Broad Clinical Manifestations of Polygenic Risk for Coronary Artery Disease in the Women's Health Initiative

Shoa L. Clarke, MD, PhD,[1,2*] Matthew Parham, MS,[3*] Joanna Lankester, PhD,[1,2] Aladdin H. Shadyab, PhD,[4] Simin Liu, MD, ScD,[5] Charles Kooperberg, PhD,[6] JoAnn E. Manson, MD, DrPH,[7] Catherine Tcheandjieu, DVM, PhD,[8,9] Themistocles L. Assimes, MD, PhD[1,2,3]

[1]VA Palo Alto Health Care System, Palo Alto, CA
[2]Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA
[3]Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, CA
[4]Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego, La Jolla, CA
[5]Department of Epidemiology, Brown University, Providence, RI
[6]Fred Hutchinson Cancer Research Center, Seattle, WA
[7]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA
[8]Gladstone Institute, San Francisco, CA
[9]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA
* These authors contributed equally

Corresponding Author
Themistocles L. Assimes, MD, PhD
1070 Arastradero Road, Suite 300, Palo Alto, CA 94304
Phone: 650-498-4154
tassimes@stanford.edu

Abstract

Background: The genetic basis for coronary artery disease (CAD) risk is highly complex. Genome-wide polygenic risk scores (PRS) can help to quantify that risk, but the broader impacts of polygenic risk for CAD are not well characterized.

Methods: We measured polygenic risk for CAD using the metaGRS, a previously validated genome-wide PRS, in a subset of genotyped participants from the Women's Health Initiative (WHI) and applied a phenome-wide association study framework to assess associations between the PRS and broad range of blood biomarkers, clinical measurements, and outcomes.

Results: Polygenic risk for CAD was associated with a variety of biomarkers, clinical measurements, behaviors, and diagnoses related to traditional risk factors, as well as risk-enhancing factors. Analysis of adjudicated outcomes showed a graded association between atherosclerosis related outcomes, with the highest odds ratios being observed for the most severe manifestations of CAD. We identified novel associations between increased polygenic risk for CAD and decreased risk for incident breast and lung cancer, with replication of the breast cancer finding in an external cohort. Genetic correlation and two-sample Mendelian randomization suggest that breast cancer association is likely due to horizontal pleiotropy, while the association with lung cancer may be causal.

Conclusion: Polygenic risk for CAD has broad clinical manifestations, reflected in biomarkers, clinical measurements, behaviors, and diagnoses. Some of these associations may represent direct pathways between genetic risk and CAD while others may reflect pleiotropic effects independent of CAD risk.

Introduction

Coronary artery disease (CAD) is a complex phenotype, and the genetic basis for CAD risk is similarly complex.[1] To date, >200 loci have been implicated in CAD risk through genome-wide association studies (GWAS).[2,3] These loci interact through a diverse set of biological pathways, and many loci have no apparent relevance to traditional risk factors for CAD. Furthermore, genetic variants that associate with CAD also associate with other phenotypes, suggesting extensive underlying pleiotropy.[2,3] The complexity of genetic risk for CAD is further highlighted by recent advances in the construction of polygenic risk scores (PRS). Contemporary scores that incorporate variants across the whole genome, including variants outside of known CAD loci, outperform scores that are constructed only from variants at known CAD loci.[4,5] Studying such genome-wide PRS for CAD may allow for improved understanding of the genetic basis for CAD risk and new insights into the implications polygenic risk beyond CAD.

One approach to assessing the impact of polygenic risk for CAD has been to measure associations between a CAD PRS and biobank-derived phenotypes.[6,7] A primary advantage of this approach is the large number of participants in such biobanks. However, a limitation of this method is lack of precision for some outcomes, particularly those inferred from electronic health records. Further, biobank studies have typically combined prevalent and incident disease and may have limited follow up after enrollment. Thus, a complementary approach to biobank analyses is to examine well-phenotyped longitudinal cohorts.

Here, we sought to identify traits and outcomes associated with polygenic risk for CAD by taking advantage of the high-quality data collected as part of the Women's Health Initiative (WHI). We aggregated data collected over ~25 years as part of either clinical trials or the observational study within WHI. We thus measured the association between polygenic risk for

CAD and blood biomarkers, clinical measurements, clinical risk scores/questionnaires, self-reported medical history, and incident adjudicated outcomes related to cardiovascular disease, cancer, and death.

Methods

Study cohort

The main study cohort was selected from WHI. The design and recruitment strategy for WHI has been previously described.[8,9] Briefly, postmenopausal women aged 50 to 79 years were enrolled at 40 sites across the United States from 1993 to 1998. Each participant was enrolled into either a clinical trial (n = 68,132) or an observational study (n = 93,676). Two successive extension studies continued follow-up of consenting participants from 2005 to 2010 and from 2010 to the present. A subset of participants who were primarily non-Hispanic white by self-report have been previously genotyped as part of 6 ancillary GWAS (**Supplementary Table 1**). Participants from those 6 GWAS were considered for inclusion in this study. Because currently validated genome-wide PRS were developed in European populations and do not transfer well to non-European populations, we did not include cohorts of primarily non-European genotyped participants in this study. Subjects with a known or likely history of atherosclerotic cardiovascular disease (ASCVD) at enrollment were excluded (**Supplementary Table 2**). We used the UK Biobank for replication of select results. The UK Biobank cohort consisted of unrelated women of European ancestry with no history of MI or stroke at enrollment.

Genotyping and imputation

Genotyping was performed with early versions of Affymetrix and Illumina gene chips for five of

the GWAS cohorts contributing to this study. For these five studies, harmonization and imputation to the 1000 Genome reference panel was previously performed as part of the WHI GWAS Harmonization and Imputation Project (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000746.v3.p3). Participants of the sixth study were genotyped with the Oncochip, and we imputed these data to the 1000 Genome reference panel using the Michigan Imputation Server.[10]

Main Exposure

We used metaGRS, a previously developed genome-wide PRS for CAD, to estimate each participant's genetic risk.[5] This score consists of ~1.7 million autosomal variants. Participants in our study cohort did not contribute to the GWAS used to construct this score. Each participant's total score was calculated using Plink 2.0, and raw scores were then scaled to mean 0 and standard deviation 1. This standardize score was used as the primary exposure.

Phenotypes

Quantitative measurements were largely collected at enrollment and included laboratory values, clinical measurements, and clinical scores. For the small number of lab measurements not collected at baseline, we used the earliest available measurement. Lab outliers were removed by excluding the top 1% of values for each biomarker. For clinical measurements such as blood pressure, the mean value was used if serial measurements were available within one research clinic visit. Self-reported medical history, medication usage, social/behavioral history, and family history was obtained through questionnaires collected primarily at enrollment but also during annual follow up mailings. Adjudicated outcomes assessed in this study include incident

cardiovascular diseases, incident cancers, and death. Annual questionnaires were completed by participants or their proxies in order to identify hospitalizations, and for each hospitalization, medical records were obtained and adjudicated by physicians using standardized criteria.[11] Deaths were further ascertained through the National Death Index. For UK Biobank analyses, cancer diagnoses were extracted from the UK cancer registry. For each cancer, only first diagnoses after enrollment were considered as incident cases, and subjects with prevalent disease at enrollment were excluded.

Statistical analysis

We selected the largest subset of subjects with similar inferred genetic ancestry using principal components analysis in order to limit confounding by population substructure. We used linear and logistic regression to estimate associations between each trait/outcome and the CAD PRS per standard deviation increase in the PRS. For each of the adjudicated outcome, we appropriately censored subjects at the end of the follow up time period where formal adjudication ended for the outcome. For death outcomes, we used Cox analysis with time zero being the time of enrollment. For each cause of death that was examined, non-cases were censored at time of death from another cause or time of last follow-up if not deceased.

Each model was adjusted for age at enrollment (or age at time of measurement for lab values), study type (clinical trial versus observational study), and genotyping platform. Associations with lipid-related labs, diabetes-related labs, and for blood pressure were additionally adjusted for self-reported cholesterol medication use, diabetes medication use, and hypertension medication use respectively. All associations with lab values were also adjusted for the assay version if more than one assay was used. The logistic regression analysis of adjudicated

cardiovascular outcomes and the Cox analysis of death outcomes were adjusted for smoking status, self-reported diabetes at baseline, systolic blood pressure, low-density lipoprotein cholesterol (LDL-C), and high-density lipoprotein cholesterol (HDL-C). For cancer outcomes, we assessed the association with and without adjustment for risk factors. The risk factor adjusted model include adjustment for smoking status, alcohol consumption, physical activity (MET-hours per week), Alternative Healthy Eating Index score, and body-mass index (BMI).

Using a phenome-wide association study framework, we consider statistical significance in three ways. Nominal significance is defined as a p-value ≤ 0.05. Where indicated, we also identify associations that are significant by a false-discovery rate (FDR) q-value ≤ 0.05, using the Benjamini and Hochberg method. Lastly, Bonferroni significance is defined as 0.05 divided by the number of association tests performed for the given analysis. For the association analysis of quantitative traits, Bonferroni significance was p-value $\leq 9.2 \times 10^{-5}$ (0.05/546). For the association analysis of incident cardiovascular diseases, Bonferroni significance was p-value ≤ 0.003 (0.05/17).

We used published summary statistics from GWAS of CAD,[2] breast cancer,[12] and lung cancer[13] to estimate genetic correlations. For CAD, only variants with INFO score >0.9 were included. For breast and lung cancer, INFO score was not available, and thus only HapMap3 variants were included, as these variants are generally well imputed. We used *ldsc* (version 1.01) to perform genetic correlation analyses.[14]

We performed two-sample Mendelian randomization using the MRBase tool with default settings.[15] We created a genetic instrument variable for CAD using the same GWAS as was used in the genetic correlation analysis.[2] We selected genome-wide significant SNPs that were determined to be independent using a clumping distance of 10 megabases with a linkage

disequilibrium R^2 threshold of 0.001. These were then harmonized to the summary statistics of each outcome, excluding palindromic SNPs and using proxies for missing SNPs only if the LD R^2 was ≥ 0.9. For lung cancer, the instrument variable consisted of 125 SNPs, of which 1 SNP was proxied. For breast cancer, the instrument variable consisted of 124 SNPs, of which none were proxied. MRBase was used to perform inverse variance weighted, weighted median, and MR Egger studies. As additional sensitivity analysis, and to test for horizonal pleiotropy, we used MR PRESSO.[16]

WHI analyses were performed using SAS 9.4 (SAS Enterprise). UK Biobank analyses, meta-analysis, Mendelian randomization, and plots were done with R version 3.5.1 (R Foundation, Vienna, Austria). All odds ratios (OR) and hazard ratios (HR) are reported as per standard deviation increase in the PRS.

Results

We identified 25,789 subjects who had undergone genotyping as part of prior GWAS within the WHI (**Supplementary Table 1**). When plotting the first two principal components, we noted a cluster of 472 subjects from the GECCO study who were clear outliers (**Supplementary Figure 1A**). The similar shape between the main cluster and the outliers suggested a batch effect leading to a systematic bias in genotyping calls. These subjects were removed. We then used the Mahalanobis distance[17] in the remaining subjects to identify a central cluster with similar genetically inferred ancestry (**Supplementary Figure 1B**). The majority of these subjects self-reported as non-Hispanic white. Lastly, we excluded 2,830 subjects (11%) with known or likely ASCVD at enrollment (**Supplementary Table 2, Supplementary Figure 2**). The remaining cohort of 21,863 subjects showed an enrichment for health traits and outcomes reflective of the

genotyping strategy of the parent WHI GWAS, which targeted genotyping for outcomes of interest (**Supplementary Table 3**). Polygenic risk for CAD was quantified in each of these participants using a validated genome-wide PRS for CAD.[5]

Association of polygenic risk for CAD and quantitative measurements

We identified 454 blood-based laboratory biomarkers for assessment with PRS after excluding biomarkers with fewer than 100 observations. Lab biomarkers were categorized as being related to lipids (n = 84), diabetes (n = 7), hormones (n = 93), inflammation (n= 62), hematology (n = 38), or other (n = 170). We further identified 48 clinical exam measurements, 31 quantitative traits reported by questionnaire, and 13 clinical scores. Associations with lipid-related labs, diabetes-related labs, and with blood pressure measurements were adjusted for cholesterol medication use, diabetes medication use, and blood pressure medication use respectively.

Polygenic risk for CAD associated with traits related to traditional risk factors and the metabolic/insulin resistance syndrome. For example, women with a higher PRS tended to have higher systolic blood pressure, larger waist-to-hip ratios, higher fasting insulin, higher LDL-C, higher triglycerides and lower HDL-C (**Figure 1, Supplementary Data**). Subjects with a higher PRS also reported less healthy diets. Among lipid measurements, lipoprotein(a) [Lp(a)] showed the most significant association with the CAD PRS. This observation may reflect the very high genetic heritability of Lp(a) levels.[18] Across multiple lab categories, we observed associations with biomarkers known or hypothesized to relate to CAD risk, including sex hormone binding globulin,[19] leptin,[20] hematocrit,[21] and hepatocyte growth factor.[22] We also observed a negative association with height, corroborating prior reports that genetically determined shorter stature is associated with a higher risk for CAD.[23] Analysis of questionnaire data demonstrated that

women with higher polygenic risk for CAD tended to report a younger age of their father's and/or mother's death, and they reported experiencing menopause at a younger age. Interestingly, higher polygenic risk for CAD was associated with lower clinically predicted risk for breast cancer using the Gail breast cancer risk model.[24]

Association of polygenic risk for CAD and self-reports

We aggregated data from structured questionnaires administered at baseline and during regular annual follow up in order to measure the association between polygenic risk for CAD and social/behavioral history, family history, medication usage, and self-reported medical history present at baseline or reported during follow-up. In order to identify prevalent associations independent of the development of CAD after enrollment, we also performed the analysis in a subset of participants with no self-reported or adjudicated CAD at the most recent follow-up (n = 18,044). Only outcomes with at least 100 cases among the CAD-free cohort were considered, resulting in a total of 128 self-reported qualitative variables. **Figure 2** shows those outcomes which are significant based on a FDR q-value ≤ 0.05 in either of the two sets. The complete results are shared in the **Supplementary Data**.

We observed associations between increased polygenic risk for CAD and known risk factors for CAD, in both the full set and the subset with no CAD. For example, among women free of CAD, a higher PRS was associated with a higher likelihood of reporting hypertension (OR 1.2, 95% CI 1.16-1.24) hypercholesterolemia (OR 1.17, 95% CI 1.12-1.23), rheumatoid arthritis (OR 1.11, 95% CI 1.03-1.19), and family history of myocardial infarction (OR 1.16, 95% CI 1.13-1.20) or stroke (OR 1.07, 95% CI 1.04-1.11). We also observed an interesting association with smoking. In both the full cohort and the CAD-free cohort, subjects with

increased polygenic risk for CAD were slightly less likely to have ever smoked. However, among women who reported having ever smoked, a higher PRS was associated with a higher likelihood of being a current smoker (**Figure 2**). Possibly related to the association with continued smoking into later adulthood, subjects with increased polygenic risk for CAD were more likely to report a diagnosis of emphysema. Beyond known risk factors, we saw evidence that the genetic drivers of CAD risk may also impact risk for other diseases, including venous thromboembolism (VTE), thyroid disease, and gallbladder-related disease.

We detected an inverse association of the CAD PRS with self-reported cancer-related outcomes. Women with higher polygenic risk for CAD were less likely to report a history of breast cancer (OR 0.81, 95% CI 0.69-0.95) or non-melanoma skin cancer (OR 0.93, 95% CI 0.89-0.98). They were also less likely to report family history of colon cancer (OR 0.95, 95% CI 0.91-0.99), which may in part explain their lower likelihood of having ever undergone a colonoscopy (OR 0.96, 95% CI 0.93-0.99). These associations did not attenuate when analyzing the subset of CAD-free women (**Figure 2**).

Association of polygenic risk for CAD and incident cardiovascular diseases

We next aimed to measure the impact of polygenic risk for CAD on incident cardiovascular disease, independent of traditional risk factors. Using high-quality adjudicated outcomes, we examined the various manifestations of CAD, non-CAD atherosclerotic cardiovascular disease, and other non-atherosclerotic cardiovascular disease. In total, we considered 17 cardiovascular outcomes with at least 100 incident cases among our study cohort. These outcomes represent first-presentation incident events. We adjusted for smoking status, diabetes, systolic blood pressure, LDL-C, and HDL-C. As expected, outcomes related to CAD showed the strongest

associations with the PRS. The more severe manifestations of CAD including myocardial infarction and the need for coronary revascularization demonstrated the largest effect sizes (**Figure 3**). A similarly strong association was seen for the first presentation of hospitalized angina ("All angina"). However, the majority of such cases were treated with coronary revascularization. Angina without revascularization demonstrated a comparably weak association that did not reach nominal statistical significance. Stroke also demonstrated a clear association with polygenic risk for CAD, though with weaker effect sizes compared to CAD-related outcomes. The association with stroke was driven by the ischemic subtype. We observed no association observed with hemorrhagic stroke. The significant association previously seen with VTE in the self-reported prevalent outcomes (**Figure 2**) was not reflected in the adjudicated incident outcomes for pulmonary embolus of deep vein thrombosis (**Figure 3**).

Results for a minimally adjusted model (adjusted only for age and genotype platform) are shown in **Supplementary Table 4** for comparison. Consistent with prior studies,[5] adjusting for ASCVD risk factors only slightly attenuates the strength of the PRS associations.

Association of polygenic risk for CAD and incident cancers

Results from our analyses of clinical scores and self-reported outcomes suggested the possibility of a protective association between polygenic risk for CAD and cancer. To better explore this finding, we tested the association between the CAD PRS and adjudicated first-occurrence incident cancers. We tested 17 cancers for which at least 100 incident cases occurred in our cohort. We found nominally significant protective associations with the aggregate outcome of any cancer (OR 0.96, 95% CI 0.93-0.99, p 0.008) and with the specific outcome of lung cancer (OR 0.91, 95% CI 0.84-0.99, p 0.02). For breast cancer, the association was borderline nominally

significant (0.96, 95% CI 0.93-1.00, p 0.05). Other cancers did not have significant associations, but the OR showed a general trend for being less than 1 (**Supplementary Figure 3**). After adjusting for cancer risk factors (smoking status, alcohol consumption, weekly physical activity, dietary health measured by the alternative healthy eating index, and BMI), the associations with any cancer (OR 0.96, 95% CI 0.93-0.99, p 0.02), lung cancer (OR 0.91, 95% CI 0.83-0.99, p 0.02), and breast cancer (0.96, 95% CI 0.92-1.00, p 0.04) were virtually unchanged. We also tested for genotyping batch effects by repeating the analysis in the subset of women who were genotyped with the Oncochip. We found consistent effect sizes. Though, given substantially decreased sample sizes, the p-values were no longer significant (**Supplementary Table 5**).

We next sough to test whether the associations with breast cancer and lung cancer replicate in an external cohort. We used data from the UK Biobank to test the association between the CAD PRS and incident first-occurrence breast cancer or lung cancer among women with no history of MI or stroke at baseline. The association replicated for breast cancer but not for lung cancer. Using a random effects meta-analysis, only the breast cancer association remained significant (**Figure 4**).

The observed inverse association between the CAD PRS and cancer outcomes may reflect several factors. One possibility is that CAD and cancer have some shared genetic architecture but with opposing effects. To test this hypothesis, we performed genetic correlations using linkage disequilibrium score regression with summary statistics of previously published GWAS for each outcome. We found a small but significant negative genetic correlation between CAD and breast cancer (rg -0.054, p 0.02) but no significant genetic correlation between CAD and lung cancer (rg 0.053, p 0.4). We next tested the hypothesis that CAD is causally protective for breast and lung cancer using two-sample Mendelian randomization. Inverse variance

weighted analysis of independent genome-wide significant SNPs suggested that CAD is causally protective for both breast cancer (OR 0.95, 95% CI 0.92-0.99, p 0.009) and lung cancer (OR 0.91, 95% CI 0.84-1.00, p 0.05). However, only the association with lung cancer was robust to sensitivity analyses with additional methods for Mendelian randomization, including weighted median, MR Egger,[25] and MR PRESSO[16]. Notably, the MR PRESSO global test detected significant horizontal pleiotropy for both the breast cancer and lung cancer analyses. After using the MR PRESSO correction for horizontal pleiotropy through outlier removal, the causal association between CAD and breast cancer was no longer significant, but the association between CAD and lung cancer remained significant (**Table**). Overall, these analyses suggest that the relationship between CAD and breast cancer is not causal, and the associations we observed in this study likely reflect shared genetic architecture. In contrast, the relationship between CAD and decreased risk for lung cancer may have a causal component.

Association of polygenic risk for CAD and causes of death

We used a time-to-event analysis, accounting for competing events, to determine the impact of polygenic risk for CAD on cause of death. In total, 11,734 women died during the follow up period, with 78 distinct causes adjudicated. We measured the association between the CAD PRS and 48 causes of death for which at least 10 cases occurred among women with sufficient data to adjust for cardiovascular disease risk factors (smoking status, self-reported diabetes at baseline, systolic blood pressure, LDL-C, and HDL-C). **Figure 5** shows all death outcomes that showed nominal significance. The complete results are available in the **Supplementary Data**. The strongest association occurred with 'definite' coronary heart disease death (HR 1.29, 95% CI 1.16-1.43). Conversely, there was no association with the outcome of

‘possible’ coronary heart disease death (HR 0.99, 95% CI 0.91-1.07), suggesting low specificity of that outcome. The magnitude of the association with unknown cause of death suggests that many of these deaths may have been secondary to ASCVD. Despite the observed inverse association with incident lung cancer and breast cancer, we did not find lower risk of death from either cancer. For lung cancer death, the HR was 0.93 (95% CI 0.85-1.02, p 0.14). For breast cancer death, the HR was 1.09 (95% CI 0.99-1.19, p 0.06). However, we did appreciate a nominally significant decreased risk for death from brain cancer and uterine cancer (**Figure 5**), and similar to our analysis of incident cancer, we observed a trend for HR<1 for cancer deaths (**Supplementary Data**).

Discussion

We have shown that polygenic risk for CAD, as quantified by a genome-wide PRS, has broad clinical manifestations in post-menopausal women. In addition to the expected association with CAD and other ASCVD outcomes, we observed associations with biomarkers, clinical measurements, behaviors, and disorders that are known to be risk factors for atherosclerosis. Recent work demonstrated an association between a 300-variant CAD PRS and traditional risk factors among participants of the UK Biobank.[6] Our analyses corroborate those findings and expand on them substaintially by leveraging a genome-wide PRS and an extensively phenotyped population that includes exquisite adjudication for multiple outcomes in the setting of prolonged follow up.

Beyond traditional risk factors, we find that polygenic risk for CAD associates with risk-enhancing features that are defined in ASCVD prevention guidelines,[26] including central adiposity, elevated Lp(a), and rheumatoid arthritis. We also highlight the heritable nature of

polygenic risk through clear associations with early age of parental death as well as family history of MI and stroke. Notably, we find that polygenic risk for CAD associates with behaviors often referred to as "lifestyle risk factors", including dietary health and persistent smoking. Some consider lifestyle risk as "environmental," but our findings indicate that genetics may influence behaviors that impact one's exposure to such risk factors. Although healthy lifestyle can help to mitigate polygenic risk for CAD,[27,28] the fact that genetic risk might impact lifestyle raises additional questions for exploration. It has been shown that interaction effects between polygenic risk and behaviorally mediated environmental exposures can exist.[29] It is possible that as PRS become more complex, such "gene by environment" interactions could become more influential in score behavior. These interactions have implications on the estimation of risk as the effect of alleles predisposing to such risk-related behaviors can only be expressed when the environmental factor is present. Subjects possessing high risk variants but never exposed to the adverse environment may have misspecification of their risk.

We found the effect sizes per standard deviation of CAD PRS for the most severe incident manifestations of CAD (i.e. myocardial infarction and coronary revascularization) to be consistent with the published literature for the same CAD PRS in validation cohorts of European-ancestry men and women combined.[5,30] One recent study using a different PRS documented heterogeneity in effects sizes between the sexes but did not report or adjust for differences in the severity of disease at presentation among males and females.[31] Given the substantially lower effect sizes we observed for an angina-only presentation, it is possible that heterogeneity of a PRS between any two groups can be influenced by differences in the case-mix of the severity/type of CAD at presentation. Collectively, the data to date suggest that a large majority of the CAD loci incorporated into the PRS affect a woman's risk of presenting with

CAD to the same degree as they do men, even if the average age of presentation may be up to a decade later for women. Importantly, our study offers the opportunity to identify associations that may be specific to women. For example, we found that a higher CAD PRS was associated with younger age of menopause. Observational data has shown that early menopause is a risk factor for CAD,[32,33] but recent Mendelian randomization suggests that this relationship is not causal.[34] Thus, our observed association between polygenic risk for CAD and age of menopause likely reflects shared heritable risk factors between CAD and early menopause.

Somewhat unexpectedly, several associations suggest that increased polygenic risk for CAD decreases the risk for cancer. Women with higher PRS had a lower Gail breast cancer risk score, and they were less likely to report prevalent breast cancer, non-melanoma skin cancer, or a family history of colorectal cancer. These findings were further corroborated by our analysis of incident adjudicated cancers, where we observed protective associations between polygenic risk for CAD and incident breast and lung cancer. The protective association with incident breast cancer replicated in the UK Biobank, and recently others have also replicated the association in other biobanks.[35] Our genetic correlation and Mendelian randomization studies suggest that this relationship between CAD and breast cancer is not causal but rather driven by horizontal pleiotropy. An overlapping genetic architecture for CAD and cancer is further supported by other observations. For example, we observed an association with higher polygenic risk for CAD and shorter stature, consistent with prior reports that genetically taller stature is associated with a lower risk for CAD and a higher risk of cance.[23,36] Despite pleiotropy, it is also plausible that CAD is causally protective against some cancers. Indeed, our Mendelian randomization studies of lung cancer consistently suggested a causal relationship. One mechanism could be that the clinical management of CAD leads to interventions that are protective for some cancers.

Engagement with the healthcare system, behavior changes, and treatments (e.g. aspirin[37,38] or statin[39]) could all potentially protect against incident cancer.

An important limitation of our analysis is that our study population is predominantly white by self-report. To date, CAD PRS have been primarily developed from GWAS in people of European ancestry, and they have been optimized for application to European-ancestry cohorts. This shortcoming remains a barrier to more broadly studying polygenic risk for CAD in diverse populations. With recent efforts to improve representation in CAD GWAS,[7] we hope that future PRS research will expand to address this limitation. A second limitation of this work is that most of our analyses are correlative are primarily hypothesis-generating and/or hypothesis-supporting. Additional work across independent cohorts and with larger sample sizes is needed to further understand the relationships between CAD and cancer outcomes.

In conclusion, polygenic risk for CAD is associated with a broad spectrum of phenotypes. Many of these associations likely reflect the complex pathophysiology of CAD risk, while others may reflect pleiotropic effects beyond CAD. Among WHI participants, higher polygenic risk for CAD was associated with increased risk for multiple forms of incident ASCVD, decreased risk for incident cancer, and increased risk of all-cause mortality. Overall, post-menopausal women with higher polygenic risk for CAD were less likely to live into old age.

Acknowledgments

The authors thank the WHI (Women's Health Initiative) participants, clinical sites, investigators, and staff for their dedicated efforts.

Funding Sources

The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts 75N92021D00001, 75N92021D00002, 75N92021D00003, 75N92021D00004, 75N92021D00005.

Disclosures

None

References

1. Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. Nat Rev Genet 2017;18:331–344. doi:10.1038/nrg.2016.160.
2. van der Harst P, Verweij N. Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. Circ Res 2018;122:433–443. doi:10.1161/CIRCRESAHA.117.312086.
3. Koyama S, Ito K, Terao C, Akiyama M, Horikoshi M, Momozawa Y, et al. Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. Nat Genet 2020;52:1169–1177. doi:10.1038/s41588-020-0705-3.
4. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet 2018;50:1219–1224. doi:10.1038/s41588-018-0183-z.
5. Inouye M, Abraham G, Nelson CP, Wood AM, Sweeting MJ, Dudbridge F, et al. Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. J Am Coll Cardiol 2018;72:1883–1893. doi:10.1016/j.jacc.2018.07.079.
6. Ntalla I, Kanoni S, Zeng L, Giannakopoulou O, Danesh J, Watkins H, et al. Genetic Risk Score for Coronary Disease Identifies Predispositions to Cardiovascular and Noncardiovascular Diseases. J Am Coll Cardiol 2019;73:2932–2942. doi:10.1016/j.jacc.2019.03.512.
7. Tcheandjieu C, Zhu X, Hilliard A, Clarke SL, Napolioni V, Ma S, et al. A large-scale multi-ethnic genome-wide association study of coronary artery disease. Preprint 2021;Research Square. doi:10.21203/rs.3.rs-275591/v1.
8. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. Control Clin Trials 1998;19:61–109. doi:10.1016/s0197-2456(97)00078-0.
9. Hays J, Hunt JR, Hubbell FA, Anderson GL, Limacher M, Allen C, et al. The Women's Health Initiative recruitment methods and results. Ann Epidemiol 2003;13:S18-77. doi:10.1016/s1047-2797(03)00042-5.
10. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet 2016;48:1284–1287. doi:10.1038/ng.3656.

11. Curb JD, McTiernan A, Heckbert SR, Kooperberg C, Stanford J, Nevitt M, et al. Outcomes ascertainment and adjudication methods in the Women's Health Initiative. Ann Epidemiol 2003;13:S122-128. doi:10.1016/s1047-2797(03)00048-6.
12. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. Nature 2017;551:92–94. doi:10.1038/nature24284.
13. Wang Y, McKay JD, Rafnar T, Wang Z, Timofeeva MN, Broderick P, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. Nat Genet 2014;46:736–741. doi:10.1038/ng.3002.
14. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet 2015;47:1236–1241. doi:10.1038/ng.3406.
15. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. ELife 2018;7:e34408. doi:10.7554/eLife.34408.
16. Verbanck M, Chen C-Y, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet 2018;50:693–698. doi:10.1038/s41588-018-0099-7.
17. Luu K, Bazin E, Blum MGB. pcadapt: an R package to perform genome scans for selection based on principal component analysis. Mol Ecol Resour 2017;17:67–77. doi:10.1111/1755-0998.12592.
18. Austin MA, Sandholzer C, Selby JV, Newman B, Krauss RM, Utermann G. Lipoprotein(a) in women twins: heritability and relationship to apolipoprotein(a) phenotypes. Am J Hum Genet 1992;51:829–840.
19. Rexrode KM, Manson JE, Lee I-M, Ridker PM, Sluss PM, Cook NR, et al. Sex hormone levels and risk of cardiovascular events in postmenopausal women. Circulation 2003;108:1688–1693. doi:10.1161/01.CIR.0000091114.36254.F3.
20. Kk K, Sm P, Mj Q. Leptin and cardiovascular disease: response to therapeutic interventions. Circulation 2008;117. doi:10.1161/CIRCULATIONAHA.107.741645.
21. Gagnon DR, Zhang TJ, Brand FN, Kannel WB. Hematocrit and the risk of cardiovascular disease--the Framingham study: a 34-year follow-up. Am Heart J 1994;127:674–682. doi:10.1016/0002-8703(94)90679-3.
22. Bell EJ, Decker PA, Tsai MY, Pankow JS, Hanson NQ, Wassel CL, et al. Hepatocyte growth factor is associated with progression of atherosclerosis: The Multi-Ethnic Study of Atherosclerosis (MESA). Atherosclerosis 2018;272:162–167. doi:10.1016/j.atherosclerosis.2018.03.040.
23. Nelson CP, Hamby SE, Saleheen D, Hopewell JC, Zeng L, Assimes TL, et al. Genetically determined height and coronary artery disease. N Engl J Med 2015;372:1608–1618. doi:10.1056/NEJMoa1404881.
24. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 1989;81:1879–1886. doi:10.1093/jnci/81.24.1879.
25. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol 2015;44:512–525. doi:10.1093/ije/dyv080.
26. Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the

American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation 2019;140:e596–e646. doi:10.1161/CIR.0000000000000678.
27. Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, et al. Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. N Engl J Med 2016;375:2349–2358. doi:10.1056/NEJMoa1605086.
28. Hasbani NR, Ligthart S, Brown MR, Heath AS, Bebo A, Ashley KE, et al. American Heart Association's Life's Simple 7: Lifestyle Recommendations, Polygenic Risk, and Lifetime Risk of Coronary Heart Disease. Circulation 2022;145:808–818. doi:10.1161/CIRCULATIONAHA.121.053730.
29. Hindy G, Wiberg F, Almgren P, Melander O, Orho-Melander M. Polygenic Risk Score for Coronary Heart Disease Modifies the Elevated Risk by Cigarette Smoking for Disease Incidence. Circ Genomic Precis Med 2018;11:e001856. doi:10.1161/CIRCGEN.117.001856.
30. Dikilitas O, Schaid DJ, Kosel ML, Carroll RJ, Chute CG, Denny JA, et al. Predictive Utility of Polygenic Risk Scores for Coronary Heart Disease in Three Major Racial and Ethnic Groups. Am J Hum Genet 2020;106:707–716. doi:10.1016/j.ajhg.2020.04.002.
31. Huang Y, Hui Q, Gwinn M, Hu Y-J, Quyyumi AA, Vaccarino V, et al. Sexual Differences in Genetic Predisposition of Coronary Artery Disease. Circ Genomic Precis Med 2021;14:e003147. doi:10.1161/CIRCGEN.120.003147.
32. Hu FB, Grodstein F, Hennekens CH, Colditz GA, Johnson M, Manson JE, et al. Age at natural menopause and risk of cardiovascular disease. Arch Intern Med 1999;159:1061–1066. doi:10.1001/archinte.159.10.1061.
33. Honigberg MC, Zekavat SM, Aragam K, Finneran P, Klarin D, Bhatt DL, et al. Association of Premature Natural and Surgical Menopause With Incident Cardiovascular Disease. JAMA 2019. doi:10.1001/jama.2019.19191.
34. Lankester J, Li J, Salfati ELI, Stefanick ML, Chan KHK, Liu S, et al. Genetic evidence for causal relationships between age at natural menopause and the risk of aging-associated adverse health outcomes. MedRxiv 2022:2022.01.26.22269835. doi:10.1101/2022.01.26.22269835.
35. Xiao B, Velez Edwards DR, Lucas A, Drivas T, Gray K, Keating B, et al. Inference of causal relationships based on the genetics of cardiometabolic traits and conditions unique to females in >50,000 participants. MedRxiv 2022:2022.02.02.22269844. doi:10.1101/2022.02.02.22269844.
36. Ong J-S, An J, Law MH, Whiteman DC, Neale RE, Gharahkhani P, et al. Height and overall cancer risk and mortality: evidence from a Mendelian randomisation study on 310,000 UK Biobank participants. Br J Cancer 2018;118:1262–1267. doi:10.1038/s41416-018-0063-4.
37. Burn J, Sheth H, Elliott F, Reed L, Macrae F, Mecklin J-P, et al. Cancer prevention with aspirin in hereditary colorectal cancer (Lynch syndrome), 10-year follow-up and registry-based 20-year data in the CAPP2 study: a double-blind, randomised, placebo-controlled trial. Lancet Lond Engl 2020;395:1855–1863. doi:10.1016/S0140-6736(20)30366-4.
38. Guo C-G, Ma W, Drew DA, Cao Y, Nguyen LH, Joshi AD, et al. Aspirin Use and Risk of Colorectal Cancer Among Older Adults. JAMA Oncol 2021;7:428–435. doi:10.1001/jamaoncol.2020.7338.
39. Carter P, Vithayathil M, Kar S, Potluri R, Mason AM, Larsson SC, et al. Predicting the effect of statins on cancer risk using genetic variants from a Mendelian randomization study in the UK Biobank. ELife 2020;9:e57191. doi:10.7554/eLife.57191.

Tables

Table. Two-sample Mendelian randomization analyses testing for a causal association between the exposure of coronary artery disease and cancer outcomes.

Outcome	Method	OR (95% CI)	P Value
Breast cancer	Inverse variance weighted	0.95 (0.92-0.99)	0.009
	Weighted median	0.96 (0.92-1.01)	0.1
	MR Egger	0.97 (0.90-1.04)	0.4
	MR PRESSO raw	0.96 (0.93-1.00)	0.04
	MR PRESSO outlier-corrected	0.98 (0.95-1.02)	0.4
Lung cancer	Inverse variance weighted	0.92 (0.84-1.00)	0.05
	Weighted median	0.84 (0.75-0.94)	0.002
	MR Egger	0.77 (0.65-0.91)	0.003
	MR PRESSO raw	0.91 (0.84-0.99)	0.04
	MR PRESSO outlier-corrected	0.92 (0.86-0.99)	0.04

Figures

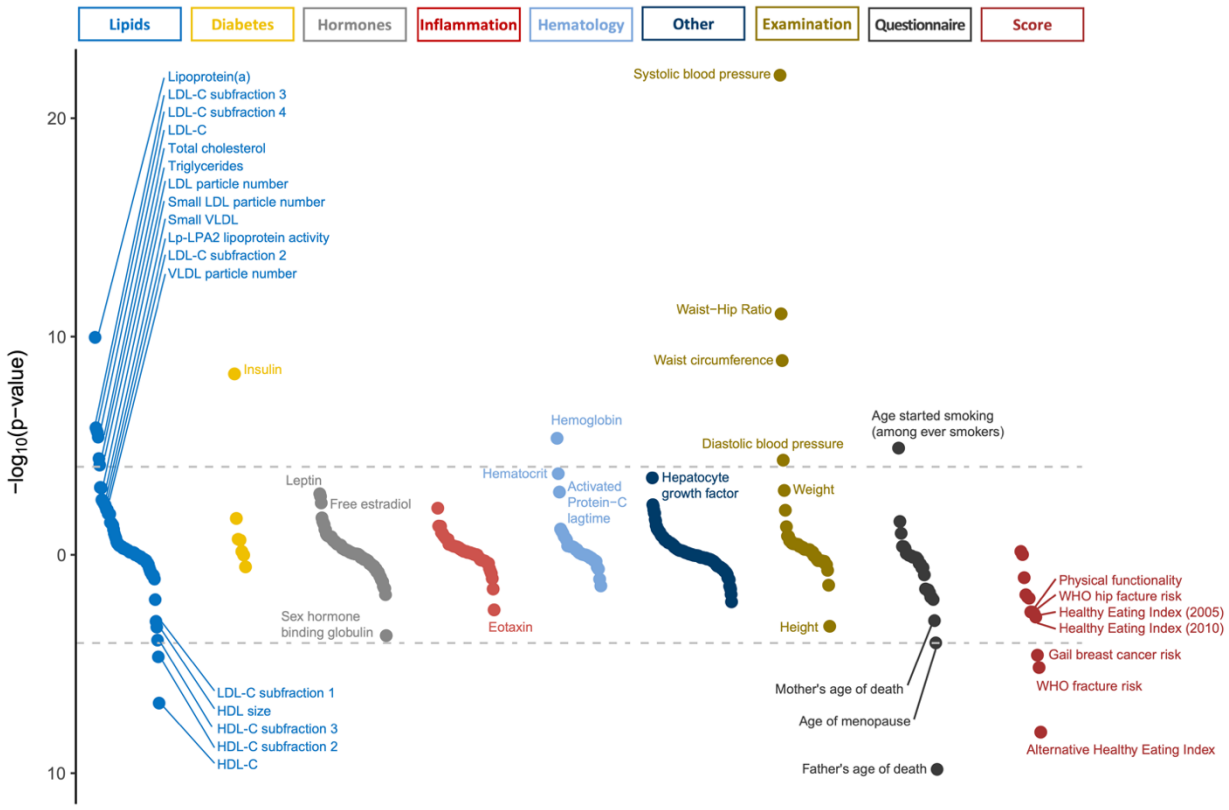

Figure 1. Associations between polygenic risk for coronary artery disease and quantitative traits derived from lab values, clinical exam, self-report, and clinical scores in the Women's Health Initiative. Positive associations are plotted in the up direction and negative associations are plotted in the down direction. Associations that are significant with FDR q-value ≤ 0.05 are labeled. The horizontal gray dashed lines represent the Bonferroni significant p-value < 9.2×10^{-5} (0.05/546).

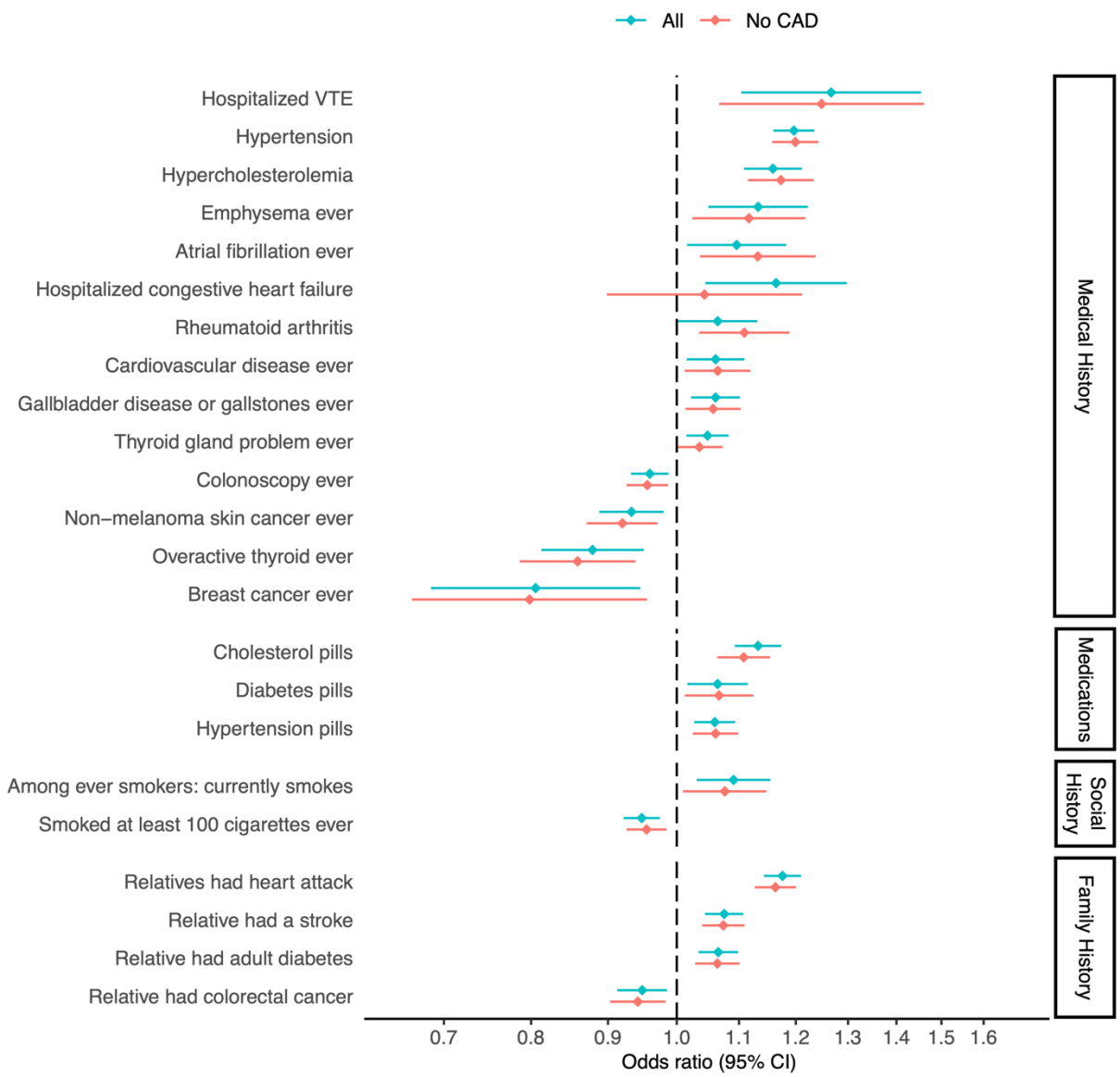

Figure 2. Associations between polygenic risk for coronary artery disease (CAD) and self-reported history, collected at baseline and throughout follow-up in the Women's Health Initiative. Associations within the full study cohort are compared to the subset with no CAD at follow up. Only outcomes with at least 100 cases in the CAD-free group were considered (128 outcomes). The plot shows all outcomes that were significant with FDR q-value ≤ 0.05 in either of the two groups.

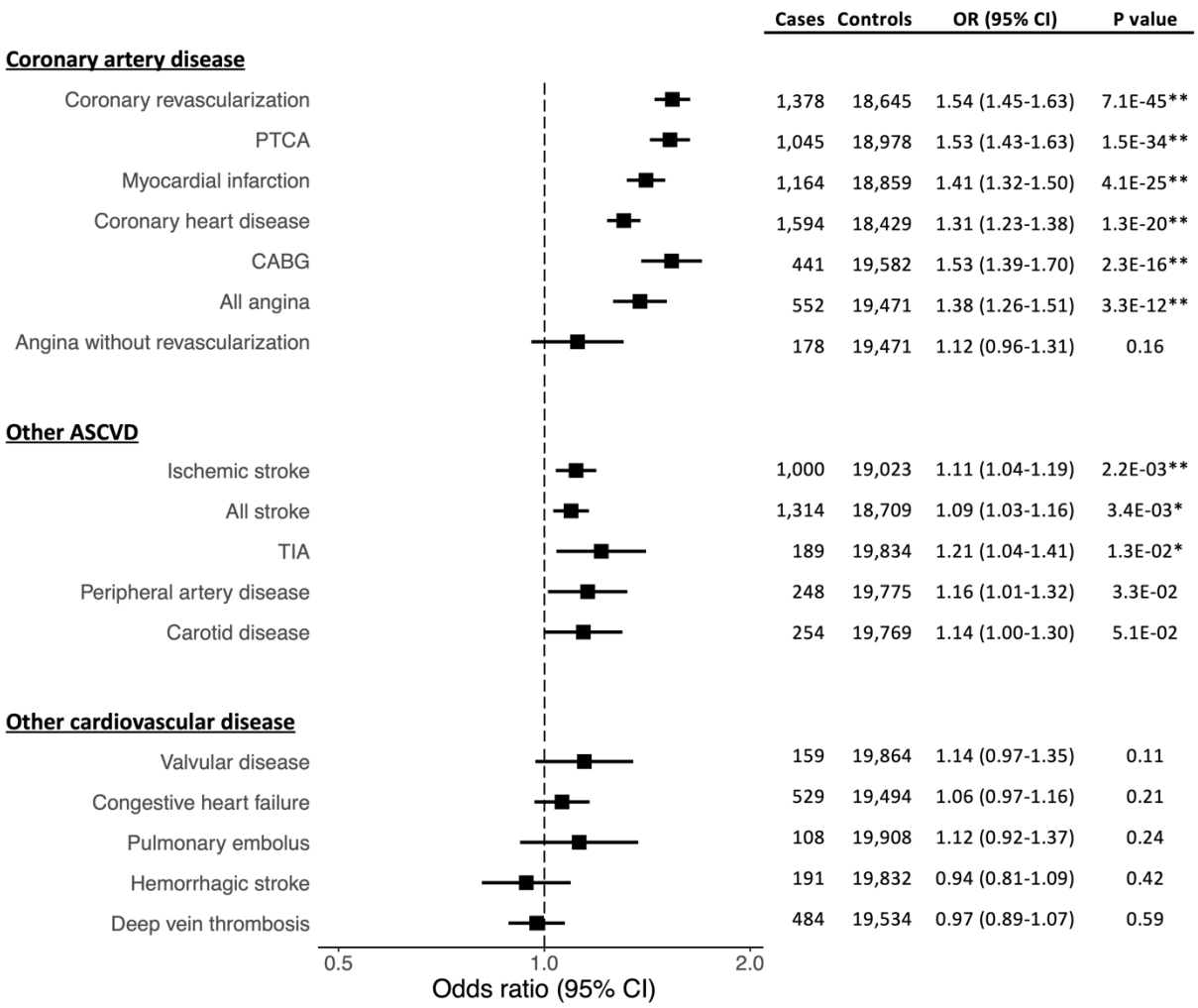

Figure 3. Associations between polygenic risk for coronary artery disease and incident adjudicated cardiovascular diseases in the Women's Health Initiative. Outcomes with at least 100 incident cases were considered. Models were adjusted for smoking status, self-reported diabetes at baseline, systolic blood pressure, low-density lipoprotein cholesterol, and high-density lipoprotein cholesterol. Odds ratios (OR) are per 1 standard deviation increase in polygenic risk score. Outcomes with a single asterisk are significant with an FDR q-value ≤ 0.05. Outcomes with double asterisks have a Bonferroni significant p-value 0.003 (0.05/17). PTCA = Percutaneous Transluminal Coronary Angioplasty; CABG = Coronary Artery Bypass Graft; ASCVD = Atherosclerotic Cardiovascular Disease; TIA = Transient Ischemic Attack

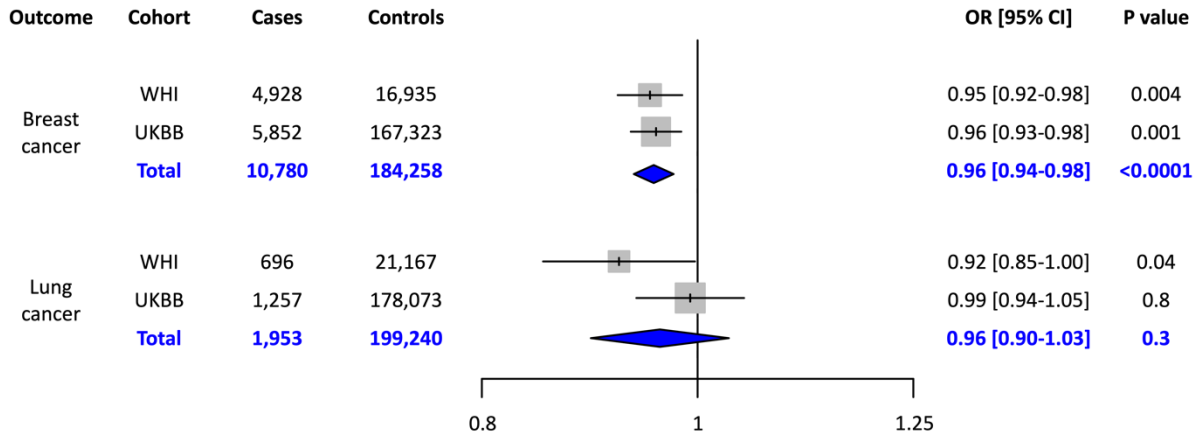

Figure 4. Associations between polygenic risk for coronary artery disease and incident breast and lung cancers in the Women's Health Initiative and the UK Biobank. Random effects meta-analysis is shown in blue. Odds ratios (OR) are per 1 standard deviation increase in polygenic risk score.

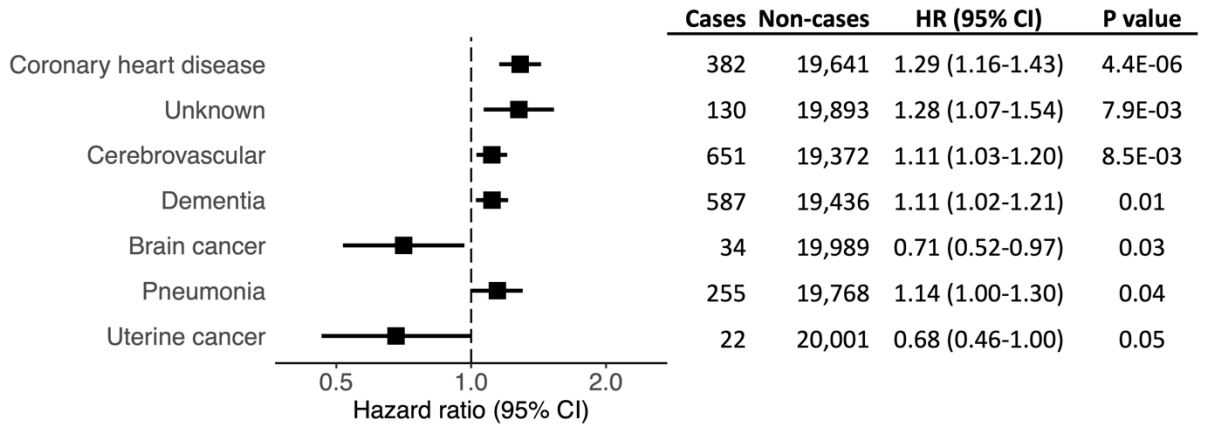

Figure 5. Association between polygenic risk for coronary artery disease and causes of death in the Women's Health Initiative. Cox analysis was performed to measure the risk of death for each given cause. Non-cases were censored at time of death from any other cause or time of last follow-up. Models were adjusted for smoking status, self-reported diabetes at baseline, systolic blood pressure, low-density lipoprotein cholesterol, and high-density lipoprotein cholesterol. Hazard ratios (HR) are per 1 standard deviation increase in the polygenic risk score.