 (0.3)	9 (0.2)	54 (0.3)	34 (0.4)	
Education, n (%)					<0.001
Grade school, high school diploma, or GED	2692 (8.9)	460 (12.8)	1358 (7.1)	874 (11.6)	
Some college, 2-year college, or vocational school	6129 (20.3)	881 (24.4)	3497 (18.3)	1751 (23.2)	
Four-year college	9427 (31.2)	1045 (29.0)	6046 (31.7)	2336 (30.9)	
Masters, doctoral, or other professional degrees	12003 (39.7)	1218 (33.8)	8183 (42.9)	2602 (34.4)	
Work schedule, n (%)					<0.001
Day worker	16616 (54.9)	1975 (54.8)	11555 (60.5)	3086 (40.8)	
Irregular	1151 (3.8)	175 (4.9)	677 (3.5)	299 (4.0)	
Rotating	659 (2.2)	99 (2.7)	385 (2.0)	175 (2.3)	
Afternoon	555 (1.8)	64 (1.8)	284 (1.5)	208 (2.8)	
Night	377 (1.2)	101 (2.8)	166 (0.9)	110 (1.5)	
Split	197 (0.7)	34 (0.9)	114 (0.6)	49 (0.6)	
Unemployed	9317 (30.8)	947 (26.3)	5050 (26.5)	3320 (43.9)	
Other	1378 (4.6)	209 (5.8)	853 (4.5)	316 (4.2)	

and increased risk of RLS (IVW OR [95% CI] = 1.018 [1.000–1.036]; $p = 0.045$), with consistent effect direction in sensitivity analyses across other MR methods (Table 5).

Discussion

Leveraging an EMR dataset, the Partners Biobank emerged as a promising tool to detect underlying phenotypic and genetic relationships between sleep duration and prevalent diseases. In our analysis of up to 43058 adult participants in the Partners Biobank with EMR, genetic data and/or self-reported sleep duration, we (1) showed that self-reported sleep duration associates with prevalent diseases identified using the EMR, in agreement with earlier findings for sleep duration or sleep disorders [9–11, 36–40], and several of these associations remained significant even following adjustment for BMI and other established risk factors; (2) validated a sleep duration PRS of 78 self-reported sleep duration genetic signals, and a genome-wide PRS, initially derived in the UK Biobank, by confirming associations with self-reported sleep duration in an independent hospital-based cohort; and (3) observed novel associations

between the PRS and five prevalent diseases determined from EMR indicating that sleep duration-increasing alleles associate with lower odds for CHF, obesity, hypertension, and insomnia, and higher odds for RLS, with obesity independent effects for associations with CHF and RLS, and causal evidence for CHF, hypertension, and RLS.

Validation of the PRS based on a focused set of SNPs or genome-wide data with self-reported sleep duration in the Partners Biobank supports the use of this score as an instrument to approximate sleep duration in further explorations. Confirmation was observed despite fundamental differences between the Partners Biobank and the discovery cohort (UK Biobank [23]), including differences in population demographics (hospital-based cohort vs. healthy population-based cohort) and ascertainment of self-reported sleep duration (based on weighted weekday/weekend sleep/wake times vs. single question on habitual sleep duration per 24 hours, including naps). The per allele effect of the association in the Partners Biobank of 0.65 (0.19) minutes longer sleep is similar to that estimated from the CHARGE consortium (0.66 minutes per allele), a meta-analysis of population-based studies [23], but

Table 2. Phenotypic associations of average weekly self-reported sleep duration with 22 prevalent diseases curated from Electronic Medical Records in the Partners Biobank (n = 30251), with additional adjustment for BMI (model 2)

	Age-, sex-, and race-adjusted (model 1)				BMI-adjusted (model 2) ^a					
	Short (<7 hours) n = 3604		Normal (7–9 hours) n = 19084		Short (<7 hours) n = 3604		Normal (7–9 hours) n = 19084			
	OR [95% CI]	P value	OR [95% CI]	P value	OR [95% CI]	P value	OR [95% CI]	P value		
Asthma (n = 2780)	1.34 [1.19–1.50]	1.53 × 10⁻⁰⁶	(ref)	1.17 [1.06–1.28]	9.67 × 10⁻⁰⁴	1.23 [1.09–1.38]	8.36 × 10⁻⁰⁴	(ref)	1.14 [1.04–1.25]	5.85 × 10⁻⁰³
Breast cancer (n = 1853)	0.82 [0.69–0.97]	0.02	(ref)	0.94 [0.84–1.05]	0.25	0.84 [0.7–1.00]	0.05	(ref)	0.94 [0.84–1.05]	0.29
Bipolar disorder (n = 74)	1.24 [0.54–2.81]	0.61	(ref)	2.80 [1.73–4.53]	2.70 × 10⁻⁰⁵	1.20 [0.53–2.74]	0.66	(ref)	2.78 [1.72–4.50]	3.16 × 10⁻⁰⁵
Chronic obstructive pulmonary disorder (n = 142)	2.08 [1.32–3.27]	1.55 × 10⁻⁰³	(ref)	1.21 [0.82–1.76]	0.34	1.96 [1.24–3.09]	3.69 × 10⁻⁰³	(ref)	1.19 [0.81–1.74]	0.37
Congestive heart failure (n = 127)	1.48 [0.89–2.46]	0.13	(ref)	1.50 [1.01–2.22]	0.04	1.38 [0.83–2.30]	0.21	(ref)	1.47 [0.99–2.18]	0.06
Coronary artery disease (n = 1267)	1.30 [1.09–1.55]	3.68 × 10⁻⁰³	(ref)	1.23 [1.08–1.40]	2.30 × 10⁻⁰³	1.23 [1.03–1.47]	0.02	(ref)	1.21 [1.06–1.39]	4.23 × 10⁻⁰³
Crohn's disease (n = 339)	1.05 [0.74–1.47]	0.79	(ref)	1.08 [0.84–1.38]	0.57	1.06 [0.75–1.50]	0.72	(ref)	1.08 [0.84–1.39]	0.55
Depression (n = 2837)	1.27 [1.13–1.44]	1.21 × 10⁻⁰⁴	(ref)	1.64 [1.50–1.78]	2.40 × 10⁻²⁹	1.21 [1.07–1.36]	3.11 × 10⁻⁰³	(ref)	1.62 [1.48–1.76]	8.37 × 10⁻²⁸
Epilepsy (n = 609)	1.04 [0.80–1.36]	0.76	(ref)	1.50 [1.26–1.80]	6.35 × 10⁻⁰⁶	1.06 [0.81–1.38]	0.69	(ref)	1.51 [1.27–1.80]	5.20 × 10⁻⁰⁶
Gout (n = 992)	1.35 [1.12–1.63]	1.54 × 10⁻⁰³	(ref)	1.02 [0.87–1.19]	0.83	1.26 [1.04–1.52]	0.02	(ref)	1.00 [0.85–1.16]	0.97
Hypertension (n = 10128)	1.29 [1.19–1.40]	7.48 × 10⁻¹⁰	(ref)	1.16 [1.09–1.24]	3.59 × 10⁻⁰⁶	1.16 [1.06–1.26]	6.14 × 10⁻⁰⁴	(ref)	1.13 [1.06–1.20]	2.73 × 10⁻⁰⁴
Insomnia (n = 3310)	1.40 [1.26–1.56]	1.23 × 10⁻⁰⁹	(ref)	1.23 [1.13–1.34]	1.81 × 10⁻⁰⁶	1.40 [1.25–1.56]	1.96 × 10⁻⁰⁹	(ref)	1.23 [1.13–1.34]	1.29 × 10⁻⁰⁶
Multiple sclerosis (n = 267)	1.55 [1.10–2.19]	0.01	(ref)	1.11 [0.83–1.47]	0.48	1.53 [1.08–2.16]	0.02	(ref)	1.10 [0.83–1.47]	0.50
Obesity (n = 8187)	1.65 [1.53–1.78]	5.89 × 10⁻³⁷	(ref)	1.11 [1.04–1.18]	9.32 × 10⁻⁰⁴	N/A	N/A	N/A	N/A	N/A
Obstructive sleep apnea (n = 3953)	1.44 [1.30–1.59]	1.05 × 10⁻¹²	(ref)	1.30 [1.20–1.40]	8.89 × 10⁻¹¹	1.23 [1.11–1.37]	8.87 × 10⁻⁰⁵	(ref)	1.26 [1.16–1.36]	5.32 × 10⁻⁰⁸
Rheumatoid arthritis (n = 640)	1.17 [0.92–1.49]	0.20	(ref)	1.08 [0.9–1.29]	0.44	1.13 [0.89–1.45]	0.32	(ref)	1.07 [0.89–1.28]	0.49
Restless legs syndrome (n = 681)	1.47 [1.17–1.84]	8.51 × 10⁻⁰⁴	(ref)	1.27 [1.07–1.51]	7.31 × 10⁻⁰³	1.36 [1.09–1.71]	7.04 × 10⁻⁰³	(ref)	1.25 [1.05–1.48]	0.01
Schizophrenia (n = 13)	-	-	(ref)	4.43 [1.45–13.58]	9.13 × 10⁻⁰³	-	-	(ref)	4.45 [1.45–13.63]	9.00 × 10⁻⁰³
Stroke (n = 159)	0.95 [0.56–1.63]	0.85	(ref)	1.65 [1.17–2.31]	3.89 × 10⁻⁰³	0.91 [0.53–1.57]	0.75	(ref)	1.63 [1.16–2.29]	4.63 × 10⁻⁰³
Type 1 diabetes (n = 175)	1.99 [1.34–2.94]	6.45 × 10⁻⁰⁴	(ref)	1.32 [0.93–1.88]	0.12	2.07 [1.39–3.07]	3.24 × 10⁻⁰⁴	(ref)	1.34 [0.94–1.90]	0.11
Type 2 diabetes (n = 1385)	1.65 [1.41–1.92]	1.70 × 10⁻¹⁰	(ref)	1.28 [1.13–1.45]	1.34 × 10⁻⁰⁴	1.44 [1.23–1.68]	6.30 × 10⁻⁰⁶	(ref)	1.22 [1.07–1.39]	2.30 × 10⁻⁰³
Ulcerative colitis (n = 269)	1.27 [0.88–1.84]	0.19	(ref)	1.24 [0.94–1.64]	0.12	1.33 [0.92–1.92]	0.13	(ref)	1.25 [0.95–1.65]	0.11

Bold denotes significant associations ($p < 2.27 \times 10^{-03}$). CI = 95% confidence interval, n = number of cases, OR = odds ratio.

^aModel 2 covariates: model 1 covariates (age, sex, and race) + further adjustment for body mass index (except for obesity outcome).

smaller than that reported from the UK Biobank (1.04 minutes per allele), which may possibly be due to consequences of higher disease prevalence and medication use in the Partners Biobank or to anticipated larger effects due to “winner’s curse” in the initial discovery sample. However, the ability to observe associations despite population and methodological differences supports the use of this instrument in sleep clinical research applications. Interestingly, the PRS association was only evident for weekday, but not weekend self-reported sleep duration, and may be due to higher heritability estimates for weekday (23.6%) compared to weekend (12.3%) sleep duration [41]. As was the

case for earlier genetic studies of limited sample sizes [41–43], there was insufficient power to detect associations between each individual locus of the PRS, including the PAX8 variant, and self-reported sleep duration. In addition to the 78 SNP PRS, we also show the promise of a PRS calculated using genome-wide data with robust associations with self-reported sleep duration.

The pleiotropic nature of some genes implicated by the PRS suggests that the PRS may also associate with related diseases. Indeed, we observed associations between the PRS and lower odds for CHF, obesity, hypertension, and insomnia, and higher odds for RLS, and associations for CHF and RLS were maintained

Table 3. Association of the 78 sleep duration signals genetic risk score with self-reported sleep duration and 22 prevalent diseases curated from Electronic Medical Records in the Partners Biobank (n = 16 033), with additional adjustment for obesity status

	Genetic risk score ^a		Genetic risk score—obesity-adjusted ^b	
	Beta (SE) per effect allele	P value	Beta (SE) per effect allele	P value
	0.65 (0.19)	7.32 × 10 ⁻⁰⁴	0.58 (0.19)	2.74 × 10 ⁻⁰³
Sleep duration (n = 7155)	OR [95% CI] per effect allele	P value	OR [95% CI] per effect allele	P value
Asthma (n = 1451)	1.006 [0.995–1.017]	0.271	1.008 [0.997–1.019]	0.148
Breast cancer (n = 737)	1.003 [0.989–1.018]	0.666	1.003 [0.989–1.018]	0.661
Bipolar disorder (n = 126)	1.005 [0.971–1.040]	0.772	1.006 [0.972–1.041]	0.731
Chronic obstructive pulmonary disorder (n = 289)	0.993 [0.971–1.016]	0.570	0.994 [0.971–1.016]	0.577
Congestive heart failure (n = 435)	0.977 (0.959–0.995)	0.015	0.977 (0.959–0.996)	0.017
Coronary artery disease (n = 2683)	1.002 [0.993–1.011]	0.608	1.003 [0.994–1.012]	0.575
Crohn's disease (n = 490)	1.006 [0.988–1.024]	0.499	1.006 [0.988–1.024]	0.529
Depression (n = 2017)	0.992 [0.983–1.001]	0.082	0.993 [0.984–1.003]	0.156
Epilepsy (n = 603)	0.993 [0.977–1.009]	0.374	0.992 [0.977–1.009]	0.356
Gout (n = 1008)	0.996 [0.983–1.009]	0.511	0.996 [0.984–1.009]	0.590
Hypertension (n cases = 7691)	0.993 [0.986–1.000]	0.039	0.995 (0.988–1.002)	0.168
Insomnia (n cases = 2123)	0.991 [0.982–1.000]	0.049	0.992 [0.983–1.001]	0.071
Multiple sclerosis (n = 160)	0.990 [0.960–1.021]	0.519	0.991 [0.961–1.021]	0.548
Obesity (n cases = 5631)	0.992 [0.986–0.999]	0.019	N/A	N/A
Obstructive sleep apnea (n = 2763)	0.996 [0.988–1.004]	0.328	0.998 [0.990–1.007]	0.706
Rheumatoid arthritis (n = 703)	1.004 [0.990–1.019]	0.557	1.005 [0.990–1.020]	0.491
Restless legs syndrome (n = 451)	1.019 [1.001–1.038]	0.041	1.020 [1.002–1.039]	0.031
Schizophrenia (n = 28)	1.006 [0.935–1.082]	0.881	1.006 [0.935–1.082]	0.873
Stroke (n = 376)	1.005 [0.985–1.025]	0.657	1.004 [0.984–1.024]	0.686
Type 1 diabetes (n = 115)	0.971 [0.937–1.006]	0.107	0.971 [0.936–1.006]	0.104
Type 2 diabetes (n = 1563)	0.990 [0.980–1.001]	0.066	0.992 [0.981–1.002]	0.121
Ulcerative colitis (n = 331)	0.998 [0.977–1.019]	0.828	0.998 [0.976–1.019]	0.821

Threshold is P = 0.05.

CI = 95% confidence interval, OR = odds ratio, SE = standard error.

^aAssociations adjusted for age, sex, genotyping array, and principal components of ancestry.

^bAssociations further adjusted for obesity status derived from EMR.

Table 4. Associations between genetic risk score quartiles and associated diseases in the Partners Biobank (n = 16 033)

	Genetic risk score								
	Weighted Q1	Weighted Q2		Weighted Q3		Weighted Q4		Weighted trend	
		OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value
Congestive heart failure (n = 435)	ref	0.800 (0.614–1.043)	0.10	0.800 (0.614–1.042)	0.10	0.703 (0.534–0.925)	0.01	0.899 (0.824–0.980)	0.02
Hypertension (n = 7691)	ref	0.973 (0.880–1.075)	0.59	0.976 (0.881–1.081)	0.64	0.887 (0.801–0.982)	0.02	0.966 (0.935–0.997)	0.03
Insomnia (n = 2123)	ref	0.941 (0.827–1.071)	0.36	0.972 (0.854–1.105)	0.66	0.894 (0.785–1.019)	0.09	0.969 (0.930–1.010)	0.13
Obesity (n = 5631)	ref	0.961 (0.876–1.054)	0.40	0.880 (0.802–0.966)	0.01	0.903 (0.823–0.992)	0.03	0.961 (0.934–0.990)	0.01
Restless legs syndrome (n = 451)	ref	1.007 (0.758–1.338)	0.96	1.401 (1.075–1.826)	0.01	1.232 (0.938–1.617)	0.13	1.098 (1.009–1.195)	0.03

CI = 95% confidence interval, OR = odds ratio, Q = quartile.

Table 5. Causal links of longer sleep duration with disease outcomes using two-sample Mendelian Randomization

	Inverse variance weighted		Weighted median		MR-Egger	
	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value
Congestive heart failure	0.978 (0.961–0.996)	0.019	0.978 (0.951–1.006)	0.12	0.991 (0.925–1.061)	0.80
Hypertension	0.993 (0.986–1.000)	0.049	0.994 (0.984–1.005)	0.28	0.989 (0.963–1.016)	0.42
Insomnia	0.992 (0.983–1.000)	0.053	0.994 (0.980–1.007)	0.34	0.985 (0.954–1.018)	0.38
Obesity	0.995 (0.987–1.003)	0.242	0.996 (0.986–1.007)	0.48	1.022 (0.992–1.052)	0.16
Restless legs syndrome	1.018 (1.000–1.036)	0.045	1.022 (0.996–1.049)	0.09	1.022 (0.957–1.091)	0.51

CI = 95% confidence interval, OR = odds ratio.

following additional adjustment for obesity status. The PRS associations with insomnia and obesity are consistent with the self-reported phenotypic cross-sectional links seen in this study, and findings from previous studies [11, 44]. Findings for insomnia are in agreement with previously observed genome-wide genetic correlations that suggest shared biological links between short sleep and insomnia, and with objective short sleep being a feature of an important insomnia subtype [44, 45]. The finding of an association with CHF, however, is novel, and further supported by our use of a genome-wide PRS. Sleep duration and sleep disorders are recognized as risk factors for cardiovascular disease [4, 46], with multiple likely causal and bi-directional associations. Our data suggest that there may be underlying genetic risk factors for both sleep duration and CHF, possibly relating to pathways influencing sympathetic nervous system activity, which is common to short sleep duration, hypertension, [47] and heart failure [48]. Discordance between the self-reported sleep duration and PRS with CHF in this study may be attributed to differences in statistical power (n cases with self-reported sleep duration: 127; n cases with genetics: 435) or differences in duration of exposure, where self-reported sleep reflects acute sleep status whereas PRS reflects chronic exposure. Dissecting this relationship is possible through comparing PRS findings with repeated self-reported sleep duration or repeated actigraphy measures. Additional differences may reflect sampling biases from differences in clinical and demographic characteristics between the Partners Biobank sub-groups that answered sleep questions and those with genetic data. Future validation of both phenotypic and PRS associations of sleep with diseases in other EMR studies will be important. Initial MR findings indicate causal effects of genetically determined longer sleep duration on lower disease prevalence, except for RLS; however, further analyses in larger samples with prospective sampling to assess associations with incident disease are needed to definitively establish causality.

Our study has several strengths worth noting. The Partners Biobank is a large study population linking EMR to genetics and sleep data, allowing for phenotypic and genetic investigation of sleep duration associations with diseases. In addition, case definitions are based on a narrative approach using natural language processing of EMR, and over 89% PPV for disease prevalence, limiting false positive cases resulting from relying on coded data that are prone to various errors [49], with the exception of sleep disorders, which were based on a codified approach using ICD-10 physician diagnoses without chart validation and thus need to be interpreted cautiously. The PRS allows approximation of sleep duration independent of questionnaire response rate, which is lower among severe (i.e. CHF) and rare diseases, enabling adequate number of cases (~200 cases) to be considered in association studies [50]. Furthermore, unlike self-reported sleep duration, a score derived by genetics should act independently of confounders that may influence the relationships between sleep duration and diseases.

Weaknesses of our study include a modest response rate to the optional sleep survey (39.7% of entire Partners Biobank population), as has been observed in previous studies [51], and possible selection bias related to survey completion (subsample with responses were younger and more female) emphasizing the need for other markers of sleep duration independent of self-report. As our analyses are primarily based on self-reported sleep duration, which is susceptible to recall

bias and other limitations [52], it is necessary to supplement biobanks with objective sleep measures and consider genetic instruments derived from objective sleep duration and quality [53]. Furthermore, residual confounding, particularly by obesity, and reverse causality are possible explanations for our cross-sectional phenotype and PRS observational associations. Our genetic analyses are conducted in individuals of European ancestry, and further investigation in other ethnicities is necessary for generalizability of our findings. In addition, given the cross-sectional design, causality cannot be assessed using the observed nongenetic phenotypic relationships, and may reflect the effects of disease onset, medication use, or other confounders, on self-reported sleep duration. Enriching future EMR datasets with disease onset timestamps and longitudinal phenotypic data through repeat assessments of self-reported and objective sleep measures as well as biomarkers of disease may enable prospective investigations. Causality may also be extended to other phenotypes and diseases using a phenome-wide Mendelian randomization (MR-PheWAS) approach to investigate causal relationships between sleep duration using genetics and a range of disease outcomes and clinical biomarkers from EMR [54]. In addition, while our PRS is estimated to capture only a modest amount of variability in sleep duration (~1.4%), the score may reflect long-term exposure. In addition, refining our PRS by delineating genetic variants associated with different aspects of sleep microarchitecture is needed to pinpoint pathways driving disease relationships.

The adoption of EMR is becoming widespread, and our results suggest the feasibility of leveraging EMR and biobank data to advance sleep clinical research. By validating the PRS for sleep duration and identifying cross-phenotype associations, we lay the groundwork for future investigations on the intersection between sleep genetics, clinical measures, and diseases, which may enable identification of robust biomarkers for sleep duration [55].

Supplementary material

Supplementary material is available at *SLEEP* online.

Acknowledgments

We thank the participants and administrators of the Partners HealthCare Biobank for their contribution to this work.

Funding

The authors are supported by NIH grants R01DK107859 (R.S., H.S.D.), R01HL113338 and R35 HL135818 (S.R.) and R01DK105072 (R.S.) and the Phyllis and Jerome Lyle Rappaport Massachusetts General Hospital Research Scholar Award (R.S.).

Conflict of interest statement. None declared.

References

1. Lane JM, et al. Genome-wide association analyses of sleep disturbance traits identify new loci and highlight shared genetics with neuropsychiatric and metabolic traits. *Nat Genet.* 2017;49(2):274–281.

2. Watson NF, et al. Recommended amount of sleep for a healthy adult: a joint consensus statement of the American academy of sleep medicine and sleep research society. *Sleep*. 2015;**38**(6):843–844.
3. Ayas NT, et al. A prospective study of self-reported sleep duration and incident diabetes in women. *Diabetes Care*. 2003;**26**(2):380–384.
4. Ayas NT, et al. A prospective study of sleep duration and coronary heart disease in women. *Arch Intern Med*. 2003;**163**(2):205–209.
5. Kripke DF, et al. Mortality associated with sleep duration and insomnia. *Arch Gen Psychiatry*. 2002;**59**(2):131–136.
6. Kronholm E, et al. Trends in self-reported sleep duration and insomnia-related symptoms in Finland from 1972 to 2005: a comparative review and re-analysis of Finnish population samples. *J Sleep Res*. 2008;**17**(1):54–62.
7. Qureshi AI, et al. Habitual sleep patterns and risk for stroke and coronary heart disease: a 10-year follow-up from NHANES I. *Neurology*. 1997;**48**(4):904–911.
8. Wingard DL, et al. Mortality risk associated with sleeping patterns among adults. *Sleep*. 1983;**6**(2):102–107.
9. Cappuccio FP, et al. Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies. *Sleep*. 2010;**33**(5):585–592.
10. Cappuccio FP, et al. Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. *Eur Heart J*. 2011;**32**(12):1484–1492.
11. Cappuccio FP, et al. Meta-analysis of short sleep duration and obesity in children and adults. *Sleep*. 2008;**31**(5):619–626.
12. Cappuccio FP, et al. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. *Diabetes Care*. 2010;**33**(2):414–420.
13. Karlson E, et al. Building the partners healthcare biobank at partners personalized medicine: informed consent, return of research results, recruitment lessons and operational considerations. *J Pers Med*. 2016;**6**(1):2.
14. Gottesman O, et al.; eMERGE Network. The electronic medical records and genomics (eMERGE) network: past, present, and future. *Genet Med*. 2013;**15**(10):761–771.
15. Denny JC, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol*. 2013;**31**(12):1102–1110.
16. Verma A, et al. PheWAS and beyond: the landscape of associations with medical diagnoses and clinical measures across 38,662 individuals from Geisinger. *Am J Hum Genet*. 2018;**102**(4). doi:10.1016/j.ajhg.2018.02.017
17. Lewis CM, et al. Prospects for using risk scores in polygenic medicine. *Genome Med*. 2017;**9**(1):96.
18. Solovieff N, et al. Pleiotropy in complex traits: challenges and strategies. *Nat Rev Genet*. 2013;**14**(7):483–495.
19. Purcell SM, et al.; International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;**460**(7256):748–752.
20. Knowles JW, et al. Cardiovascular disease: the rise of the genetic risk score. *PLoS Med*. 2018;**15**(3):e1002546.
21. Khera AV, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;**50**(9):1219–1224.
22. de Castro JM. The influence of heredity on self-reported sleep patterns in free-living humans. *Physiol Behav*. 2002;**76**(4–5):479–486.
23. Dashti H, et al. GWAS in 446,118 European adults identifies 78 genetic loci for self-reported habitual sleep duration supported by accelerometer-derived estimates. *bioRxiv*. 2018. <http://biorxiv.org/content/early/2018/04/19/274977.abstract>
24. Smoller JW, et al. An eMERGE clinical center at partners personalized medicine. *J Pers Med*. 2016;**6**(1):5.
25. Das S, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;**48**(10):1284–1287.
26. McCarthy S, et al.; Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;**48**(10):1279–1283.
27. O'Connell J, et al. A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet*. 2014;**10**(4):e1004234.
28. Wang C, et al. Improved ancestry estimation for both genotyping and sequencing data using projection procrustes analysis and genotype imputation. *Am J Hum Genet*. 2015;**96**(6):926–937.
29. Cann HM, et al. A human genome diversity cell line panel. *Science*. 2002;**296**(5566):261–262.
30. Euesden J, et al. PRSice: polygenic risk score software. *Bioinformatics*. 2015;**31**(9):1466–1468.
31. Jones SE, et al. Genome-wide association analyses in 128,266 individuals identifies new morningness and sleep duration loci. Shi J, ed. *PLoS Genet*. 2016;**12**(8):e1006125.
32. Lawlor DA. Commentary: two-sample Mendelian randomization: opportunities and challenges. *Int J Epidemiol*. 2016;**45**(3):908–915.
33. Burgess S, et al. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. 2013;**37**(7):658–665.
34. Bowden J, et al. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;**44**(2):512–525.
35. Bowden J, et al. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol*. 2016;**40**(4):304–314.
36. Singh M, et al. The association between obesity and short sleep duration: a population-based study. *J Clin Sleep Med*. 2005;**1**(4):357–363.
37. Devinsky O, et al. Epilepsy and sleep apnea syndrome. *Neurology*. 1994;**44**(11):2060–2064.
38. Cappuccio FP, et al. Gender-specific associations of short sleep duration with prevalent and incident hypertension: the Whitehall II Study. *Hypertension*. 2007;**50**(4):693–700.
39. Li W, et al. Sleep duration and risk of stroke events and stroke mortality: a systematic review and meta-analysis of prospective cohort studies. *Int J Cardiol*. 2016;**223**:870–876.
40. Perlman CA, et al. The prospective impact of sleep duration on depression and mania. *Bipolar Disord*. 2006;**8**(3):271–274.
41. Gottlieb DJ, et al. Novel loci associated with usual sleep duration: the CHARGE consortium genome-wide association study. *Mol Psychiatry*. 2015;**20**(10):1232–1239.
42. Cade BE, et al. Common variants in DRD2 are associated with sleep duration: the CARE consortium. *Hum Mol Genet*. 2016;**25**(1):167–179.
43. Gottlieb DJ, et al. Genome-wide association of sleep and circadian phenotypes. *BMC Med Genet*. 2007;**8**(Suppl 1):S9.

44. Lane JM, et al. Biological and clinical insights from genetics of insomnia symptoms. *bioRxiv*. 2018. <http://biorxiv.org/content/early/2018/02/02/257956.abstract>
45. Vgontzas AN, et al. Insomnia with objective short sleep duration: the most biologically severe phenotype of the disorder. *Sleep Med Rev*. 2013;17(4):241–254.
46. Laugsand LE, et al. Insomnia and the risk of incident heart failure: a population study. *Eur Heart J*. 2014;35(21):1382–1393.
47. DiBona GF. Sympathetic nervous system and hypertension. *Hypertens (Dallas, Tex 1979)*. 2013;61(3):556–560.
48. Nagai M, et al. Sleep duration as a risk factor for cardiovascular disease- a review of the recent literature. *Curr Cardiol Rev*. 2010;6(1):54–61.
49. O'Malley KJ, et al. Measuring diagnoses: ICD code accuracy. *Health Serv Res*. 2005;40(5 Pt 2):1620–1639.
50. Verma A, et al. A simulation study investigating power estimates in phenome-wide association studies. *BMC Bioinf*. 2018;19(1):120.
51. Holliday EG, et al. Short sleep duration is associated with risk of future diabetes but not cardiovascular disease: a prospective study and meta-analysis. *Arias-Carrion O, ed. PLoS One*. 2013;8(11):e82305.
52. Bianchi MT, et al. An open request to epidemiologists: please stop querying self-reported sleep duration. *Sleep Med*. 2017;35:92–93.
53. Jones SE, et al. Genetic studies of accelerometer-based sleep measures in 85,670 individuals yield new insights into human sleep behaviour. *bioRxiv*. 2018. <http://biorxiv.org/content/early/2018/09/13/303925.abstract>
54. Li X, et al. MR-PheWAS: exploring the causal effect of SUA level on multiple disease outcomes by using genetic instruments in UK Biobank. *Ann Rheum Dis*. 2018. doi:10.1136/annrheumdis-2017-212534.
55. Mullington JM, et al. Developing biomarker arrays predicting sleep and circadian-coupled risks to health. *Sleep*. 2016;39(4):727–736.