

ORIGINAL ARTICLE

Genetics of Cholesterol and Coronary Disease Risk across Six Global Ancestries

Sarah M. Ubut, M.D., Ph.D.,^{1,2,3} Qiuli Chen, B.S.,^{4,5,6} Yang Sui, Ph.D.,^{1,2} Anika Misra, B.S.,^{1,2} Satoshi Koyama, M.D., Ph.D.,^{1,2} Whitney Hornsby, Ph.D.,^{1,2} and Gina M. Peloso, Ph.D.,⁷ for the Genes and Health Research Team⁸ and Sarah Finer, M.B.B.S., Ph.D.,⁸ David A. van Heel, D.Phil.,⁸ Kaoro Ito, M.D., Ph.D.,⁹ Fowzan Alkuraya, M.B.B.S.,^{10,11} Akl C. Fahed, M.D., M.P.H.,^{1,2} Minxian Wang, Ph.D.,^{4,5,6} and Pradeep Natarajan, M.D., M.M.Sc.^{1,2,3}

Abstract

BACKGROUND Clinical trials of low-density lipoprotein cholesterol (LDL-C)-lowering medicines have shown improvement in cardiovascular disease risk. However, representation across diverse communities in such trials is limited. We set out to study whether or not more diverse human genetic studies could provide an opportunity to test whether these trial results can be generalized to populations poorly represented in drug clinical trials.

METHODS We included six cohorts across 967,325 individuals (54.5% female), including 65,258 African, 45,393 admixed American, 179,521 East Asian, 616,045 European, 4686 Middle Eastern, and 56,422 South Asian individuals. Genetic ancestry was determined using genetic principal components and k-nearest neighbors. We constructed ancestry-specific LDL-C polygenic risk scores (PGS) using genome-wide association study (GWAS) summary statistics trained and externally validated on each ancestry. We associated each PGS with population LDL-C, scaled for a 40 mg/dl increase, and tested these scores against coronary artery disease (CAD) risk using hierarchical Bayesian meta-analyses.

RESULTS The associations between genetically predicted LDL-C levels and observed LDL-C were consistent across ancestries, supporting the validity of our genetic proxy. When scaled to mimic a 40 mg/dl increase in LDL-C, the PGS showed positive but heterogeneous associations with CAD. Posterior odds ratios for CAD ranged from 1.35 (95% credible interval, 1.05 to 1.68) in African populations to 1.82 (95% credible interval, 1.33 to 2.97) in Middle Eastern populations. Bayesian analysis revealed that 97.9% of posterior samples showed all population means greater than 0 on a log scale with a between-ancestry posterior variance (τ^2) of 0.062 (95% credible interval, 0.001 to 0.352).

CONCLUSIONS Our findings demonstrated a consistent association between LDL-C and CAD risk across diverse ancestry groups that are often poorly represented in clinical trials, although effect magnitudes varied. (Funded by the National Institutes of Health and others.)

The author affiliations are listed at the end of the article.

Introduction

Cardiovascular disease remains a leading cause of morbidity and death worldwide, with coronary artery disease (CAD) being one of the most prevalent manifestations.¹ Low-density lipoprotein cholesterol (LDL-C) has long been established as

Pradeep Natarajan can be contacted at pnatarajan@mgh.harvard.edu or at the Center for Genomic Medicine, Massachusetts General Hospital, 185 Cambridge Street, CPZN 5.238, Boston, MA 02114.

a key modifiable risk factor for CAD, with numerous studies demonstrating that interventions that lower LDL-C levels, such as statin therapy, significantly reduce CAD incidence.²⁻⁴ This established relationship between LDL-C and CAD has been the cornerstone of cardiovascular prevention strategies for decades.

However, the majority of evidence supporting the relationship between LDL-C and CAD has been derived from studies predominantly involving individuals of European ancestry. This raises a critical question: Is the relationship between LDL-C and CAD risk generalizable across diverse ancestral populations? The number of statin prescriptions issued to lower LDL-C is strikingly low in non-European and non-North American countries, which may be influenced by this evidence gap.⁵⁻⁷ The global burden of cardiovascular disease is not uniformly distributed, with substantial variations in prevalence and outcomes observed across different ancestry groups.¹

Genetic analyses of LDL-C and CAD may begin to bridge this gap. Given the inherently fixed nature of germline genotypes, genetic variants associated with LDL-C can serve as genetic proxies that mimic the random assignment of a clinical trial; individuals are essentially randomly assigned at conception to higher or lower LDL-C levels based on their genetic makeup.⁸ This natural randomization allows us to assess relationships between LDL-C and CAD risk. Importantly, genetic data sets are often larger than those of clinical trials and include far more diverse populations, enabling robust assessment across ancestry groups that are typically underrepresented in drug trials but bear a substantial burden of cardiovascular disease globally.

The primary aim of this study is to determine whether the impact of genetically predicted LDL-C on CAD risk is generalizable across multiple ancestries. We employ Bayesian hierarchical meta-analysis, which allows for the estimation of population-specific effects while “borrowing strength” across ancestries through partial pooling.⁹ This approach provides direct probability estimates for clinically meaningful effect sizes and naturally handles the nested design of multiple studies within ancestry groups.

Methods

STUDY POPULATION

The study included a total of 967,325 participants, with individual-level data from six biobanks. In the United States, we utilized data from two biobanks: the nationwide

All of Us (AOU)¹⁰ biobank, which enrolled 226,513 participants; and the New England-based Mass General Brigham Biobank (MGBB),¹¹ which enrolled 53,682 participants. The study also included 469,709 participants enrolled in the UK Biobank (UKB) and 170,830 participants enrolled in BioBank Japan (BBJ),¹² which focuses on individuals of East Asian descent. We further incorporated data from 44,390 participants enrolled in the Genes and Health (G&H)¹³ study of individuals of South Asian descent living in London and 2201 participants of Arab descent from a cohort based in Saudi Arabia. Written informed consent was obtained from all participants in each contributing study, and local institutional review board protocols approved the scope of the presented analyses. A description of the included studies and demographic characteristics is provided by ancestry in [Table 1](#) (by cohort in Tables S1–S4 in the Supplementary Appendix) and in Table S12. Sex was determined as chromosomal sex, given access to genetic data.

ANCESTRY CALCULATION

Ancestry classification was performed using a nearest-neighbor approach based on genetic principal components (PCs). Reference populations from the 1000 Genomes Project were used as training data, with ancestry labels confirmed through previous validation. Additional details are described in the Supplementary Appendix.

POLYGENIC RISK SCORE CALCULATION

We aggregated LDL-C GWAS summary statistics for diverse ancestries from studies conducted by the Global Lipids Genetics Consortium, the Million Veteran Program, BBJ, the Korean Genome and Epidemiology Study, and the Mexican Biobank.¹⁴ Scores derived from ancestry-specific GWAS data were then linearly combined to generate a multi-ancestry LDL-C polygenic risk score (PGS) (Supplementary Appendix). These scores were validated among individuals of each ancestry identified within the UKB, while the test data sets comprised separate UKB participants not involved in the validation data set of the multi-ancestry LDL-C PGSs. Similarly, for individuals of East Asian ancestry, the BBJ was excluded from the training set. Finally, scores were regressed on principal components of ancestry by study, and residuals were scaled to a normal distribution and standardized between populations. These scores were scaled (Table S7) to reflect a 40 mg/dl increase in LDL-C within each ancestry group, per study. Additional details on each cohort are provided in the Supplementary Appendix. Notably, each ancestry received an ancestry-specific score; cohorts used in this analysis were not

Table 1. Baseline Characteristics by Ancestry Population.*

Variable	African Population (N=65,258)	Admixed American Population (N=45,393)	East Asian Population (N=179,521)	European Population (N=616,045)	Middle Eastern Population (N=4686)	South Asian Population (N=56,422)	Overall (N=967,325)
Age, years	52.7 (14.1)	47.8 (15.8)	62.3 (14.4)	57.4 (11.2)	53.1 (14.3)	43.3 (14.2)	55.7 (12.6)
Female sex	38,056 (58.3%)	30,511 (67.2%)	84,237 (46.9%)	342,195 (55.5%)	1892 (40.4%)	30,174 (53.5%)	527,065 (54.5%)
Type 2 DM	6395 (9.8%)	4630 (10.2%)	36,263 (20.2%)	59,756 (9.7%)	1485 (31.7%)	11,849 (21.0%)	120,378 (12.4%)
Hypertension	23,950 (36.7%)	16,478 (36.3%)	54,574 (30.4%)	239,025 (38.8%)	2605 (55.6%)	16,588 (29.4%)	353,220 (36.5%)
CAD outcome (complete ascertainment)	5022 (7.7%)	2922 (6.4%)	28,013 (15.6%)	54,069 (8.8%)	1665 (35.5%)	3745 (6.6%)	95,436 (9.9%)
Lipid-lowering medication	14,357 (22.0%)	10,350 (22.8%)	35,545 (19.8%)	111,504 (18.1%)	1884 (40.2%)	15,460 (27.4%)	189,100 (19.6%)
Antihypertensive medication	22,645 (34.7%)	17,159 (37.8%)	57,267 (31.9%)	109,040 (17.7%)	1954 (41.7%)	22,907 (40.6%)	230,972 (23.9%)
Total cholesterol, mg/dl	198.2 (45.9)	194.5 (44.9)	208.1 (43)	216.9 (41.9)	194.8 (46.1)	210.1 (31.1)	203.85 (37)
Triglycerides, mg/dl	145 (92.6)	142.7 (94.3)	130.5 (66.6)	152.9 (91.1)	147.3 (95.6)	175.7 (68.7)	135.96 (71)
HDL-C, mg/dl	57.5 (13.7)	57.7 (13.8)	55.4 (15)	56.1 (14.9)	51.3 (15.4)	51.5 (10.3)	55.47 (14.2)
LDL-C, mg/dl	124.8 (39.1)	123 (40.6)	128.8 (40.3)	138.6 (34.1)	133.9 (46.3)	127.2 (38.5)	135.6 (35.8)
Contributing LDL-C n (%)	31,818 (48.8%)	19,510 (43.0%)	175,013 (97.5%)	501,081 (81.3%)	3879 (82.8%)	38,261 (67.8%)	768,419 (79.4%)
Contributing TG, n (%)	31,811 (48.7%)	18,825 (41.5%)	174,919 (97.4%)	498,047 (80.8%)	3861 (82.4%)	38,205 (67.7%)	765,088 (79.1%)
Contributing HDL-C, n (%)	29,331 (44.9%)	18,102 (39.9%)	174,646 (97.3%)	459,219 (74.5%)	3752 (80.1%)	37,420 (66.3%)	722,452 (74.7%)

*Baseline characteristics by ancestry population were used for the variables used in genetic association analysis. Values represent sample-size weighted averages across contributing cohorts. Comprehensive cohort-specific characteristics are provided in Table S1. Continuous values are means±SD. To convert the values for cholesterol to mmol/l, multiply by 0.02586. To convert the values for triglycerides to mmol/l, multiply by 0.01129. In the AOU database, any use of statin medication was recorded for lipid-lowering medications; in the UKB, lipid-lowering medications were as reported at baseline; and in MGBB, lipid-lowering medications were as reported at intake. Data on lipid-lowering medications as reported at baseline for the UKB and MGBB, as retrieved from the duration of health record for AOU, and included the all lipid-lowering G&H cohort were ascertained from the linked electronic health record medication record. CAD represents total cases in each ancestry group (outcome variable with complete ascertainment, 0% missing). Cohorts are further described in the Supplementary Appendix. CAD denotes coronary artery disease; DM, diabetes mellitus; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; and TG, triglycerides.

used to inform the training of these ancestry-specific scores to avoid overlap of training and validation sets.

QUANTITATIVE PHENOTYPES

LDL-C was either directly measured or calculated using the Friedewald equation. When lipid-lowering medicines were prescribed, we adjusted LDL-C by dividing by 0.7 to account for the average effect of lipid-lowering medicines, which is consistent with previous studies.¹⁵⁻¹⁷ We harmonized the CAD phenotype across each cohort (Table S5). Table S1 provides the distributions of LDL-C by cohort and ancestral group. For the UKB, we curated the first instance of the four lipids (data field numbers LDL-C-30780). The lipid measurements were converted from mmol/l to mg/dl. We executed similar steps of phenotype harmonization and normalization for the replication cohorts. Because information on lipid-lowering medication may not have

been available at enrollment for some participants in the AOU cohort, LDL-C was defined as the maximum recorded LDL-C, and it was assumed that no LDL-C-lowering treatments were in use. In the MGBB, Saudi, BBJ, and Genes and Health cohorts, the LDL-C level adjusted for statins at enrollment was used. In the UKB and MGBB, LDL-C was collected at baseline enrollment in the cohort. For the AOU, BBJ, Saudi, and Genes and Health cohorts, LDL-C was obtained from the patient’s medical record.

STUDY END POINTS

Phenotype definitions of CAD across various biobanks, including the UKB, AOU, MGBB, BBJ, Genes and Health, and Saudi cohorts, have been established based on a combination of electronic health records, direct physical measurements, and self-reported data. The presence of CAD incorporates both prevalent and/or incident

disease depending on the cohort (see Table S6). In the UKB,^{18,19} CAD was defined using *International Statistical Classification of Diseases and Related Health Problems*, Tenth Revision (ICD-10) codes. Of note, we used a restrictive “hard” CAD threshold to allow for the clearest harmonization among cohorts (see Table S6). In the UKB, myocardial infarction (MI) was based on self-report or hospital admission diagnosis as performed centrally and recorded with *International Classification of Diseases* billing codes. Coronary revascularization was assessed based on an Office of Population Censuses and Surveys Classification of Interventions and Procedures version 4–coded procedure for coronary artery bypass grafting or coronary angioplasty with or without stenting. The AOU and Genes and Health cohorts curated CAD phenotypes similarly, using ICD-10 and procedure codes, and used Current Procedural Terminology codes for procedures including percutaneous coronary intervention and coronary artery bypass graft, alongside self-reported personal history of MI or CAD.²⁰ In MGBB, MI curation referred to Research Patient Data Registry domains — including “ICD-9” (International Classification of Diseases, 9th Revision), “DSM-4” (Diagnostic and Statistical Manual of Mental Disorders, 4th Edition), “ICD-10” (International Classification of Diseases, 10th Revision), “LMR” (Longitudinal Medical Record), “Oncall,” “APDRG” (All Patient Diagnosis Related Groups), “Phenotype,” “APDRG” (All Patient Refined Diagnosis Related Groups), and “DRG” (Diagnosis Related Groups) — corresponding to ICD-10 codes I21 to I23.²¹ The BBJ and the Saudi cohorts relied on clinical diagnoses. The Saudi cohort used the presence of obstructive coronary atherosclerosis on cardiac catheterization or an MI diagnosis based on clinical documentation.²² The BBJ cohort relied on clinical diagnosis of obstructive CAD or MI. To address the competing risk of death, we required definitive CAD outcome ascertainment for all participants. Deaths without documented CAD were conservatively classified as CAD negative. Mortality rates across databases were less than 6%.

Missing data rates for LDL-C varied across cohorts (Table 1), ranging from 0.35% (BBJ) to 59% (AOU). All participants had available genomic and disease data. We adopted a two-stage analytical approach to limit potential bias from missing LDL-C data. Specifically, we used complete cases to estimate PGS-LDL-C relationships for scaling (stage 1), then applied these scaling factors (Table S7) to test PGS-CAD associations in all individuals with genetic and CAD data (stage 2). Although missing LDL-C data could bias scaling factors if missingness is related to both

LDL-C levels and CAD risk, the primary PGS-CAD analysis uses all available participants.

STATISTICAL ANALYSIS

We adjusted the LDL-C PGS for population stratification by regressing the PGS on PCs, and then normalized the PGS within each ancestry group per cohort. To ensure direct comparability of effect sizes across studies and populations, we employed a two-stage analytical approach. In stage 1, we scaled each PGS within each study-ancestry combination after controlling for sex, age, and 10 ancestral PCs to represent a standardized 40 mg/dl increase in LDL-C. In stage 2, we associated these scaled PGS values with CAD risk, again controlling for covariates (Supplementary Appendix). This sought to ensure that all odds ratios entering the meta-analysis represented the effect of an identical exposure magnitude, enabling a direct comparison across populations despite differences in genetic architecture and study characteristics.

To address potential differences in follow-up time across cohorts, we conducted sensitivity analyses comparing results from logistic regression (odds ratios) with Cox proportional hazards models.

We employed a Bayesian hierarchical meta-analysis as our primary approach to account for heterogeneity across populations while allowing for borrowing strength across ancestries (Fig. S1). This approach models each study’s effect as arising from a population-specific mean, with population means themselves arising from a shared global effect. The hierarchical structure naturally handles the nested design, in which multiple studies contribute to each ancestry group. The model estimates three levels of effects: study-specific effects within each population; population-specific means; and a global mean across all populations.

We specified a hierarchical model in which the observed log-odds ratios from each study are modeled as arising from these true effects with known sampling error (SE):

$$\log \text{Odds Ratio}_{\text{observed}} \sim \text{Normal}(\beta_{ij}, \text{SE}^2). \quad (1)$$

The true effect in each cohort (j) for ancestry (i), β_{ij} , arises from an ancestry mean μ_i ; the ancestry means μ_i then arise from a shared global mean μ_{global} (Fig. S1):

$$\beta_{ij} \sim \text{Normal}(\mu_i, \sigma^2) \quad (2)$$

$$\mu_i \sim \text{Normal}(\mu_{\text{global}}, \tau^2_{\text{global}}) \quad (3)$$

We used weakly informative priors and estimated posterior distributions using Markov chain Monte Carlo sampling with 2000 iterations across four chains and report credible intervals (CrIs) and posterior means (Fig. S1). Statistical analyses were performed using R (version 4.1.3; the R Foundation, Vienna, Austria). We assessed the robustness of our findings to prior specifications by comparing four approaches: main analysis with Normal(0,1) priors for the global mean parameter (μ_{global}), half-Normal(0,1) for between-population variance (τ), and half-Cauchy(0,1) for within-population variance (σ); conservative priors with Normal(0,0.5) for μ_{global} and half-Normal(0,0.5) for both variance parameters (τ and σ); diffuse priors with Normal(0,2) for μ_{global} , half-Normal(0,2) for τ , and half-Cauchy(0,2) for σ ; and alternative variance priors using Normal(0,1) for μ_{global} , half-Cauchy(0,1) for τ , and half-Normal(0,1) for σ .

Results

The UKB (N=469,709) and MGBB (N=53,682) cohorts are primarily made up of individuals with European ancestry, with over half being female participants (54.1 and 55.6%, respectively). The AOU cohort (N=226,513) provides broader ancestry diversity, with a higher proportion of female individuals (61.4%). The BBJ (N=170,830; 46.0% female), Genes and Health (N=44,390; 55.2% female), and Saudi (N=2201; 34.2% female) cohorts offer insight into East Asian, South Asian, and Middle Eastern populations, respectively (Tables S1–S3).

Across these biobanks, key variables such as age, LDL-C level, and CAD prevalence varied, reflecting the distinct population or study characteristics. The mean CAD incidence is varied, with the Middle Eastern population demonstrating 35.5% CAD and the South Asian population 6.6% CAD (by ancestry in [Table 1](#); by cohort in [Table S4](#)), reflecting population- and cohort-specific variation.

LOW-DENSITY LIPOPROTEIN CHOLESTEROL ASSOCIATION

The association between the LDL-C PGS and observed LDL-C levels was consistent in direction across ancestries, consistent with the validity of the ancestry-tailored PGS. After residualizing for population stratification, we scaled the PGS to estimate the effect of a 40 mg/dl increase in LDL-C — a threshold chosen because it approximates the average LDL-C reduction achieved by traditional statin therapy. Although scaling factors varied across

study-ancestry combinations ([Table S7](#)), the PGS-LDL-C relationship remained consistently positive across populations, albeit with different magnitudes of effect.

CORONARY ARTERY DISEASE ASSOCIATION

When scaled to represent a 40 mg/dl increase in LDL-C, the PGS showed a consistently positive association with CAD across ancestries. The Bayesian hierarchical analysis yielded a global odds ratio of 1.51 (95% credible interval, 1.21 to 1.91) for a 40 mg/dl increase in genetically predicted LDL-C. Population-specific effects ranged from 1.35 (95% credible interval, 1.05 to 1.68) in African populations to 1.82 (95% CrI, 1.33 to 2.97) in Middle Eastern populations ([Fig. 1](#) and [Table S9](#)). In the sensitivity analysis, associations were similar in time-to-event models ([Table S11](#)).

The analysis estimated a 97.9% posterior probability that all populations share the same effect direction, with a between-population variance (μ^2) of 0.062 (95% CrI, 0.001 to 0.352) ([Fig. 2](#)).

We further quantified the probability of clinically meaningful effects ([Fig. 3](#)). There was a 99% probability that genetically predicted LDL-C increases CAD odds by at least 10% (odds ratio > 1.1), a 97.7% probability of at least a 20% increase (odds ratio > 1.2), a 92.7% probability of increasing odds by 30% or more (odds ratio > 1.3), and a 53.6% chance of at least a 50% increase (odds ratio > 1.5). The sensitivity analysis demonstrated the robustness of our findings across prior specifications ([Table S8](#)). The between-population variance τ^2 ranged from 0.058 to 0.065, and the probability of shared positive direction remained greater than 97% across all specifications.

When examining ancestry-specific effects, all populations showed a greater than 95% probability of the LDL-C PGS conferring at least a 10% increase in CAD odds ([Figs. 2 and 3](#)). However, the magnitude of the effect varied across ancestries. The Middle Eastern population exhibited the strongest effect, with a 100% probability of at least a 10% increase in odds and an 83.5% chance of a 50% or greater increase. South Asian and East Asian populations showed a greater than 99% probability of at least a 10% increase and 67.1 and 57.6% probabilities, respectively, of a 50% or greater increase.

The European and admixed American populations demonstrated similar patterns, with greater than 99 and 97.5% probability, respectively, for an odds ratio greater than 1.1, but lower probabilities of larger effects (38.4 and 34.5% for an odds ratio greater than 1.5, respectively). The African population showed a 95.3% probability for

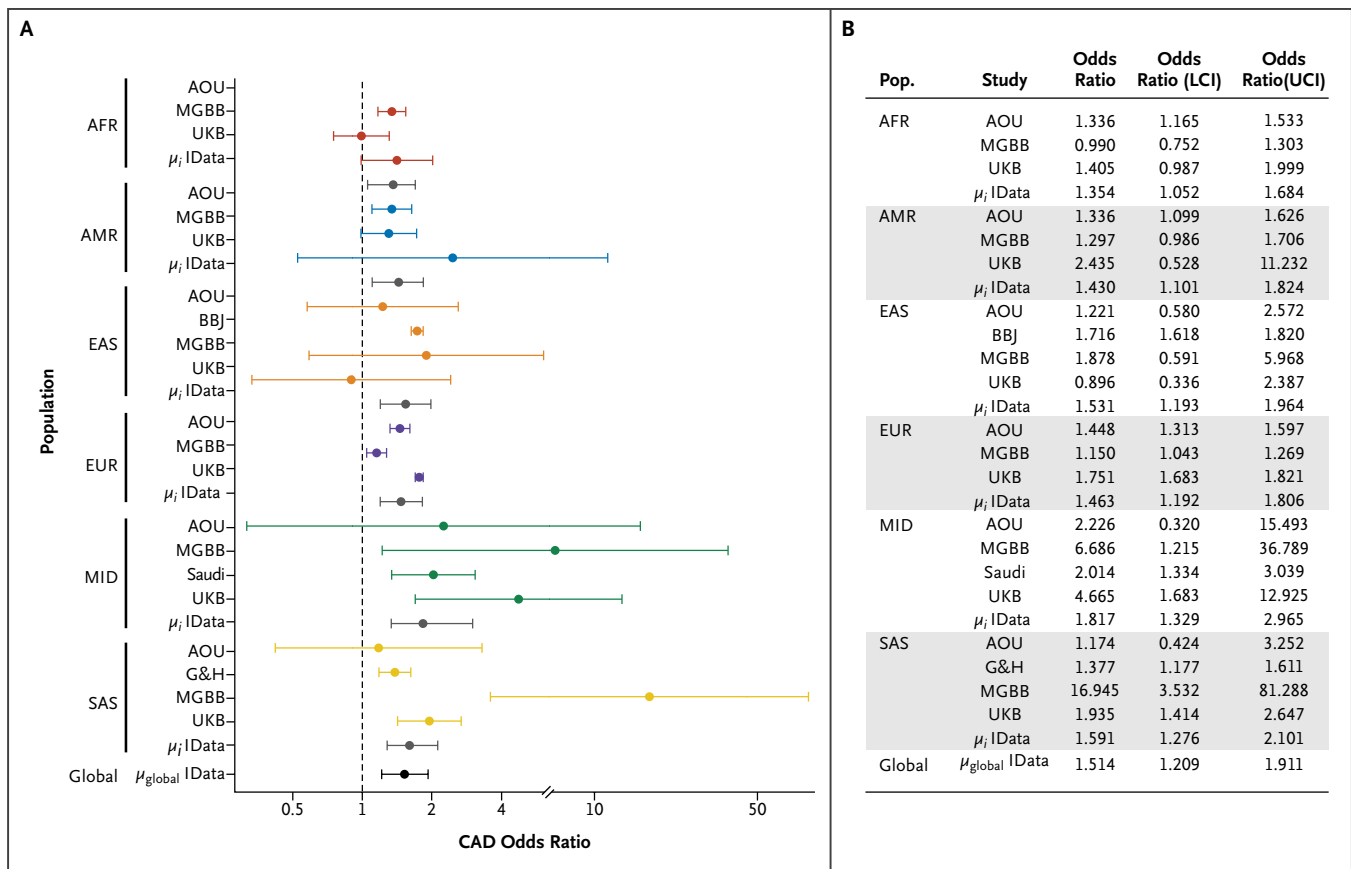


Figure 1. Association between Low-Density Lipoprotein Cholesterol Polygenic Scores and Coronary Artery Disease Risk across Populations and Cohorts.

Panel A shows a forest plot showing Bayesian hierarchical meta-analysis results for the association between genetically predicted low-density lipoprotein cholesterol levels and coronary artery disease. Each point represents the odds ratio and 95% credible interval for coronary artery disease risk per 40 mg/dl increase in polygenic score–predicted low-density lipoprotein cholesterol within each population–cohort combination. The overall pooled estimate (μ_j IData) reflects the posterior distribution from the Bayesian hierarchical model. In both Panel A and Panel B, data include six ancestral populations (African, admixed American, East Asian, European, Middle Eastern, and South Asian) across six cohorts (All of Us, Mass General Brigham Biobank, UK Biobank, BioBank Japan, the Saudi cohort, and Genes and Health). Between-ancestry heterogeneity: $\tau^2=0.062$ (95% credible interval, 0.001 to 0.352). AFR denotes African; AMR, admixed American; AOU, All of Us; BBJ, BioBank Japan; CAD, coronary artery disease; EAS, East Asian; EUR, European; G&H, Genes and Health; LCI, lower credible interval; MGBB, Mass General Brigham Biobank; MID, Middle Eastern; Pop., population; SAS, South Asian; UCI, upper credible interval; and UKB, UK Biobank.

an odds ratio greater than 1.1, but the lowest probability of large effects among all groups (19.0% for an odds ratio greater than 1.5).

Discussion

Our Bayesian hierarchical meta-analysis demonstrated a robust genetic association between LDL-C PGS and CAD risk across diverse populations, with a 51% increase in the odds of CAD per 40 mg/dl increase in genetically predicted

LDL-C. In our evaluation of ancestry-specific effects, all populations demonstrated a high probability (>95%) of at least a modest effect (>10% increase in CAD) associated with higher LDL-C, although the magnitude of effect varied across ancestries. Our study suggests that most of this heterogeneity is likely related to study-level differences rather than fundamental biological differences, supporting the generalizability of LDL-C-lowering interventions across populations.

Previous studies evaluating the association of LDL-C and CAD risk have either been restricted to individuals of

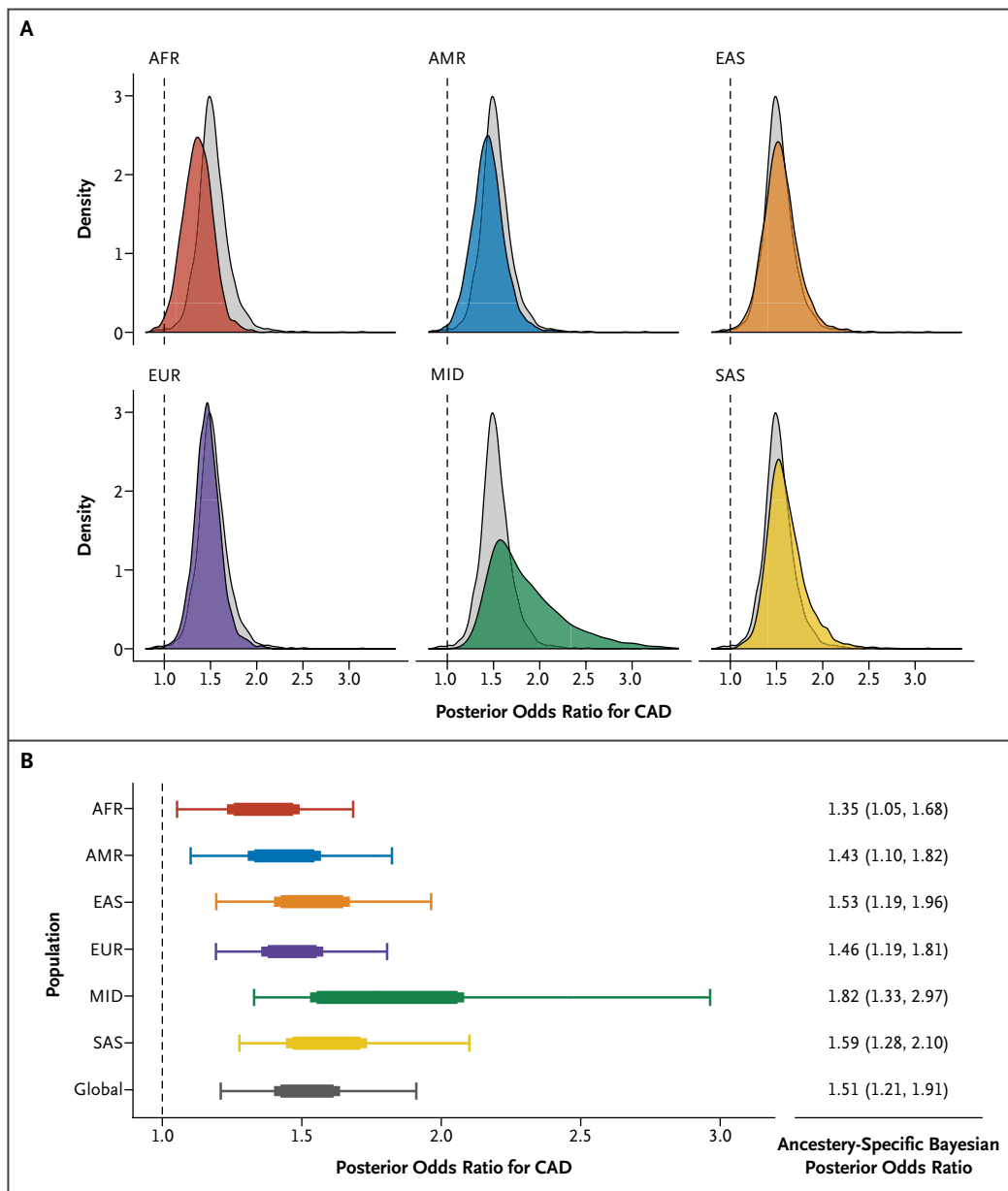


Figure 2. Bayesian Hierarchical Meta-Analysis of Low-Density Lipoprotein Cholesterol Polygenic Score Effects on Coronary Artery Disease across Ancestries.

Panel A shows posterior distributions of study-specific log odds ratios (β_{ij}) for the association between a 40 mg/dl increase in low-density lipoprotein cholesterol polygenic risk score and coronary artery disease risk. Each density curve represents the pooled posterior distribution of true study effects within an ancestry group, estimated via Bayesian hierarchical modeling that accounts for study-level sampling uncertainty and population-level heterogeneity. The gray area represents global distribution. Panel B shows summary odds ratios and 95% credible intervals derived from the posterior distributions in Panel A, along with the global pooled estimate across all ancestries. Thick lines represent 95% credible intervals and thin lines represent 90% credible intervals of the posterior distribution. AFR denotes African; AMR, admixed American; CAD, coronary artery disease; EAS, East Asian; EUR, European; MID, Middle Eastern; and SAS, South Asian.

European descent or were likely underpowered to evaluate the association in individuals of other ancestral backgrounds.^{3,4} For example, clinical trials of rosuvastatin

versus placebo among intermediate-risk individuals recruited across multiple countries demonstrated overall cardiovascular benefit.²³ Although treatment effects

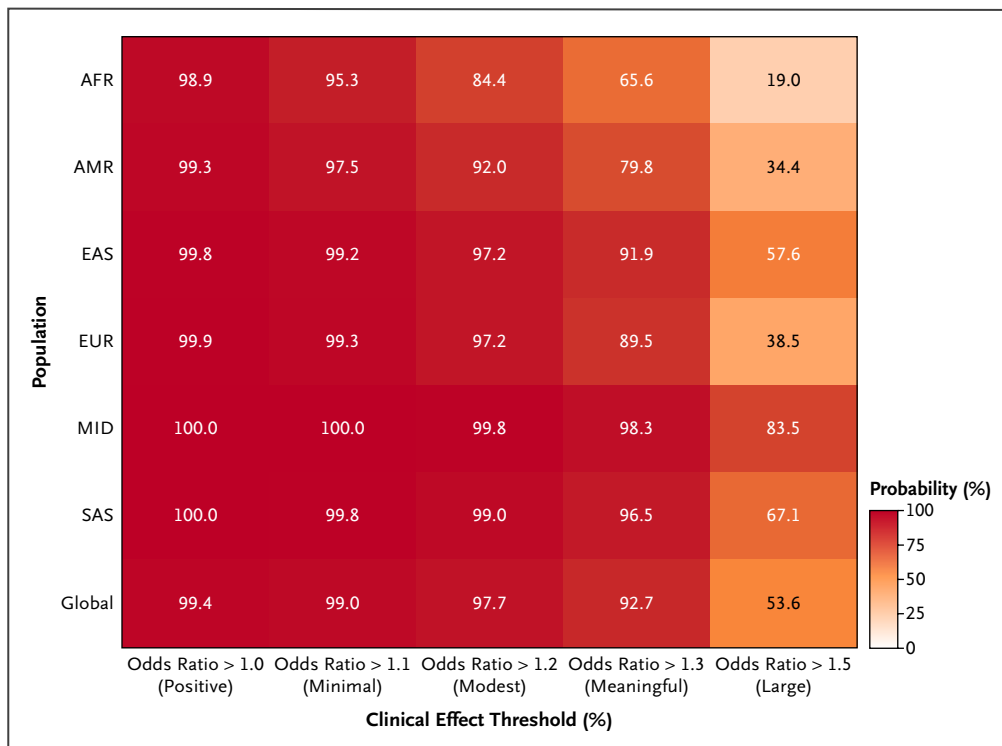


Figure 3. Bayesian Analysis of Low-Density Lipoprotein Cholesterol Effects on Coronary Artery Disease Risk across Global Populations.

Panel A shows posterior probabilities of clinically meaningful effects. Each cell shows the probability that genetically predicted low-density lipoprotein cholesterol increases coronary artery disease risk above specified odds ratio thresholds in each population. Values represent the proportion of 8000 posterior samples (from 2000×4 Markov chain Monte Carlo chains) exceeding each threshold, corresponding to 10% (odds ratio>1.1), 20% (odds ratio>1.2), 30% (odds ratio>1.3), and 50% (odds ratio>1.5) increases in coronary artery disease risk per 40 mg/dl increase in low-density lipoprotein cholesterol. AFR denotes African; AMR, admixed American; EAS, East Asian; EUR, European; MID, Middle Eastern; and SAS, South Asian.

appeared consistent by self-reported race and ancestry, a protective effect was only observed among those of European descent.^{24,25} In this context, increasingly diverse human genetic data sets that are larger than clinical trials may allow for effective assessments across much broader groups.²⁶ Our findings extend these previously limited clinical observations with improved power to show an association of LDL-C and CAD across individuals of diverse ancestries.

Our findings have potential implications for multiple stakeholders in cardiovascular medicine. For individuals from diverse ancestral backgrounds, these results provide reassurance that LDL-C-lowering interventions are likely to be beneficial regardless of ancestry, with a high probability (>95% across all populations) of at least modest cardiovascular risk reduction. This is particularly relevant for patients who may question whether clinical

trial results from predominantly European populations apply to them.

For clinicians treating patients, our findings support the broad application of LDL-C-lowering guidelines across diverse patient populations. The consistent direction of effect across all ancestries, combined with high probabilities of clinically meaningful benefit, provides evidence-based support for prescribing statins and other LDL-C-lowering therapies to patients from underrepresented populations. However, the observed heterogeneity in effect magnitude suggests that individualized risk assessment may be valuable, particularly for populations where genetic effects appear more modest.

For clinical trialists, these findings provide a stronger justification for including diverse populations in cardiovascular trials, as the biological rationale for LDL-C lowering appears consistent across ancestries; they also suggest that

results from existing trials may have broader generalizability than previously assumed, although effect sizes may vary across populations.

For public health entities and governments, particularly in regions with predominantly non-European populations, these findings provide a rationale for the implementation of population-wide LDL-C-lowering strategies. The 97.9% probability that all populations share the same positive effect direction provides strong evidence for global cardiovascular prevention policies targeting LDL-C reduction. However, the heterogeneity in effect magnitude may inform resource allocation and the intensity of intervention strategies across different populations.

As well as our observation of the global association between genetic LDL-C and CAD, we also observed the heterogeneity of this association across ancestries. There are several factors that could account for this. First, population-specific genetic architectures may differentially modulate CAD risk, with non-LDL-C pathways potentially playing larger roles in some ancestries. For example, variants conferring protection against LDL-C-mediated atherosclerosis may be more prevalent in certain populations, or different linkage-disequilibrium patterns may cause our PGS to capture distinct causal variants across ancestries.^{11,27} Second, environmental and lifestyle factors that modify cardiovascular risk independent of LDL-C — such as dietary patterns, physical activity levels, or psychosocial stressors — may vary systematically across populations and exert differential effects on CAD development.²⁸ Third, biological factors, including inflammatory profiles, high-density lipoprotein cholesterol metabolism, or vascular reactivity, may show population-specific variation that influences how LDL-C translates to clinical outcomes.²⁹ Finally, social determinants of health, including health care access, medication adherence, and structural inequities, which we have previously shown to correlate with genetic ancestry, may create differential nongenetic exposures that modulate the relationship between LDL-C and CAD risk.²⁹ Future studies incorporating environmental, lifestyle, and biomarker data will be essential to disentangle these potential mechanisms.

Several limitations should be noted. First, heterogeneity in the association between genetic LDL-C and CAD could be partially attributed to differences in CAD definitions across cohorts. Second, individuals of European ancestry contributed the largest weights to the analysis, influencing the global estimates the most. Third, the varying sample sizes across ancestry groups may affect the precision of ancestry-specific estimates.

In conclusion, ancestry-specific LDL-C PGSs were associated with CAD risk across diverse populations, though with important heterogeneity in the strength of associations. These findings support the association of LDL-C and CAD across diverse populations.

Disclosures

Author disclosures and other supplementary materials are available at evidence.nejm.org.

Drs. Natarajan and Peloso are supported by the National Institutes of Health/National Heart, Lung, and Blood Institute (grant number: R01HL127564), and Dr. Natarajan is also supported by the National Institutes of Health/National Human Genome Research Institute (grant number: U01HG011719). Dr. Urbut is supported by the National Institutes of Health/National Human Genome Research Institute (grant number: T32HG010464). Genes and Health is/has recently been co-funded by the Wellcome Trust (grant numbers: WT102627 and WT210561), the Medical Research Council (UK) (grant numbers: M009017, MR/X009777/1, and MR/X009920/1), the Higher Education Funding Council for England Catalyst Project, Barts Charity (grant number: 845/1796), and Health Data Research UK (for the London substantive site), and received research delivery support from the National Health Service, National Institute for Health Research Clinical Research Network (North Thames).

We would like to thank the participants of the UK Biobank, All of Us, Mass General Brigham Biobank, the Saudi, the Japan BioBank, and Genes and Health cohorts. We thank Social Action for Health, the Centre of The Cell, the members of our community advisory group, and the staff who recruited and collected data from volunteers. We thank the National Institute for Health Research National Biosample Centre (UK Biocentre), the Social, Genetic and Developmental Psychiatry Centre (King's College London), the Wellcome Sanger Institute, and the Broad Institute for their sample processing, genotyping, sequencing, and variant annotation.

Author Affiliations

¹Division of Cardiovascular Disease, Massachusetts General Hospital, Boston

²Cardiovascular Disease Initiative, Broad Institute, Cambridge, MA

³Center for Genomic Medicine, Massachusetts General Hospital, Boston

⁴National Genomics Data Center, China National Center for Bioinformatics, Beijing

⁵Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing

⁶University of Chinese Academy of Sciences, Beijing

⁷Department of Biostatistics, Boston University, Boston

⁸Queen Mary University of London, London

⁹RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

¹⁰Genomic Medicine Center of Excellence, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia

¹¹College of Medicine, Alfaisal University, Riyadh, Saudi Arabia

References

1. Martin SS, Aday AW, Almarzooq ZI, et al. 2024 heart disease and stroke statistics: a report of US and global data from the American Heart Association. *Circulation* 2024;149:e347-e913. DOI: [10.1161/CIR.0000000000001209](https://doi.org/10.1161/CIR.0000000000001209).
2. Ference BA, Ginsberg HN, Graham I, et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. *Eur Heart J* 2017;38:2459-2472. DOI: [10.1093/eurheartj/ehx144](https://doi.org/10.1093/eurheartj/ehx144).
3. Domanski MJ, Tian X, Wu CO, et al. Time course of LDL cholesterol exposure and cardiovascular disease event risk. *J Am Coll Cardiol* 2020;76:1507-1516. DOI: [10.1016/j.jacc.2020.07.059](https://doi.org/10.1016/j.jacc.2020.07.059).
4. Collaborators CTT (CTT). The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. *Lancet* 2012;380:581-590. DOI: [10.1016/S0140-6736\(12\)60367-5](https://doi.org/10.1016/S0140-6736(12)60367-5).
5. Marcus ME, Manne-Goehler J, Theilmann M, et al. Use of statins for the prevention of cardiovascular disease in 41 low-income and middle-income countries: a cross-sectional study of nationally representative, individual-level data. *Lancet Global Health* 2022;10:e369-e379. DOI: [10.1016/S2214-109X\(21\)00551-9](https://doi.org/10.1016/S2214-109X(21)00551-9).
6. Vallejo-Vaz AJ, Stevens CAT, Lyons ARM, et al. Global perspective of familial hypercholesterolaemia: a cross-sectional study from the EAS Familial Hypercholesterolaemia Studies Collaboration (FHSC). *Lancet* 2021;398:1713-1725. DOI: [10.1016/S0140-6736\(21\)01122-3](https://doi.org/10.1016/S0140-6736(21)01122-3).
7. Guadamuz JS, Shooshtari A, Qato DM. Global, regional and national trends in statin utilisation in high-income and low/middle-income countries, 2015–2020. *BMJ Open* 2022;12:e061350. DOI: [10.1136/bmjopen-2022-061350](https://doi.org/10.1136/bmjopen-2022-061350).
8. Smith GD, Ebrahim S. “Mendelian randomization”: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;32:1-22. DOI: [10.1093/ije/dyg070](https://doi.org/10.1093/ije/dyg070).
9. Bayesian Way. *NEJM Evid* 2023;2(6). DOI: [10.1056/EVIDx2300121](https://doi.org/10.1056/EVIDx2300121).
10. Bick AG, Metcalf GA, Mayo KR, et al. Genomic data in the All of Us research program. *Nature* 2024;627:340-346. DOI: [10.1038/s41586-023-06957-x](https://doi.org/10.1038/s41586-023-06957-x).
11. Koyama S, Wang Y, Paruchuri K, et al. Decoding genetics, ancestry, and geospatial context for precision health. October 25, 2023 (<https://www.medrxiv.org/content/10.1101/2023.10.24.23297096>). Preprint.
12. Yokomichi H, Noda H, Nagai A, et al. Cholesterol levels of Japanese dyslipidaemic patients with various comorbidities: BioBank Japan. *J Epidemiol* 2017;27:S77-S83. DOI: [10.1016/j.je.2016.12.014](https://doi.org/10.1016/j.je.2016.12.014).
13. Burki T. Research focus: the genes & health study. *Lancet* 2024;404:1915. DOI: [10.1016/S0140-6736\(24\)02264-5](https://doi.org/10.1016/S0140-6736(24)02264-5).
14. Sohail M, Palma-Martínez MJ, Chong AY, et al. Mexican Biobank advances population and medical genomics of diverse ancestries. *Nature* 2023;622:775-783. DOI: [10.1038/s41586-023-06560-0](https://doi.org/10.1038/s41586-023-06560-0).
15. Fahed AC, Wang M, Homburger JR, et al. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat Commun* 2020;11:3635. DOI: [10.1038/s41467-020-17374-3](https://doi.org/10.1038/s41467-020-17374-3).
16. Nam K, Kim J, Lee S. Genome-wide study on 72,298 individuals in Korean biobank data for 76 traits. *Cell Genom* 2022;2:100189. DOI: [10.1016/j.xgen.2022.100189](https://doi.org/10.1016/j.xgen.2022.100189).
17. Klarin D, Damrauer SM, Cho K, et al. Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nat Genet* 2018;50:1514-1523. DOI: [10.1038/s41588-018-0222-9](https://doi.org/10.1038/s41588-018-0222-9).
18. Khera AV, Chaffin M, Aragam KG, 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](https://doi.org/10.1038/s41588-018-0183-z).
19. Patel AP, Wang M, Ruan Y, et al. A multi-ancestry polygenic risk score improves risk prediction for coronary artery disease. *Nat Med* 2023;29:1793-1803. DOI: [10.1038/s41591-023-02429-x](https://doi.org/10.1038/s41591-023-02429-x).
20. Truong B, Hull LE, Ruan Y, et al. Integrative polygenic risk score improves the prediction accuracy of complex traits and diseases. *Cell Genom* 2024;4:100523. DOI: [10.1016/j.xgen.2024.100523](https://doi.org/10.1016/j.xgen.2024.100523).
21. Mars N, Kerminen S, Feng Y-CA, et al. Genome-wide risk prediction of common diseases across ancestries in one million people. *Cell Genom* 2022;2:100118. DOI: [10.1016/j.xgen.2022.100118](https://doi.org/10.1016/j.xgen.2022.100118).
22. Shim I, Kuwahara H, Chen N, et al. Clinical utility of polygenic scores for cardiometabolic disease in Arabs. *Nat Commun* 2023;14:6535. DOI: [10.1038/s41467-023-41985-1](https://doi.org/10.1038/s41467-023-41985-1).
23. Michos ED, Van Spall HGC. Increasing representation and diversity in cardiovascular clinical trial populations. *Nat Rev Cardiol* 2021;18:537-538. DOI: [10.1038/s41569-021-00583-8](https://doi.org/10.1038/s41569-021-00583-8).
24. Yusuf S, Hawken S, Ounpuu S, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* 2004;364:937-952. DOI: [10.1016/S0140-6736\(04\)17018-9](https://doi.org/10.1016/S0140-6736(04)17018-9).
25. Albert MA, Glynn RJ, Fonseca FAH, et al. Race, ethnicity, and the efficacy of rosuvastatin in primary prevention: the Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) trial. *Am Heart J* 2011;162:106-114.e2. DOI: [10.1016/j.ahj.2011.03.032](https://doi.org/10.1016/j.ahj.2011.03.032).
26. Khan MS, Shahid I, Siddiqi TJ, et al. Ten-year trends in enrollment of women and minorities in pivotal trials supporting recent US

- Food and Drug Administration approval of novel cardiometabolic drugs. *J Am Heart Assoc* 2020;9:e015594. DOI: [10.1161/JAHA.119.015594](https://doi.org/10.1161/JAHA.119.015594).
27. Tcheandjieu C, Zhu X, Hilliard AT, et al. Large-scale genome-wide association study of coronary artery disease in genetically diverse populations. *Nat Med* 2022;28:1679-1692. DOI: [10.1038/s41591-022-01891-3](https://doi.org/10.1038/s41591-022-01891-3).
28. Nettleton JA, Polak JF, Tracy R, Burke GL, Jacobs DR. Dietary patterns and incident cardiovascular disease in the Multi-Ethnic Study of Atherosclerosis. *Am J Clin Nutr* 2009;90:647-654. DOI: [10.3945/ajcn.2009.27597](https://doi.org/10.3945/ajcn.2009.27597).
29. Hackler E, Lew J, Gore MO, et al. Racial differences in cardiovascular biomarkers in the general population. *JAHA* 2019;8:e012729. DOI: [10.1161/JAHA.119.012729](https://doi.org/10.1161/JAHA.119.012729).