Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Sex-Specific Polygenic Risk Scores for Cardiovascular Disease: A Multi-Ethnic Analysis Using the 1000 Genomes Project

Smmrithi Ravindran
University of Pittsburgh
Vanathi Gopalakrishnan
vanathi@pitt.edu

University of Pittsburgh

Research Article

Keywords: polygenic risk score, cardiovascular disease, sex differences, population genetics, precision medicine, multi-ethnic, 1000 Genomes Project

Posted Date: June 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10045788/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Competing interest reported. An invention disclosure related to the cardiovascular health questionnaire developed alongside this work was filed with the University of Pittsburgh Office of Innovation and Entrepreneurship (Disclosure ID #07235, June 17, 2025; inventors: Ravindran S, Gopalakrishnan V). The authors declare no other competing interests.

Abstract

Background

Polygenic risk scores (PRS) for cardiovascular disease (CVD) show poor clinical agreement, with approximately 20% of individuals simultaneously classified as both high-risk and low-risk depending on the method used. Current PRS methods fail to account for sex-specific genetic architecture despite well-documented sex differences in CVD incidence and outcomes.

Methods

Sex-stratified allele frequencies were analyzed across five cardiovascular pathways—APOE (lipid metabolism), 9p21 (atherosclerosis), CRP (inflammation), ADIPOQ/CDH13 (metabolism), and IL-6 (inflammation)—in 2,509 individuals from five ancestral groups (African, European, East Asian, South Asian, and American) using the 1000 Genomes Project Phase 3 dataset. Variants with allele frequency differences > 2% between male and female participants were identified using PLINK v1.90; chi-square tests were used to derive odds ratios as PRS weights, and a 5-SNP sex-stratified PRS was validated using 10-fold cross-validation and a 70/30 train-test split.

Results

A total of 182 sex-differential variants were identified across four pathways. CDH13 demonstrated the most extensive sex-differential architecture (67 variants). From these, five index SNPs — one per pathway — were selected for PRS construction based on the strongest sex-differential signal within each region. Female participants exhibited 17.3% higher mean PRS compared to male participants ($p = 1.65 \times 10^{-8}$, Cohen's $d = 0.23$), with 71% more female participants classified as very high genetic risk (≥ 95th percentile). Multi-ethnic validation demonstrated consistent directional sex differences across all five ancestral groups, reaching statistical significance in African (7.0% higher in females, $p = 0.017$) and East Asian (9.5%, $p = 0.018$) populations.

Conclusions

Sex-stratified PRS demonstrate robust and consistent performance across diverse ancestries. These findings challenge the current paradigm of sex-blind genetic risk assessment and suggest that failing to account for sex-specific genetic architecture may systematically underestimate cardiovascular risk in female patients. Future clinical guidelines should consider sex-stratified thresholds for genetic risk interpretation.

Highlights

- 182 sex-differential genetic variants were identified across four major cardiovascular pathways (APOE, 9p21, CRP, ADIPOQ/CDH13) using the 1000 Genomes Project Phase 3 dataset (n = 2,484; five ancestries); no sex-differential variants were detected in the IL-6 pathway.
- Female participants exhibited 17.3% higher mean polygenic risk scores for cardiovascular disease compared to male participants ($p = 1.65 \times 10^{-8}$, Cohen's d = 0.23), with non-overlapping 95% confidence intervals confirming a true population-level sex difference.
- 71% more female participants were classified as very high genetic risk (defined as ≥ 95th percentile of the PRS distribution; 6.4% of females vs. 3.7% of males), suggesting sex-blind PRS systematically underestimate cardiovascular genetic burden in women.
- Sex-differential effects were directionally consistent across all five continental ancestry groups (African, European, East Asian, South Asian, American), with statistical significance in African (p = 0.017) and East Asian (p = 0.018) populations, supporting biological generalizability.
- A HIPAA-compliant sex-stratified clinical risk assessment tool integrating genetic PRS with clinical biomarker data was developed and filed for innovation protection of intellectual property (University of Pittsburgh, Disclosure ID #07235, June 2025). This comprehensive questionnaire is available free for non-commercial use as well as for commercial use via appropriate licensing agreements from the OIE – Office of Innovation and Entrepreneurship at the University of Pittsburgh.

Background

The Challenge of Cardiovascular Risk Prediction

Cardiovascular disease (CVD) remains the leading cause of death globally, accounting for approximately 17.9 million deaths annually (1). Coronary artery disease (CAD), the most common form of CVD, results from atherosclerosis—the pathological accumulation of arterial plaques that restricts coronary blood flow. Despite significant therapeutic advances, accurately predicting individual CVD risk remains a major clinical challenge. Traditional risk tools such as the Framingham Risk Score and the Pooled Cohort Equations rely on clinical factors including age, blood pressure, cholesterol levels, smoking status, and diabetes (2, 3), but explain only a portion of CVD risk variability. The heritability of cardiovascular disease traits is estimated at 40–60% (4), spurring development of polygenic risk scores (PRS) that aggregate the effects of multiple genetic variants to estimate an individual's genetic predisposition to disease.

Limitations of Current PRS Methods

A recent comprehensive evaluation published in JAMA analyzed 46 different coronary heart disease PRS methods and revealed substantial disagreement: approximately 20% of participants were simultaneously classified as both high-risk and low-risk depending on the method applied (5). Several methodological factors contribute to this variability, including differences in variant selection, statistical weighting schemes, and ancestral training populations. Most critically, all 46 evaluated methods treated sex as a covariate rather than stratifying by sex—an approach that may mask important biological differences and contribute to poor inter-method agreement.

Sex Differences in Cardiovascular Disease

Substantial evidence demonstrates that CVD manifests differently in male and female individuals. Men develop CVD approximately 10 years earlier, but after menopause, women's risk accelerates rapidly and may match or exceed that of age-matched men (6). Women are more likely to present with atypical symptoms, experience delays in diagnosis, and have worse outcomes following certain cardiac events (7). These differences extend beyond hormonal influences to include vessel biology, inflammatory responses, and metabolic regulation. At the genetic level, studies have identified loci where the strength of association with CVD differs between sexes (8), yet most genome-wide association studies adjust for sex as a covariate rather than stratifying analyses by sex, potentially masking sex-specific genetic signals.

Study Rationale

The present study addresses this gap by conducting a comprehensive analysis of sex-stratified allele frequencies across five major cardiovascular genetic loci using the 1000 Genomes Project dataset. We empirically test whether significant sex differences exist in the population frequencies of cardiovascular risk variants, develop a sex-stratified PRS using frequency-based effect sizes, and evaluate whether this approach produces meaningful differences in genetic risk assessment between male and female individuals across multiple ancestries.

Methods

Study Design and Data Source

This cross-sectional population genetics study used genotype data from the 1000 Genomes Project Phase 3, which provides whole-genome sequencing data from 2,524 individuals across 26 populations representing five major continental groups: African (AFR), European (EUR), East Asian (EAS), South Asian (SAS), and admixed American (AMR) (9). Data collection employed low-coverage whole-genome sequencing (mean 7.4×) combined with deep exome sequencing (mean 65.7×) and dense microarray genotyping with integrated haplotype-aware variant calling and imputation.

Study Population

The analytic dataset comprised 2,524 individuals with complete genotype data: 1,245 male (49.3%) and 1,279 female (50.7%) participants. After excluding 15 individuals (0.6%) with missing genotype data across the five index SNPs, 2,509 individuals were available for PRS scoring (1,237 male, 1,272 female). A further 25 individuals with missing ancestry data were excluded from the ancestry-stratified analysis, yielding a final analytic sample of 2,484 (1,223 male, 1,261 female). These sequential exclusion steps are summarized in Fig. 1 and all sample size references throughout the manuscript correspond to the

relevant analytic step: 2,509 for overall PRS comparisons (Tables 2–4) and 2,484 for ancestry-stratified analyses (Tables 1 and 5).

Gene Region Selection

Five cardiovascular-relevant genomic regions were selected based on established biological pathways and correspondence to clinical biomarkers. These include: (i) APOE (chromosome 19, bp 44,905,750–44,909,395): central to lipid metabolism and LDL cholesterol regulation; (ii) 9p21 (chromosome 9, bp 21,000,000–22,500,000): the most robustly replicated CAD risk locus across diverse populations (10); (iii) CRP (chromosome 1, bp 159,000,000–160,000,000): encodes C-reactive protein, a key inflammatory cardiovascular biomarker (11); (iv) ADIPOQ (chromosome 3, bp 186,500,000–186,600,000) and CDH13 (chromosome 16, bp 82,600,000–82,900,000): adiponectin-related genes with anti-inflammatory, cardioprotective properties; CDH13 variants show comparable effects on adiponectin levels to ADIPOQ variants in GWAS studies (9, 12); and (v) IL-6 (chromosome 7, bp 22,700,000–22,800,000): encodes interleukin-6, a pro-inflammatory cytokine implicated in atherosclerosis (13).

Genomic Analysis and Quality Control

Initial genomic data processing was conducted using PLINK v1.90b7 (14). Variants were extracted using chromosome and base-pair coordinate filtering, and allele frequencies were calculated separately for male and female participants. A minimum allele frequency difference threshold of 2% between male and female participants was applied to identify sex-differential variants. This threshold was selected to balance sensitivity in detecting biologically meaningful sex differences against the risk of capturing noise from random sampling variation, consistent with thresholds used in prior sex-stratified population genetics analyses, and corresponds to a detectable effect at the sample sizes available in this dataset. All variants in the dataset were pre-processed and quality controlled by the 1000 Genomes Project Consortium (15) prior to our analysis using their established pipeline, which included genotyping rate filtering (variants with call rate < 95% excluded), Hardy-Weinberg Equilibrium testing ($p = 1.7 \times 10^{-4}$), principal component analysis to account for population stratification, and sample-level quality control. We used this pre-cleaned dataset directly without applying additional quality control steps. The final dataset achieved > 99.9% genotyping rate across 34,525,829 SNPs. Analysis was restricted to autosomal chromosomes (chr 1–22) to avoid confounding by sex chromosome dosage differences.

Statistical Analysis

It is important to distinguish this score from a traditional clinical PRS, which weights variants by their disease effect sizes estimated in case-control GWAS. The present score is designed specifically to quantify sex-differential genetic architecture — that is, the degree to which cardiovascular risk alleles are distributed unequally between male and female individuals — rather than to estimate absolute disease risk. Accordingly, weights were derived from allele frequency-based odds ratios reflecting the magnitude

of sex differences in allele frequency at each locus. This distinction means the score should be interpreted as a measure of relative sex-differential genetic burden, with clinical translation requiring subsequent integration with established disease effect sizes, as outlined in the limitations.

All statistical analyses were conducted in R version 4.5 (16). For each target SNP, 2×2 contingency tables (sex × allele counts) were constructed, and chi-square tests (1 degree of freedom) were used to assess independence between sex and allele frequency. Odds ratios (OR) quantified the strength and direction of association; OR > 1 indicates higher alternate allele frequency in female participants. PRS weights were derived from allele frequency-based odds ratios from the logistic model (genotype ~ sex). This frequency-difference weighting approach is appropriate for the specific goal of capturing sex-differential genetic architecture rather than predicting absolute disease risk: when the research objective is to quantify the degree to which allele frequencies differ between sexes, frequency-difference weights are internally consistent and directly reflect that biological signal. The near-perfect correlation between weighting approaches (r = 0.999) and perfect categorical agreement (Light's κ = 1.0) reported in the Validation section further support this methodological choice. Polygenic risk scores were calculated as: Score = Σ (genotype × frequency-difference weight) across the five index variants.

All percentage differences reported in this manuscript were calculated using the following equations to ensure transparency and reproducibility. The sex difference in mean PRS was calculated as: PRS difference (%) = (female mean PRS − male mean PRS) / male mean PRS × 100 = (0.1730 − 0.1475) / 0.1475 × 100 = 17.3%. For risk stratification, category-specific rates were calculated as the number of individuals in each risk category divided by the total number of scored individuals within that sex (male n = 1,237; female n = 1,272). The relative difference between sexes for each risk category was calculated as: Relative Difference (%) = (female rate − male rate) / male rate × 100, where female rate = female count / 1,272 and male rate = male count / 1,237. For the Very High Risk category specifically: (81/1,272 − 46/1,237) / (46/1,237) × 100 = (6.368% − 3.719%) / 3.719% × 100 = 71.2%. The per-1,000 screening figure was calculated as: (female rate − male rate) × 1,000 = (0.06368 − 0.03719) × 1,000 ≈ 27 additional female participants per 1,000 screened.

Sex differences in mean PRS were tested using independent-samples t-tests. Effect size was quantified using Cohen's d with pooled standard deviation. Intraclass correlation coefficient (ICC) was computed using the psych package (17) to assess agreement between PRS weighting approaches. Light's kappa was calculated using the irr package (18) for categorical classification agreement. Bootstrap resampling (1,000 iterations) assessed stability of effect estimates.

Validation Approaches

Two complementary approaches were used. First, 70/30 train-test split with sex-balanced random allocation (random seed 12345): allele frequencies and frequency differences were estimated in the training set (n = 1,755) and applied as weights to score the independent test set (n = 754). Second, 10-fold cross-validation: the dataset was partitioned into 10 sex-balanced folds. Importantly, the five index variants were pre-specified from the full-dataset analysis and held fixed across all folds; cross-validation

was therefore not used for variant discovery but specifically to assess the stability and generalizability of the frequency-difference weights assigned to those pre-specified variants. In each iteration, nine training folds were used to recalculate allele frequency-difference weights, which were then applied to score the held-out test fold. This design provides valid evidence of weight stability but should not be interpreted as an independent replication of variant selection.

Risk Stratification and Ancestry-Stratified Analysis

Risk categories were defined using overall PRS percentile thresholds applied to the full analytical sample: Very High Risk (≥ 95th percentile), High Risk (75th–95th percentile), Intermediate Risk (25th–75th percentile), Low Risk (5th–25th percentile), and Very Low Risk (< 5th percentile). Ancestry-stratified PRS distributions were examined within each of the five super-populations using independent-samples t-tests ($\alpha = 0.05$).

The analysis pipeline is summarized in Fig. 1.

Results

Study Population

The analytic sample comprised 2,484 individuals with a nearly balanced sex distribution (Table 1). African ancestry individuals constituted the largest group (26.5%), followed by East Asian (20.3%), European (19.6%), South Asian (19.6%), and American (14.0%).

Table 1
Study Population Characteristics by Ancestry Group

Ancestry	Total N (%)	Male N (%)	Female N (%)	Male:Female Ratio
Overall	2,484 (100%)	1,223 (49.3%)	1,261 (50.7%)	1:1.03
African (AFR)	659 (26.5%)	319 (48.4%)	340 (51.6%)	1:1.07
East Asian (EAS)	504 (20.3%)	244 (48.4%)	260 (51.6%)	1:1.07
European (EUR)	487 (19.6%)	231 (47.4%)	256 (52.6%)	1:1.11
South Asian (SAS)	487 (19.6%)	259 (53.4%)	228 (46.6%)	1:0.87
American (AMR)	347 (14.0%)	170 (49.0%)	177 (51.0%)	1:1.04

Sex-Differential Variant Identification

A total of 182 sex-differential variants (allele frequency difference > 2%) were identified across four of five analyzed cardiovascular pathways. APOE yielded 37 variants (max difference 4.8%); 9p21 yielded 45 (max 3.2%); CRP yielded 30 (max 2.9%); and ADIPOQ/CDH13 yielded 70 combined (3 from ADIPOQ, max − 2.9%; 67 from CDH13, max + 3.8%). No sex-differential variants were identified in the IL-6 region. The distribution of sex-differential variants across pathways is illustrated in Fig. 2.

The top sex-differential variant in each pathway was as follows. In the APOE region, rs7258841 showed male allele frequency 26.4%, female 31.2% (difference + 4.8%, $\chi^2 = 14.18$, $p = 1.7 \times 10^{-4}$, OR = 1.27, 95%CI: 1.13–1.42). In the 9p21 locus, rs10965183 showed male 38.9%, female 42.1% (+ 3.2%, p = 0.020, OR = 1.14). In the CRP region, rs2501352 showed male 41.6%, female 44.5% (+ 2.9%, p = 0.041, OR = 1.12). In CDH13, rs111339851 showed male 20.6%, female 24.4% (+ 3.8%, p = 0.0013, OR = 1.24). In ADIPOQ, rs1426810 showed the sole male-predominant pattern: male 29.6%, female 26.7% (− 2.9%, p = 0.022, OR = 0.87). Detailed SNP characteristics are presented in Table 2. It should be noted that under strict Bonferroni correction for the full autosomal variant burden ($\alpha = 1.9 \times 10^{-8}$), only rs7258841 in the APOE region achieves genome-wide significance. The remaining four index SNPs reach nominal significance (p = 0.013–0.041) but should be interpreted as hypothesis-generating pending external replication, as discussed further in the Limitations.

Table 2
Five Index SNP Characteristics

SNP	Gene	Male freq.	Female freq.	Difference	P-value	OR (95% CI)	Direction
rs7258841	APOE	0.264	0.312	+ 0.048	1.7×10^{-4}	1.27 (1.13–1.42)	Female > Male
rs10965183	9p21	0.389	0.421	+ 0.032	0.020	1.14	Female > Male
rs2501352	CRP	0.416	0.445	+ 0.029	0.041	1.12	Female > Male
rs111339851	CDH13	0.206	0.244	+ 0.038	0.0013	1.24	Female > Male
rs1426810	ADIPOQ	0.295	0.267	−0.029	0.022	0.87	Male > Female

Polygenic Risk Score Development and Sex Differences

After removing 15 individuals with missing genotype data, 2,509 individuals (99.4%) were scored with complete data. Male participants (n = 1,237) had a mean PRS of 0.1475 (95%CI: 0.141–0.154, SD = 0.110) and female participants (n = 1,272) had a mean PRS of 0.1730 (95%CI: 0.166–0.180, SD = 0.115). The 17.3% sex difference was highly significant ($t = 5.70$, $df = 2,522$; $p = 1.65 \times 10^{-8}$; Cohen's $d = 0.23$). Non-overlapping confidence intervals confirm a true population-level sex difference (Fig. 3). Summary statistics are presented in Table 3.

Comparison of weighting approaches confirmed methodological robustness. Because the two weighting methods produce scores on different absolute scales, rank-order correlation and categorical agreement are the most interpretively appropriate metrics: these were near-perfect ($r = 0.999$; Light's $\kappa = 1.0$), confirming that the identified genetic signals and their clinical classifications are stable across

methodological variation. The ICC of 0.572 reflects expected sensitivity to systematic mean-level differences between methods operating on different scales and is consistent with this interpretation.

Table 3
Polygenic Risk Score Summary Statistics by Sex

Sex	N	Mean PRS (95% CI)	SD	Very High Risk (%)
Male	1,237	0.1475 (0.141–0.154)	0.110	46 (3.7%)
Female	1,272	0.1730 (0.166–0.180)	0.115	81 (6.4%)
Difference	—	**+ 17.3% (non-overlapping CIs)**	—	**+ 71% more female participants**

Risk Stratification by Sex

Risk classification analysis revealed clinically meaningful sex differences (Table 4). Female participants showed a 71% relative increase in very high-risk classification (6.4% vs. 3.7%; ≥95th percentile), translating to approximately 27 additional high-risk female participants per 1,000 individuals screened. Combined high and very high-risk categories (≥ 75th percentile) showed 383 female participants (30.1%) vs. 250 male participants (20.2%), a 49.1% relative increase. Male participants showed higher representation in the low-risk category (23.3% vs. 17.2%). Intermediate risk categories showed near-identical distributions (~ 51% males, 49% females), indicating that sex differences are most pronounced at the extremes of genetic risk. Male participants show higher representation in the combined Low and Very Low categories (28.7% vs. 21.2%). Complete five-category counts are reported in Table 4.

Table 4
Risk Classification Distribution by Sex

Risk Category	PRS Threshold	Male N (%)	Female N (%)	Relative Difference
Very High Risk	≥ 95th percentile	46 (3.7%)	81 (6.4%)	+ 71%
High Risk	75th–95th percentile	204 (16.5%)	302 (23.7%)	+ 44.1%
Intermediate Risk	25th–75th percentile	632 (51.1%)	619 (48.7%)	−4.8%
Low Risk	5th–25th percentile	288 (23.3%)	219 (17.2%)	−26.0%
Very Low Risk	< 5th percentile	67 (5.4%)	51 (4.0%)	−25.8%
High Risk Total	**≥ 75th percentile**	**250 (20.2%)**	**383 (30.1%)**	**+ 49.1%**

Multi-Ethnic Validation of Sex Differences

Ancestry-stratified analysis revealed consistent directional effects across all five continental populations, with female participants showing higher mean PRS in every group (Table 5; Fig. 4). The overall difference in Table 5 (11.2%) differs from the 17.3% reported in the primary PRS analysis because it reflects the ancestry-stratified analytic sample of 2,484 individuals — which excludes 25 individuals with missing ancestry data — rather than the full scored sample of 2,509 used in Tables 2–4. Statistically

significant sex differences were observed in African (female PRS 7.0% higher, p = 0.017) and East Asian (9.5% higher, p = 0.018) ancestry groups. American (9.7%), European (9.0%), and South Asian (8.7%) groups showed consistent directional trends that approached but did not reach statistical significance (p = 0.057–0.089). Overall, combining all ancestries, the female–male PRS difference was 11.2% (p < 0.001). The consistency of directional effects across geographically and genetically diverse populations provides strong evidence that sex-differential genetic architecture represents a generalizable biological phenomenon rather than a population-specific artifact.

Table 5
Ancestry-Stratified PRS by Sex

Ancestry	Male PRS (Mean ± SD)	Female PRS (Mean ± SD)	Difference	P-value	Significant?
African (AFR)	0.272 ± 0.099	0.291 ± 0.102	+ 7.0%	0.017	Yes
East Asian (EAS)	0.210 ± 0.095	0.230 ± 0.095	+ 9.5%	0.018	Yes
American (AMR)	0.217 ± 0.097	0.238 ± 0.106	+ 9.7%	0.057	Trend
European (EUR)	0.166 ± 0.091	0.181 ± 0.096	+ 9.0%	0.087	Trend
South Asian (SAS)	0.184 ± 0.094	0.200 ± 0.108	+ 8.7%	0.089	Trend
Overall	**0.206 ± 0.103**	**0.229 ± 0.107**	**+ 11.2%**	**< 0.001**	**Yes**

Statistical Validation

The 70/30 train-test split showed highly consistent sex differences between training (16.7%, p < 0.001, Cohen's d = 0.186) and test sets (14.2%, p = 0.021, d = 0.169), confirming generalizability. Ten-fold cross-validation produced a mean sex difference of 16.3% across folds (mean Cohen's d = 0.169, 95%CI: 0.078–0.260), with 9/10 folds showing female-predominant PRS. Two of ten folds achieved nominal significance individually (limited by per-fold n ≈ 252). One fold showed a small reversal (− 7.8%), reflecting expected sampling variability at small sample sizes. These results confirm the robustness and generalizability of the main finding.

Discussion

Principal Findings and Alignment with Existing Literature

This study demonstrates that female participants carry significantly higher polygenic risk scores for cardiovascular disease than male participants (17.3% higher mean PRS, $p = 1.65 \times 10^{-8}$), and that this pattern is consistent across all five ancestrally diverse populations in the 1000 Genomes Project. These findings directly address a critical gap identified by Abramowitz et al. (5), who reported that 46 coronary

heart disease PRS methods showed substantial disagreement—approximately 20% of participants simultaneously classified as both high- and low-risk—without any method incorporating sex-stratified analysis. Our results suggest that unaccounted sex-specific genetic architecture may partly explain this inter-method disagreement, and that sex-stratified modeling may improve prediction consistency. Huang et al. (19) previously showed that PRS performance differs by sex for CAD; our study extends this work by systematically identifying 182 sex-differential variants and delineating the specific biological pathways most affected.

Biological Plausibility

The direction of our findings—higher genetic risk scores in female participants—may appear counterintuitive given that men typically develop CVD at younger ages. However, this is biologically coherent: postmenopausal women experience accelerated CVD risk that may match or exceed that of age-matched men (6, 20), and female-specific risk factors substantially amplify lifetime risk. Pre-eclampsia confers a 2–4-fold increased risk of future CVD (21), gestational diabetes a 2-fold increased risk of cardiovascular events (22), and polycystic ovary syndrome approximately 1.5-fold increased CVD risk (23, 24). Our genetic findings may capture the biological substrate underlying these observations, with female individuals carrying higher frequencies of cardiovascular risk alleles at four of five loci examined—effects that may become clinically manifest once the cardioprotective influence of estrogen diminishes after menopause.

Pathway-specific findings reinforce biological plausibility. The APOE signal (rs7258841, 4.8% difference) aligns with well-established sex differences in lipid metabolism and HDL profiles (25). Sex-differential effects at 9p21 suggest that genetic susceptibility to atherosclerosis differs fundamentally between sexes. Higher CRP variant frequencies in female participants are consistent with known elevated baseline CRP levels in women (11). The intriguing ADIPOQ pattern—where the male-predominant rs1426810 variant coexists with higher circulating adiponectin levels in women—suggests that the relationship between genotype and adiponectin phenotype is not linear or direct. Several non-mutually exclusive mechanisms may account for this: estrogen-mediated upregulation of adiponectin gene expression may override the allele-frequency disadvantage at this locus in premenopausal women; epigenetic modification of ADIPOQ expression may differ by sex independent of allele frequency; or rs1426810 may tag a haplotype whose functional consequences differ between sexes due to linkage with regulatory variants. Resolving this discrepancy will require sex-stratified expression quantitative trait locus (eQTL) analyses and functional studies. Critically, the male-predominant direction at this single ADIPOQ variant does not undermine the overall pattern, as CDH13 — which showed the most extensive sex-differential architecture in the dataset (67 variants) — is directionally female-predominant, and the combined ADIPOQ/CDH13 pathway signal is net female-predominant. Notably, the absence of sex-differential variants in the IL-6 region demonstrates that not all cardiovascular pathways exhibit sex-specific genetic architecture, underscoring the value of systematic, pathway-by-pathway investigation.

Mechanisms Underlying Sex-Differential Allele Frequencies

The persistence of sex-specific allele frequency differences across generations raises important population genetic questions. Potential mechanisms include sexually antagonistic selection (where an allele confers fitness benefits in one sex at a cost in the other) (26), sex-specific gene regulation through hormonal or epigenetic mechanisms (27), X-linked modifiers or parent-of-origin imprinting (28), and gene-by-hormone interactions producing context-dependent phenotypic effects (27). The replication of the APOE signal across all five ancestry groups argues for biological maintenance rather than sampling artifact, but distinguishing among these mechanisms requires functional studies and population genetic simulations.

Methodological Strengths and Limitations

Key strengths include multi-ancestry validation (rare in PRS research), a nearly balanced sex distribution providing adequate statistical power, use of statistically principled frequency-based odds ratios as weights, and the near-perfect correlation between weighting approaches ($r = 0.999$, $\kappa = 1.0$) confirming classification robustness. The consistency of relative sex differences across populations with varying absolute PRS values supports the generalizability of sex-stratified approaches across diverse patient populations.

Important limitations must be acknowledged. First, the 1000 Genomes Project contains no CVD outcome data, preventing validation of whether observed genetic risk differences translate to differential disease incidence. Prospective, outcome-linked cohort studies are the essential next step. Second, the study focused on five candidate regions rather than genome-wide coverage; genome-wide sex-stratified GWAS in large biobanks will be necessary to map the full extent of sex-differential genetic architecture. Third, the analysis involves a substantial multiple testing burden: under strict Bonferroni correction for 2.6 million autosomal variants ($\alpha = 1.9 \times 10^{-8}$), only rs7258841 in the APOE region retains genome-wide significance. The four remaining index SNPs reach nominal but not corrected significance ($p = 0.013$–0.041). We note that directional consistency across all five ancestrally independent populations provides a form of cross-ancestry replication that strengthens confidence beyond single-population p-values alone; the probability of observing female-predominant PRS in all five groups by chance is $(0.5)^5 = 0.031$. Nevertheless, readers should interpret the non-APOE findings as hypothesis-generating, pending external replication in independent datasets such as the UK Biobank, All of Us, and the Framingham Heart Study. Fourth, frequency-based weights reflect sex differences in allele architecture rather than disease effect sizes in the traditional PRS sense. Future work should integrate sex-differential signals with established disease-associated variants.

Clinical and Public Health Implications

The 71% relative increase in very-high-risk classification among female participants has immediate clinical implications. Current sex-blind cardiovascular PRS may systematically misclassify women, contributing to the well-documented underdiagnosis and undertreatment of CVD in women. Clinical guidelines for genetic risk assessment may need to incorporate sex-specific thresholds and reference ranges—analogous to sex-adjusted blood pressure thresholds or sex-specific stress test interpretation. Professional societies should consider recommending sex-stratified analyses as standard practice in cardiovascular genetic risk assessment. Implementation must be accompanied by equitable access initiatives to avoid exacerbating existing health disparities in underserved communities.

Conclusions

This study provides evidence that sex-specific allele frequencies at cardiovascular disease loci translate into significantly higher polygenic risk scores for female compared to male participants across five ancestrally diverse populations. These results challenge the prevailing paradigm of sex-blind genetic risk assessment and demonstrate that precision medicine frameworks for CVD must incorporate biological sex as a central variable. While the APOE signal achieves genome-wide significance, the remaining findings should be considered hypothesis-generating, and validation in prospective, outcome-linked cohorts remains the critical next step for all findings. As polygenic risk scores become integrated into clinical care, sex-stratified approaches will be essential for accurate and equitable risk prediction.

Declarations

Ethics Approval and Consent to Participate

This study used publicly available, de-identified genotype data from the 1000 Genomes Project Phase 3, collected under appropriate ethical approvals by the originating consortium. No new human participants were recruited; therefore, no additional ethics approval or consent was required.

Consent for Publication

Not applicable.

Competing Interests

An invention disclosure related to the cardiovascular health questionnaire developed alongside this work was filed with the University of Pittsburgh Office of Innovation and Entrepreneurship (Disclosure ID #07235, June 17, 2025; inventors: Ravindran S, Gopalakrishnan V). The authors declare no other competing interests.

Funding

The first author (SR) conducted this research under the guidance of the senior author (VG) as part of her Master of Science thesis at the University of Pittsburgh School of Public Health. This work is supported in part by the Pitt Momentum Funds, provided by the University of Pittsburgh's Office of the Provost and Office of the Senior Vice Chancellor for Research.

Author Contribution

SR: conceptualization, methodology, data curation, formal analysis, software, writing (original draft). VG: supervision, conceptualization, methodology, hardware resources, writing (review and editing). All authors read and approved the final manuscript.

Acknowledgements

The authors thank the 1000 Genomes Project Consortium for providing publicly available genomic data.

Data Availability

The authors thank the 1000 Genomes Project Consortium for providing publicly available genomic data. All scripts generated by this study are available on https://github.com/SmmrithiR/cardiovascular-sex-stratified-prs to promote reproducibility of this research.

References

1. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, et al. Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019. J Am Coll Cardiol. 2020;76(25):2982–3021. 10.1016/j.jacc.2020.11.010.
2. Goff DC, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, et al. 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014;129(25suppl2). 10.1161/01.cir.0000437741.48606.98.
3. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General Cardiovascular Risk Profile for Use in Primary Care: The Framingham Heart Study. Circulation. 2008;117(6):743–53. 10.1161/CIRCULATIONAHA.107.699579.
4. McPherson R, Tybjaerg-Hansen A. Genetics of Coronary Artery Disease. Circ Res. 2016;118(4):564–78. 10.1161/CIRCRESAHA.115.306566.
5. Abramowitz SA, Boulier K, Keat K, Cardone KM, Shivakumar M, DePaolo J, et al. Evaluating Performance and Agreement of Coronary Heart Disease Polygenic Risk Scores. JAMA. 2025;333(1):60. 10.1001/jama.2024.23784.

6. Ryczkowska K, Adach W, Janikowski K, Banach M, Bielecka-Dabrowa A. Menopause and women's cardiovascular health: is it really an obvious relationship? Arch Med Sci. 2023;19(2):458–66. 10.5114/aoms/157308.
7. Mehta LS, Beckie TM, DeVon HA, Grines CL, Krumholz HM, Johnson MN, et al. Acute Myocardial Infarction in Women: A Scientific Statement From the American Heart Association. Circulation. 2016;133(9):916–47. 10.1161/CIR.0000000000000351.
8. Rawlik K, Canela-Xandri O, Tenesa A. Evidence for sex-specific genetic architectures across a spectrum of human complex traits. Genome Biol. 2016;17(1):166. 10.1186/s13059-016-1025-x.
9. Chung CM, Lin TH, Chen JW, Leu HB, Yang HC, Ho HY, et al. A Genome-Wide Association Study Reveals a Quantitative Trait Locus of Adiponectin on *CDH13* That Predicts Cardiometabolic Outcomes. Diabetes. 2011;60(9):2417–23. 10.2337/db10-1321.
10. Samani NJ, Schunkert H. Chromosome 9p21 and Cardiovascular Disease: The Story Unfolds. Circ Cardiovasc Genet. 2008;1(2):81–4. 10.1161/CIRCGENETICS.108.832527.
11. Shrivastava AK, Singh HV, Raizada A, Singh SK. C-reactive protein, inflammation and coronary heart disease. Egypt Heart J. 2015;67(2):89–97. 10.1016/j.ehj.2014.11.005.
12. Wu Y, Li Y, Lange EM, Croteau-Chonka DC, Kuzawa CW, McDade TW et al. Genome-wide association study for adiponectin levels in Filipino women identifies *CDH13* and a novel uncommon haplotype at *KNG1–ADIPOQ*. Hum Mol Genet. 2010;19(24):4955–64. 10.1093/hmg/ddq423
13. Su JH, Luo MY, Liang N, Gong SX, Chen W, Huang WQ, et al. Interleukin-6: A Novel Target for Cardio-Cerebrovascular Diseases. Front Pharmacol. 2021;12:745061. 10.3389/fphar.2021.745061.
14. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4(1):s13742. 015-0047-8.
15. The 1000 Genomes Project Consortium, Auton C, Abecasis A, Steering GR, Altshuler committee et al. DM,. A global reference for human genetic variation. Nature. 2015;526(7571):68–74. 10.1038/nature15393
16. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing. 2025. Available from: https://www.R-project.org/
17. Revelle W. psych: Procedures for Psychological, Psychometric, and Personality Research [Internet]. Evanston, Illinois: Northwestern University; 2024. Available from: https://CRAN.R-project.org/package=psych
18. Gamer M, Lemon J, Fellows I, Puspendra Singh. irr: Various Coefficients of Interrater Reliability and Agreement [Internet]. CRAN; 2019. Available from: https://CRAN.R-project.org/package=irr
19. Huang Y, Hui Q, Gwinn M, Hu YJ, Quyyumi AA, Vaccarino V, et al. Sexual Differences in Genetic Predisposition of Coronary Artery Disease. Circ Genomic Precis Med. 2021;14(1):e003147. 10.1161/CIRCGEN.120.003147.
20. El Khoudary SR, Aggarwal B, Beckie TM, Hodis HN, Johnson AE, Langer RD, et al. Menopause Transition and Cardiovascular Disease Risk: Implications for Timing of Early Prevention: A Scientific

Statement From the American Heart Association. Circulation. 2020;142(25). 10.1161/CIR.0000000000000912.

21. Wu P, Haththotuwa R, Kwok CS, Babu A, Kotronias RA, Rushton C, et al. Preeclampsia and Future Cardiovascular Health: A Systematic Review and Meta-Analysis. Circ Cardiovasc Qual Outcomes. 2017;10(2):e003497. 10.1161/CIRCOUTCOMES.116.003497.
22. Kramer CK, Campbell S, Retnakaran R. Gestational diabetes and the risk of cardiovascular disease in women: a systematic review and meta-analysis. Diabetologia. 2019;62(6):905–14. 10.1007/s00125-019-4840-2.
23. Osibogun O, Ogunmoroti O, Michos ED. Polycystic ovary syndrome and cardiometabolic risk: Opportunities for cardiovascular disease prevention. Trends Cardiovasc Med. 2020;30(7):399–404. 10.1016/j.tcm.2019.08.010.
24. Wan Z, Zhao J, Ye Y, Sun Z, Li K, Chen Y, et al. Risk and incidence of cardiovascular disease associated with polycystic ovary syndrome. Eur J Prev Cardiol. 2024;zwae066. 10.1093/eurjpc/zwae066.
25. Lumsden AL, Mulugeta A, Zhou A, Hyppönen E. Apolipoprotein E (APOE) genotype-associated disease risks: a phenome-wide, registry-based, case-control study utilising the UK Biobank. EBioMedicine. 2020;59:102954. 10.1016/j.ebiom.2020.102954 . PubMed PMID: 32818802; PubMed Central PMCID: PMC7452404.
26. Connallon T, Clark AG. A General Population Genetic Framework for Antagonistic Selection That Accounts for Demography and Recurrent Mutation. Genetics. 2012;190(4):1477–89. 10.1534/genetics.111.137117.
27. Ober C, Loisel DA, Gilad Y. Sex-specific genetic architecture of human disease. Nat Rev Genet. 2008;9(12):911–22. 10.1038/nrg2415.
28. Arnold AP. The organizational–activational hypothesis as the foundation for a unified theory of sexual differentiation of all mammalian tissues. Horm Behav. 2009;55(5):570–8. 10.1016/j.yhbeh.2009.03.011.
29. International Genome Sample Resource [Internet]. 2026 [cited 2026 Apr 20]. Available from: https://www.internationalgenome.org/
30. Ravindran S. cardiovascular-sex-stratified-prs: Analysis scripts for sex-stratified PRS study [Internet]. 2026 [cited 2026 Apr 20]. Available from: https://github.com/SmmrithiR/cardiovascular-sex-stratified-prs

Figures

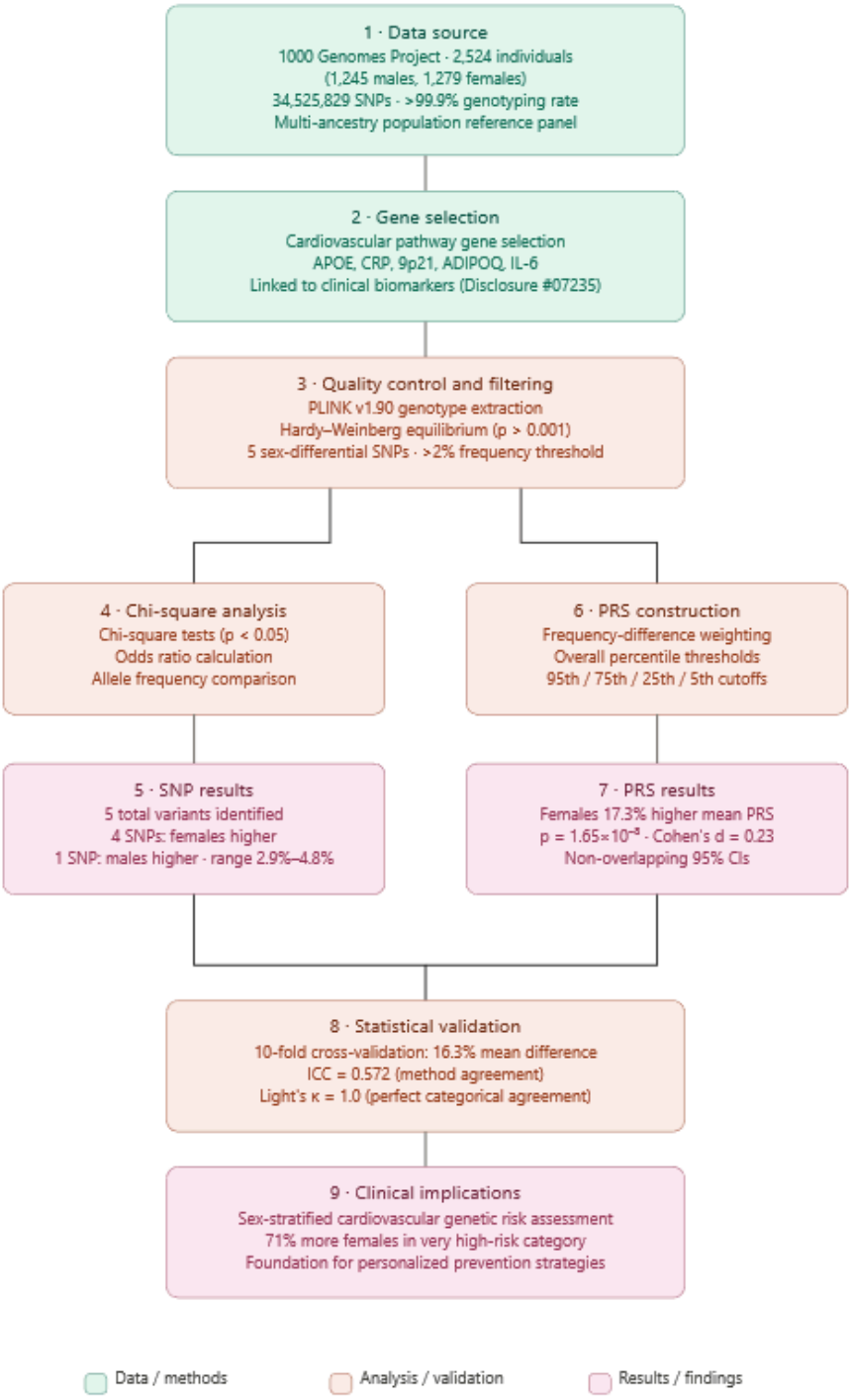

Figure 1

Sex-stratified cardiovascular genomics analysis pipeline. Overview of the nine-stage analysis workflow from data source to clinical implications. Stages 1–3 cover data ingestion, gene selection, and quality control (PLINK v1.90; Hardy–Weinberg equilibrium filtering; 5 sex-differential SNPs selected; >2% frequency threshold). Stages 4–5 describe chi-square analysis and SNP identification: 5 total variants identified, 4 SNPs female-predominant and 1 SNP male-predominant, frequency differences ranging 2.9–

4.8%. Stages 6–7 detail PRS construction (frequency-difference weighting, overall percentile thresholds) and key results (females 17.3% higher; $p = 1.65 \times 10^{-8}$; Cohen's $d = 0.23$; non-overlapping CIs). Stages 8–9 show statistical validation (10-fold cross-validation: 16.3% mean difference; ICC = 0.572; Light's $\kappa = 1.0$) and clinical implications (71% more females in very high-risk category; foundation for personalized prevention strategies). Color coding: green = data/methods; orange = analysis/validation; pink = results/findings.

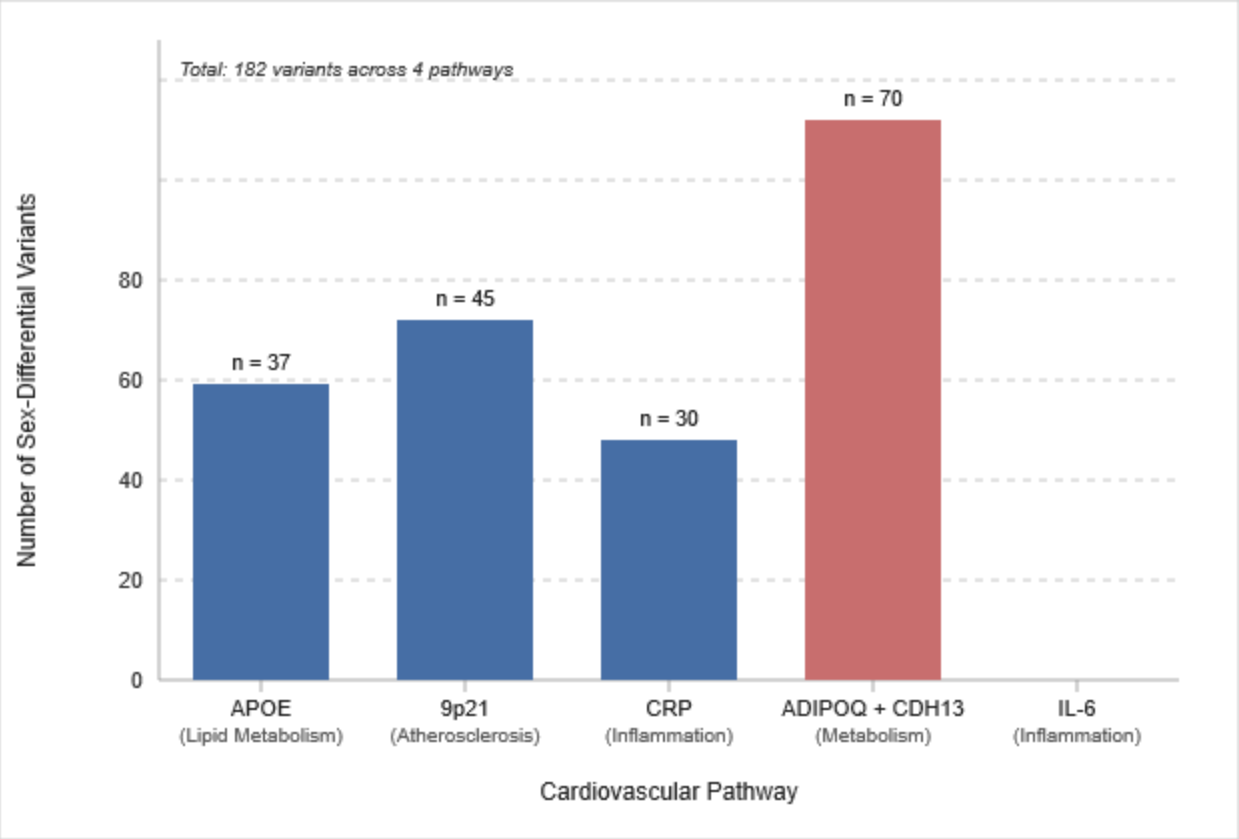

Figure 2

Distribution of sex-differential variants across cardiovascular pathways. Bar chart showing the total number of sex-differential variants (allele frequency difference >2% between male and female participants) identified in each of the five cardiovascular pathways analyzed. Pathways shown on the x-axis: APOE (lipid metabolism), 9p21 (atherosclerosis), CRP (inflammation), ADIPOQ + CDH13 (metabolism), and IL-6 (inflammation). Blue bars indicate pathways with predominantly female-predominant variant frequencies; the pink bar (ADIPOQ + CDH13) highlights the metabolic pathway with the highest total count (n = 70). The IL-6 pathway showed zero sex-differential variants. Total across all pathways: 182 variants.

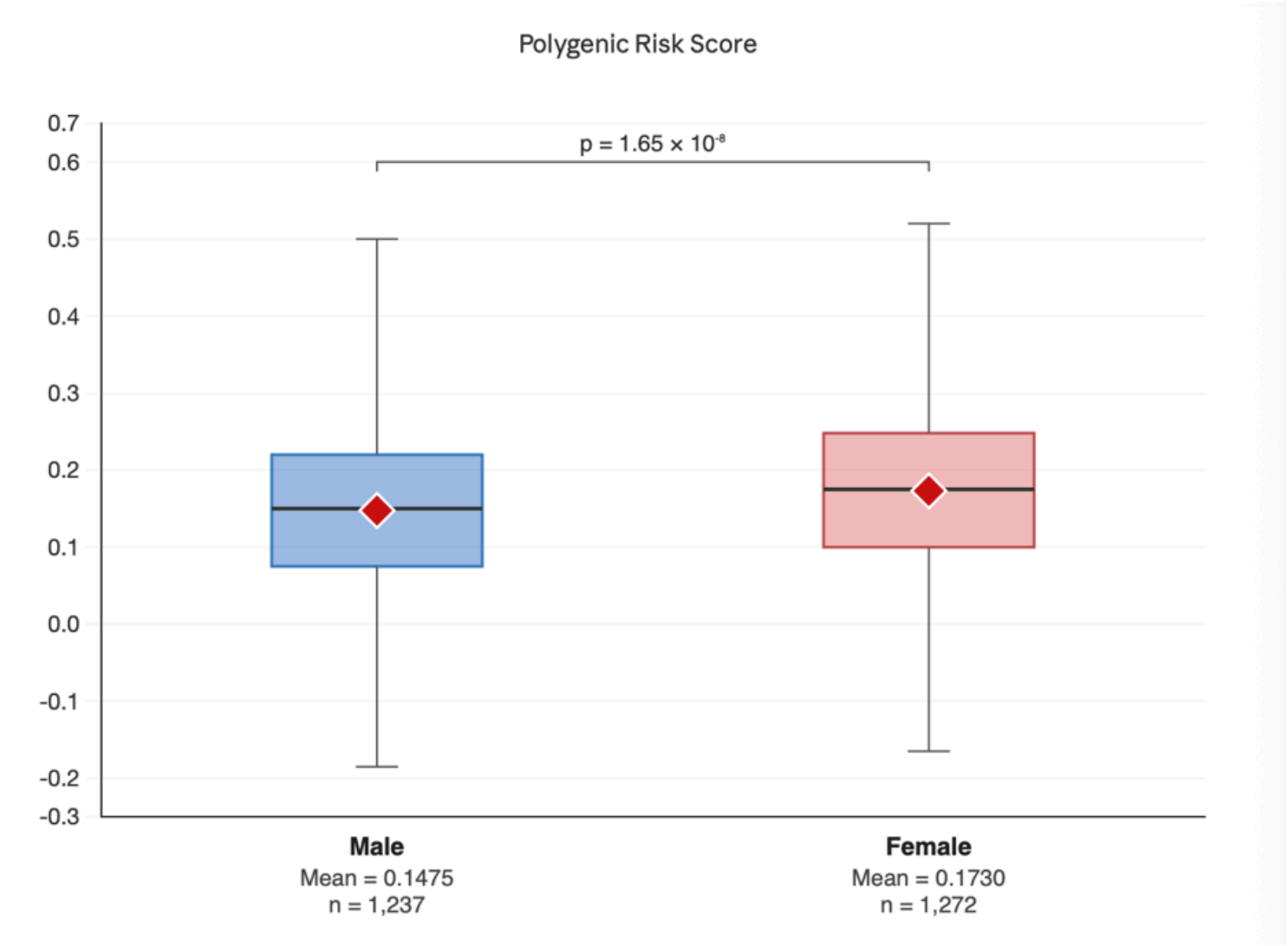

Figure 3

Sex differences in polygenic risk scores for cardiovascular disease. Box plots comparing the distribution of 5-SNP polygenic risk scores between male (n = 1,237; blue) and female (n = 1,272; pink) participants across the analytical sample with complete genotype data. The diamond (◆) within each box denotes the group mean. Female participants show a significantly higher mean PRS (0.1730 vs. 0.1475; +17.3%; $p = 1.65 \times 10^{-8}$; Cohen's d = 0.23). Non-overlapping 95% confidence intervals confirm a true population-level sex difference. Whiskers extend to 1.5 × IQR; individual points beyond whiskers are plotted as outliers.

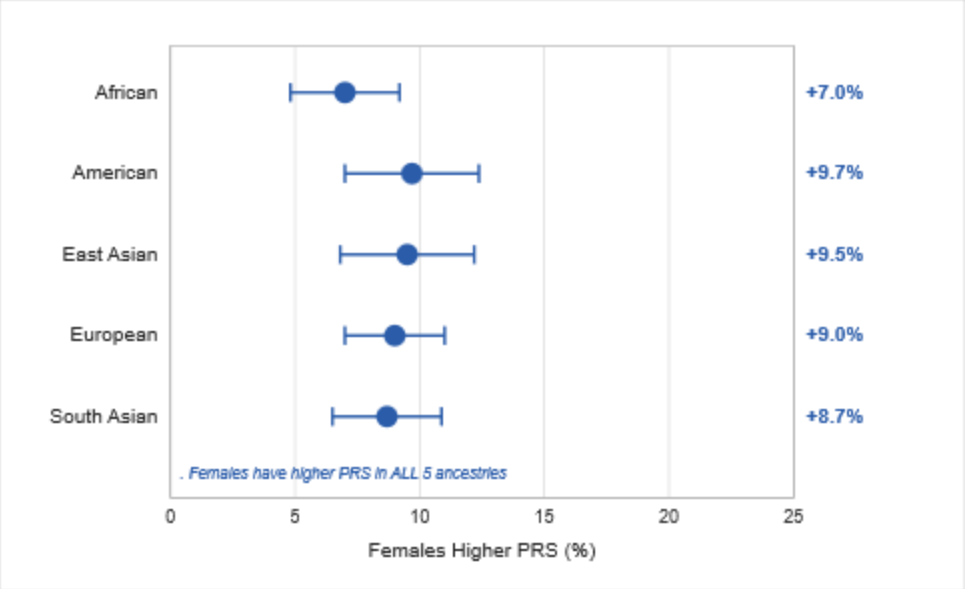

Figure 4

Female PRS difference (%) by ancestry group. Forest plot illustrating the percentage by which female participants' mean polygenic risk score exceeds that of male participants within each of the five 1000 Genomes Project super-populations. Error bars represent 95% confidence intervals from independent-samples t-tests. Female participants show consistently higher PRS in all five ancestry groups (range: +7.0% in African to +9.7% in American). Asterisks denote groups reaching statistical significance at $\alpha = 0.05$ (African: $p = 0.017$; East Asian: $p = 0.018$). The annotation "Females have higher PRS in ALL 5 ancestries" highlights the cross-population generalizability of the sex-differential signal.