Low depression frequency is associated with decreased risk of cardiometabolic disease

Michael C. Honigberg, Yixuan Ye, Lillian Dattilo, Amy A. Sarma, Nandita S. Scott, Jordan W. Smoller, Hongyu Zhao, Malissa J. Wood and Pradeep Natarajan

Polygenic risk scores (PRS) are an increasingly available tool to refine risk prediction for cardiometabolic diseases. Favorable lifestyle behaviors might offset increased polygenic risk, but whether frequency of depressed mood stratifies PRS-associated risk is unknown. Here, we calculated individual-level 3-million variant PRS for coronary artery disease (CAD), type 2 diabetes (T2D) and atrial fibrillation among 328,152 genotyped, individuals of European ancestry in the UK Biobank. After adjustment for clinical/lifestyle factors and PRS, low versus high frequency of depressed mood was associated with lower risks of incident CAD by 34%, T2D by 33% and atrial fibrillation by 20%. Frequency of depressed mood stratified risk for CAD and T2D across low (lowest quintile), intermediate (middle three quintiles) and high (highest quintile) PRS strata. Depression was associated more strongly with CAD in women compared with men (Pinteraction < 0.001). Overall, lower burden of depressed mood was independently associated with lower risk of CAD and T2D across the cardiometabolic polygenic risk spectrum.

While diet, exercise and smoking have long been recognized as important behavioral contributors to cardiovascular health, mounting evidence also implicates depression as an independent risk factor for cardiovascular disease (CVD). In the INTERHEART study of approximately 25,000 individuals across 52 countries, depression was independently associated with a 1.55-fold risk of myocardial infarction. In a recent study of 563,255 individuals from 22 pooled cohorts, each additional standard deviation (s.d.) of greater depressive symptoms was independently associated with a 0.6–1.1-fold increased risk for a composite of CAD and stroke.

PRS are the quantitative representation of heritable risk for a particular trait or condition conferred by many common genetic variants. PRS are predictive of cardiometabolic disease and may identify individuals most likely to benefit from preventive therapies and are increasingly available in clinical and direct-to-consumer settings. PRS may be particularly useful for risk assessment early in life, before the development of overt cardiometabolic risk factors. Although optimal management of high polygenic risk for cardiometabolic diseases remains to be defined, previous observational studies suggest favorable lifestyle behaviors may mitigate heightened polygenic risk, implying lifestyle modification may be an efficient strategy to address elevated polygenic risk. This analysis aimed to assess whether frequency of depressed mood further stratifies polygenic risk of cardiometabolic conditions, independent of lifestyle and conventional CVD risk factors.

Here, we tested the association of frequency of depressed mood with incident CAD, T2D and atrial fibrillation across the spectrum of polygenic risk, and secondarily in those at high polygenic risk, in the UK Biobank.

Results

Description of the study cohort. The study cohort included 328,152 genotyped, unrelated, European ancestry participants from the UK Biobank (Extended Data Fig. 1). Excluded individuals were older and had a greater burden of cardiometabolic risk factors and prevalent CVD (Supplementary Table 1). Mean (s.d.) age of the study sample was 56.8 (7.9) years, and 173,333 (52.8%) were female. Overall, 255,078 individuals (77.7%) reported no episodes of depressed mood in the past 2 weeks (low frequency of depressed mood), 59,950 (18.3%) reported depression on several days in the past 2 weeks (moderate frequency) and 13,124 (4.0%) reported depression more than half of the days or nearly every day (high frequency).

Individuals with greater frequency of depressed mood were younger at baseline, more likely to be female and more likely to be current smokers; reported less vegetable and fresh fruit intake, exercise and sleep; and had higher body mass index (BMI), higher C-reactive protein and higher baseline prevalence of CAD, hypertension, hypercholesterolemia and T2D (Table 1).

Participants were followed up for outcomes over a median (interquartile range (i.q.r.)) of 11.1 (10.4–11.8) years. Among individuals without each condition at baseline in the full cohort, 17,880 (6.8%) developed CAD, 14,345 (4.5%) developed T2D and 15,397 (4.8%) developed atrial fibrillation.

Polygenic risk and incident outcomes. PRS for CAD, T2D and atrial fibrillation were derived from genome-wide association studies performed in consortia external to the UK Biobank (CARDIoGRAMplusC4D, DIAGRAM and AFGen, respectively) using AnnoPred, as previously described. Briefly, AnnoPred is a genome-wide Bayesian method that leverages genomic and epigenomic functional annotations to adjust variant weights. The AnnoPred method recently yielded superior risk prediction compared with other current state-of-the-art polygenic risk scoring approaches.

1Cardiology Division, Massachusetts General Hospital, Boston, MA, USA. 2Program in Medical and Population Genetics and Cardiovascular Disease Initiative, Broad Institute of Harvard and MIT, Cambridge, MA, USA. 3Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA. 4Harvard Medical School, Boston, MA, USA. 5Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA. 6Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. 7Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. e-mail: pnatarajan@mgh.harvard.edu
Table 1 | Baseline characteristics of study cohort

| Characteristic                                      | Depression burden                                      | P value          |
|-----------------------------------------------------|--------------------------------------------------------|-----------------|
|                                                      | High (n = 13,124)                                      | Low (n = 255,078) |
| Age, years                                          | 54.4 (7.7)                                             | 57.3 (7.9)       | <2.2x10^{-16} |
| Female                                              | 7,762 (59.1%)                                         | 129,701 (50.8%)  | <2.2x10^{-16} |
| Country of enrolling UK Biobank assessment centre   |                                                        |                 |
| • England                                           | 11,575 (88.2%)                                        | 225,590 (88.4%)  | 2.0x10^{-10}  |
| • Scotland                                          | 867 (6.6%)                                            | 18,713 (7.3%)    |                 |
| • Wales                                             | 682 (5.2%)                                            | 10,775 (4.2%)    |                 |
| Townsend deprivation index (median (i.q.r.))        | −1.4 (−3.3, 1.7)                                      | −2.5 (−3.8, −0.4)| <2.2x10^{-16}  |
| Smoking status                                      |                                                        |                 |
| • Current                                           | 2,418 (18.4%)                                        | 21,547 (8.4%)    | <2.2x10^{-16}  |
| • Former                                            | 4,276 (32.6%)                                        | 90,811 (35.6%)   |                 |
| • Never                                             | 6,430 (49.0%)                                        | 142,720 (56.0%)  |                 |
| Total pack-years of smoking among current or former smokers (median (i.q.r.)) | 15.9 (0, 31.5)                                      | 9.3 (0, 23.6)    | <2.2x10^{-16}  |
| Alcohol intake (median (i.q.r.))                    | 1–2 times weekly (1–3 times monthly, daily or almost daily) |                 |
| Total portions of vegetables and fresh fruit daily  | 6.6 (4.3)                                             | 71 (3.7)         | <2.2x10^{-16}  |
| Days of moderate physical activity ≥10 min          | 3.3 (2.5)                                             | 3.6 (2.3)        | <2.2x10^{-16}  |
| Days of vigorous physical activity ≥10 min          | 1.6 (2.0)                                             | 1.9 (1.9)        | <2.2x10^{-16}  |
| Average sleep duration, h                           | 6.9 (1.3)                                             | 7.2 (1.0)        | <2.2x10^{-16}  |
| BMI, kg/m²                                          | 28.4 (5.7)                                            | 27.2 (4.5)       | <2.2x10^{-16}  |
| Systolic blood pressure, mmHg                       | 136.3 (19.0)                                         | 140.9 (19.6)     | <2.2x10^{-16}  |
| Antihypertensive medication use                     | 2,782 (21.2%)                                        | 51,802 (20.3%)   | <2.2x10^{-16}  |
| Cholesterol-lowering medication use                 | 2,470 (18.8%)                                        | 43,202 (16.9%)   | <2.2x10^{-16}  |
| Antiplatelet medication use                         | 2,095 (16.0%)                                        | 36,274 (14.2%)   | 3.7x10^{-14}   |
| Antihyperglycemic medication use                    | 673 (5.1%)                                           | 7,761 (3.0%)     | <2.2x10^{-16}  |
| Antidepressant medication use                       | 3,481 (26.5%)                                        | 9,965 (3.9%)     | <2.2x10^{-16}  |
| Non-HDL cholesterol, mg ml^-1                       | 1.65 (0.43)                                          | 1.65 (0.41)      | 0.13           |
| C-reactive protein, mg L^-1 (median (i.q.r.))       | 1.6 (0.7, 3.4)                                       | 1.3 (0.6, 2.6)   | <2.2x10^{-16}  |
| CAD                                                 | 724 (5.5%)                                           | 10,160 (4.0%)    | 1.0x10^{-11}   |
| Hypertension                                        | 4,266 (32.5%)                                        | 72,196 (28.3%)   | <2.2x10^{-16}  |
| Hypercholesterolemia                                | 2,119 (16.1%)                                        | 37,400 (14.7%)   | <2.2x10^{-16}  |
| T2D                                                 | 501 (3.8%)                                           | 5,007 (2.0%)     | <2.2x10^{-16}  |
| Atrial fibrillation                                 | 197 (1.5%)                                           | 4,526 (1.8%)     | 6.8x10^{-4}    |

For each condition under study, polygenic risk was independently associated with the corresponding outcome. After complete covariate adjustment including frequency of depressed mood, genetically increased risk was associated with higher risk of CAD (hazard ratio (HR) per s.d. of CAD polygenic risk: 1.32, 95% confidence interval (CI) 1.30–1.34, P < 2.2x10^{-16}, T2D (HR per s.d.: 1.38, 95% CI 1.35–1.40, P < 2.2x10^{-16}) and atrial fibrillation (HR per s.d.: 1.46, 95% CI 1.44–1.49, P < 2.2x10^{-16}).

Greater frequency of depressed mood was modestly associated with higher CAD PRS (higher CAD PRS by 0.012 s.d. per depression category, 95% CI 0.005–0.019 s.d., P = 3.6x10^{-14}) and higher T2D PRS (higher T2D by 0.009 s.d., 95% CI 0.003–0.016 s.d., P = 0.006), but not with atrial fibrillation PRS (P = 0.29).

Frequency of depressed mood and incident outcomes. The primary study exposure was self-reported frequency of depressed mood, ascertained at study enrollment. After adjustment for age, age-squared, sex, the first 20 principal components of ancestry (PC 1–20), genotyping array, country of study enrollment (England, Scotland or Wales), Townsend deprivation index, smoking status, pack-year smoking history, alcohol intake, vegetable and fresh fruit intake, days per week of moderate and vigorous exercise, sleep duration, systolic blood pressure, antihypertensive medication use, non-high-density lipoprotein (HDL) cholesterol, cholesterol-lowering medication use, antiplatelet medication use, antihyperglycemic medication use, prevalent T2D (models for CAD and atrial fibrillation only), BMI, C-reactive protein and polygenic risk, low
association between depression and CAD was slightly smaller among individuals with high polygenic risk, each additional point on the PHQ-2 score was associated with greater risk of CAD (HR 1.10, 95% CI 1.08−1.11, P < 2.2×10^{-16}), T2D (HR 1.11, 95% CI 1.09−1.13, P < 2.2×10^{-16}) and atrial fibrillation (HR 1.05, 95% CI 1.03−1.07, P = 5.7×10^{-4}).

Assessment of interactions. We observed a significant interaction between polygenic risk and frequency of depressed mood for incident CAD (after full covariate adjustment, HR(interaction): 0.95, 95% CI 0.93−0.98, P = 0.003; Supplementary Table 8), whereby the association between depression and CAD was slightly smaller among individuals at high versus low polygenic CAD risk (Supplementary Table 9). To test whether this finding might be driven by a higher proportion of individuals with prevalent CAD excluded from analysis among those with high polygenic risk, we tested interations for total (prevalent and incident) CAD using logistic regression and observed identical interactions (Supplementary Table 10). By contrast, interactions between polygenic risk and frequency of depressed mood were not observed for other outcomes (Supplementary Table 8).

Sex differences. We observed an interaction between sex and frequency of depressed mood for incident CAD (fully adjusted HR(interaction of depression × female sex): 1.11, 95% CI 1.05−1.18, P < 0.001). In fully adjusted models including polygenic risk, a low versus high burden of depression was independently associated with greater CAD risk reduction in women (HR 0.57, 95% CI 0.51−0.63, P < 2.2×10^{-16}) versus men (HR 0.74, 95% CI 0.67−0.81, P = 3.1×10^{-4}; Supplementary Table 11). We did not observe significant sex × depression interactions in fully adjusted models for T2D or atrial fibrillation.

Comparing depression frequency and polygenic risk. A high frequency of depressed mood was associated with the same crude cumulative incidence as the top 27th percentile of CAD PRS and the top 22nd percentile of T2D PRS. Using the magnitude of associations derived from fully adjusted models, low versus high frequency of depressed mood was equivalent to having a lower CAD PRS by 1.51 s.d., lower T2D PRS by 1.24 s.d. and lower atrial fibrillation PRS by 0.57 s.d.

Discussion
In a prospective cohort of >325,000 British adults, a lower burden of depressed mood was associated with decreased risk of CAD, T2D and atrial fibrillation across the polygenic risk spectrum. Among individuals at high polygenic risk, who have the highest absolute risk of developing disease, a low versus high burden of depression.
was independently associated with lower risk of CAD and T2D by 30–32%. Although absolute risk differences associated with frequency of depressed mood were modest in this relatively low-risk cohort, the association of depressed mood with outcomes was independent of lifestyle factors known to associate with both mental health and CVD (diet, exercise, sleep, tobacco and alcohol use) and other conventional CVD risk factors. In addition, frequency of depressed mood was more strongly linked to incident CAD in women versus men.

This analysis extends previous work on depression and incident cardiometabolic disease in several respects. Our study now demonstrates additive effects of depression across strata of polygenic risk, as assessed using a recently described PRS approach. Although lifestyle behaviors also stratify polygenic risk of cardiometabolic disease\(^{2,11,12}\), frequency of depressed mood was associated with cardiometabolic outcomes independent of these lifestyle factors. The large cohort size and detailed phenotyping of UK Biobank participants permitted a more comprehensive set of covariates (including vital signs and laboratory biomarkers) than previous large studies of depression and CVD\(^{3,9,18,19}\) and enabled examination of outcomes beyond atherosclerotic CVD alone.

We identified an interaction between frequency of depressed mood and CAD PRS, whereby the association between depression and CAD was slightly smaller among individuals at high versus low polygenic CAD risk. To our knowledge, no such interaction has been previously reported. Given the correlation we observed between CAD PRS and frequency of depressed mood, one potential explanation for this interaction may be that a genome-wide CAD PRS incorporating millions of variants may be already capturing some of the association between depression and CAD in the PRS itself.

The findings of this study may have implications for improving individual- and population-level cardiometabolic health. First, a reduced burden of depressed mood is associated with lower risk
| Coronary artery disease | Type 2 diabetes mellitus | Atrial fibrillation |
|------------------------|--------------------------|--------------------|
| **Polygenic risk**     |                          |                    |
| Low                    | Model 1                  | Model 2            |
| 1.89 (1.56–2.28)       | 2.10 (1.79–2.49)         | 2.20 (2.10–2.30)   |
| Moderate               | 1.61 (1.3–1.98)          | 1.90 (1.63–2.21)   |
| High                   | 2.86 (2.42–3.39)         | 2.88 (2.42–3.42)   |
| **Frequency of depressed mood** |                      |                    |
| Low                    | Model 1                  | Model 2            |
| 2.49 (2.19–2.85)       | 2.51 (2.12–2.96)         | 2.53 (2.12–2.97)   |
| Moderate               | 1.79 (1.51–2.12)         | 1.81 (1.53–2.16)   |
| High                   | 2.32 (2.00–2.68)         | 2.33 (2.01–2.70)   |

Values (unadjusted for multiple comparisons) were calculated using multivariable-adjusted Cox proportional hazard models. Model 1: adjusted for age, age-squared, sex, PC 1-20, genotyping array, country, Townsend deprivation index. Model 2: adjusted for age, age-squared, sex, PC 1-20, genotyping array, country, Townsend deprivation index, prevalent T2D, BMI, C-reactive protein, fresh fruit intake, days per week of moderate and vigorous exercise, sleep duration, systolic blood pressure, antihypertensive medication use, non-HDL cholesterol, cholesterol-lowering medication use, antiplatelet medication use, antihyperglycemic medication use, Townsend deprivation index. Model 2 was additionally adjusted for sex differences by restricting the analysis to women. Two-sided P-values calculated for multiple comparisons were derived using the Bonferroni correction. Covariates in the proportional hazards model were age, age-squared, sex, PC 1-20, genotyping array, country, Townsend deprivation index, prevalent T2D, BMI, C-reactive protein, fresh fruit intake, days per week of moderate and vigorous exercise, sleep duration, systolic blood pressure, antihypertensive medication use, non-HDL cholesterol, cholesterol-lowering medication use, antiplatelet medication use, antihyperglycemic medication use, Townsend deprivation index, prevalent T2D, BMI, C-reactive protein, fresh fruit intake, days per week of moderate and vigorous exercise, sleep duration, systolic blood pressure, antihypertensive medication use, non-HDL cholesterol, cholesterol-lowering medication use, antiplatelet medication use, antihyperglycemic medication use, Townsend deprivation index. Model 2 was additionally adjusted for sex differences by restricting the analysis to women.
the time of recruitment between 2006 and 2010 (ref. 13). The current study was conducted under UK Biobank application number 7089. The UK Biobank was approved by the North West Multi-centre Research Ethics Committee, and all subjects provided written informed consent. Participants were not compensated for study participation, although reimbursement was available for expenses incurred through participation. The Massachusetts General Hospital institutional review board approved secondary analysis of these data.

At enrollment, participants provided detailed information on medical history, medication use, lifestyle factors and mental health, and underwent physical assessment and phlebotomy for laboratory analysis and genotyping. Participants were followed for the development of incident diagnoses through linkage to national health records and follow-up study visits 14. The current analysis included genotyped, unrelated European ancestry participants with complete available data on self-reported frequency of depressed mood, socioeconomic status, sleep and health behaviors. Follow-up for the primary analyses occurred through March 2020.

**Study cohort, genotyping and imputation.** A subset of 49,950 participants were genotyped using the UK Biobank BILEVE Axiom Array 15, and the remaining 4,382 individuals were genotyped using the UK Biobank Axiom Array (bead array by Affymetrix); these arrays share >95% of marker content 16. Imputation was performed centrally by the UK Biobank using the Haplotype Reference Consortium, UK10K and 1000 Genomes Project phase 3 reference panels as described previously 17. The current analysis considered individuals with white British ancestry, concordance between genetic and reported sex and no sex chromosome aneuploidy. Participants within three degrees of relatedness were identified using the Kinship-Based Inference for Genome-Wide Association studies tool 18, and one from each pair of related individuals was excluded at random.

Polygenic risk scores and risk strata. PRS for CAD, T2D and atrial fibrillation were derived from genome-wide association studies performed in consortia external to the UK Biobank (CARDIoGRAMplusC4D14, DIAGRAM15 and AFGen16), respectively using AnnoPred 19–21. AnnoPred is a genome-wide Bayesian method that leverages genomic and epigenomic functional annotations, all external to the UK Biobank, to adjust variant weights 19. Ye et al. recently identified optimal AnnoPred PRS for CAD, T2D and atrial fibrillation by testing 88 candidate PRS for each trait using 4 levels of functional annotations (including diverse genome annotations 22, GenoCanyon functionality scores, GenoSkyline tissue-specific functionality scores, and cell-specific functionality scores), 2 different assumptions about the effects and 11 different tuning parameters, selecting the score which maximized area-under-the-curve in training and testing datasets 23. The AnnoPred method yielded superior risk prediction compared with other current state-of-the-art polygenic risk scoring approaches 24. The optimal AnnoPred-derived polygenic scores (using variants with minor allele frequency ≥0.05) included 2,994,054 variants for CAD, 2,996,767 variants for T2D and 2,996,792 variants for atrial fibrillation. Polygenic risk for each condition was designated to be high (highest quintile of PRS), intermediate (middle three quintiles) or low (lowest quintile).

**Exposures.** The primary study exposure was self-reported frequency of depressed mood in the previous two weeks, ascertained at study enrollment. Participants were prompted by touchscreen to complete the PHQ-2 screening instrument for depression, which includes the question: “Over the past two weeks, how often have you felt down, depressed or hopeless?” (UK Biobank field ID 2050). Available answers included ‘not at all’, ‘several days’, ‘more than half of days’ and ‘nearly every day’. Frequency of depressed mood was categorized as low (no depressed mood in the previous two weeks), moderate (several days of depressed mood in the previous two weeks) or high (depressed mood more than half of days or nearly every day). In primary analyses, individuals with a high burden of depressed mood constituted the reference group.

**Covariates.** Medical conditions at baseline were captured by participant self-report, with verification by a trained study nurse and/or by the presence of qualifying ICD code in the participant’s medical record (Supplementary Appendix). Health behaviors, including dietary intake, exercise frequency, sleep, alcohol and tobacco use, were systematically ascertained by self-report. Pack-years of tobacco use were coded by the UK Biobank as 0 if subject quit smoking before age 16 or total duration of smoking was <6 months. We extracted medication use at baseline, including all cholesterol-lowering medications, antihypertensive medications, antplatelet medications (aspirin, clopidogrel, dipyridamole) and antihyperglycemic medications (metformin, sulfonylureas, thiazolidinediones and insulin). In addition, we extracted data on baseline use of antidepressant medications (selective serotonin reuptake inhibitors, serotonin/norepinephrine reuptake inhibitors, tricyclic antidepressants, monoamine oxidase inhibitors, bupropion, mirtazapine and trazodone). Participants underwent measurement of anthropomorphic data and vital signs, as well as phlebotomy for laboratory analysis, including total and HDL cholesterol and high-sensitivity C-reactive protein (AU5800 platform, Beckman Coulter).

**Outcomes.** The three coprimary study outcomes were incident (newly diagnosed) CAD, T2D and atrial fibrillation. Outcomes were ascertained by the appearance of a qualifying International Classification of Diseases or procedure code (Supplementary Appendix) in the patient’s medical record.

**Statistical analysis.** Participants’ baseline characteristics were compared across depression categories using analysis of variance or the Kruskal–Wallis test, as appropriate, for continuous variables, and using the Pearson chi-squared test for categorical ones. Absolute differences in incidence rates and incidence rate differences across strata of frequency of depressed mood and polygenic risk were calculated using the ‘fmsb’ package in R v 3.6.0 per 1,000 person-years of follow-up time.

Multivariable Cox proportional hazard models were fitted to test the association of frequency of depressed mood with each cardiometabolic disease outcome. Follow-up began at study enrollment, and time to censoring was determined by the appearance of a qualifying diagnosis in the participant’s medical record or the date of last follow-up. Participants were excluded from each model for which they had an established diagnosis at baseline (for example, participants with prevalent CAD were excluded from models for incident CAD). The proportional hazards assumption was verified using Schoenfeld residuals. Cox models for each outcome were performed with two levels of covariate adjustment: Model 1, adjusted for age, age-squared, sex, the first 20 principal components of ancestry, genotyping array, country of UK Biobank enrollment (England, Scotland or Wales) and the inverse-normalized Townsend deprivation index; and Model 2, Model 1 plus smoking status, pack-years of smoking, alcohol intake, vegetable and fresh fruit intake, days per week of moderate and vigorous exercise, nightly sleep duration, systolic blood pressure, antihypertensive medication use, non-HDL cholesterol, cholesterol-lowering medication use, antplatelet medication use, antihyperglycemic medication use, prevalent T2D (models for CAD and atrial fibrillation only), BMI, and log-transformed C-reactive protein. Sensitivity analyses (1) probed possible reverse causation by excluding individuals with any prevalent study outcome (CAD, T2D, and/or atrial fibrillation) as well as ischemic stroke, peripheral artery disease or heart failure at baseline, (2) further adjusted for antidepressant use and excluded those taking antidepressant medications, (3) tested associations between PHQ-2 score and outcomes and (4) truncated follow-up at 31 December 2019, to ensure observations were not influenced by the SARS-CoV2 pandemic or associated lockdowns.

We evaluated interactions between frequency of depressed mood, modeled as an ordinal term (0, 1, 2), and polygenic risk as a continuous quantitative variable. To avoid residual confounding, interactions terms were included between frequency of depressed mood and each covariate. We further assessed interactions between frequency of depressed mood and, sex-stratified models were performed for outcomes with significant sex x depression interactions.

Given three outcomes studied, two-sided \( P < 0.05/3 = 0.0167 \) was considered statistically significant. Secondary analyses were considered supportive and hypothesis-generating at \( P < 0.05 \). Analyses were performed using R V 3.6.0 (R Foundation for Statistical Computing).

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**

UK Biobank data are available to researchers by application (https://www.ukbiobank.ac.uk/). PRS for CAD, T2D and atrial fibrillation have been submitted to the Polygenic Score Catalog (https://www.pgscatalog.org/).

**Code availability**

Code for the ‘survival’, ‘fmsb’ and ‘ggplot2’ R packages are publicly available (https://cran.r-project.org/).

Received: 6 August 2021; Accepted: 13 December 2021; Published online: 14 February 2022

**References**

1. Aragam, K. G. & Natarajan, P. Polygenic scores to assess atherothrombotic cardiovascular disease risk: clinical perspectives and basic implications. Circ. Res. 126, 1159–1177 (2020).
2. Ye, Y. et al. Interactions between enhanced polygenic risk scores and lifestyle for cardiovascular disease, diabetes mellitus and lipid levels. Circ. Genom. Precis. Med. 14, e003128 (2021).
3. Stampler, M. J., Hu, F. B., Manson, J. E., Rimm, E. B. & Willett, W. C. Primary prevention of coronary heart disease in women through diet and lifestyle. N. Engl. J. Med. 343, 16–22 (2000).
4. Xanthakis, V. et al. Ideal cardiovascular health: associations with biomarkers and subclinical disease and impact on incidence of cardiovascular disease in the Framingham Offspring Study. Circulation 130, 1676–1683 (2014).
5. Chiueh, S. E., McCulloch, M. L., Sacks, F. M. & Rimm, E. B. Healthy lifestyle factors in the primary prevention of coronary heart disease among men: benefits among users and nonusers of lipid-lowering and antihypertensive medications. Circulation 114, 160–167 (2006).
6. Corlin, L., Short, M. I., Vasan, R. S. & Xanthakis, V. Association of the duration of ideal cardiovascular health through adulthood with cardiometabolic outcomes and mortality in the Framingham Offspring Study. JAMA Cardiol. 5, 549–556 (2020).

7. Liu, N. et al. Association of major depression with risk of ischemic heart disease in a mega-cohort of Chinese adults: the China Kadoorie Biobank Study. J. Am. Heart Assoc. 12, 5 (2016).

8. Rosengren, A. et al. Association of psychosocial risk factors with risk of acute myocardial infarction in 11191 cases and 13648 controls from 52 countries (the INTERHEART study): case–control study. Lancet 364, 953–962 (2004).

9. Harshfield, E. L. et al. Association between depressive symptoms and incident cardiovascular diseases. JAMA 324, 2396–2405 (2020).

10. Aragam, K. G. et al. Limitations of contemporary guidelines for managing patients at high genetic risk of coronary artery disease. J. Am. Coll. Cardiol. 75, 2769–2780 (2020).

11. Khera, A. V. et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. N. Engl. J. Med. 375, 2349–2358 (2016).

12. Natarajan, P. et al. Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting. Circulation 135, 2891–2901 (2017).

13. Said, M. A., Verweij, N. & van der Harst, P. Associations of combined genetic and lifestyle risks with incident cardiovascular disease and diabetes in the UK Biobank Study. JAMA Cardiol. 3, 693–702 (2018).

14. Nikpay, M. et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat. Genet. 47, 1121–1130 (2015).

15. Scott, R. A. et al. An expanded genome-wide association study of type 2 diabetes in Europeans. Diabetes 66, 2888–2902 (2017).

16. Christophersen, I. E. et al. Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. Nat. Genet. 49, 946–952 (2017).

17. Hu, Y. et al. Leveraging functional annotations in genetic risk prediction for human complex diseases. PLoS Comput. Biol. 13, e1005589 (2017).

18. Rajan, S. et al. Association of symptoms of depression with cardiovascular disease and mortality in low-, middle-, and high-income countries. JAMA Psychiatry 77, 1052–1063 (2020).

19. Pan, A., Sun, Q., Okereke, O. I., Ruxode, K. M. & Hu, F. B. Depression and risk of stroke morbidity and mortality: a meta-analysis and systematic review. JAMA Cardiol. 306, 1241–1249 (2011).

20. Lu, Y., Wang, Z., Georgakis, M. K., Lin, H. & Zheng, L. Genetic liability to depression and risk of coronary artery disease, myocardial infarction, and other cardiovascular outcomes. J. Am. Heart Assoc. 10, e017986 (2021).

21. Li G. H. et al. Evaluation of bi-directional causal association between depression and cardiovascular diseases: a Mendelian randomization study. Psychol. Med. 1–12, https://doi.org/10.1017/S0033291720003566 (2020).

22. Vaccarino, V. et al. Depression and coronary heart disease: 2018 position paper of the ESC working group on coronary pathophysiology and microcirculation. Eur. Heart J. 41, 1687–1696 (2020).

23. Penninx, B. W. Depression and cardiovascular disease: epidemiological evidence on their linking mechanisms. Neurosci. Biobehav. Rev. 74, 277–286 (2017).

24. Miller, A. H. & Raison, C. L. The role of inflammation in depression: from evolutionary imperative to modern treatment target. Nat. Rev. Immunol. 16, 22–34 (2016).

25. Khandaker, G. M. et al. Shared mechanisms between coronary heart disease and depression: findings from a large UK general population-based cohort. Mol. Psychiatry 25, 1477–1486 (2020).

26. Smoldersen, K. G. et al. Depressive symptoms in younger women and men with acute myocardial infarction: insights from the VIRGO study. J. Am. Heart Assoc. 4, e01424 (2015).

27. Gillespie, S. L. et al. Allostatic load in the association of depressive symptoms with incident coronary heart disease: The Jackson Heart Study. Psychoneuroendocrinology 109, 104369 (2019).

28. Shah, A. J. et al. Sex and age differences in the association of depression with obstructive coronary artery disease and adverse cardiovascular events. J. Am. Heart Assoc. 3, e000741 (2014).

29. Sun, W. J., Xu, L., Chan, W. M., Lam, T. H. & Schooling, C. M. Are depressive symptoms associated with cardiovascular mortality among older Chinese: a cohort study of 64,000 people in Hong Kong? Am. J. Geriatr. Psychiatry 21, 1107–1115 (2013).

30. Bryant, K. B. et al. Time-varying depressive symptoms and cardiovascular all-cause mortality: does the risk vary by age or sex? J. Am. Heart Assoc. 9, e016661 (2020).

31. Ramsey, J. M. et al. Sex differences in serum markers of major depressive disorder in the Netherlands Study of Depression and Anxiety (NESDA). PLoS ONE 11, e0156624 (2016).

32. Labaka, A., Goñi-Balentziaga, O., Lebeña, A. & Pérez-Tejada, J. Biological sex differences in depression: a systematic review. Biol. Res. Nurs. 20, 383–392 (2018).

33. Fernandes, N. et al. The impact of SSRs on mortality and cardiovascular events in patients with coronary artery disease and depression: systematic review and meta-analysis. Clin. Res. Cardiol. 110, 183–193 (2020).

34. Klarin, D. et al. Genetic analysis in UK Biobank participants with those of the general population. Am. J. Epidemiol. 186, 1026–1034 (2017).

35. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203–209 (2018).

36. Wain, L. V. et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. Lancet Respir. Med. 3, 769–781 (2015).

37. Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. Bioinformatics 26, 2867–2873 (2010).

38. Finucane, H. K. et al. Partitioning heritability by functional annotation in human complex diseases. Nat. Genet. 39, 1052–1063 (2017).

39. Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. Bioinformatics 26, 2867–2873 (2010).

40. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47, 1228–1235 (2015).

Acknowledgements
This research was conducted under UK Biobank application number 7089. Authors were supported by a CRICO Patient Safety Grant (A.A.S.); the U.S. National Heart, Lung and Blood Institute R01HL133149-04 (A.A.S.), R01HL142711 (P.N.), R01HL148050 (P.N.), R01HL118565 (P.N.), and R01HL151283 (P.N.); Fondation Leducq TNE-18CVD04 (P.N.); and a Hassenfeld Scholar Award from Massachusetts General Hospital (P.N.).

Author contributions
M.C.H., M.J.W. and P.N. conceived and designed the project. Y.Y. and H.Z. provided Annopred-derived polygenic risk scores. M.C.H., Y.Y., L.D., N.S.S., J.W.S., H.Z., M.J.W., and P.N. made substantial contributions to the data analysis and/or interpretation, drafted and/or revised the manuscript, and approved the final version for publication.

Competing interests
P.N. reports grant support from Amgen, Apple, AstraZeneca, Boston Scientific and Novartis, personal fees from Apple, AstraZeneca, Blackstone Life Sciences, Foresite Labs, Genentech and Novartis, and spousal employment at Vertex, all unrelated to the present work. All other authors do not report any disclosures.

Additional information
Extended data The online version contains supplementary material available at https://doi.org/10.1038/s44161-021-00011-7.
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44161-021-00011-7.
Correspondence and requests for materials should be addressed to Pradeep Natarajan.

Peer review information Nature Cardiovascular Research thanks James Brophy, David Hare and the other, anonymous, reviewer for their contribution to the peer review of this work. Andrew Robson was the primary handling editor.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2022.
Extended Data Fig. 1 | Creation of the study cohort. Genotyped, unrelated European ancestry participants in the UK Biobank with complete available data on self-reported frequency of depressed mood, socioeconomic status, sleep, and health behaviors were included. CAD = coronary artery disease.
Extended Data Fig. 2 | Additive associations of frequency of depressed mood and polygenic risk for incident coronary artery disease and type 2 diabetes mellitus. Spline plots of crude cumulative incidence (right) were generated with the ‘ggplot2’ package in R version 3.6.0 using a cubic spline with 3 knots. The colored lines represent the modeled cumulative incidence at each polygenic risk percentile, and the shaded bands represents the modeled 95% confidence bands.
Extended Data Fig. 3 | Reduction in cardiometabolic risk associated with lower frequency of depressed mood among individuals at high polygenic risk. Data are presented as hazard ratios (black squares) and 95% confidence intervals (error bars). Multivariable-adjusted hazard ratios are calculated among 60,849 individuals without prevalent coronary artery disease (2,510 with high, 11,305 with moderate, and 47,034 with low frequency of depressed mood); 62,974 individuals without prevalent type 2 diabetes (2,520 with high, 11,622 with moderate, and 48,832 with low frequency of depressed mood); and 63,414 individuals without prevalent atrial fibrillation (2,573 with high, 11,683 with moderate, and 49,158 with low frequency of depressed mood). Hazard ratios (left) are adjusted for age, age², sex, the first 20 principal components of ancestry, genotyping array, country, socioeconomic deprivation, smoking status, pack-year smoking history, alcohol intake, vegetable and fresh fruit intake, days per week of moderate and vigorous exercise, sleep duration, systolic blood pressure, antihypertensive medication use, non-HDL cholesterol, cholesterol-lowering medication use, antiplatelet medication use, antihyperglycemic medication use, prevalent type 2 diabetes mellitus (models for coronary artery disease and atrial fibrillation only), body-mass index, and C-reactive protein.
Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a | Confirmed

☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement

☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly

☐ ☒ The statistical test(s) used AND whether they are one- or two-sided

Only common tests should be described solely by name; describe more complex techniques in the Methods section.

☒ ☒ A description of all covariates tested

☒ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons

☒ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)

☒ ☒ For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted

Give P values as exact values whenever suitable.

☒ ☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

☒ ☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes

☒ ☒ Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about: availability of computer code

Data collection

No software was used to collect data for this study.

Data analysis

R version 3.6.0 was used to analyze data for this study. Base R functionalities and the ‘survival’, ‘fmsb’, and ‘ggplot2’ packages were used to perform analyses. Code for these packages is publicly available.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about: availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

UK Biobank data are available to researchers by application [https://www.ukbiobank.ac.uk]. Polygenic risk scores used in these analyses for coronary artery disease, type 2 diabetes, and atrial fibrillation have been submitted to the Polygenic Score Catalog.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

**Sample size**
Sample size (N=328,152) was determined by the number of genotyped UK Biobank participants who met inclusion criteria. Extended Data Figure 1 summarizes the creation of the study cohort.

**Data exclusions**
Based this study involved polygenic risk scores derived from genetic discovery efforts performed in large European-ancestry cohorts, and because the vast majority of the UK Biobank is of European ancestry, the study was restricted to European-ancestry participants. Participants were further excluded if they were missing key behavioral/lifestyle covariates or already had previously developed all study outcomes at the time of study enrollment.

**Replication**
In keeping with the epidemiologic nature of the analysis and given the unique size and phenotyping of the UK Biobank not available elsewhere, replication in external cohorts was not performed.

**Randomization**
This observational study did not involve randomization.

**Blinding**
As this observational study did not involve an intervention, there was no blinding.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | Methods |
|---------------------------------|---------|
| n/a                             | Involved in the study |
| ☑ Antibodies                    | ☑ ChIP-seq |
| ☑ Eukaryotic cell lines         | ☑ Flow cytometry |
| ☑ Palaeontology and archaeology | ☑ MRI-based neuroimaging |
| ☑ Animals and other organisms   |         |
| ☑ Human research participants   |         |
| ☑ Clinical data                 |         |
| ☑ Dual use research of concern  |         |

Human research participants

Policy information about studies involving human research participants

**Population characteristics**
The UK Biobank involved >500,000 UK adult residents aged 40-69 years old at the time of recruitment from 22 regional centers between 2006-2010.

**Recruitment**
Recruitment occurred at 22 centers throughout the UK between 2006-2010. Previous work has documented a "healthy-participant bias" whereby the UK Biobank population is healthier on average than the UK population as a whole; however, despite this known limitation, the very large size of the cohort is felt to yield valid exposure-outcome associations (Fry et al., Am J Epidemiol 2017).

**Ethics oversight**
The North West Multi-centre Research Ethics Committee approved the UK Biobank, and all subjects provided written informed consent. The Massachusetts General Hospital institutional review board approved secondary analysis of UK Biobank for the present analysis.

Note that full information on the approval of the study protocol must also be provided in the manuscript.
### Clinical data

Policy information about clinical studies: All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

| Clinical trial registration | Not applicable. |
|----------------------------|------------------|
| Study protocol             | Details of the UK Biobank are available at [https://www.ukbiobank.ac.uk](https://www.ukbiobank.ac.uk). |
| Data collection            | Baseline data were collected by the UK Biobank between 2006-2010. Follow-up data has been generated on an ongoing basis largely via linkage to UK NHS records as well as follow-up study visits. |
| Outcomes                   | The 3 co-primary study outcomes were incident (newly diagnosed) coronary artery disease, type 2 diabetes mellitus, and atrial fibrillation. Outcomes were ascertained by the appearance of a qualifying International Classification of Diseases of procedural code in the subject's medical record. |