**Genetic Risk Score for Coronary Disease Identifies Predispositions to Cardiovascular and Noncardiovascular Diseases**

Brief title: Coronary artery disease genetic risk score

Ioanna Ntalla PhD[a,b], Stavroula Kanoni PhD[a,b], Lingyao Zeng PhD[c], Olga Giannakopoulou MRes[a,b], John Danesh MD[d,e,f], Hugh Watkins MD[g,h], Nilesh J. Samani MD[i,j], Panos Deloukas PhD[a,b, k,*], Heribert Schunkert MD[c,l,*], for the UK Biobank CardioMetabolic Consortium CHD Working Group

a. Clinical Pharmacology, William Harvey Research Institute, Barts & the London Medical School, Queen Mary University of London, London EC1M 6BQ, UK

b. Centre for Genomic Health, Queen Mary University of London, London EC1M 6BQ, UK

c. Deutsches Herzzentrum München, Klinik für Herz- und Kreislauferkrankungen, Technische Universität München, 80636 Munich, Germany

d. MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK

e. NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK

f. Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1RQ, UK

g. Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford OX3 9DU, UK

h. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK

i. Department of Cardiovascular Sciences, University of Leicester, Leicester LE3 9QP, UK

j. National Institute for Health Research Leicester Cardiovascular Biomedical Research Centre, Leicester LE3 9QP, UK

k. Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia

l. Deutsches Zentrum für Herz- und Kreislauferkrankungen (DZHK), partner site Munich Heart Alliance, 80636 Munich, Germany

*These authors contributed equally to this work

**Funding:**

This work was also supported by the National Institute of Health Research (NIHR) Barts Biomedical Research Centre, which is funded by the NIHR (IS-BRC-1215-20022) as well as by grants from the Fondation Leducq (CAD genomics, 12CVD02), the Deutsche Forschungsgemeinschaft (DFG) as part of the Sonderforschungsbereich CRC 1123 (B2), and the German Federal Ministry of Education and Research (BMBF) (ERA-CVD: grant JTC2017_21-040), within the frame work of target validation (Block CAD: 16GW0198K and AbCD-Net: grant 01ZX1706C). Dr. Deloukas was supported by the British Heart Foundation (BHF) (grant RG/14/5/30893). Ms. Giannakopoulou was supported by BHF (award FS/14/66/3129).

**Disclosures:**

Dr. Danesh has received personal fees and non-financial support from Merck Sharp and Dohme UK Atherosclerosis, Novartis Cardiovascular and Metabolic Advisory Board, and from the Pfizer Population Research Advisory Panel; and has received grants from the British Heart Foundation, European Research Council, Merck Sharp and Dohme, NIHR, NHS Blood and Transplant, UK MRC Wellcome, AstraZeneca, Merck, Novartis, and Pfizer. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

**Address for correspondence**

Panos Deloukas, PhD

William Harvey Research Institute

Queen Mary University of London, Charterhouse Square

London, EC1M 6BQ

United Kingdom

E-mail: p.deloukas@qmul.ac.uk

**Abstract**

**Background:** The taxonomy of cardiovascular (CV) diseases is divided into a broad spectrum of clinical entities. Many such diseases coincide in specific patient groups suggesting shared predisposition.

**Objectives:** Focussing on coronary artery disease (CAD), we investigated the genetic relatedness to CV and non-CV diseases with reported CAD comorbidity.

Methods: This study examined 425,196 UK Biobank participants to determine a genetic risk score (GRS) based on 300 CAD associated variants (CAD-GRS). This score was associated with 22 traits, including risk factors, diseases secondary to CAD, as well as comorbid and non-CV conditions. Sensitivity analyses were performed in individuals free from CAD or stable angina diagnosis.

**Results:** Hypercholesterolemia (odds ratio [OR]: 1.27; 95% CI: 1.26 to 1.29) and hypertension (OR: 1.11; 95% CI: 1.10 to1.12) were strongly associated with the CAD-GRS, which indicated that the score contained variants predisposing to these conditions. However, the CAD-GRS was also significant in patients with CAD who were free of CAD risk factors (OR: 1.37; 95% CI: 1.30 to 1.44). The study observed significant associations between the CAD-GRS and peripheral arterial disease (OR: 1.28; 95% CI: 1.23 to 1.32), abdominal aortic aneurysms (OR: 1.28; 95% CI: 1.20 to 1.37), and stroke (OR: 1.08; 95% CI: 1.05 to 1.10), which remained significant in sensitivity analyses that suggested shared genetic predisposition. The score was also associated with heart failure (OR: 1.25; 95% CI: 1.22 to 1.29), atrial fibrillation (OR: 1.08; 95% CI: 1.05 to1.10), and premature death (OR: 1.04; 95% CI: 1.02 to 1.06). These associations were abolished in sensitivity analyses that indicated that they were secondary to prevalent CAD. Finally, an inverse association was observed between the score and migraine headaches (OR: 0.94; 95% CI: 0.93 to 0.96).

**Conclusions:** A wide spectrum of CV conditions, including premature death, might develop consecutively or in parallel with CAD for the same genetic roots. In conditions like heart failure, the study found evidence that the CAD-GRS could be used to stratify patients with no or limited genetic overlap with CAD risk. Increased genetic predisposition to CAD was inversely associated with migraine headaches.

**Condensed Abstract:** Focussing on coronary artery disease (CAD), we investigated the genetic relatedness to cardiovascular and non-cardiovascular diseases in 425,196 UK Biobank participants. We determined a genetic risk score based on 300 CAD associated variants. Sensitivity analyses were performed in individuals free from a CAD or stable angina diagnosis. We found that genetic predisposition to CAD is associated with increased risk of other atherogenic diseases, while for conditions like heart failure our findings indicated that a large proportion of clinically appearing disease manifestations are secondary to CAD. Finally, the observed genetic overlap between CAD and migraine is driven by genetic loci with opposing effects.

**Key words:** Coronary artery disease, genetic risk score, migraine, heart failure, cardiovascular diseases, UK Biobank

**Abbreviation list**

CAD = coronary artery disease

GRS = genetic risk score

CV = cardiovascular

HES = Hospital Episode Statistics

GWAS = genome-wide association study

SNV = single nucleotide variant

OR = odds ratio

SD = standard deviation

CI = confidence interval

FDR = False discovery rate

## INTRODUCTION

It is clinically evident that coronary artery disease (CAD) and a broad spectrum of other cardiovascular (CV) conditions often coincide in a given patient. CAD is caused by a number of factors that may also precipitate atherosclerosis in other vascular beds, leading to ischemic stroke, abdominal aortic aneurysms (1), peripheral arterial disease (2), or aortic stenosis (3). In contrast, CAD and its main complication, myocardial infarction (MI), increase the risk for developing conditions such as congestive heart failure (4), atrial fibrillation (5), or sudden cardiac death (6). It is currently unclear to which extent the genetic roots of CAD and these conditions overlap and contribute to the coincidence of CV diseases.

Through large-scale, genome-wide association studies and targeted approaches (e.g., exome array) we and others have identified >243 CAD risk loci (7–12). Several of these known CAD risk loci have also been associated with other CV diseases (12). For example, the strongest association signal for CAD located at the 9p21 locus (11) has also been associated with carotid atherosclerosis (13), stroke (14), aneurysms (15), congestive heart failure (16), and CV mortality (17).

The cumulative effect of CAD risk loci can be best depicted by a genetic risk score (GRS), which integrates the number and effect sizes of all known risk alleles of any given individual, which allows an aggregated and continuous quantitative measure of genetic susceptibility. Such GRSs may be better suited than single risk alleles to investigate the full genetic overlap between CAD and other heart and vascular diseases (18, 19). Tragante et al. (20) showed that a GRS based on 30 CAD risk variants was associated with peripheral arterial disease and ischemic stroke. However, due to the lack of comprehensively characterized and genotyped cohorts, there is limited information on how such a CAD-GRS relates to the full spectrum of CV diseases.

UK Biobank is a large longitudinal biobank study in the United Kingdom (21). In addition to self-reported disease outcomes and extensive health and lifestyle questionnaire data, UK Biobank participants are being tracked through their National Health Service records and national registries. In 2017, UK Biobank released the genotypes of 488,377 participants, in whom we assessed the overlap in genetic risk between CAD and 13 CV disease phenotypes, mainly vascular, using CAD-GRS. We aggregated risk alleles of all CAD associated variants

(genome-wide significant and a 5% false discovery rate). In addition, we tested the overlap of this score with traditional CAD risk factors, as well as 3 nonvascular diseases with some previous evidence of shared genetic predisposition to CAD, namely, renal disease (which is often associated with CV complications) (22), migraine headaches (23), and rheumatoid arthritis (24).

## METHODS

### STUDY POPULATION

The UK Biobank project is a large prospective cohort study of approximately 500,000 individuals aged 40 to 69 years at recruitment (21). Following informed consent, a rich variety of phenotypic and health-related information was collected for each participant. Biochemical data for each of the 4 main blood lipids were not avail-able at the time of analyses. Health-related outcome records included death notifications and cancer diagnoses through linkage to national death and cancer registries, and hospital inpatient episode statistics, which contained coded data on admissions, operations, and procedures (primary and secondary).

### GENOTYPING QUALITY CONTROL

Using sample quality control information provided by UK Biobank, we excluded 968 heterozygosity and/or missingness outliers, 373 individuals with sex discordance, 471 individuals with putative sex chromosome aneuploidy, and individuals who withdrew from the study, which resulted in 486,553 individuals (25). We restricted analyses to individuals of European ancestry by combining self-reported ethnicity and principal components (PCs) provided by UK Biobank. We used the kmeans algorithm in R 3.4.0 (R Foundation, Vienna, Austria) to perform a 4-means clustering for each of the first 5 PCs separately corresponding to 4 ethnic groups (white, black, Asian, Chinese). An intersection of these 2 clusterings resulted in 5 PC-derived clusters (white, black, Asian, Chinese, mixed/other). After intersecting self-reported ethnicity information with the derived PC ancestry clusters, we identified 460,312 Europeans. Related individuals were excluded from analyses (second degree of relationship or closer; kinship coefficient <0.088), which left 425,196 individuals

with nonmissing data for most of the outcomes. When possible, we prioritized exclusion of control subjects for all investigated traits.

**DEFINITION OF CAD AND OTHER HEART AND VASCULAR DISEASES**

Hospital Episode Statistics (HES) data incorporates National Health Service data, clinical commissioning groups, and local area teams in the United Kingdom from 1997 to March 2015, with the Scottish data dating back as early as 1981. HES uses International Classification of Diseases-9 and -10to record diagnosis information, and Office of Population, Censuses and Surveys: Classification of Interventions and Procedures (version 4) to code operative procedures. Death registries include all deaths in the United Kingdom up to January 2016, with both primary and secondary causes of death coded in International Classification of Diseases-10.Only participants admitted to the hospital had a HES record. To define disease cases, we leveraged all available data sources, i.e., self-reported data from baseline and follow-up assessments for diseases and operations, HES records, and death registry data (Online Table 1). As cases, we defined all individuals who had a prevalent outcome or developed the outcome in the observational period. Control subjects for each disease were defined as individuals who were not disease cases. Based on clinical reasoning, we divided the investigated traits relative to CAD in to underlying risk factors, comorbid conditions, conditions secondary to CAD, and diseases that did not primarily affect the heart and vasculature.

**CONSTRUCTION OF THE CAD-GRS**

The set of genome-wide significant single nucleotide variants for CAD (N=164) was expanded by including uncorrelated variants at a false discovery rate of 5% (7-12). For the SLC22A3/LPAL2/LPA/PLG and COL4A1/COL4A2 loci, we performed conditional analysis in the CARDIoGRAMplusC4D data (10) with GCTA software (26), and selected 7 and 4 conditionally independent variants, respectively. In estimating the additive effects of all variants, we did not include the 2 known recessive loci (rs11830157, KSR2; rs12976411, ZNF507-LOC400684). We included 319 variants (249 CAD loci), 282 of which were available in UK Biobank; 18had a

proxy (r2>0.8), and 19 were not available. In total, we extracted 300 variants representing 240 loci (Online Table 2).

With these 300 variants, we calculated a weighted GRS (CAD-GRS) for all individuals. A value ranging from 0 to 2 was given to each individual for every variant according to the sum of the posterior probabilities of the imputed genotypes to indicate the number of CAD increasing alleles. We multiplied the number of risk alleles by their reported CAD effect sizes in Nelson et al. (12). Per individual, we totalled these values across all 300 variants to generate a weighted CAD-GRS, which we modeled as a continuous variable and standardized into Z-scores (centered and scaled to have a mean of 0 and standard deviation [SD] of 1). We then calculated 5 quantiles of the standardized CAD-GRS and categorized individuals into 3 groups: Q1; Qreference (individuals from Q2 to 4), which served as the reference group; and Q5.

**GRS ASSOCIATION ANALYSES**

Logistic regression models implemented in R 3.4.0 (R Foundation) we reused to determine the association between CAD-GRS and other conditions. All phenotypes were coded as 0 (control subjects) or 1 (cases). All analyses were adjusted for age, sex, first 5 PCs, and genotyping array (UK Biobank vs.UK-BiLEVE). Regression coefficients of the CAD-GRS on each trait were calculated for an increase of 1 SD in the score and were converted to odds ratios (ORs). Sensitivity analyses were performed to investigate the effect of the CAD-GRS on trait prevalence, including CAD risk factors, comorbid conditions, diseases secondary to CAD, and non-CV diseases, in individuals free from CAD and/or angina diagnosis (n=389,952). We also performed sex-stratified analyses. Statistical significance was set to $p<0.002$ to account for the number of traits tested (a <0.05 divided by 23 traits).

**POWER CALCULATIONS**

We performed power calculations with QUANTO (27), assuming an average genetic frequency of 0.45 (mean frequency of all effect alleles included in the CAD-GRS), and trait prevalence was calculated as number of cases/(number of cases + number or control subjects). We had sufficient statistical power (>80%) for all out-comes, except for congenital

heart disease, cardiomyopathy, pulmonary embolism, pericardial problem, sudden cardiac death, and vein thrombosis (Online Table 3).

## RESULTS

We studied 425,196 unrelated UK Biobank participants of European ancestry, 21,051 of whom had CAD (Table 1). The prevalence of all tested CAD risk factors and comorbid conditions is shown in Online Table 4. As expected, CAD risk factors and other CV diseases were significantly more prevalent among CAD cases compared with CAD-free individuals ($p<0.05$). We selected 300 previously published uncorrelated CAD associated variants (Online Table 2) to construct a CAD-GRS score.

## CAD-GRS AND CAD

We first confirmed that the CAD-GRS was robustly associated with CAD prevalence (OR: 1.65 per 1 SD of the CAD-GRS; 95% confidence interval [CI]: 1.62 to 1.65) (Table 2, Figure 1). CAD prevalence in the top quantile (Q5) of the CAD-GRS compared with the reference group (Qreference) was significantly higher (OR: 2.05; 95% CI: 1.99 to 2.12), whereas CAD prevalence in the lowest quantile (Q1) was significantly lower (OR: 0.53; 95% CI: 0.51 to 0.56) (Central Illustration). We then estimated the effect of the score on CAD prevalence after exclusion of all CAD cases with a known CAD risk factor (including hypercholesterolemia, hypertension, type 2 diabetes, and obesity). Despite the lower number of analysed CAD cases (n=1,543), the effect of the score on CAD prevalence remained significant (OR: 1.37; 95% CI: 1.30 to 1.44). Although it was possible that some risk factor cases (e.g., patients with CAD with hyperlipidemia) were unreported, and serum lipid levels were unavailable at the time of analysis to corroborate findings, our results supported the wider view that the current set of known CAD variants captured more biological pathways than those corresponding to traditional risk factors. Exclusion of CAD cases with comorbid conditions (OR: 1.65; 95% CI: 1.63 to 1.68 in the remaining 17,602 CAD cases) or exclusion of CAD cases with diseases secondary to CAD (OR: 1.67; 95% CI: 1.64 to1.70 in the remaining 13,533 CAD cases) did not significantly affect the association between the score and CAD prevalence.

**CAD-GRS AND CAD RISK FACTORS**

We observed highly significant associations of the CAD-GRS with CAD risk factors (Table 2 and Figure 1). Repeating the preceding analyses after exclusion of all individuals with CAD and/or chronic angina (n=35,199) (Figure 2, Online Table 5) further demonstrated the robustness of the CAD-GRS association with hypercholesterolemia (OR: 1.04; 95% CI: 1.03 to 1.06) and hypertension (OR: 1.04; 95% CI: 1.03 to 1.06). The signal observed for type 2 diabetes was attenuated in the sensitivity analyses, but remained significant (OR: 1.04; 95% CI: 1.03 to 1.06). The lack of any effect of the score on smoking status after exclusion of CAD cases (OR: 1.00; 95% CI: 0.99 to 1.00) supported the view that it was mediated via the exposure, rather than by shared genetic underpinnings (Online Table 5).

**CAD-GRS AND COMORBID CONDITIONS**

We next investigated the extent to which there was a shared genetic component with other heart and vascular diseases. We observed significant associations (Table 2) between CAD-GRS and peripheral arterial disease (OR: 1.28; 95% CI: 1.23 to 1.32), abdominal aortic aneurysms (OR: 1.28; 95% CI: 1.20 to 1.37), aortic stenosis (OR: 1.21; 95% CI: 1.15 to 1.26), and stroke (OR: 1.08; 95% CI: 1.05 to 1.10). Although all these associations, apart from aortic stenosis, were attenuated in sensitivity analyses, they remained significant (Figure 2, Online Table 5). Stratification in disease prevalence was observed across CAD-GRS quantiles. In individuals in Q5, peripheral arterial disease was significantly more prevalent (OR: 1.35; 95% CI: 1.25 to 1.46) compared with Qreference, whereas prevalence in individuals in Q1 was significantly lower compared with Qreference (OR: 0.71; 95% CI: 0.64 to 0.79) (Central Illustration, Online Table 6). Although the same trend was observed in sensitivity analyses, the test did not reach our significance threshold, probably due to the smaller sample size (the coincidence of CAD and peripheral arterial disease was high) and a reduction in power (Online Table 3). Similar association trends were observed for abdominal aortic aneurysms and stroke (Central Illustration, Online Table 6).

**CAD-GRS AND DISEASES SECONDARY TO CAD**

The most significant association of the CAD-GRS was with heart failure (OR: 1.25; 95% CI: 1.22 to 1.29) (Table 2). However, after exclusion of CAD and/or stable angina cases from the analysis, the association was no longer significant (OR: 1.02; 95% CI: 0.97 to 1.07) (Figure 1, Online Table 5). The lack of any signal for heart failure in the sensitivity analysis might suggest that the risk for heart failure among individuals with a high CAD-GRS was best explained by patients with prevalent CAD. Furthermore, there was a sizable subgroup of patients who developed heart failure with no or a limited overlap with CAD genetic risk factors. Atrial fibrillation and premature death gave similar results for heart failure because the strong association observed in the main analysis was statistically nonsignificant after exclusion of CAD cases (Figure 2, Online Table 5). After Bonferroni correction (a=0.05 divided by 23 traits), no significant associations were found for congenital heart disease, cardiomyopathy, pulmonary embolism, pericardial problems, sudden cardiac death, and vein thrombosis (Table 2), which might be due to insufficient statistical power (<80%) (Online Table 3).

**CAD-GRS AND NON-CV CONDITIONS**

We next tested the CAD-GRS for association with diseases that are not primarily related to heart and vasculature but postulated to have a shared genetic component with CAD, including migraine headaches, rheumatoid arthritis, and renal disease. Prevalence of migraine headaches was lower among CAD cases compared with control subjects (2.7% vs.3.5%) (Online Table 4). Likewise, a higher CAD-GRS was associated with a lower prevalence of migraine (OR: 0.94; 95% CI: 0.93 to 0.96) (Table 2), and this signal was not attenuated in sensitivity analysis (OR: 0.94; 95% CI: 0.93 to 0.96) (Figure 2, Online Table 5). Prevalence of migraine headaches was significantly lower in individuals in the highest CAD-GRS quantile compared with the reference group (OR: 0.90; 95% CI: 0.87 to 0.95) (Central Illustration, Online Table 6), and this association was only slightly attenuated but remained significant in sensitivity analyses. Three CAD loci (*LRP1, PHACTR1, FHL5*) reached genome-wide significance ($p<5\times10^{-8}$), and another 11 loci were associated with migraine headaches at $p<10^{-3}$ (Online Table 7). After removal of the top variants in these loci from the CAD-GRS, we observed an

attenuation in the association signal that nonetheless remained significant (Online Table 8). This indicated that the inverse association was based on a wide spectrum of genetic loci represented by the score.

We observed that increased genetic predisposition for CAD, as indicated by the CAD-GRS, was associated with an increased prevalence of rheumatoid arthritis (OR: 1.54; 95% CI: 1.03 to 1.08) (Table 2); however, this did not remain significant in the sensitivity analysis (Figure 2, Online Table 5).

The positive association observed between the score and either acute renal disease (OR: 1.06; 95% CI: 1.03 to 1.09) or chronic renal disease (OR: 1.06; 95% CI: 1.03 to 1.09) (Table 2) did not remain significant in the sensitivity analysis (Figure 2, Online Table 5).

The results of the sex-stratified analyses for all risk factors and disease phenotypes are summarized in Online Table 9. We did not observe any major sex differences for the CAD-GRS in the tested outcomes in the sensitivity analysis, except for peripheral arterial disease (males: OR: 1.17; 95% CI: 1.10 to 1.26; females: OR: 1.05; 95% CI: 0.98 to 1.13).

## DISCUSSION

In this study, we made 3 principal observations. First, we demonstrated that genetic variants that affected CAD also contributed to the development of other diseases brought about by atherosclerosis (e.g., peripheral arterial disease), which suggested a shared inherited component. Second, in a large population sample such as UK Biobank, highly significant association signals of the CAD-GRS were observed in individuals with congestive heart failure, atrial fibrillation, and those who experienced premature death. These findings underscored an etiological role of genetics and the (often) preceeding manifestation of coronary disease in the development of these conditions. Third, the CAD-GRS was inversely associated with migraine and this signal was driven by a broad range of genetic variants with opposing effects on the 2 diseases.

As expected, CAD-GRS was significant for all traditional CAD risk factors, but not for smoking, in which the effect was exerted through exposure. However, the significant association of CAD-GRS with CAD in individuals free of traditional risk factors suggested that

there might be clinical usefulness in such a genetic score to identify people who are at risk, despite being free of known predisposing conditions.

Applying a comprehensive CAD-GRS to >400,000 individuals revealed extensive sharing of genetic risk between CAD and multiple CV diseases. Because many such diseases are triggered by the same etiological substrate, an atherosclerotic lesion, this is not a surprise. It is conceivable that variants within the CAD-GRS that affect mechanisms that increase susceptibility to atherosclerosis might lead to manifestation of 1 or the other vascular bed, or their joint appearance (18, 28). This was shown specifically for the 9p21 locus, which had the strongest effect among common variants of CAD and/or MI risk because it was also associated with large artery stroke, abdominal aneurysms, and peripheral arterial disease (13–15). Looking at UK Biobank alone, the risk variant at the 9p21 locus reached genome-wide significance for abdominal aneurysms and peripheral arterial disease, but not with stroke. The CAD-GRS was significantly associated with prevalence of abdominal aneurysms, peripheral arterial disease, and stroke, which suggested a broader shared genetic component.

Aortic stenosis, a manifestation of atherosclerosis in most patients, was also associated with CAD-GRS. However, the signal became insignificant when patients with known CAD or angina were excluded from the analyses. Beside the reduced statistical power in the sensitivity analysis (64% of aortic stenosis cases were excluded due to comorbid CAD or angina), the lack of association might be due to the fact that patients with severe aortic stenosis often underwent coronary angiography, which leads to a high sensitivity for diagnosing CAD. Helgadottir et al. (29) also reported that a weighted CAD-GRS based on 71 CAD-associated variants was associated with aortic stenosis in an Icelandic population and individuals from UK Biobank, but this association did not remain significant after controlling for CAD diagnosis, which agreed with our results.

The effect of the CAD-GRS on abdominal aneurysm, aortic stenosis, and peripheral arterial disease was 73% to 78% of the effect observed for CAD, whereas for stroke, it was 65%. Thus, the observed attenuation in the association signals for the CAD-GRS implied some heterogeneity in the genetic component across these diseases. The effect of the CAD-GRS on CAD after exclusion of all comorbid conditions or diseases secondary to CAD suggested that some of the CAD risk alleles might be specific for CAD.

The most significant association of the CAD-GRS with any of the other CV comorbidities was observed for congestive heart failure. This agreed with the clinical observation that ischemic heart disease is often accompanied by heart failure (30). MI is a well-established risk factor of heart failure (31). This association was no longer significant in the sensitivity analysis. In conjunction with the strength of the association between the CAD-GRS and heart failure, this observation might also suggest that a large proportion of clinically appearing manifestations of heart failure were secondary or functionally related to CAD. Moreover, from a genetic point of view, it could be concluded that heart failure pathogenesis could be stratified based on CAD genetic risk, as well as other different mechanisms. It is clinically evident that multiple causes of heart failure coexist in a patient population with this syndrome (32).

The significant association of CAD-GRS with premature death was not only clinically relevant, but was also a measure for the relevance of the genetic markers that constitute the CAD-GRS. The studies that led to identification of the CAD risk alleles used here were largely based on clinically identified patients (i.e., survivors of a MI) (11). This led to the criticism that these alleles might indicate markers for survival rather than disease. The observation that the CAD-GRS was related to worse outcome in the UK Biobank might dispute this line of thought and was in agreement with smaller studies on the topic (33, 34).

Single variant and pathway analysis implicated genes involved in inflammation in both CAD and inflammatory conditions, such as rheumatoid arthritis (12). Likewise, clinical studies reported increased risk of CAD among rheumatoid arthritis patients (35), although results at the molecular level were not conclusive (24). In our analysis, CAD-GRS was significantly associated with rheumatoid arthritis, but this association did not remain significant in the sensitivity analysis.

Finally, we observed a strong negative association between the CAD-GRS and migraine headaches, that is, an increased genetic predisposition to CAD was related to a lower prevalence of migraine headaches. Our results were in agreement with findings from another study that also indicated that a CAD poly-genic risk score was associated with a reduced risk of migraine (23). Specifically, it was shown that a CAD risk score was associated only with a reduced risk of migraine without aura, whereas no association was observed for the risk of migraine with aura; this indicated that the signal was mainly driven by a limited number of

loci (23). Unfortunately, migraine was not defined in such detail in UK Biobank participants, and we were able to define only 29 and 194 patients without and with aura, respectively, using HES records. Three CAD-associated loci, LRP1, PHACTR1, and FHL5, are also known migraine risk loci (23, 36). In our study, the inverse association be-tween CAD-GRS and migraine could not be explained exclusively by variants in these loci and another 11 CAD loci that gave an inverse signal for migraine at p<10-3 in UK Biobank. This might indicate that there is a rather broad opposing genetic component be-tween CAD and migraine beyond the established loci. This might furthermore suggest that the related mechanisms influence crucially and delicately balanced traits in vascular biology in an opposing fashion. In the light of such inverse and apparently pleiotropic effects, it might be important to mention that the PHACTR1 risk allele also affects other vascular phenotypes, including fibromuscular dysplasia and arterial dissections in the opposite direction to CAD risk (37).

**STUDY LIMITATIONS**

First, despite the large sample size of UK Biobank, our study was underpowered for adequately studying a number of CV disease phenotypes, such as sudden cardiac death (n=198) or congenital heart disease (n=671). This was in part due to low prevalence, but underreporting might have contributed as well. The latter was exemplified by familial hypercholesterolemia, which was documented in only 50 patients, whereas approximately 2,000 would have been expected because of the prevalence of respective mutations. The same applies to biochemical measures, like the lipid profile, that were not available for inclusion in our analyses. Another limitation might be seen in the precision of the diagnoses reported to the UK Biobank. For example, we could not differentiate between heart failure with reduced or preserved ejection fraction. Additionally, to increase our power, we estimated the effects of the CAD-GRS on overall disease prevalence, which did not allow us to distinguish any differences in the effects of the score between prevalent and incident disease cases. Furthermore, our CAD-GRS did not include all the variants that underlaid CAD risk. Recently, Khera et al. (38) showed that a polygenic risk score of 6.6 million common variants demonstrated the best predictive ability compared with scores derived by the aggregation of smaller number of variants. In addition, it was shown that the predictive capacity of genetic risk scores would be likely to improve further with the identification of

more associated variants and more precise report of their effect sizes from larger genome-wide association studies (39). Our study provided evidence of association between CAD-GRS and disease prevalence, but further studies are required to show whether this relationship is causal. The associations were obtained by studying individuals of mainly European descent (7–12), and disease prevalence in individuals of European ancestry in UK Biobank. Therefore, generalizability of our findings in other ancestral groups and their clinical usefulness remains to be tested.

## CONCLUSIONS

Our results suggested that genetic predisposition to CAD also affected the prevalence of other CV diseases, including premature death. Genetic variants associated with increased risk of CAD were also associated with increased risk of other atherogenic diseases, whereas for conditions like heart failure, our findings indicated that a large proportion of clinically appearing manifestations of the disease were secondary to CAD. Finally, the genetic overlap between CAD and migraine was based on a range of genetic loci with opposing effects to the 2 diseases.

## ACKNOWLEDGMENT

The authors are grateful to UK Biobank for access to data to undertake our study (Project #9922).

## PERSPECTIVES

COMPETENCY IN MEDICAL KNOWLEDGE: Genetic factors related to risk of coronary disease are associated with multiple CV conditions and premature death.

TRANSLATIONAL OUTLOOK: Further studies are needed to elucidate the mechanisms that link genetic variants associated with coronary risk and how they contribute to the risk of developing other diseases as well.

**REFERENCES**


1. Forsdahl SH, Singh K, Solberg S, Jacobsen BK. Risk factors for abdominal aortic aneurysms: a 7-year prospective study: the Tromso Study, 1994-2001. Circulation 2009;119:2202-8.

2. Selvin E, Erlinger TP. Prevalence of and risk factors for peripheral arterial disease in the United States: results from the National Health and Nutrition Examination Survey, 1999-2000. Circulation 2004;110:738-43.

3. Pibarot P, Dumesnil JG. Improving assessment of aortic stenosis. J Am Coll Cardiol 2012;60:169-80.

4. Hellermann JP, Goraya TY, Jacobsen SJ et al. Incidence of heart failure after myocardial infarction: is it changing over time? Am J Epidemiol 2003;157:1101-7.

5. Schmitt J, Duray G, Gersh BJ, Hohnloser SH. Atrial fibrillation in acute myocardial infarction: a systematic review of the incidence, clinical features and prognostic implications. Eur Heart J 2009;30:1038-45.

6. Chugh SS, Reinier K, Teodorescu C et al. Epidemiology of sudden cardiac death: clinical and research implications. Prog Cardiovasc Dis 2008;51:213-28.

7. Deloukas P, Kanoni S et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet 2013;45:25-33.

8. Howson JMM, Zhao W, Barnes DR et al. Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. Nat Genet 2017;49:1113-1119.

9. Myocardial Infarction G, Investigators CAEC, Stitziel NO et al. Coding Variation in ANGPTL4, LPL, and SVEP1 and the Risk of Coronary Disease. N Engl J Med 2016;374:1134-44.

10. Nikpay M, Goel A, Won HH et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet 2015;47:1121-1130.

11. Samani NJ, Erdmann J, Hall AS et al. Genomewide association analysis of coronary artery disease. N Engl J Med 2007;357:443-53.

12. Nelson CP, Goel A, Butterworth AS et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. Nat Genet 2017.

13. Ye S, Willeit J, Kronenberg F, Xu Q, Kiechl S. Association of genetic variation on chromosome 9p21 with susceptibility and progression of atherosclerosis: a population-based, prospective study. J Am Coll Cardiol 2008;52:378-84.

14. Anderson CD, Biffi A, Rost NS, Cortellini L, Furie KL, Rosand J. Chromosome 9p21 in ischemic stroke: population structure and meta-analysis. Stroke 2010;41:1123-31.

15. Helgadottir A, Thorleifsson G, Magnusson KP et al. The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. Nat Genet 2008;40:217-24.

16. Yamagishi K, Folsom AR, Rosamond WD, Boerwinkle E, Investigators A. A genetic variant on chromosome 9p21 and incident heart failure in the ARIC study. Eur Heart J 2009;30:1222-8.

17. Newton-Cheh C, Cook NR, VanDenburgh M, Rimm EB, Ridker PM, Albert CM. A common variant at 9p21 is associated with sudden and arrhythmic cardiac death. Circulation 2009;120:2062-8.

18. Dichgans M, Malik R, Konig IR et al. Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. Stroke 2014;45:24-36.

19. Smith JA, Ware EB, Middha P, Beacher L, Kardia SL. Current Applications of Genetic Risk Scores to Cardiovascular Outcomes and Subclinical Phenotypes. Curr Epidemiol Rep 2015;2:180-190.

20. Tragante V, Doevendans PA, Nathoe HM et al. The impact of susceptibility loci for coronary artery disease on other vascular domains and recurrence risk. Eur Heart J 2013;34:2896-904.

21. Sudlow C, Gallacher J, Allen N et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015;12:e1001779.

22. Olden M, Teumer A, Bochud M et al. Overlap between common genetic polymorphisms underpinning kidney traits and cardiovascular disease phenotypes: the CKDGen consortium. Am J Kidney Dis 2013;61:889-98.

23. Winsvold BS, Bettella F, Witoelar A et al. Shared genetic risk between migraine and coronary artery disease: A genome-wide analysis of common variants. PLoS One 2017;12:e0185663.

24. Jansen H, Willenborg C, Lieb W et al. Rheumatoid Arthritis and Coronary Artery Disease: Genetic Analyses Do Not Support a Causal Relation. J Rheumatol 2017;44:4-10.

25. Bycroft C, Freeman C, Petkova D et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018;562:203-209.

26. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 2011; 88:76-82.

27. Gauderman WJ, Morrison JM. QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies.http://hydra.usc.edu/gxe. 2006.

28. Calling S, Ji J, Sundquist J, Sundquist K, Zoller B. Shared and non-shared familial susceptibility of coronary heart disease, ischemic stroke, peripheral artery disease and aortic disease. Int J Cardiol 2013;168:2844-50.

29. Helgadottir A, Thorleifsson G, Gretarsdottir S et al. Genome-wide analysis yields new loci associating with aortic valve stenosis. Nat Commun 2018;9:987.

30. Kenchaiah S, Narula J, Vasan RS. Risk factors for heart failure. Med Clin North Am 2004;88:1145-72.

31. Poulsen SH, Jensen SE, Egstrup K. Longitudinal changes and prognostic implications of left ventricular diastolic function in first acute myocardial infarction. Am Heart J 1999;137:910-8.

32. Velagaleti RS, Vasan RS. Heart failure in the twenty-first century: is it a coronary artery disease or hypertension problem? Cardiol Clin 2007;25:487-95; v.

33. Christiansen MK, Nyegaard M, Larsen SB et al. A genetic risk score predicts cardiovascular events in patients with stable coronary artery disease. Int J Cardiol 2017;241:411-416.

34. Hernesniemi JA, Lyytikainen LP, Oksala N et al. Predicting sudden cardiac death using common genetic risk variants for coronary artery disease. Eur Heart J 2015;36:1669-75.

35. Avina-Zubieta JA, Thomas J, Sadatsafavi M, Lehman AJ, Lacaille D. Risk of incident cardiovascular events in patients with rheumatoid arthritis: a meta-analysis of observational studies. Ann Rheum Dis 2012;71:1524-9.

36. Gormley P, Anttila V, Winsvold BS et al. Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. Nat Genet 2016;48:856-66.

37. Debette S, Kamatani Y, Metso TM et al. Common variation in PHACTR1 is associated with susceptibility to cervical artery dissection. Nat Genet 2015;47:78-83.

38. Khera AVC, M.; Aragam, K.; Emdin, C. A.; Klarin, D.; Haas, M.; Roselli, C.; Natarajan, P.; Kathiresan, S. . Genome-wide polygenic score to identify a monogenic risk-equivalent for coronary disease. bioRxiv 218388, 2017.

39. Inouye M, Abraham G, Nelson CP et al. Genomic risk prediction of coronary artery disease in nearly 500,000 adults: implications for early screening and primary prevention. J Am Coll Cardiol 2018;72:1884-1893.

**Table 1** Characteristics of all UK Biobank participants and of selected disease cases.

| | N | Age, years (Mean ± standard deviation) | Sex (Male, %) |
|---|---|---|---|
| All individuals | 425,196 | 56.65 ± 8.00 | 45.86 |
| Coronary artery disease (CAD) cases | 21,051 | 61.62 ± 8.28 | 68.10 |
| CAD and/or stable angina cases | 35,199 | 61.46 ± 6.17 | 21.82 |
| **Cases of CAD risk factors** | | | |
| Hypercholesterolemia | 79,552 | 61.17 ± 6.13 | 60.51 |
| Hypertension | 209,513 | 59.17 ± 7.15 | 53.16 |
| Obesity (BMI ≥ 30kg/$m^2$) | 101,403 | 57.04 ± 7.74 | 48.49 |
| Ever smokers | 138,236 | 57.63 ± 7.78 | 53.76 |
| Type 2 diabetes | 19,274 | 60.51 ± 6.62 | 62.69 |
| **Cases of comorbid conditions with CAD** | | | |
| Abdominal aortic aneurysm | 949 | 64.5 ± 4.17 | 87.98 |
| Aortic stenosis | 1,799 | 62.58 ± 5.78 | 65.31 |
| Peripheral arterial disease | 3,305 | 61.02 ± 6.74 | 65.35 |
| Stroke | 9,664 | 59.6 ± 6.72 | 59.06 |
| **Cases of diseases secondary to CAD** | | | |
| Atrial fibrillation | 15,319 | 62.17 ± 5.86 | 67.72 |
| Congenital heart disease | 671 | 56.16 ± 7.99 | 47.39 |
| Cardiomyopathy | 1,457 | 59.6 ± 7.01 | 67.74 |
| Heart failure | 5,735 | 62.03 ± 6.06 | 70.98 |

| | | | |
|---|---|---|---|
| Pulmonary embolism | 6,068 | 59.52 ± 7.18 | 50.52 |
| Pericardial problem | 1,561 | 59.42 ± 7.27 | 65.59 |
| Premature death | 11,110 | 61.14 ± 6.32 | 61.19 |
| Sudden cardiac death | 198 | 61.09 ± 6.5 | 69.69 |
| Vein thrombosis | 10,561 | 59.6 ± 7.12 | 44.54 |
| **Cases of non-cardiovascular conditions** | | | |
| Migraine | 14,718 | 55.47 ± 7.7 | 23.52 |
| Rheumatoid arthritis | 6,808 | 59.43 ± 7 | 32.87 |
| Acute renal disease | 4,591 | 61.38 ± 6.43 | 64.95 |
| Chronic renal disease | 5,393 | 61.86 ± 6.32 | 57.92 |

**Table 2** Effect of the genetic risk score for coronary artery disease (CAD-GRS) on CAD prevalence and prevalence of selected diseases.

| | All individuals | | |
|---|---|---|---|
| | **Prevalence (%)** | **OR** | **95% CI** |
| Coronary artery disease (CAD) | 5.1 | 1.65 | 1.65-1.62 |
| **CAD risk factors** | | | |
| Hypercholesterolemia | 18.7 | 1.27 | 1.26-1.29 |
| Hypertension | 49.3 | 1.11 | 1.10-1.11 |
| Obesity | 23.8 | 1.02 | 1.02-1.03 |
| Smoking (ever vs never smokers) | 37.7 | 1.01 | 1.01-1.02 |
| Type 2 diabetes | 4.5 | 1.11 | 1.09-1.13 |
| **Comorbid conditions with CAD** | | | |
| Abdominal aortic aneurysm | 0.2 | 1.28 | 1.20-1.37 |
| Aortic stenosis | 0.4 | 1.21 | 1.15-1.26 |
| Peripheral arterial disease | 0.8 | 1.28 | 1.23-1.32 |
| Stroke | 2.3 | 1.075 | 1.05-1.10 |
| **Diseases secondary to CAD** | | | |
| Atrial fibrillation | 3.6 | 1.07 | 1.05-1.08 |
| Congenital heart disease | 0.2 | 1.03 | 0.95-1.11 |
| Cardiomyopathy | 0.3 | 1.03 | 0.97-1.08 |
| Heart failure | 1.3 | 1.25 | 1.22-1.29 |
| Pulmonary embolism | 1.4 | 1.01 | 0.99-1.04 |
| Pericardial problem | 2.6 | 1.07 | 1.02-1.13 |
| Premature death | 2.6 | 1.04 | 1.02-1.06 |
| Sudden cardiac death | 0.05 | 1.18 | 1.02-1.35 |
| Vein thrombosis | 2.5 | 1.02 | 1.00-1.04 |
| **Non-cardiovascular conditions** | | | |
| Migraine | 3.5 | 0.94 | 0.93-0.96 |
| Rheumatoid arthritis | 1.6 | 1.05 | 1.03-1.08 |
| Acute renal disease | 1.1 | 1.06 | 1.03-1.09 |

| | All individuals | | |
|---|---|---|---|
| | **Prevalence (%)** | **OR** | **95% CI** |
| Chronic renal disease | 1.3 | 1.06 | 1.03-1.09 |

The genetic risk score was derived using the 300 SNPs associated with CAD, was standardized into Z-scores (centred and scaled to have a mean of 0 and standard deviation (SD) of 1), and modelled as a continuous variable. All estimates were derived in UK Biobank using logistic regression models (adjusting for age, sex, genotyping array and the first 5 principal components of ancestry). Odds ratios (OR) are reported for a one SD increase in the weighted genetic risk score of CAD. OR: odds ratio; CI: confidence interval.

Figures

Figure 1 Association of CAD-GRS with selected diseases

Association of coronary artery disease genetic risk score (CAD-GRS), consisting of 300 CAD risk alleles, with traditional CAD risk factors, CAD comorbid conditions, and diseases secondary to CAD including premature death in the UK Biobank. All arrows indicate significant associations. * Association between the CAD-GRS and disease prevalence was no longer significant after exclusion of CAD and/or angina cases from the analyses.

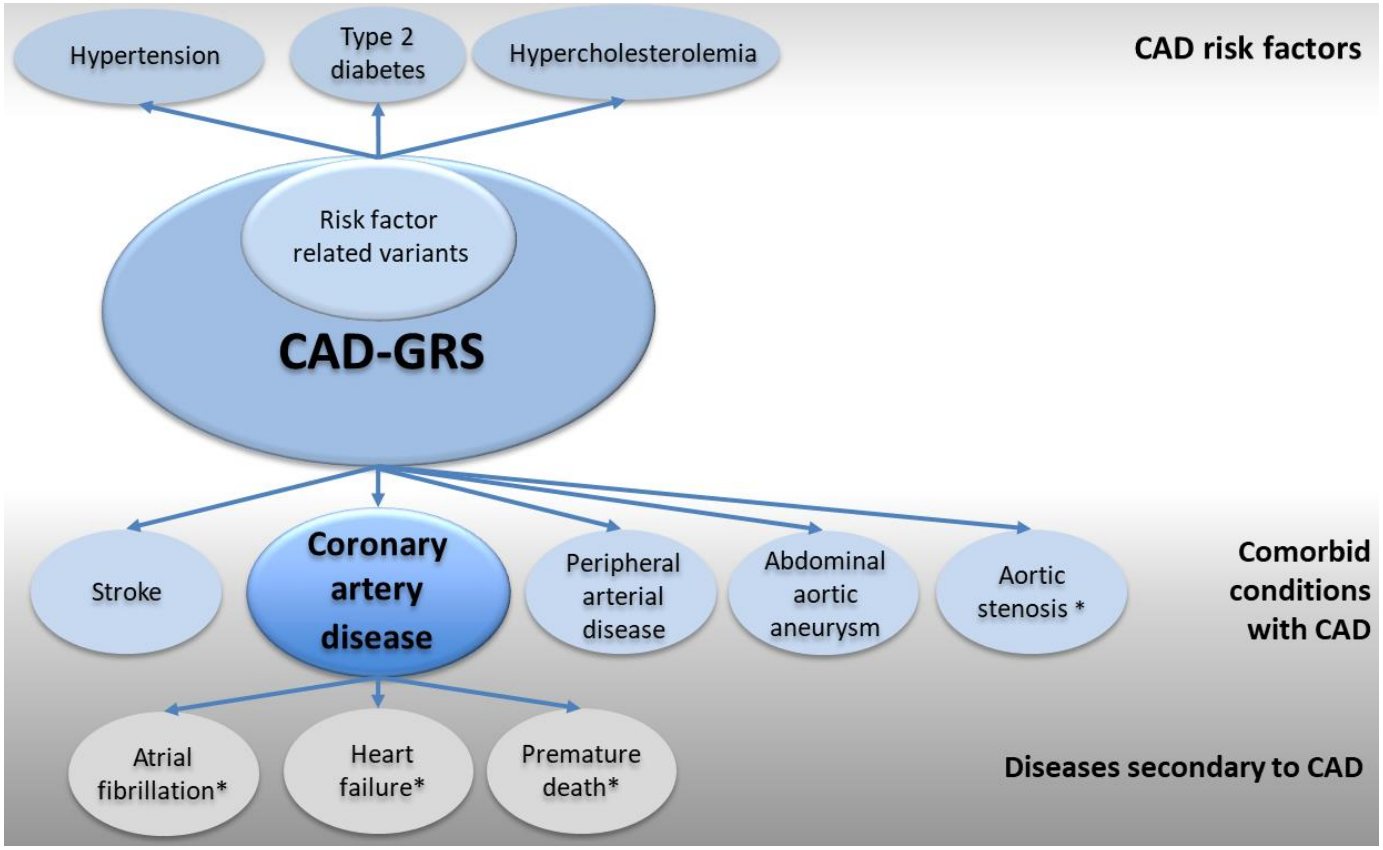
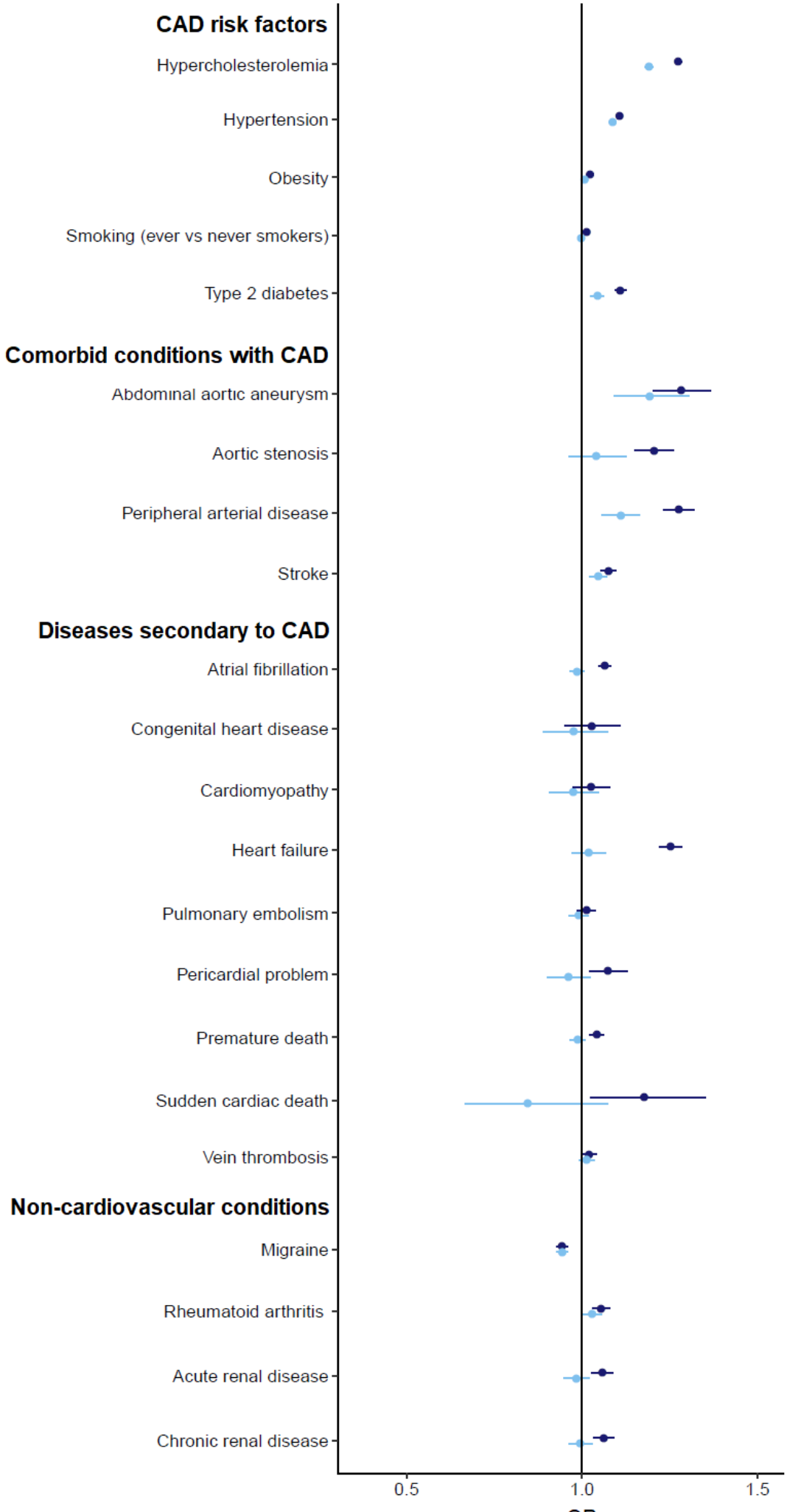


Figure 2 Sensitivity analysis of CAD-GRS effect on prevalence of selected traits

Effect of the genetic risk score for coronary artery disease (CAD-GRS) on prevalence of selected traits in all individuals (dark blue) and in individuals without CAD and/or angina diagnosis (light blue). Odds ratios (ORs) are reported for a one SD increase in the weighted CAD-GRS. Lines show the 95% CI.


CAD risk factors
Hypercholesterolemia
Hypertension
Obesity
Smoking (ever vs never smokers)
Type 2 diabetes
Comorbid conditions with CAD
Abdominal aortic aneurysm
Aortic stenosis
Peripheral arterial disease
Stroke
Diseases secondary to CAD
Atrial fibrillation
Congenital heart disease
Cardiomyopathy
Heart failure
Pulmonary embolism
Pericardial problem
Premature death
Sudden cardiac death
Vein thrombosis
Non-cardiovascular conditions
Migraine
Rheumatoid arthritis
Acute renal disease
Chronic renal disease
0.5
1.0
1.5

Central Illustration Stratified analysis of coronary artery disease genetic risk score

Bars show prevalence of disease within each group. The odds ratios (ORs) for individuals in Quintile 1 and Quintile 5 of the CAD-GRS were compared with these of individuals in the intermediate quantiles (Quintiles 2-4, reference). Results are presented only for diseases that passed the significance threshold. Detailed results are provided in Online Table 6.

* Association did not reach our significance threshold.