

Polygenic prediction of coronary heart disease among 130 000 Mexican adults

Tianshu Liu ^{1†}, Jaime Berumen ^{2†}, Jason Torres ^{1†}, Jesus Alegre-Díaz ²,
Paulina Baca ², Carlos González-Carballo ², Raul Ramirez-Reyes ²,
Fernando Rivas ², Diego Aguilar-Ramirez ¹, Fiona Bragg ^{1,3},
William Herrington ¹, Michael Hill ¹, Eirini Trichia ¹, Alejandra Vergara-Lope ¹,
Rachel Wade ¹, Rory Collins ¹, Pablo Kuri-Morales ^{4,5},
Jonathan R. Emberson ^{1*‡}, Roberto Tapia-Conyer ^{5‡}, and
Louisa Gnatiuc Friedrichs ^{1‡}

¹Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health (NDPH), University of Oxford, Oxford, UK; ²Experimental Research Unit, Faculty of Medicine, National Autonomous University of Mexico (UNAM), Mexico City, Mexico; ³Health Data Research UK Oxford, University of Oxford, Oxford, UK; ⁴Instituto Tecnológico y de Estudios Superiores de Monterrey, Monterrey, Mexico; and ⁵Faculty of Medicine, UNAM, Mexico City, Mexico

Received 13 February 2025; revised 3 July 2025; accepted 4 November 2025; online publish-ahead-of-print 13 November 2025

Aims

Most polygenic risk scores (PRSs) for coronary heart disease (CHD) are derived from European ancestry populations. This paper evaluates the performance of CHD PRSs in a Mexican population.

Methods and results

133 207 participants aged 35–79 years from the Mexico City Prospective Study were included. Eight PRSs (comprising 44–6 472 620 polymorphisms) were selected for prediction of CHD (defined as self-reported prior heart attack or angina, or CHD death before age 80 years). Logistic regression was adjusted for age, sex, and the first seven genetic principal components, before and after adjustment for other cardiovascular risk factors. The area under the receiving operating characteristic curve (‘C-statistic’) was also estimated. Of the included participants, 67% were women, the mean (\pm SD) age was 51 ± 12 years, and Indigenous American ancestry averaged 67%. CHD was documented for 5163 participants (3.9%), including 1901 prevalent and 3479 fatal cases. All eight PRSs were positively and log-linearly associated with CHD, with odds ratios (ORs) per 1 SD PRS increase ranging from 1.05 (95% CI, 1.03–1.08) to 1.29 (95% CI, 1.25–1.33). Associations were consistent across strata of age and ancestry and were independent of other vascular risk factors. For six PRSs, however, associations were substantially stronger in men than women. Multi-ancestry PRSs outperformed Eurocentric-ancestry PRSs. Despite remaining predictive of risk independently of established non-genetic risk factors, inclusion of a PRS into a risk model did not increase the C-statistic noticeably.

Conclusion

In this Mexican population, existing PRSs predicted CHD independently of established vascular risk factors, particularly for men. PRSs better capturing genetic variation in Latin American people may further enhance risk prediction in such populations.

Lay summary

This study looked at how well eight previously developed genetic risk scores for coronary heart disease (CHD) apply to people in Mexico, a population that has mixed ancestry.

- We studied over 133 000 Mexican adults aged 35–79. All eight genetic scores were linked to a higher risk of CHD, even after accounting for the traditional risk factors of smoking, adiposity, blood pressure, diabetes, and education.
- In general, the scores that were developed using data from multiple ancestral groups were better at predicting CHD risk than those based mostly on people of European descent. Six of the eight scores were also more strongly linked to CHD in men than in women.

* Corresponding author. Tel: +44 1865 743906, Email: jonathan.emberson@ndph.ox.ac.uk

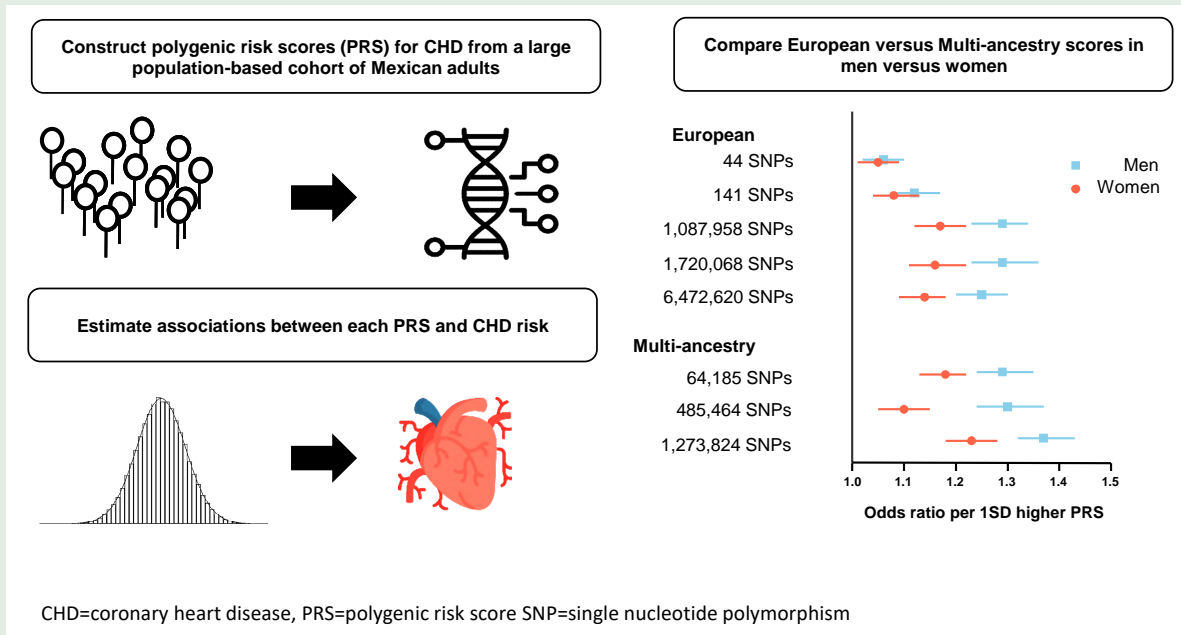
† The lead authors contributed equally to the paper.

‡ The last authors contributed equally to the paper.

© The Author(s) 2025. Published by Oxford University Press on behalf of the European Society of Cardiology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Graphical Abstract



Keywords

Mexico • Prospective study • Coronary heart disease • Polygenic risk scores

Introduction

Coronary heart disease (CHD) is a major cause of death and disability worldwide. It was estimated that in 2021, CHD was responsible for 9 million deaths (13% of deaths globally).¹ Despite reductions in age-specific CHD mortality rates in many developed countries over the past few decades, CHD remains the leading cause of death among adults globally and therefore its prevention remains a global public health priority.

CHD is estimated to have a heritability up to 50%,^{2,3} and several studies have shown that accounting for genetic predisposition can enhance CHD risk stratification beyond conventional vascular risk factors.^{4–6} Polygenic CHD risk scores (PRSs) that combine multiple single nucleotide polymorphism (SNP) effects into a single genetic score have gained particular interest and have been shown to identify individuals with CHD risk equivalent to (or even higher than) those with rare monogenic mutations with large effects (e.g. familial hypercholesterolemia).⁷ Increasingly advanced methods have been developed to improve the predictive power of PRSs among European populations,^{8–10} but the transferability of these PRSs to populations of other ancestries (with different linkage disequilibrium patterns and allele frequencies) is unclear.¹¹ Since participants of non-European descent comprise only ~5% of those in existing GWAS studies,¹² PRS evaluation and development in other populations has been limited.

In Mexico, the CHD mortality rate at ages 35–79 years has increased over the past 70 years¹ (in men from ~60 per 100 000 in 1955 to 210 per 100 000 in 2020, and in women from ~40 per 100 000 in 1955 to 90 per 100 000 in 2020),¹³ in large part because of substantial population increases in major CHD risk factors over this period (in particular obesity and diabetes). The impact of varying prevalences (and effects) of major CHD risk factors on the performance of a PRS in a population is uncertain, especially for scores derived in one population but applied to

another. Using data from the Mexico City Prospective Study (MCPS), the aim of this paper is to evaluate the shape and strength of the association of eight existing CHD polygenic risk scores with CHD in an admixed population of Mexican adults with high levels of obesity and diabetes.

Methods

Study design and population

Details of the MCPS design, methods, and population have been described previously.¹⁴ Briefly, between 1998 and 2004, households in two districts of Mexico City (Coyoacán and Iztapalapa) were visited and household members aged 35 years or older were invited to participate. Of 112 333 households with eligible inhabitants, one or more individuals from 106 059 (95%) households consented to participate. Ethics approval was obtained from the Mexican Ministry of Health, the Mexican National Council for Science and Technology, and the University of Oxford. All participants provided written informed consent.

Data collection

During household visits, trained nurses administered electronic questionnaires collecting information on sociodemographic and lifestyle factors, medication, and disease history. Height, weight, hip circumference, waist circumference, and sitting blood pressure were measured using calibrated instruments and standard protocols. A non-fasting venous blood sample was collected into an EDTA vacutainer and separated into two plasma and one buffy coat aliquots for long-term storage at -150°C . Variants were genotyped on the Illumina GSAv2 BeadChip and mapped to genome build GRCh38 (hg38).¹⁵ Variant quality control of genotype and individual-level missingness, departures from Hardy–Weinberg equilibrium, Mendel errors, genotype imputation (to TOPMed version 2), and per-individual proportions of Indigenous, European, African, and East Asian ancestry were estimated as previously described.¹⁵

CHD polygenic risk scores

From a systematic search of PubMed, Embase, and Medline, we selected eight published CHD-specific PRSs based either on their relevance to the ethnicities represented in the MCPS cohort, their extensive evaluation in the existing literature, or the methodology they employed (see [Supplementary material online, Table S1](#) for details) to assess their transferability for predicting CHD risk within the MCPS cohort. Five of the selected PRSs (Khera *et al.*,⁷ Inouye *et al.*,¹⁶ Tamlander *et al.*,¹⁷ Oni-Orisan *et al.*,¹⁸ and Tada *et al.*¹⁹) were derived from European ancestry populations using a variety of methods. The remaining three PRSs selected (Koyama *et al.*,²⁰ Patel *et al.*,²¹ and Tcheandjieu *et al.*²²) involved multiple ancestries during training (i.e. GWAS source or PRS tuning). The eight CHD-specific PRSs were recreated for each MCPS participant using individual genotyped data and the pipeline developed by the PGS Catalog team.^{23,24} The SNP variant matching rates for all the selected PRSs were >85%, based on exact matching with no use of proxy SNPs. The PGS Catalog program `pgsc_calc` (v1.3.2) was performed using Nextflow-23.04.0,²⁵ PLINK-2.0,²⁶ and Anaconda Distribution 3.

Follow-up for mortality

Participants are followed up for cause-specific mortality through probabilistic linkage (based on name, including phonetic coding of names, age, and sex) to the Mexican System for Epidemiologic Death Statistics (Subsistema Epidemiológico y Estadístico de Defunciones or SEED) electronic death registry in Mexico City, administered by the Ministry of Health. Field validation of more than 7000 matched deaths confirmed the reliability of the matching algorithm in over 95% of cases. Death registration in Mexico City is reliable and complete, with causes of almost all deaths certified medically.²⁷ Diseases recorded on death certificates are coded using the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10), with subsequent review by study clinicians (who are unaware of baseline information) to recode, when necessary, the underlying cause of death.²⁸ Participants deaths were tracked until September 2022.

Statistical analyses

The main analyses define CHD as either a self-reported history of myocardial infarction or angina at the baseline assessment (for individuals aged <80 years at recruitment) or death before age 80 with CHD (ICD-10 codes I20-I25) listed anywhere on the death certificate. For individuals aged <80 years at recruitment, logistic regression was then used to estimate the association between each PRS and the odds of CHD, first by categorizing each PRS into fifths [with the lowest fifth treated as the reference category and group-specific confidence intervals around each odds ratio (OR), including the reference group with OR of 1.0]²⁹ and second by treating each PRS as a continuous variable (per 1SD increase). For consistency with most of the PRS source studies, these regression models were adjusted initially just for age at recruitment, sex, and the first seven genetic principal components to account for populational stratification ('partial adjustment'). Subsequently, we additionally adjusted for educational attainment (university or high school, middle school, elementary/other, none), waist-to-hip ratio (WHR), systolic and diastolic blood pressure (SBP and DBP), smoking status (never, former, current), and diabetes at recruitment (previously diagnosed or HbA1c $\geq 6.5\%$, vs. not) ('full adjustment'). Specification of the fully adjusted model was based on prior knowledge of CHD epidemiology plus their availability within the study dataset. For covariates included in the fully adjusted model, the few participants with missing data had values imputed using the median for numeric variables and the most frequently selected category for categorical variables. The area under the receiver operating characteristic curve (or 'C-statistic') was estimated for each model.

Analyses subsequently assessed for potential effect modification by baseline age, sex, proportion of Indigenous American ancestry, as well as other factors included in the fully adjusted model, with tests of heterogeneity or

trend performed to assess whether the ORs in these subgroups varied significantly around the overall odds ratio. Sensitivity analyses included redefining the primary CHD outcome excluding deaths where CHD was not the underlying (i.e. the primary) cause, restricting to non-fatal self-reported myocardial infarction, angina or both at baseline, and restricting to the CHD death component (as listed anywhere on the death certificate and as listed as the underlying cause). Analyses of the primary CHD outcome were also repeated limited to participants who were unrelated to the 3rd degree [based on identity-by-descent (IBD) estimates¹⁵] to investigate the potential effect of familial relatedness. Finally, additional analyses extended the age range studied (participants and CHD events) up to age 90 years.

Analyses were done using R (version 4.3.2).

Results

Included participants

Of the 159 755 participants recruited, 22 364 (14%) were excluded. These comprised 18 924 (11.8%) without genetic data available or genetic data meeting QC thresholds, a further 2221 (1.4%) with uncertain mortality linkage, and a further 1219 (0.7%) aged ≥ 90 years at recruitment. Of the remaining 137 391 participants, 133 207 were aged 35–79 years at recruitment and 4184 were aged 80–89 years.

Baseline characteristics

Of the 133 207 participants aged 35–79 years at recruitment, the mean age was 51 years (SD 12 years), 33% were men, and 57% were unrelated to the 3rd family degree. Average ancestry proportions were 67% Indigenous American, 28% European, 3% African, and 1% East Asian ([Table 1](#)). The highest level of educational attainment was university or high school for 16%, middle school for 25%, and elementary school for 47%. Women had a lower proportion of university or high school attainment than men. Overall, 32% were current smokers and 20% former smokers, with a much higher proportion of men than women being current or former smokers. Mean SBP and DBP were 127 mmHg (SD 17 mmHg) and 83 mmHg (SD 10 mmHg), respectively. Self-reported history of CHD (myocardial infarction or angina) was reported by 1901 (1.4%) participants and stroke was reported by 1414 (1.1%). A total of 24 796 participants (19%) had self-reported diabetes or an HbA1c concentration indicative of undiagnosed diabetes ($\geq 6.5\%$), while 8% reported having at least one other chronic disease. The baseline characteristics by study district are shown in [Supplementary material online, Table S2](#), while baseline characteristics of all 137 391 participants aged 35–89 years at recruitment are shown in [Supplementary material online, Table S3](#).

Fatal and non-fatal CHD cases

During a median (IQR) follow-up in survivors of 20.3 years (19.4 to 21.5 years), there were 24 596 deaths at ages 35–79 years including 3479 where CHD was listed on the death certificate (2927 as the underlying cause and 552 elsewhere on the certificate). Of these 3479 participants, 217 reported having CHD at recruitment and 3262 did not. Consequently, in total, there were 5163 participants who either had self-reported CHD prior to age 80 or died before age 80 with CHD listed on their death certificate (i.e. 3479 with CHD death before age 80 plus 1901 with pre-existing CHD at recruitment minus 217 with both). Extending the population to the 137 391 participants aged 35–89 at recruitment, 7155 had either self-reported CHD prior to age 90 or died before age 90 with CHD listed on their death certificate.

Table 1 Baseline characteristics of 133 207 participants aged 35–79 years

	Men n = 43 338 (33%)	Women n = 89 869 (67%)	All n = 133 207
Age, years	52.0 (12.1)	50.8 (11.7)	51.2 (11.8)
Resident of Coyoacán	17 958 (41%)	33 057 (37%)	51 015 (38%)
Unrelated participants	24 145 (56%)	51 730 (58%)	75 875 (57%)
Ancestry admixture percentage			
Indigenous American	66.5 (18.0)	67.0 (17.8)	66.8 (17.9)
African	3.4 (2.8)	3.5 (2.8)	3.4 (2.8)
East Asian	1.4 (2.0)	1.4 (1.7)	1.4 (1.8)
European	28.7 (16.3)	28.1 (16.1)	28.3 (16.2)
Highest attained educational level			
University/high school	10 340 (24%)	10 369 (12%)	20 709 (16%)
Middle school	11 468 (26%)	21 964 (24%)	33 432 (25%)
Elementary	17 902 (41%)	45 005 (50%)	62 907 (47%)
Other	3616 (8%)	12 475 (14%)	16 091 (12%)
Missing	12 (<0.5%)	56 (<0.5%)	68 (<0.5%)
Smoking status			
Never	8719 (20%)	55 872 (62%)	64 591 (48%)
Former	13 060 (30%)	13 116 (15%)	26 176 (20%)
Current	21 521 (50%)	20 809 (23%)	42 330 (32%)
Missing	38 (<0.5%)	72 (<0.5%)	110 (<0.5%)
Alcohol intake			
Never	2660 (6%)	23 443 (26%)	26 103 (20%)
Former	7598 (18%)	10 653 (12%)	18 251 (14%)
Current	33 062 (76%)	55 741 (62%)	88 803 (67%)
Missing	18 (<0.5%)	32 (<0.5%)	50 (<0.5%)
Physical measures			
SBP, mmHg	128.5 (15.8)	126.6 (16.9)	127.2 (16.6)
DBP, mmHg	84.4 (9.9)	82.5 (10.2)	83.1 (10.2)
BMI, kg/m ²	28.0 (4.3)	29.6 (5.3)	29.1 (5.1)
Waist-to-hip Ratio	0.95 (0.07)	0.88 (0.07)	0.90 (0.08)
Laboratory measurements			
HDL-C, mmol/L	0.93 (0.19)	1.03 (0.22)	1.00 (0.21)
LDL-C, mmol/L	2.39 (0.79)	2.49 (0.79)	2.46 (0.79)
Triglycerides, mmol/L	1.65 (0.68)	1.54 (0.64)	1.57 (0.66)
HbA1c, %	6.10 (1.72)	6.10 (1.71)	6.10 (1.71)
eGFR, mL/min/1.73m ^{2a}	101.0 (15.9)	102.0 (16.1)	101.7 (16.0)
Prior disease^b			
Coronary heart disease	848 (2%)	1053 (1%)	1901 (1%)
Stroke	485 (1%)	929 (1%)	1414 (1%)
Cancer	266 (1%)	1314 (1%)	1580 (1%)
Diabetes ^c	8250 (19%)	16 546 (18%)	24 796 (19%)
Other ^d	2276 (5%)	8837 (10%)	11 113 (8%)

Numbers are n (%) or mean (SD). SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index; HDL-C, high-density lipoprotein cholesterol; LDL-C, low density lipoprotein cholesterol; HbA1c, glycosylated haemoglobin A1c; eGFR, estimated glomerular filtration rate.

^aCalculated using the 2021 CKD-EPI equation, based on NMR-measured creatinine levels.

^bSelf-reported previous diagnoses unless otherwise stated.

^cSelf-reported previously-diagnosed diabetes or glycosylated haemoglobin $\geq 6.5\%$.

^dOther diseases include self-reported emphysema, chronic kidney disease, peptic ulcer, liver cirrhosis, and peripheral arterial disease.

Relationship of each PRS to the odds of CHD

The eight evaluated CHD PRSs included between 44 and 6 472 620 SNPs; [Supplementary material online, Figure S1](#) displays the pairwise

correlations between the eight PRSs. The three European PRSs that included thousands of SNPs^{7,16,17} were all strongly correlated with each other. Two of the multi-ancestry PRSs^{20,21} were also strongly correlated with each other and with the European PRSs. Baseline characteristics of those aged 35–79 years at recruitment by fifth of

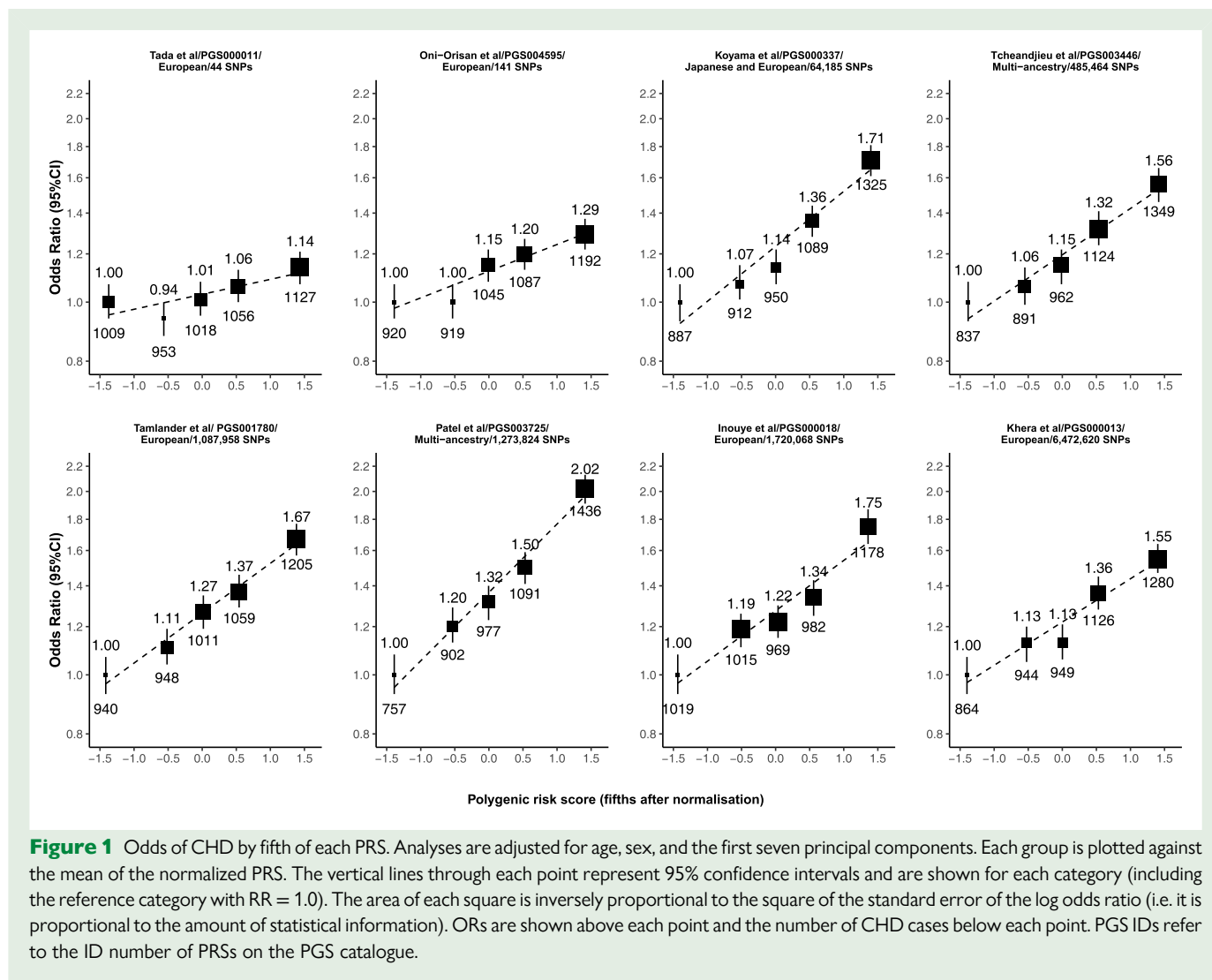


Figure 1 Odds of CHD by fifth of each PRS. Analyses are adjusted for age, sex, and the first seven principal components. Each group is plotted against the mean of the normalized PRS. The vertical lines through each point represent 95% confidence intervals and are shown for each category (including the reference category with $RR = 1.0$). The area of each square is inversely proportional to the square of the standard error of the log odds ratio (i.e. it is proportional to the amount of statistical information). ORs are shown above each point and the number of CHD cases below each point. PGS IDs refer to the ID number of PRSs on the PGS catalogue.

each of the eight PRSs are shown in [Supplementary material online, Tables S4–S11](#).

For all eight PRSs, the associations between the PRS and the odds of CHD were positive and approximately log-linear (*Figure 1*). Those with the highest fifth of genetic predisposition to CHD had 1.14–2.02 times the odds of CHD compared with those in the lowest (reference) fifth (*Figure 1*). The PRSs with fewer SNPs included tended to show weaker associations compared with the PRSs which included thousands of SNPs. When adjusted for age, sex, and the first seven genetic principal components, the PRS by Patel *et al.*²¹ showed the strongest overall association with CHD, with each 1 SD higher level of genetic predisposition associated with a 29% increase in the odds of CHD ($OR = 1.29$, 95% CI 1.25–1.33) (*Figure 2*). In contrast, the weakest association was seen for the PRS by Tada *et al.*,¹⁹ with each 1 SD higher level of genetic predisposition associated with only a 5% increase in the odds of CHD ($OR = 1.05$, 1.03–1.08). For all eight PRSs, the strengths of association were attenuated only marginally after further adjustment for conventional cardiovascular risk factors (*Figure 2*). The multi-ancestry sourced PRSs displayed relatively stronger associations with CHD risk compared with the European-sourced PRSs (ORs ranged from 1.19 to 1.29 vs. 1.05 to 1.22, respectively).

The C-statistic based on a logistic model including just age, sex, and the first seven genetic principal components was 0.716 (0.709–0.722). This increased to 0.745 (0.739–0.751) on further inclusion of waist-to-hip ratio, systolic and diastolic blood pressure, smoking status, level of education, and diabetes. In both cases, adding a PRS into the models increased the C-statistic only marginally. For example, for the Patel *et al.*²¹ PRS, the C-statistic in these models was increased to 0.724 (0.717–0.730) and 0.749 (0.743–0.755), respectively (*Figure 2*).

Differences in genetic predisposition to CHD risk by sex

Two of the evaluated PRSs with only around a hundred genome-wide significant SNPs included [Tada *et al.*¹⁹ (44 SNPs) and Oni-Orisan *et al.*¹⁸ (141 SNPs)] showed similar strengths of associations in men and women (*Figure 3*). However, for the other six PRSs evaluated, each of which included thousands or hundreds of thousands of SNPs (Koyama *et al.*,²⁰ Tcheandjieu *et al.*,²² Tamlander *et al.*,¹⁷ Patel *et al.*,²¹ Inouye *et al.*,¹⁶ and Khera *et al.*,⁷), genetic predisposition to subsequent CHD risk was significantly stronger for men compared with women. The (multi-ancestry sourced) Tcheandjieu *et al.*¹⁷ PRS exhibited the greatest heterogeneity by sex, with an OR per 1 SD higher genetic

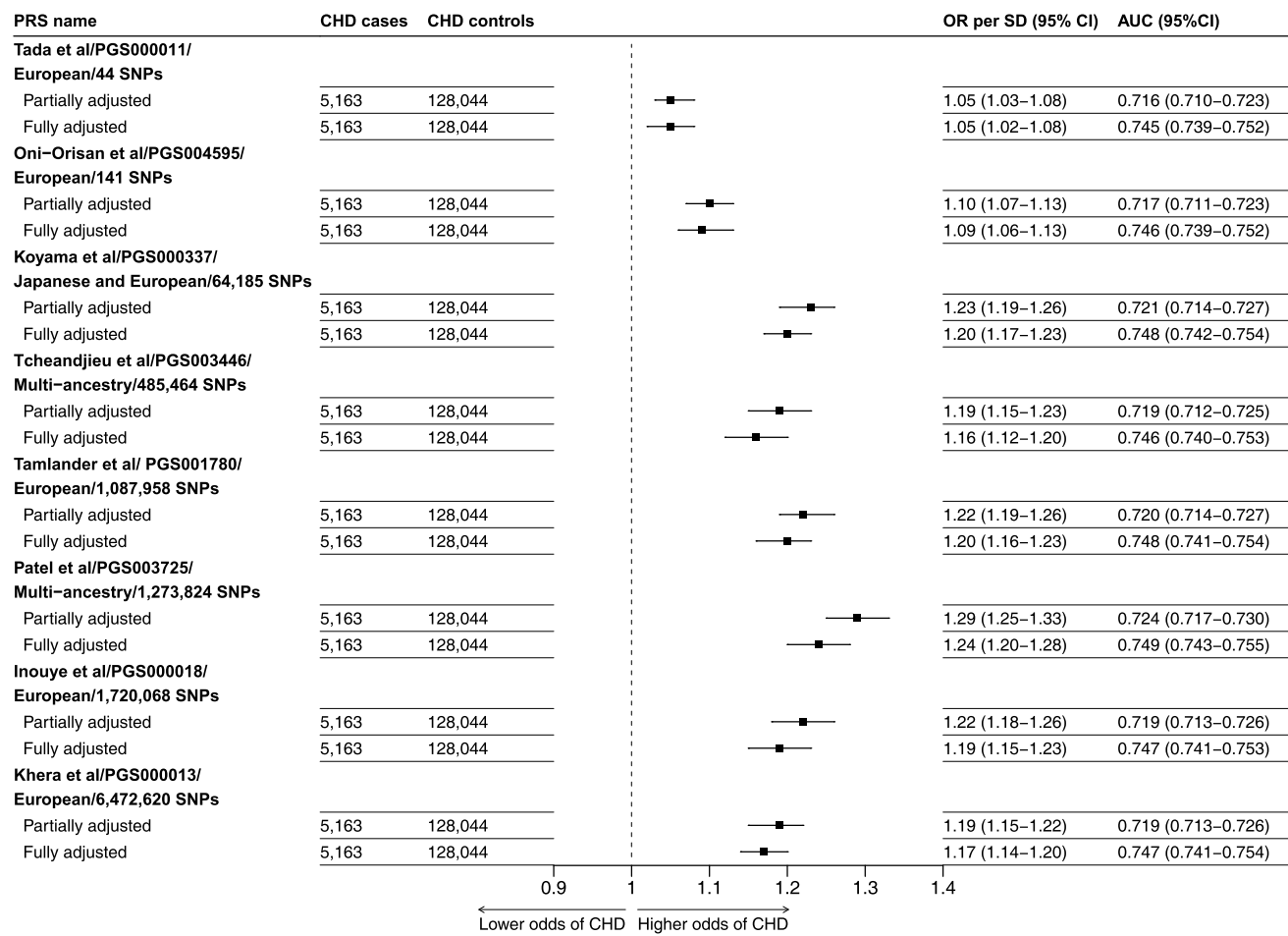


Figure 2 Odds of CHD per 1 SD increase in each PRS. In the partially adjusted model adjustment is for age, sex, and the first seven genetic principal components. The fully adjusted model is further adjusted for baseline waist-to-hip ratio, systolic and diastolic blood pressure, smoking status, level of education, and diabetes. The model C-statistic for a model including only age, sex, and the first seven genetic principal components was 0.716 (0.709–0.722). This increased to 0.745 (0.739–0.751) on further addition of waist-to-hip ratio, systolic and diastolic blood pressure, smoking status, level of education, and diabetes. Analyses are restricted to eligible participants with complete data on all covariates. PGS IDs refer to the ID number of PRSs on the PGS catalogue.

predisposition of 1.30 (95% CI 1.24–1.37) in men and 1.10 (95% CI 1.05–1.15) in women. In both men and women, the Patel *et al.*²¹ PRS displayed the strongest association with CHD risk, with an OR of 1.37 (1.32–1.43) in men and 1.23 (1.18–1.28) in women. Further adjustments for conventional vascular risk factors did not explain the sex differences observed.

Sensitivity analyses

Subgroup analyses of each PRS by age, education, waist-hip ratio, blood pressure, smoking status, history of diagnosed or undiagnosed diabetes, and indigenous ancestry proportion are shown in [Supplementary material online, Figures S2–S9](#). Overall, results for each PRS were broadly consistent by levels of these factors. However, for the strongest overall performing PRSs, associations appeared slightly stronger at younger ages, among those with a higher waist-to-hip ratio, among those with a higher level of education, and among those with lower blood pressure. Alternative definitions of CHD gave broadly similar results, with the weakest association generally seen for self-reported

angina (albeit with wide confidence intervals) (see [Supplementary material online, Figures S10–S17](#)). For the PRS with the strongest overall association with CHD (Patel *et al.*²¹), the interaction with sex was apparent for all definitions of CHD but these interactions were most apparent for CHD definitions based on self-reported CHD at recruitment (see [Supplementary material online, Figure S18](#)). Restriction to those unrelated to the 3rd degree or extension to those aged 35–89 years at recruitment (and CHD deaths before age 90) reinforced the main findings (see [Supplementary material online, Figures S19–S21](#)).

Discussion

In this large Mexican prospective study, we evaluated the relevance of genetic predisposition to CHD risk, using eight external PRSs derived from large European or multi-ancestry source populations. Overall, we found reasonable transferability and potential for predicting CHD risk in a Mexican population, with all of the evaluated PRSs showing

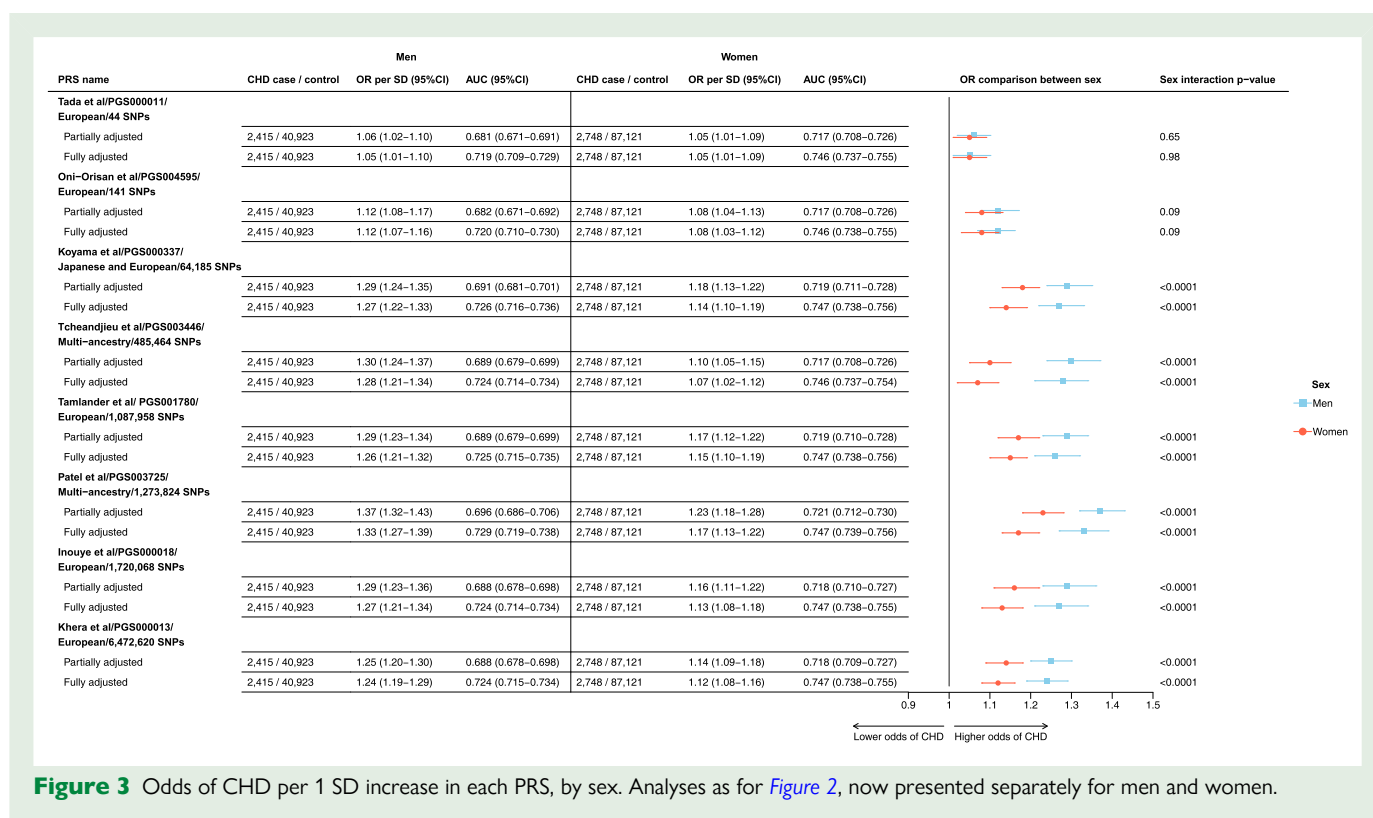


Figure 3 Odds of CHD per 1 SD increase in each PRS, by sex. Analyses as for [Figure 2](#), now presented separately for men and women.

clear positive log-linear associations with the odds of CHD at ages 35–79 years.

For six of the eight polygenic risk scores (specifically those including tens of thousands to millions of SNPs), the association with coronary heart disease was significantly stronger in men than in women. This may reflect the greater ability of larger PRSs to capture sex-specific genetic predisposition to CHD. These findings align with previous studies in European populations.^{16,21,30} For instance, among 317 509 unrelated individuals of European ancestry in the UK Biobank (UKB),³⁰ a 1 SD increase in genetic risk was associated with a 38% higher risk of incident coronary artery disease (CAD) in men, compared with a 25% increase in women. In that study, three trait-based subscores (for blood pressure, blood lipids, and body mass index) were examined, and only the blood pressure-mediated score showed a significant sex difference in its association with CAD. This difference was partly, but not entirely, driven by a novel locus at *21q22.11*, which was significantly associated with increased CAD in men but not in women. This locus mapped to genes expressed in arterial tissues and has previously been found to be associated with bone mineral density, waist-hip ratio, and pulse pressure.

In addition to sex-specific genetic factors, the weaker predictive performance of PRSs in women may also reflect inherent sex biases in the clinical definitions of CHD (such as differences in disease progression, presentation, and age at onset), which are dominated by studies involving men.³¹ To address these gaps, more sex-specific analyses (e.g. GWAS and PRS studies) are needed to better understand genetic risk for CHD in both sexes and to further investigate the biological pathways underlying observed sex differences.

With the exception of sex, the associations of the PRSs with CHD were broadly similar according to other participant characteristics and were comparable when using different definitions of CHD.

Associations were largely independent of established vascular risk factors, which is consistent with previous reports from European populations.^{32–34} Indeed, several recent developments that employed state-of-the-art methods for PRS construction^{8–10,35} have shown that when integrated with existing clinical risk-score tools for CHD, genetic information can enhance screening in addition to existing clinical tools, and inform precision medicine efforts for better stratification of CHD risk among otherwise ‘healthy’ individuals.^{5,32,36}

Polygenic risk scores derived from European populations have demonstrated strong performance within European cohorts (as shown in [Supplementary material online, Table S1](#)) achieving C-statistic ranging from 0.75 to 0.81. However, their performance was somewhat weaker when applied to the current Mexican population. Consistent with our findings, several other studies reported weaker yet positive associations in genetic predisposition to CHD risk when European GWAS-sourced PRSs were subsequently applied to Hispanic populations,^{22,37–39} as well as other ethnically diverse groups, including populations of both Middle Eastern and African ancestries.^{22,38–40} The performance of a PRS depends on several factors, including sample size of the input GWAS and the accurate estimation of genetic effects for causal variants. These genetic effects can be influenced by population-specific environmental factors, which vary significantly across ancestries. Ideally, PRS for Hispanic individuals would use effect sizes derived from large, population-specific cohorts to capture both shared and unique genetic risk factors. Unfortunately, Hispanics remain significantly underrepresented in GWAS, leaving a critical gap in the development of PRS. Leveraging diverse population sources for GWAS SNP inputs may enhance transferability across diverse populations worldwide, as the true causal variants for CHD risk are likely shared across populations.³⁴ Indeed, in our analyses, the multi-ancestry GWAS-sourced PRSs

from Patel et al.²¹ and Koyama et al.,²⁰ which were derived from some of the largest GWAS conducted to date outperformed most of the European-dominant sourced-PRSs. However, the multi-ancestry PRSs also have limitations as they do not fully account for population-specific environmental influences and may overlook risk variants that are unique to underrepresented populations. The findings underscore the need for large-scale GWAS in Hispanic populations to ensure their fair representation in genetic research and to enhance the accuracy and applicability of PRSs in diverse populations. There is also a specific need for further GWAS in women.

The major strength of the present analysis is that it includes a large number of people from a previously understudied population of Hispanic adults. The high-quality genetic data allowed CHD-specific PRSs to be reconstructed with careful consideration of genetic misclassification, genetic ancestry admixture, and relatedness. As for many prospective studies, previous diagnoses of angina and myocardial infarction were based on self-reports, so are subject to misclassification bias. A further limitation was the lack of data on non-fatal CHD incidence, although the large sample size and prolonged follow-up for mortality resulted in a reasonable number of CHD cases for analysis. The study population arises from just two districts of Mexico City, and so the participants are not representative of adults throughout Mexico (or even Mexico City). However, prospective studies of non-representative cohorts of individuals can provide reliable evidence about the associations of risk factors with disease that are widely generalizable.^{41,42} Either way, the findings from the current study clearly illustrate the potential for CHD polygenic risk scores to predict CHD risk in Mexican adults.

In summary, in this large Mexican population, existing polygenic risk scores were able to predict CHD risk independently of established vascular risk factors. However, current PRSs do not fully reflect the genetic architecture of CHD in Mexico. Ancestry-specific instruments that more closely represent genetic variation in Mexico may further enhance polygenic prediction of CHD risk in Mexican adults.

Supplementary material

Supplementary material is available at [European Journal of Preventive Cardiology](https://www.ajph.org/doi/10.1093/ajph/2024.14.1000000000000000).

Acknowledgements

We thank the study participants and staff.

Previous presentations: This work was presented as an oral presentation at the 2024 European Society of Cardiology meeting.

Rights retention statement

For the purposes of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

Author contributions

Tianshu Liu (Conceptualization, Formal analysis, Methodology, Visualization, Writing—original draft, Writing—review and editing), Jaime Berumen (Conceptualization, Writing—review and editing), Jason Torres (Conceptualization, Methodology, Supervision, Writing—review and editing), Jesus Alegre-Díaz (Conceptualization, Funding acquisition,

Writing—review and editing), Paulina Baca (Writing—review and editing), Carlos González-Carballo (Writing—review and editing), Fernando Rivas (Writing—review and editing), Diego Aguilar-Ramirez (Writing—review and editing), Fiona Bragg (Writing—review and editing), Michael Hill (Resources, Writing—review and editing), Eirini Trichia (Writing—review and editing), Rory Collins (Conceptualization, Funding acquisition, Writing—review and editing), Pablo Kuri-Morales (Conceptualization, Funding acquisition, Writing—review and editing), Jonathan R. Emberson (Conceptualization, Funding acquisition, Methodology, Supervision, Writing—review and editing), Roberto Tapia-Conyer (Conceptualization, Funding acquisition, Writing—review and editing), Louisa Gnatiuc Friedrichs (Conceptualization, Methodology, Supervision, Writing—review and editing), Raul Ramirez-Reyes (Data curation, Writing—review and editing), Will Herrington (Data curation, Writing—review and editing), Alejandra Vergara (Data curation, Writing—review and editing), and Rachel Wade (Data curation, Writing—review and editing).

Funding

This work was supported by the Mexican Health Ministry; the National Council of Science and Technology for Mexico; Wellcome Trust [058299/Z/99]; Cancer Research UK; Kidney Research UK [MR/R007764/1]; the British Heart Foundation [RE/13/1/30181]; and the UK Medical Research Council [MC_UU_00017/2, MR/Z504543/1]. Genotyping was funded through an academic partnership between the National Autonomous University of Mexico, the University of Oxford, Regeneron Pharmaceuticals, and AstraZeneca UK. The computational aspects of this research were supported by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z and the NIHR Oxford BRC. The funding sources had no role in the design, conduct or analysis of the study or the decision to submit the manuscript for publication.

Conflict of interest: J.R.E. and R.C. report grants to the University of Oxford from AstraZeneca UK and Regeneron Pharmaceuticals. R.C. reports having a patent for a statin-related myopathy genetic test licenced to the University of Oxford from Boston Heart Diagnostics (R.C. has waived any personal reward with any share in royalty and other payments waived in favour of the Nuffield Department of Population Health, University of Oxford) and being deputy chair of not-for-profit clinical trial company PROTAS, Chief Executive of UK Biobank, and Chair of the steering committee of the ORION-4 clinical trial of inclisiran. All other authors declare no competing interests.

Data availability

Data from the Mexico City Prospective Study are available to bona fide researchers. The study's Data and Sample Sharing policy is available in English and Spanish on the study website (<https://www.ctsu.ox.ac.uk/research/mcps>). Available study data can be examined in detail through the study's Data Showcase (<https://datashare.ndph.ox.ac.uk/mexico/>). MCPS ancestry-specific allele frequencies are available in a public browser (<https://rgc-mcps.regeneron.com/>).

References

- Ferrari AJ, Santomauro DF, Aali A, Abate YH, Abbafati C, Abbastabar H, et al. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the global burden of disease study 2021. *The Lancet* 2024;**403**:2133–2161.
- McPherson R, Tybjaerg-Hansen A. Genetics of coronary artery disease. *Circ Res* 2016;**118**:564–578.
- Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, De Faire U. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J Intern Med* 2002;**252**:247–254.
- Cooper RS, Kaufman JS, Ward R. Race and genomics. *N Engl J Med* 2003;**348**:1166–1170.

5. Riveros-Mckay F, Weale ME, Moore R, Selzam S, Krapohl E, Sivley RM, et al. Integrated polygenic tool substantially enhances coronary artery disease prediction. *Circ Genom Precis Med* 2021;**14**:e003304.
6. King A, Wu L, Deng HW, Shen H, Wu C. Polygenic risk score improves the accuracy of a clinical risk score for coronary artery disease. *BMC Med* 2022;**20**:385.
7. Khara AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* 2018;**50**:1219–1224.
8. Privé F, Arbel J, Vilhjálmsson BJ. LDpred2: better, faster, stronger. *Bioinformatics* 2021;**36**:5424–5431.
9. Mak TSH, Porsch RM, Choi SW, Zhou X, Sham PC. Polygenic scores via penalized regression on summary statistics. *Genet Epidemiol* 2017;**41**:469–480.
10. Ge T, Chen CY, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* 2019;**10**:1776.
11. Patel AP, Fahed AC. Pragmatic approach to applying polygenic risk scores to diverse populations. *Curr Protoc* 2023;**3**:e911.
12. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS catalog: knowledgebase and deposition resource. *Nucleic Acids Res* 2022;**51**:D977–D985.
13. Richard Doll Consortium, www.richarddollconsortium.org (Hosted by Oxford Population Health) using World Health Organization mortality and United Nations Population Division data. [Internet]. Richard Doll Consortium. [Accessed Jun 2024]. Available from: <https://www.richarddollconsortium.org/projects/mortality-trends>.
14. Tapia-Conyer R, Kuri-Morales P, Alegre-Díaz J, Whitlock G, Emberson J, Clark S, et al. Cohort profile: the Mexico city prospective study. *Int J Epidemiol* 2006;**35**:243–249.
15. Ziyatdinov A, Torres J, Alegre-Díaz J, Backman J, Mbatchou J, Turner M, et al. Genotyping, sequencing and analysis of 140,000 adults from Mexico city. *Nature* 2023;**622**:784–793.
16. Inouye M, Abraham G, Nelson CP, Wood AM, Sweeting MJ, Dudbridge F, et al. Genomic risk prediction of coronary artery disease in 480,000 adults: implications for primary prevention. *J Am Coll Cardiol* 2018;**72**:1883–1893.
17. Tamlander M, Mars N, Pirinen M, FinnGen, Widen E, Ripatti S. Integration of questionnaire-based risk factors improves polygenic risk scores for human coronary heart disease and type 2 diabetes. *Commun Biol* 2022;**5**:158.
18. Oni-Orisan A, Haldar T, Cayabyab MAS, Ranatunga DK, Hoffmann TJ, Iribarren C, et al. Polygenic risk score and statin relative risk reduction for primary prevention of myocardial infarction in a real-world population. *Clin Pharmacol Ther* 2022;**112**:1070–1078.
19. Tada H, Melander O, Louie JZ, Catanese JJ, Rowland CM, Devlin JJ, et al. Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. *Eur Heart J* 2016;**37**:561–567.
20. Koyama S, Ito K, Terao C, Akiyama M, Horikoshi M, Momozawa Y, et al. Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. *Nat Genet* 2020;**52**:1169–1177.
21. Patel AP, Wang M, Ruan Y, Koyama S, Clarke SL, Yang X, et al. A multi-ancestry polygenic risk score improves risk prediction for coronary artery disease. *Nat Med* 2023;**29**:1793–1803.
22. Tcheandjieu C, Zhu X, Hilliard AT, Clarke SL, Napolioni V, Ma S, et al. Large-scale genome-wide association study of coronary artery disease in genetically diverse populations. *Nat Med* 2022;**28**:1679–1692.
23. Lambert SA, Wingfield B, Gibson JT, Gil L, Ramachandran S, Yvon F, et al. The Polygenic Score Catalog: new functionality and tools to enable FAIR research. medRxiv 24307783, <http://doi.org/10.1101/2024.05.29.24307783>, 31 May 2024, preprint: not peer reviewed.
24. Lambert SA, Gil L, Jupp S, Ritchie SC, Xu Y, Buniello A, et al. The polygenic score catalog as an open database for reproducibility and systematic evaluation. *Nat Genet* 2021;**53**:420–425.
25. Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. *Nat Biotechnol* 2017;**35**:316–319.
26. 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**:s13742-015.
27. Mikkelsen L, Phillips DE, AbouZahr C, Setel PW, de Savigny D, Lozano R, et al. A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. *Lancet* 2015;**386**:1395–1406.
28. Alegre-Díaz J, Herrington W, López-Cervantes M, Gnatiuc L, Ramirez R, Hill M, et al. Diabetes and cause-specific mortality in Mexico city. *N Engl J Med* 2016;**375**:1961–1971.
29. Plummer M. Improved estimates of floating absolute risk. *Stat Med* 2004;**23**:93–104.
30. Huang Y, Hui Q, Gwinn M, Hu YJ, Quyyumi AA, Vaccarino V, et al. Sexual differences in genetic predisposition of coronary artery disease. *Circ Genom Precis Med* 2021;**14**:e003147.
31. Byars SG, Inouye M. Genome-wide association studies and risk scores for coronary artery disease: sex biases. *Adv Exp Med Biol* 2018;**1065**:627–642.
32. Li L, Pang S, Starnecker F, Mueller-Myhsok B, Schunkert H. Integration of a polygenic score into guideline-recommended prediction of cardiovascular disease. *Eur Heart J* 2024;**45**:1843–1852.
33. Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores. *Hum Mol Genet* 2019;**28**:R133–R142.
34. Xiang R, Kelemen M, Xu Y, Harris LW, Parkinson H, Inouye M, et al. Recent advances in polygenic scores: translation, equitability, methods and FAIR tools. *Genome Med* 2024;**16**:33.
35. Choi SW, O'Reilly FP. PRSice-2: polygenic risk score software for biobank-scale data. *GigaScience* 2019;**8**:giz082.
36. de La Harpe R, Thorball CW, Redin C, Fournier S, Müller O, Strambo D, et al. Combining European and U.S. Risk prediction models with polygenic risk scores to refine cardiovascular prevention: the CoLausPsyCoLaus study. *Eur J Prev Cardiol* 2023;**30**:561–571.
37. Vassy JL, Posner DC, Ho YL, Gagnon DR, Galloway A, Tanukonda V, et al. Cardiovascular disease risk assessment using traditional risk factors and polygenic risk scores in the million veteran program. *JAMA Cardiol* 2023;**8**:564–574.
38. Dikilitas O, Schaid DJ, Kosel ML, Carroll RJ, Chute CG, Denny JA, et al. Predictive utility of polygenic risk scores for coronary heart disease in three major racial and ethnic groups. *Am J Hum Genet* 2020;**106**:707–716.
39. Aragam KG, Dobbyn A, Judy R, Chaffin M, Chaudhary K, Hindy G, et al. Limitations of contemporary guidelines for managing patients at high genetic risk of coronary artery disease. *J Am Coll Cardiol* 2020;**75**:2769–2780.
40. Saad M, El-Menyar A, Kunji K, Ullah E, Al Suwaidi J, Kullo IJ. Validation of polygenic risk scores for coronary heart disease in a middle eastern cohort using whole genome sequencing. *Circulation Genom Precis Med* 2022;**15**:e003712.
41. Chen Z, Emberson J, Collins R. Strategic need for large prospective studies in different populations. *JAMA* 2020;**323**:309–310.
42. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. *Int J Epidemiol* 2013;**42**:1012–1014.