

Minimal improvement in coronary artery disease risk prediction in the Chinese population using polygenic risk scores: Evidence from the China Kadoorie Biobank

Songchun Yang, Dong Sun, Zhijia Sun, *et al.*

Supplementary Material

| | |
|---|-----------|
| Appendix. Members of the China Kadoorie Biobank collaborative group..... | 1 |
| Supplementary Methods | 2 |
| Definition of the training sets | 2 |
| Selection of previous PRSs..... | 2 |
| Clumping & thresholding (C+T) method | 2 |
| LDpred method..... | 3 |
| Supplementary Figures | 4 |
| Figure S1. Flowchart for the study population | 4 |
| Figure S2. High-quality variants in CKB | 5 |
| Figure S3. Quality control processes of GWAS summary statistics files | 6 |
| Figure S4. ORs of PRSs from previous studies in training sets | 7 |
| Figure S5. ORs of PRSs from the "C+T" method in training sets | 8 |
| Figure S6. ORs of PRSs from the LDpred method in training sets | 9 |
| Figure S7. Correlation plot of different PRSs in training sets | 10 |
| Figure S8. Correlation plot of the optimal PRSs for hard CAD and soft CAD in the testing set | 11 |
| Figure S9. Adjusted HRs for CAD associated with the optimal PRS for soft CAD..... | 12 |
| Figure S10. Adjusted HRs for CAD associated with the optimal PRS for hard CAD, stratified by different baseline characteristics | 13 |
| Figure S11. Calibration plots before and after the addition of PRS | 14 |
| Supplementary Tables | 15 |
| Table S1. The detailed process of case-control matching..... | 15 |
| Table S2. The ascertainment methods and definitions of predictors | 16 |
| Table S3. Quality control processes of PRS files from previous studies | 17 |
| Table S4. Baseline characteristics of the training sets | 18 |
| Table S5. Baseline characteristics of the testing set | 20 |
| Table S6. Adjusted HRs for CAD associated with the optimal PRS for hard CAD in different models..... | 21 |
| Table S7. Reclassification based on the continuous NRI and relative IDI | 22 |
| Supplementary References | 23 |

Appendix. Members of the China Kadoorie Biobank collaborative group

International Steering Committee: Junshi Chen, Zhengming Chen (PI), Robert Clarke, Rory Collins, Yu Guo, Liming Li (PI), Jun Lv, Richard Peto, Robin Walters. **International Co-ordinating Centre, Oxford:** Daniel Avery, Ruth Boxall, Derrick Bennett, Yumei Chang, Yiping Chen, Zhengming Chen, Robert Clarke, Huaidong Du, Simon Gilbert, Alex Hacker, Mike Hill, Michael Holmes, Andri Iona, Christiana Kartsonaki, Rene Kerosi, Garry Lancaster, Sarah Lewington, Kuang Lin, John McDonnell, Iona Millwood, Qunhua Nie, Jayakrishnan Radhakrishnan, Paul Ryder, Sam Sansome, Dan Schmidt, Paul Sherliker, Rajani Sohoni, Becky Stevens, Iain Turnbull, Robin Walters, Jenny Wang, Lin Wang, Neil Wright, Ling Yang, Xiaoming Yang. **National Co-ordinating Centre, Beijing:** Yu Guo, Xiao Han, Can Hou, Jun Lv, Pei Pei, Chao Liu, Canqing Yu, Qingmei Xia. **10 Regional Co-ordinating Centres:** **Qingdao CDC:** Zengchang Pang, Ruqin Gao, Shanpeng Li, Shaojie Wang, Yongmei Liu, Ranran Du, Yajing Zang, Liang Cheng, Xiaocao Tian, Hua Zhang, Yaoming Zhai, Feng Ning, Xiaohui Sun, Feifei Li. **Licang CDC:** Silu Lv, Junzheng Wang, Wei Hou. **Heilongjiang Provincial CDC:** Mingyuan Zou, Ge Jiang, Xue Zhou. **Nangang CDC:** Liqui Yang, Hui He, Bo Yu, Yanjie Li, Qinai Xu, Quan Kang, Ziyang Guo. **Hainan Provincial CDC:** Dan Wang, Ximin Hu, Jinyan Chen, Yan Fu, Zhenwang Fu, Xiaohuan Wang. **Meilan CDC:** Min Weng, Zhendong Guo, Shukuan Wu, Yilei Li, Huimei Li, Zhifang Fu. **Jiangsu Provincial CDC:** Ming Wu, Yonglin Zhou, Jinyi Zhou, Ran Tao, Jie Yang, Jian Su. **Suzhou CDC:** Fang Liu, Jun Zhang, Yihe Hu, Yan Lu, Liangcai Ma, Aiyu Tang, Shuo Zhang, Jianrong Jin, Jingchao Liu. **Guangxi Provincial CDC:** Zhenzhu Tang, Naying Chen, Ying Huang. **Liuzhou CDC:** Mingqiang Li, Jinhui Meng, Rong Pan, Qilian Jiang, Jian Lan, Yun Liu, Liuping Wei, Liyuan Zhou, Ningyu Chen Ping Wang, Fanwen Meng, Yulu Qin, Sisi Wang. **Sichuan Provincial CDC:** Xianping Wu, Ningmei Zhang, Xiaofang Chen, Weiwei Zhou. **Pengzhou CDC:** Guojin Luo, Jianguo Li, Xiaofang Chen, Xunfu Zhong, Jiaqiu Liu, Qiang Sun. **Gansu Provincial CDC:** Pengfei Ge, Xiaolan Ren, Caixia Dong. **Maiji CDC:** Hui Zhang, Enke Mao, Xiaoping Wang, Tao Wang, Xi zhang. **Henan Provincial CDC:** Ding Zhang, Gang Zhou, Shixian Feng, Liang Chang, Lei Fan. **Huixian CDC:** Yulian Gao, Tianyou He, Huarong Sun, Pan He, Chen Hu, Xukui Zhang, Huifang Wu. **Zhejiang Provincial CDC:** Min Yu, Ruying Hu, Hao Wang. **Tongxiang CDC:** Yijian Qian, Chunmei Wang, Kaixu Xie, Lingli Chen, Yidan Zhang, Dongxia Pan, Qijun Gu. **Hunan Provincial CDC:** Yuelong Huang, Biyun Chen, Li Yin, Huilin Liu, Zhongxi Fu, Qiaohua Xu. **Liuyang CDC:** Xin Xu, Hao Zhang, Huajun Long, Xianzhi Li, Libo Zhang, Zhe Qiu

Supplementary Methods

Definition of the training sets

Among the potential training set ($n=28,490$), every hard CAD event (including nonfatal I21-I23 and fatal I20-I25) observed during follow-up was recorded. Only the first event was considered in the present analysis. All hard CAD events were sorted in order of age at onset from largest to smallest ("potential case group"). The potential control group consisted of participants who had neither CAD at baseline nor incident CAD event (including any fatal or nonfatal I20-I25) during follow-up. In turn, each case was 1:1 matched with control for study area, sex, and year of birth. A control participant should not be a close relative of the corresponding case. The KING software ^[1] was used to calculate the kinship matrix among the study population. The kinship coefficient $\phi > 0.125$ was considered as a close relative. The censored age of the control participant(s) should be larger than the age of the case. When multiple potential controls met the above criteria, only the participant with the youngest censored age was retained as control. Each participant could only be selected as a control once. If no control was identified, we expanded the year of birth selection by ± 1 year, ± 2 year, and ± 3 year. If the above procedure still failed to match a case with appropriate control, the case was excluded from the subsequent analysis.

Finally, 3513 (98.0%) of 3585 hard CAD cases were successfully matched with controls ("training set for hard CAD"). Following a similar procedure, 7142 (89.4%) of 7993 soft CAD cases (including nonfatal and fatal I20-I25) were successfully matched with controls ("training set for soft CAD") (**Figure S1, Table S1**).

Selection of previous PRSs

We searched the PGS Catalog ^[2] to obtain CAD PRS directly from previous studies (Date of searching: 2022-04-30). The detailed search strategy was: coronary artery disease OR myocardial infarction OR angina

The inclusion criteria of PRS in the current study were as follows:

- Newly developed.
- The PRS should integrate the information of multiple genetic variants across the whole genome and calculate individual genetic risk by weighted sum.
- The target trait of PRS should be CAD (including myocardial infarction and angina).
- The primary purpose of the original study was to examine the strength of association between PRS and CAD or to evaluate the effect of PRS on improving a traditional CAD risk prediction model.

The exclusion criteria of PRS were as follows:

- The base data of PRS did not include GWAS of CAD. For example, the PRS developed using blood pressure-related genetic associations.
- Variants in PRS were selected directly based on genome-wide significant associations.
- The training set of PRS was a population with a certain disease (such as individuals with diabetes, familial hypercholesterolemia, breast cancer, etc.).
- The information used to construct a PRS (i.e., chromosome, position, effect allele, weight, etc.) was not publicly available from the PGS Catalog website or the Supplementary files of the original study.

Following the above search strategy and inclusion and exclusion criteria, ten previously reported CAD PRSs were identified. Standard quality control for genetic variants was conducted before subsequent analysis (**Table S3**).

Clumping & thresholding (C+T) method

This approach involves taking the estimated single nucleotide polymorphism (SNP) effects from the largest available GWAS as the SNP weights. In the current study, a grid search strategy was used to construct multiple sets of PRS: we applied the r^2 threshold as 0 (=no pruning), 0.2, 0.4, 0.6, and 0.8, the P-value threshold from 5×10^{-8} to 1 (40 values in total, see **Figure S5** for details). For each r^2 threshold, we used PLINK (v1.90) to prune variants separately for the 22 autosomes

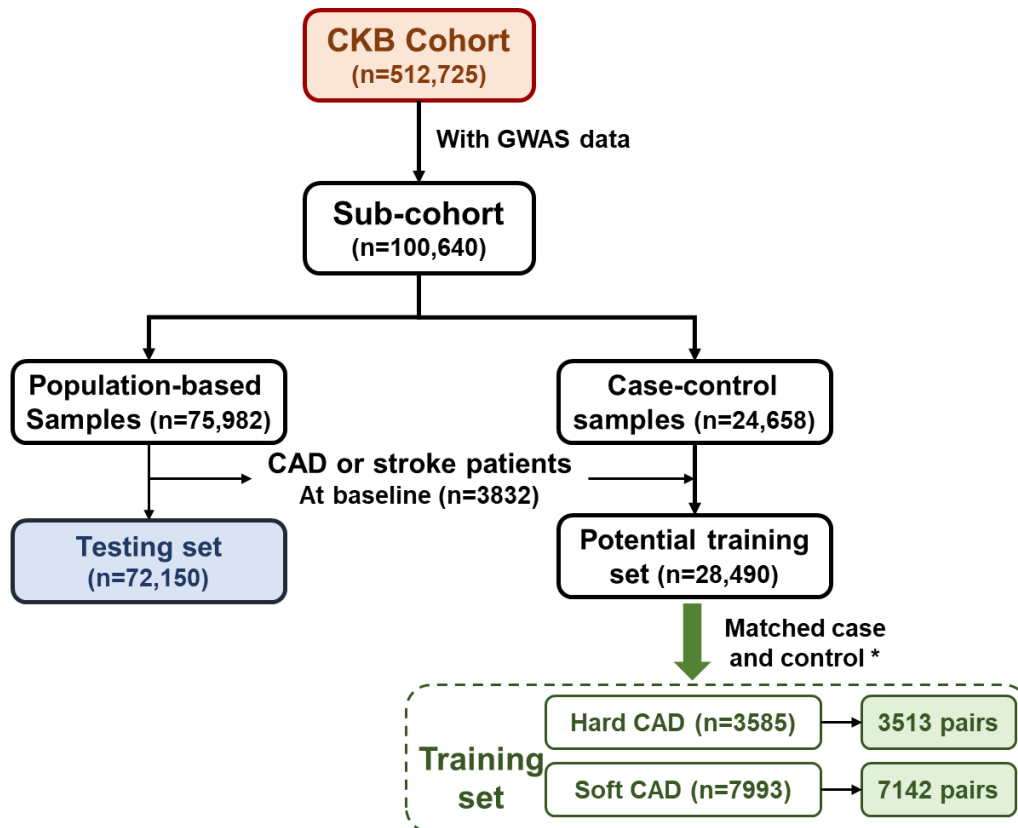
(--clump-kb 250). The threshold on P-value was not applied during linkage disequilibrium (LD) pruning (--clump-p1 1). The reference panel used for LD pruning was 1595 unrelated participants from CKB. We then applied different thresholds on the P-value for associations from the original GWAS. The PRS was computed by a weighted sum of the SNP dosages. After the above process, a GWAS summary statistics file could produce $5 \times 40 = 200$ PRSs with different r^2 thresholds and different P-value thresholds.

LDpred method

This Bayesian approach calculates a posterior mean effect for each variant based on a prior and subsequent shrinkage based on the extent to which this variant is correlated with similarly associated variants in the reference population [3]. Three steps are involved to develop PRS by LDpred (v1.0.10): (1) coordination of SNPs; (2) calculation of SNP posterior effects; (3) calculation of PRS. The variants were restricted to HapMap3 SNPs in the current analysis. Two parameters were required to run LDpred. The first parameter is the LD radius, i.e., the number of SNPs that we adjust for on each side of a given SNP. We used $M/3000$, the default value recommended by the software, where M is the total number of SNPs used in the analysis. This corresponds to a 2 Mb LD window on average in the genome. The second parameter is the fraction p of non-zero effects in the prior. A range of p values recommended by the software were used: 1, 0.3, 0.1, 0.03, 0.01, 0.003, and 0.001. In addition, the LD reference panel was required to compute the correlations between genetic variants. Previous analyses have suggested that this LD reference panel mimic the primary ancestral background of the original GWAS, rather than the target population [3, 4]. Therefore, East Asians ($n=504$) and Europeans ($n=503$) in 1000 Genomes Project Phase 3 were used as LD reference panels, respectively for Biobank of Japan (BBJ) and "UKB-CARDIoGRAMplusC4D meta-analysis (UCM)".

Supplementary Figures

Figure S1. Flowchart for the study population

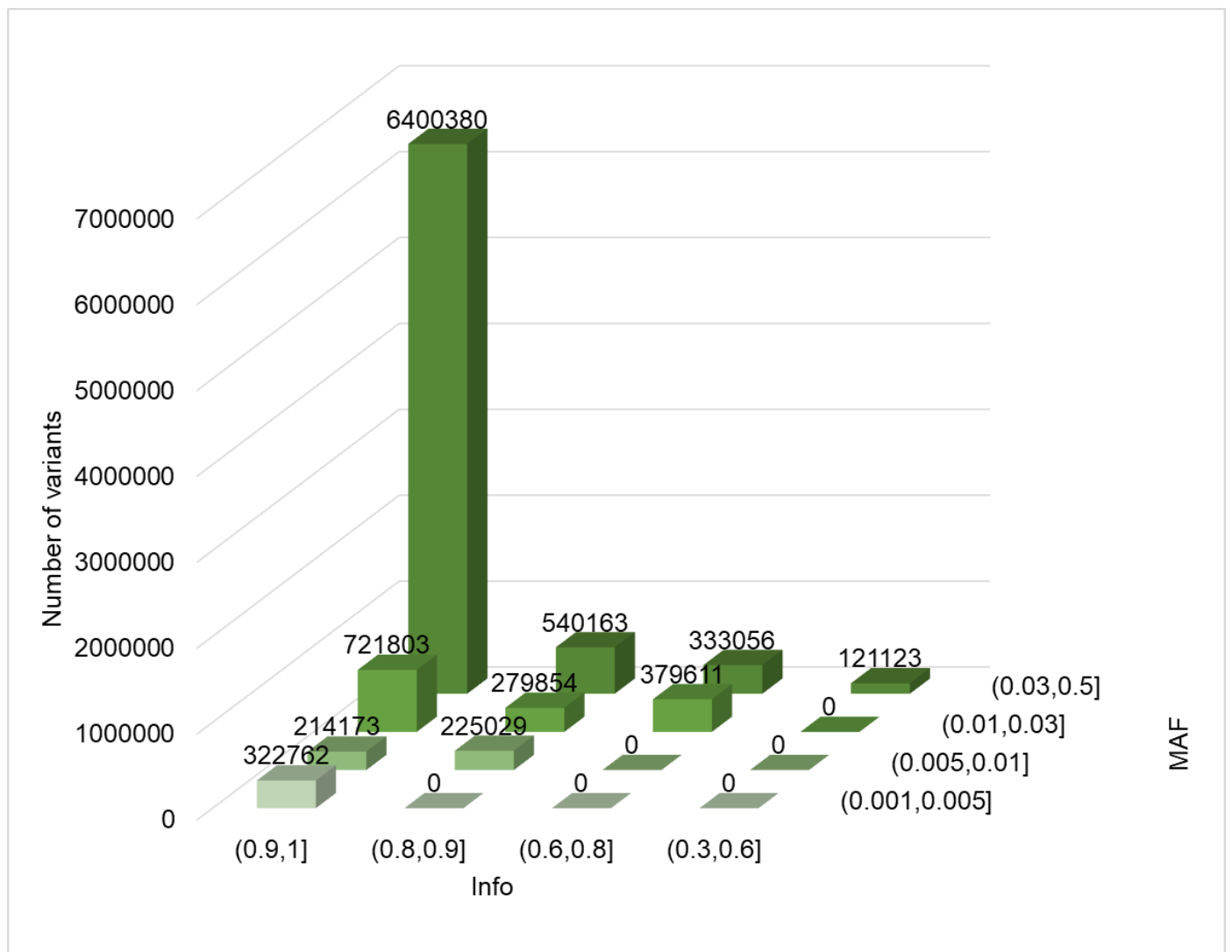


Abbreviations: CAD, coronary artery disease; CKB, China Kadoorie Biobank.

* Please refer to *Supplementary Methods* for detailed procedures of case-control matching.

Hard CAD events included nonfatal myocardial infarction (I21–I23) and fatal CAD (I20–I25); soft CAD events included all fatal or nonfatal CAD (I20–I25).

Figure S2. High-quality variants in CKB

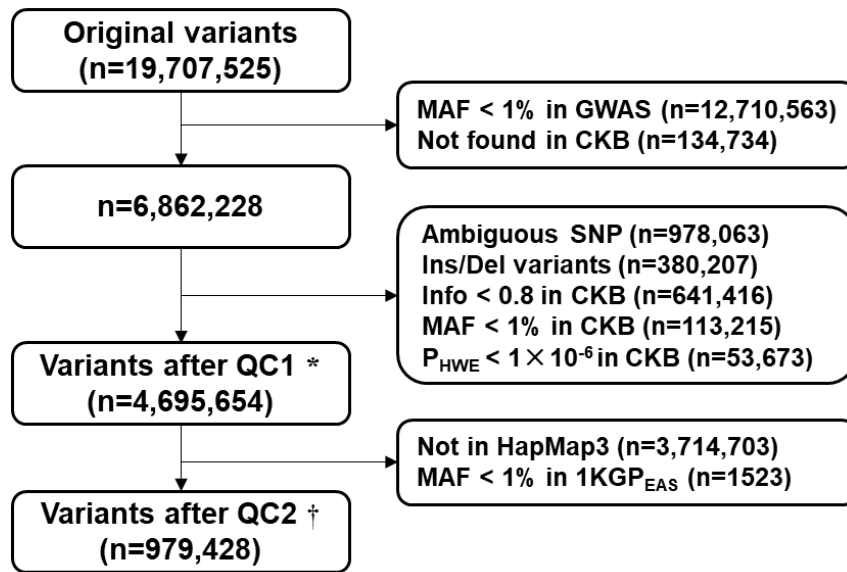


Abbreviations: Info, imputation quality score; MAF, minor allele frequency.

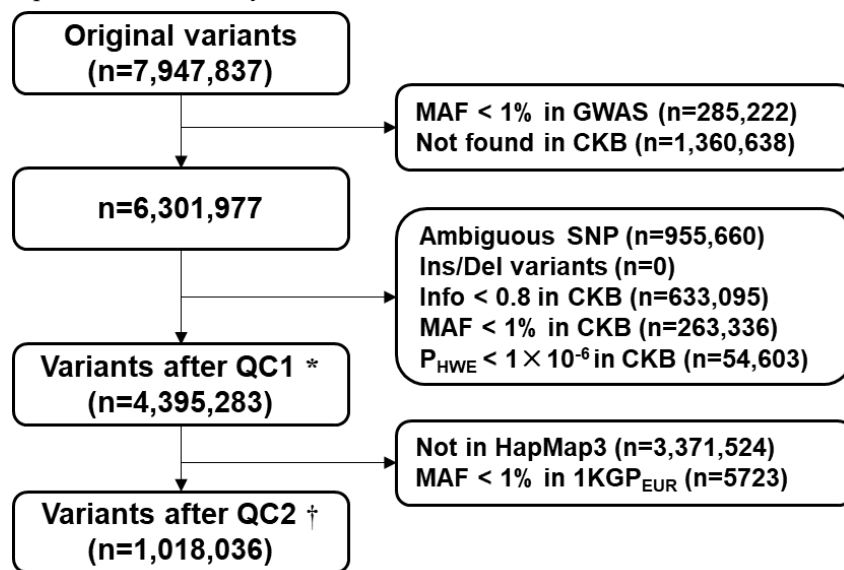
High-quality variants were defined as: (1) $\text{Info} > 0.3$ & $\text{MAF} > 0.03$; or (2) $\text{Info} > 0.6$ & $\text{MAF} > 0.01$; or (3) $\text{Info} > 0.8$ & $\text{MAF} > 0.005$; or (4) $\text{Info} > 0.9$ & $\text{MAF} > 0.001$.

Figure S3. Quality control processes of GWAS summary statistics files

(A) Biobank of Japan



(B) UKB-CARDIoGRAMplusC4D meta-analysis

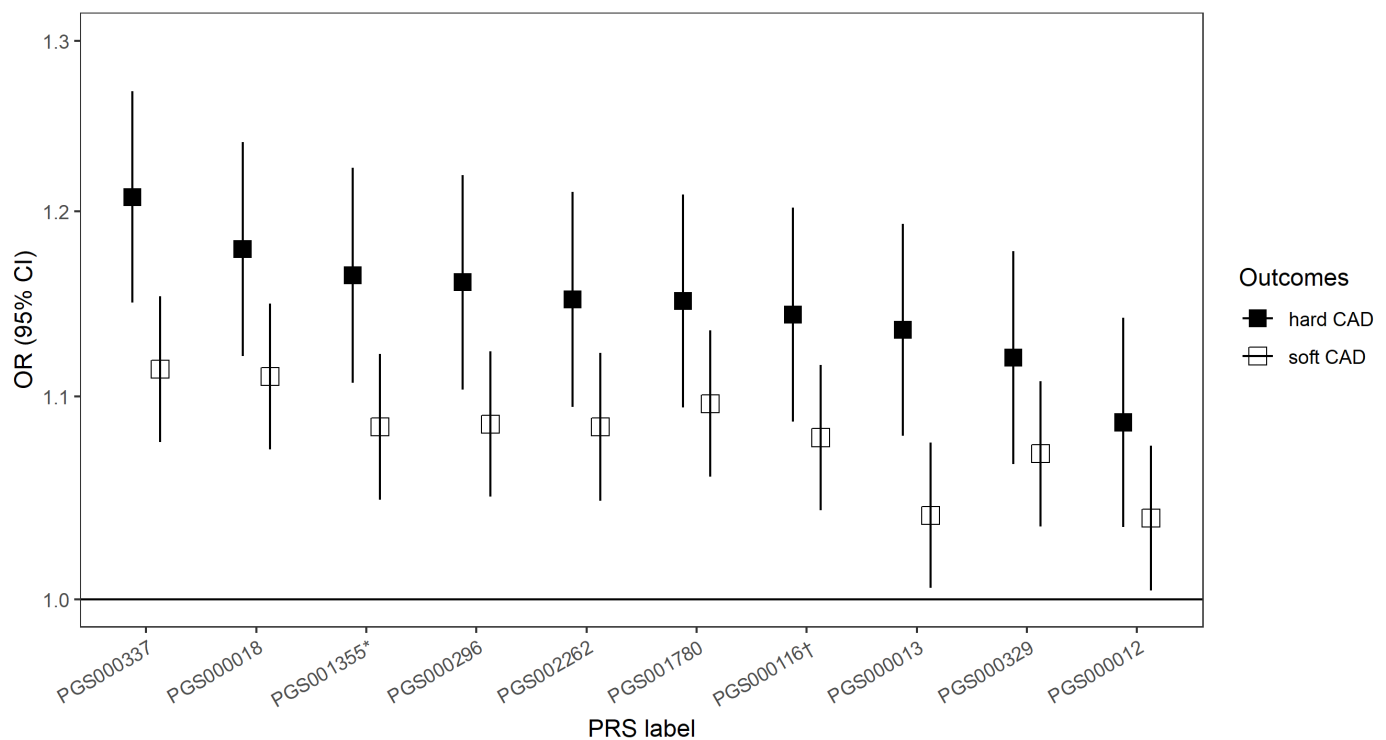


Abbreviations: 1KGP, 1000 Genomes Project (Phase 3); CAD, coronary artery disease; CKB, China Kadoorie Biobank; EAS, East Asian; EUR, European; HapMap3, the International HapMap Project Phase 3; HWE, Hardy-Weinberg Equilibrium; Info, imputation quality score; Ins/Del, insertion/deletion; MAF, minor allele frequency; PRS, polygenic risk score; QC, quality control; SNP, single nucleotide polymorphism.

* Variants after QC1 were used in the clumping & thresholding method.

† Variants after QC2 were used in the LDpred method.

Figure S4. ORs of PRSs from previous studies in training sets



Abbreviations: CAD, coronary artery disease; CI, confidence interval; OR, odds ratio; PRS, polygenic risk score.

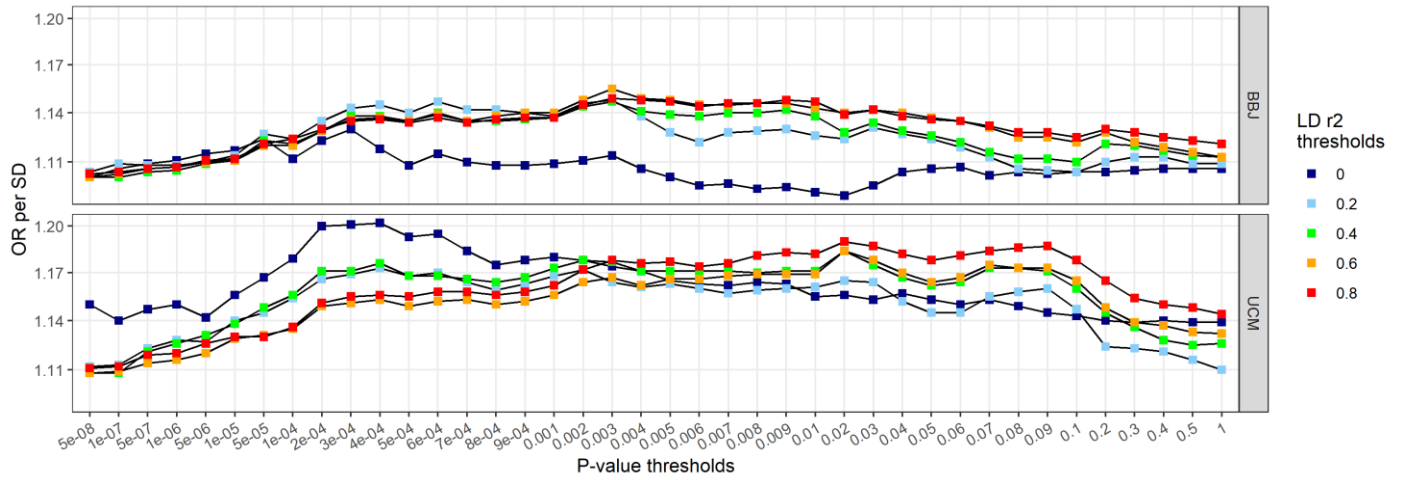
Standard quality control for genetic variants was conducted before subsequent analysis (**Table S3**). Hard CAD events included nonfatal myocardial infarction (I21–I23) and fatal CAD (I20–I25); soft CAD events included all fatal or nonfatal CAD (I20–I25).

* The effect allele in PGS001355 provided by the PGS Catalog was confused with the non-effect allele. We have confirmed this with the original author. This was corrected in the current analysis.

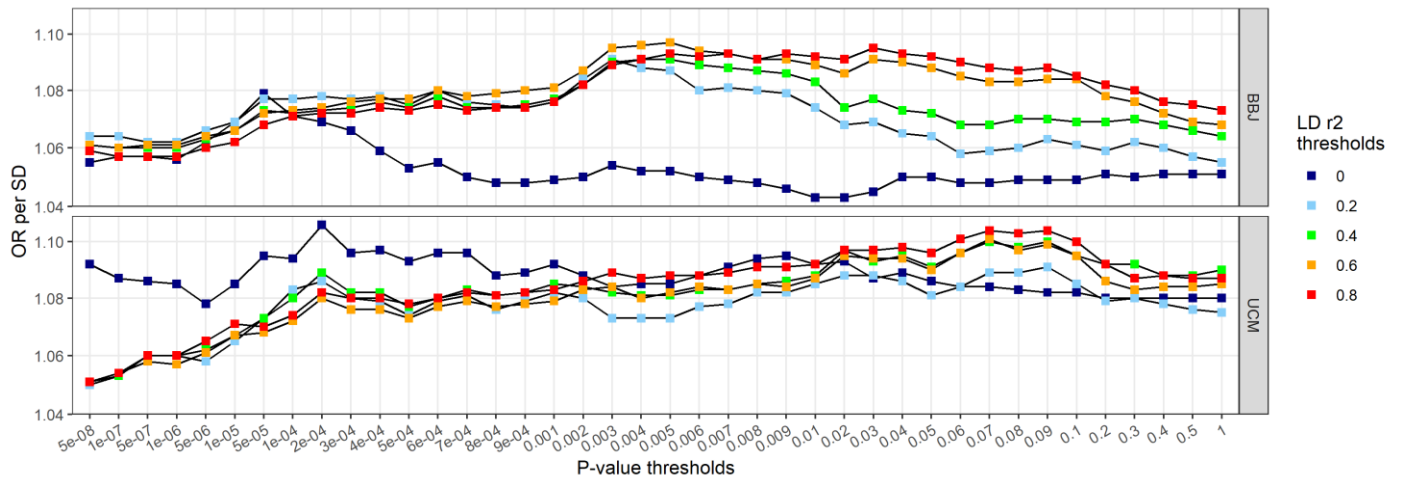
† The original PGS000116 provided by the PGS Catalog was negatively associated with CAD. The effect allele might be confused with the non-effect allele. This was also corrected in the current analysis.

Figure S5. ORs of PRSs from the "C+T" method in training sets

(A) Hard CAD



(B) Soft CAD

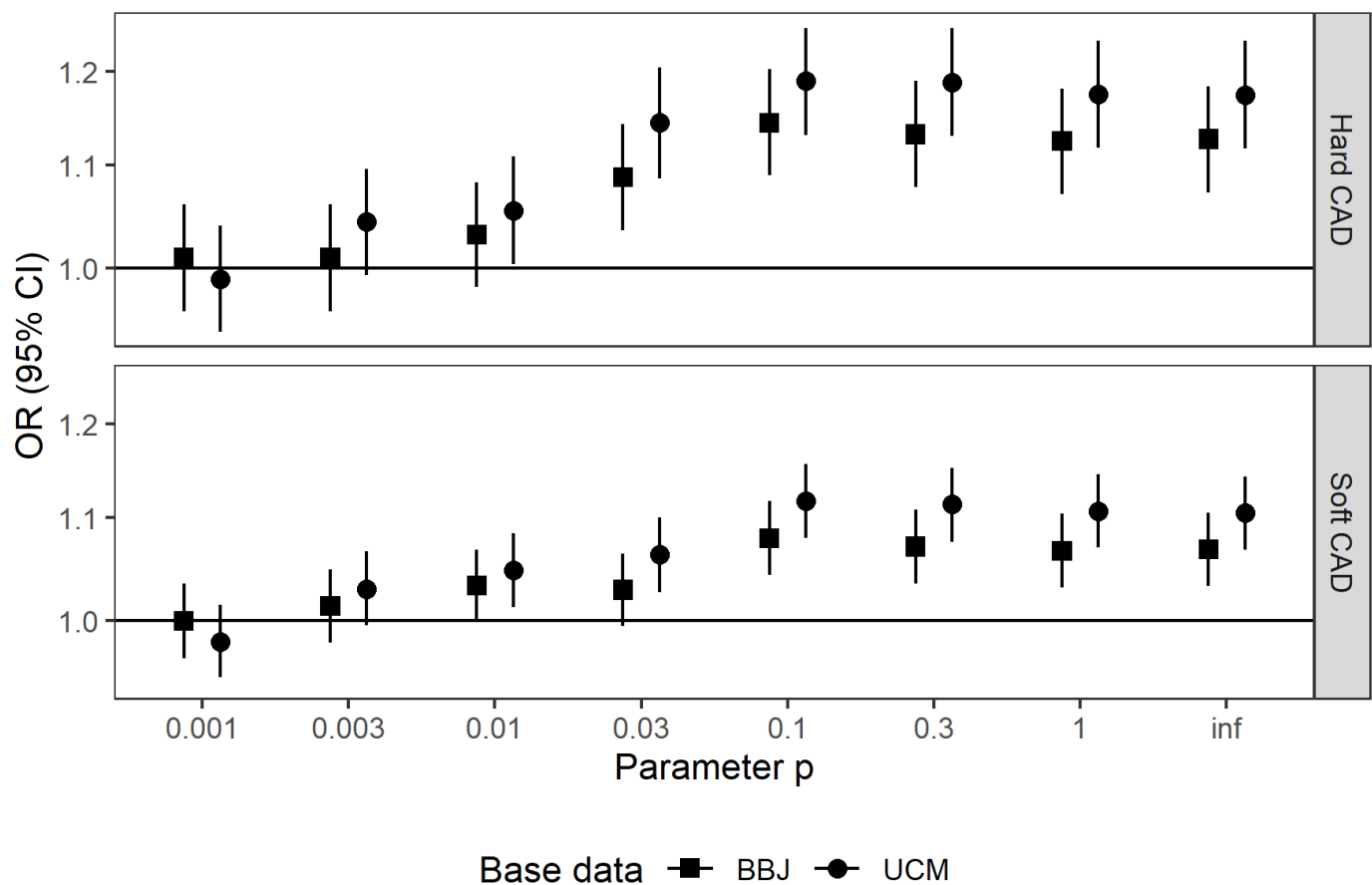


Abbreviations: BBJ, Biobank of Japan; CAD, coronary artery disease; LD, linkage disequilibrium; OR, odds ratio; SD, standard deviation; UCM, UKB-CARDIoGRAMplusC4D meta-analysis.

Hard CAD events included nonfatal myocardial infarction (I21–I23) and fatal CAD (I20–I25); soft CAD events included all fatal or nonfatal CAD (I20–I25).

(A) Hard CAD. Optimal PRS from BBJ: $r^2=0.6$, $P=0.003$, including 7436 variants, $OR_{SD} = 1.16$ (95% CI: 1.10–1.22). Optimal PRS from UCM: $r^2=0$, $P=0.0004$, including 1403 variants, $OR_{SD} = 1.20$ (95% CI: 1.43–1.26). **(B) Soft CAD.** Optimal PRS from BBJ: $r^2=0.6$, $P=0.005$, including 10,352 variants, $OR_{SD} = 1.10$ (95% CI: 1.06–1.14). Optimal PRS from UCM: $r^2=0$, $P=0.0002$, including 1093 variants, $OR_{SD} = 1.11$ (95% CI: 1.07–1.14).

Figure S6. ORs of PRSs from the LDpred method in training sets

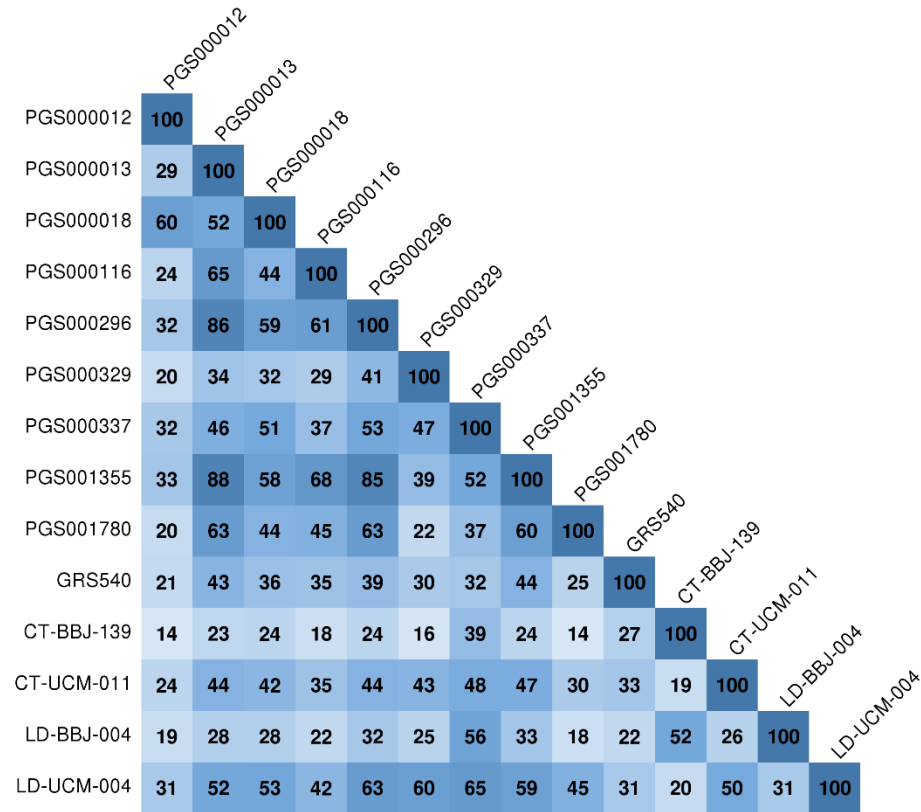


Abbreviations: BBJ, Biobank of Japan; CAD, coronary artery disease; CI, confidence interval; inf, infinite; LD, linkage disequilibrium; OR, odds ratio; SD, standard deviation; UCM, UKB-CARDIoGRAMplusC4D meta-analysis.

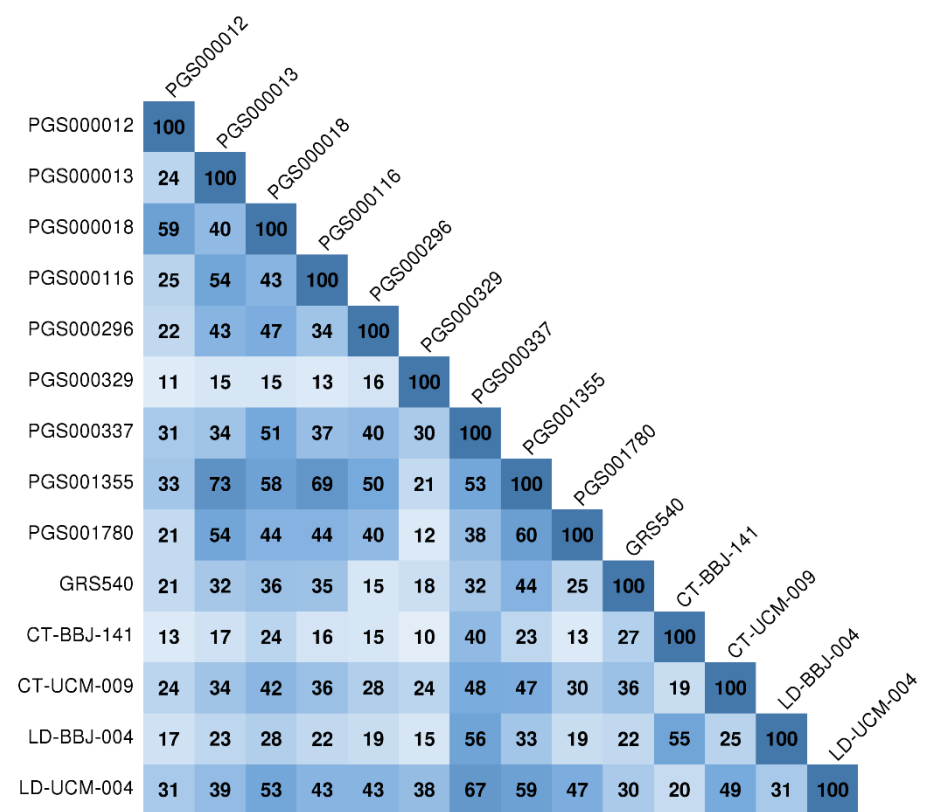
The parameter p is the fraction of non-zero effects in the prior. For both BBJ and UCM and both hard CAD and soft CAD, the optimal parameter p was 0.1 in the current analysis. Hard CAD events included nonfatal myocardial infarction (I21–I23) and fatal CAD (I20–I25); soft CAD events included all fatal or nonfatal CAD (I20–I25).

Figure S7. Correlation plot of different PRSs in training sets

(A) Hard CAD



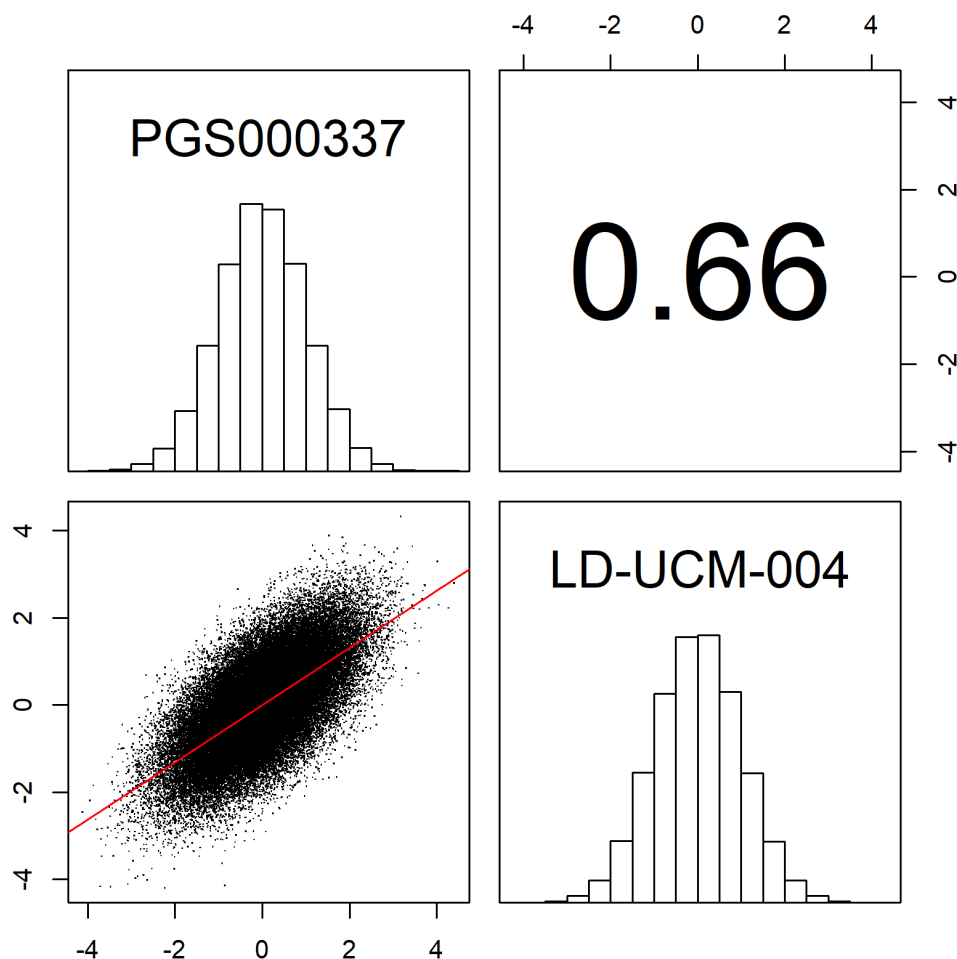
(B) Soft CAD



Abbreviations: BBJ, Biobank of Japan; CAD, coronary artery disease; UCM, UK Biobank (UKB) - CARDIoGRAMplusC4D meta-analysis.

Hard CAD events included nonfatal myocardial infarction (I21–I23) and fatal CAD (I20–I25); soft CAD events included all fatal or nonfatal CAD (I20–I25). The numbers in the grid of the matrix are Pearson correlation coefficients $\times 100$. In the column/row tag, "PGS####" is the index of the PRS file from the PGS Catalog. GRS540 represents the PRS developed by Lu X, et al (2022) (PGS002262). "CT" represents the "C+T" method and "LD" represents the LDpred method. Only the optimal PRSs of each base data for the two methods were displayed: CT-BBJ-139 (for hard CAD only, $r^2=0.6$, P value=0.003); CT-BBJ-141 (for soft CAD only, $r^2=0.6$, P value=0.005), CT-UCM-011 (for hard CAD only, $r^2=0$, P value=0.0004), CT-UCM-009 (for soft CAD only, $r^2=0$, P value=0.0002), LD-BBJ-004 ($p=0.1$), LD-UCM-004 ($p=0.1$).

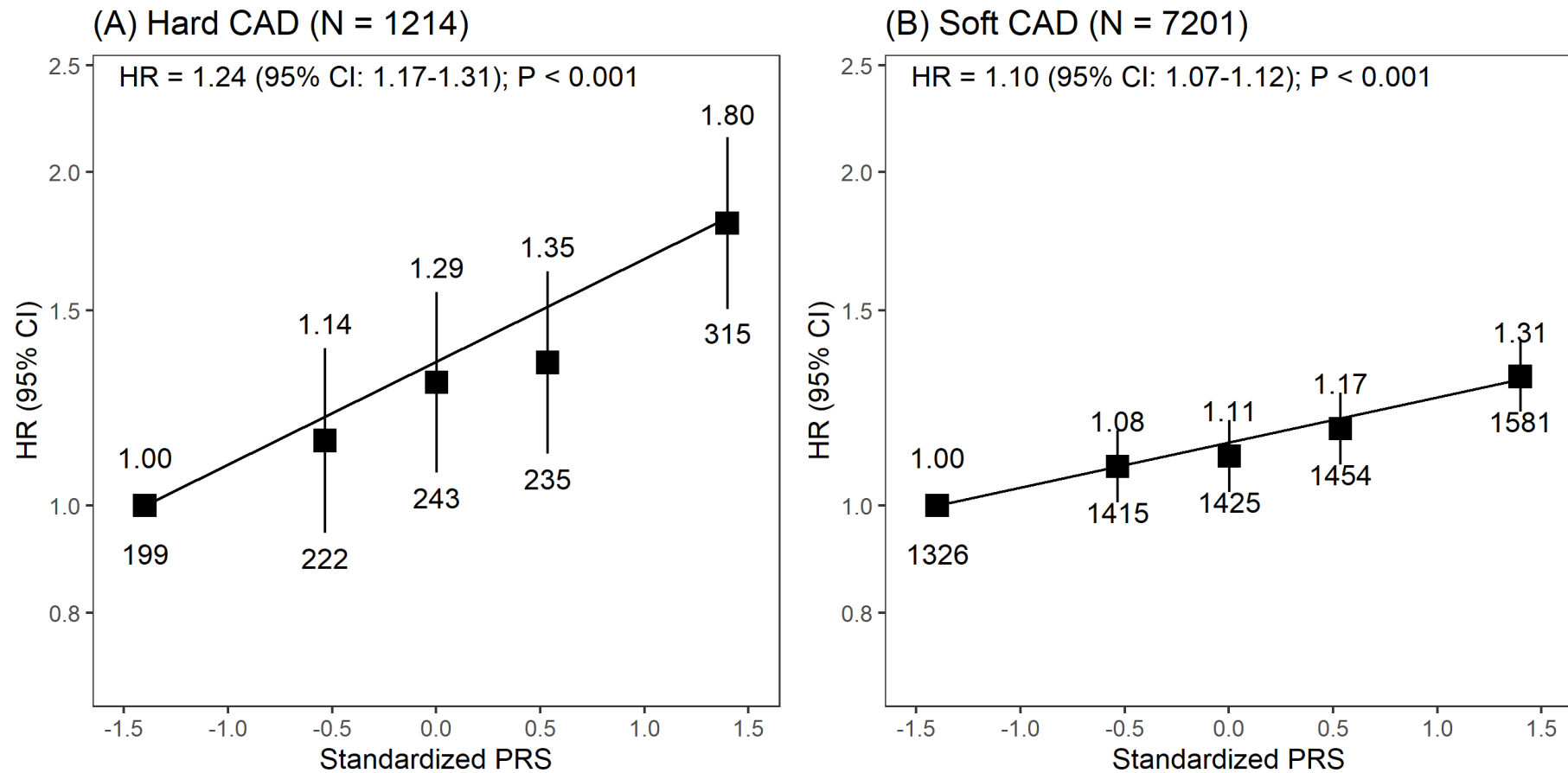
Figure S8. Correlation plot of the optimal PRSs for hard CAD and soft CAD in the testing set



Abbreviations: CAD, coronary artery disease.

Hard CAD events included nonfatal myocardial infarction (I21–I23) and fatal CAD (I20–I25); soft CAD events included all fatal or nonfatal CAD (I20–I25). PGS000337 and LD-UCM-004 are the optimal PRS for hard CAD and soft CAD, respectively (see **Table 1** for details). The two PRSs were standardized in the testing set ($n=72,150$) before plotting. The number in the upper-right square of the plot represents the Pearson correlation coefficient. The red line in the lower-left square represents the regression line.

Figure S9. Adjusted HRs for CAD associated with the optimal PRS for soft CAD

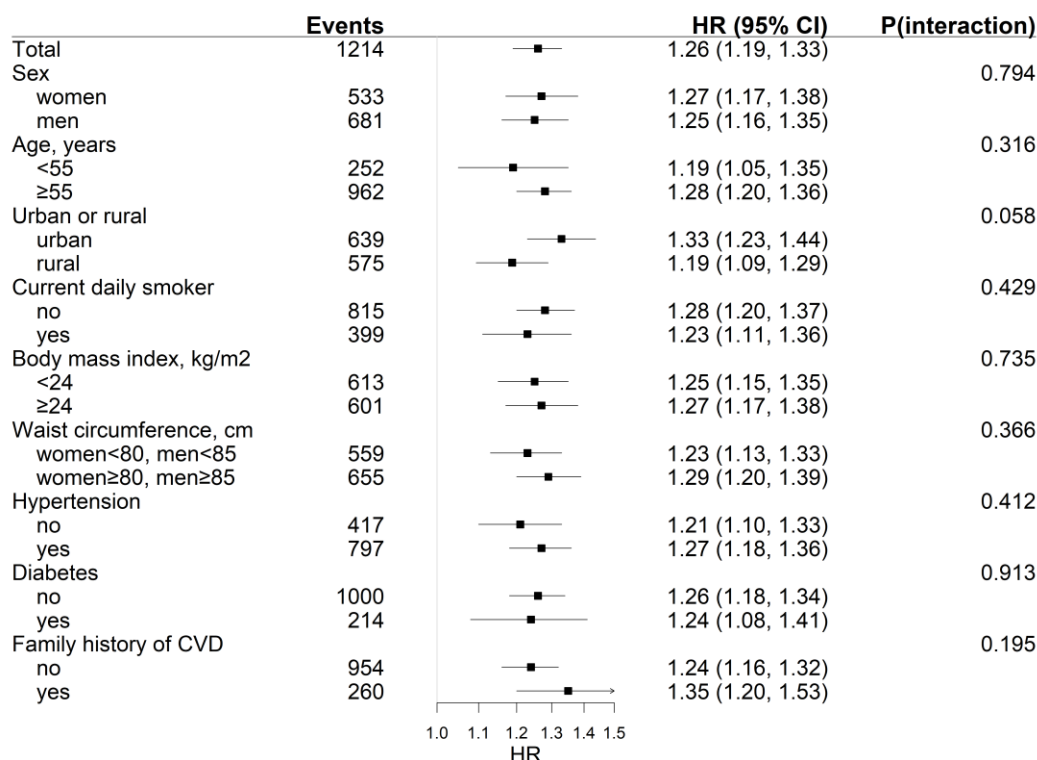


Abbreviations: CAD, coronary artery disease; CI, confidence interval; HR, hazard ratio; PRS, polygenic risk score.

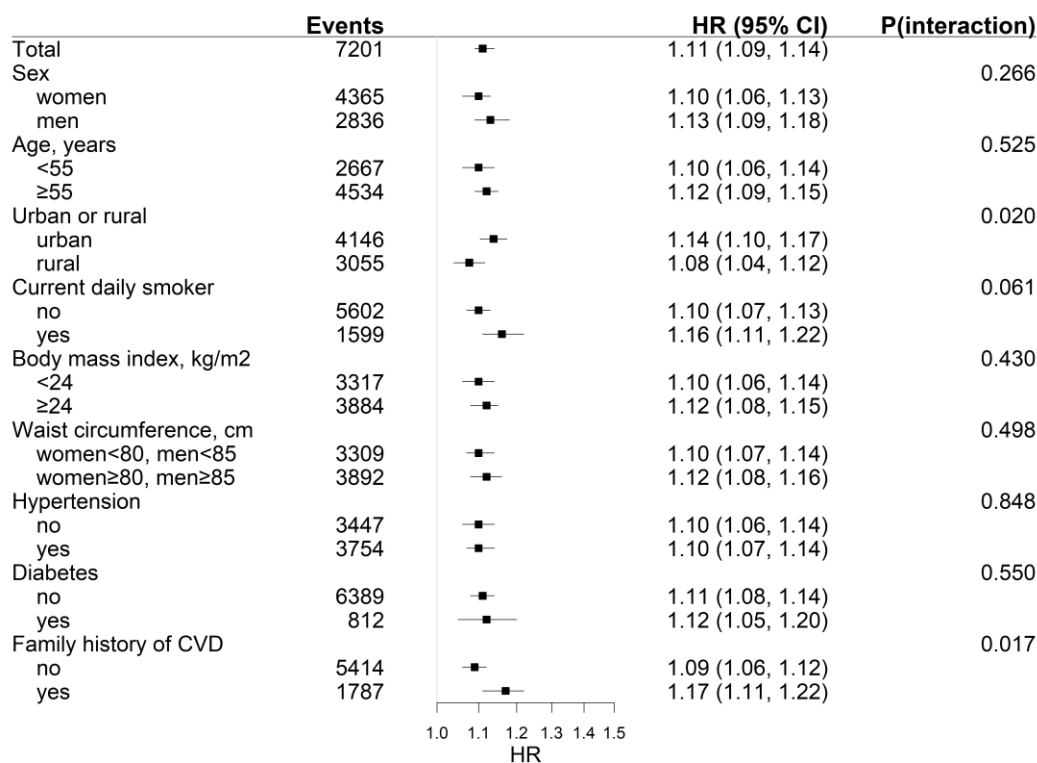
The PRS reported here is the optimal PRS for soft CAD (LD-UCM-004, see **Table 1** for details). Hard CAD events included nonfatal myocardial infarction (I21–I23) and fatal CAD (I20–I25); soft CAD events included all fatal or nonfatal CAD (I20–I25). The models were stratified by sex and ten study regions and adjusted simultaneously for the top 10 principal components of ancestry and array versions, with age as the time scale. In each subgraph, "N" in the title represents the number of events. HRs and P values on the upper left were derived from linear trend tests. The oblique line represents the trend line. The abscissa of each closed square represents the mean value of the standardized PRS in the corresponding quintile group. The number above the closed square represents the HR. The number below the closed square represents the number of events in this group. The vertical lines indicate 95% CIs.

Figure S10. Adjusted HRs for CAD associated with the optimal PRS for hard CAD, stratified by different baseline characteristics

(A) Hard CAD



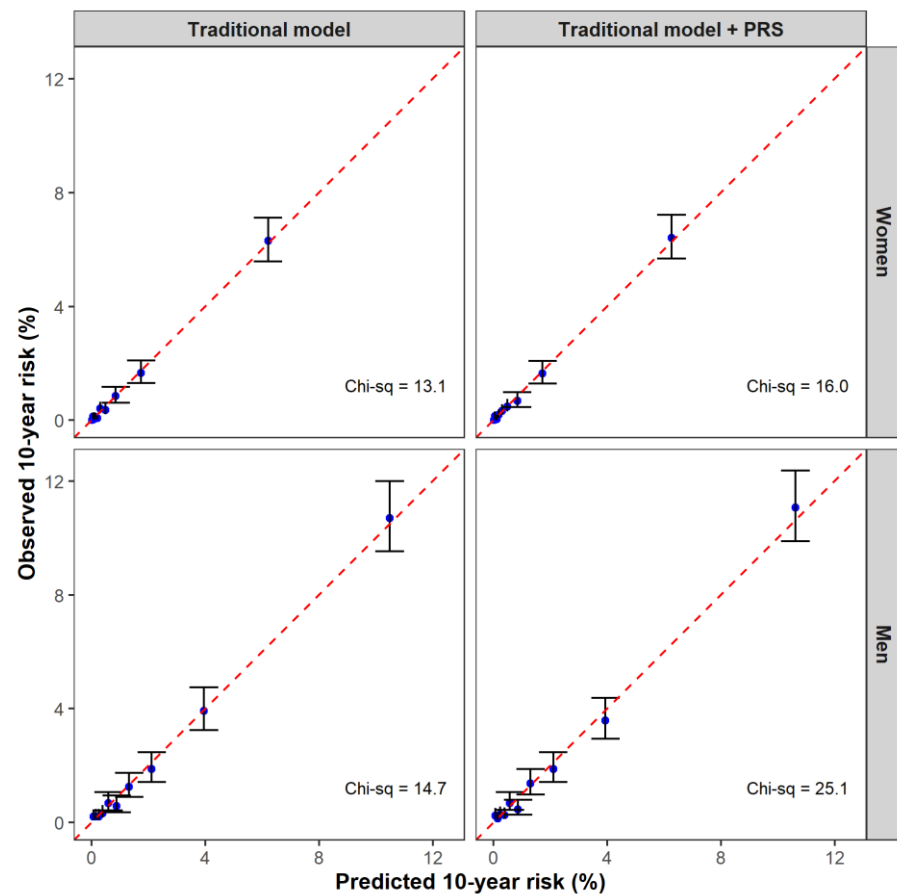
(B) Soft CAD



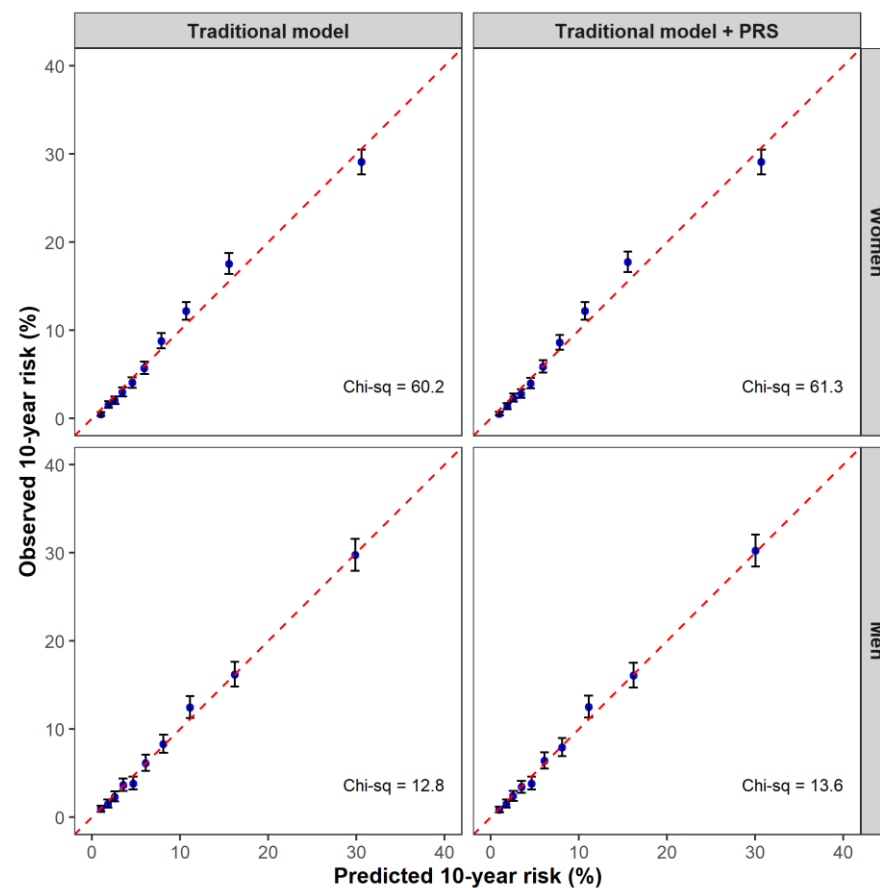
The PRS reported here is the optimal PRS for hard CAD (PGS000337, see **Table 1** for details). Hard CAD events included nonfatal myocardial infarction (I21–I23) and fatal CAD (I20–I25); soft CAD events included all fatal or nonfatal CAD (I20–I25). The models were stratified by sex and ten study regions and adjusted simultaneously for the top 10 principal components of ancestry and array versions, with age as the time scale. The tests for multiplicative interaction were performed using likelihood ratio tests by comparing models with and without cross-product terms.

Figure S11. Calibration plots before and after the addition of PRS

(A) Hard CAD



(B) Soft CAD



Abbreviations: CAD, coronary artery disease; PRS, polygenic risk score. The PRS reported here is the optimal PRS for hard CAD (PGS000337, see **Table 1** for details). Hard CAD events included nonfatal myocardial infarction (I21–I23) and fatal CAD (I20–I25); soft CAD events included all fatal or nonfatal CAD (I20–I25). Traditional models for CAD were defined as sex-specific Cox models stratified by ten study regions, with time-on-study as the time scale, including models for hard CAD and models for soft CAD. Predictors included in traditional models were the same as the "CKB-CVD models", including age, systolic and diastolic blood pressure, use of blood pressure-lowering treatment, current daily smoker, self-reported diabetes, and waist circumference (**Table S2**). Interactions between age and the other six predictors were also included. The vertical lines indicate 95% CIs. Chi-sq represents the Nam-D'Agostino test chi-square with nine degrees of freedom.

Supplementary Tables

Table S1. The detailed process of case-control matching

| Steps | Difference between the year of birth (control - case) | Number of matched cases | |
|---------|--|-------------------------|----------|
| | | Hard CAD | Soft CAD |
| 1 | 0 | 3426 | 7003 |
| 2 | -1 | 55 | 79 |
| 3 | 1 | 24 | 38 |
| 4 | -2 | 1 | 6 |
| 5 | 2 | 3 | 8 |
| 6 | -3 | 1 | 2 |
| 7 | 3 | 3 | 6 |
| Summing | | 3513 | 7142 |

Abbreviations: CAD, coronary artery disease.

Hard CAD events included nonfatal myocardial infarction (I21–I23) and fatal CAD (I20–I25); soft CAD events included all fatal or nonfatal CAD (I20–I25).

Table S2. The ascertainment methods and definitions of predictors

| Predictors | Methods of measurement and definitions |
|--|--|
| Sex | Questionnaire Women or men |
| Age | Questionnaire Unit: years |
| Systolic blood pressure (SBP) | Measured using a UA-779 digital sphygmomanometer. Measurements were made after a minimum of five minutes of sitting and repeated twice. If the two measurements for SBP were more than 10 mmHg apart, a third measurement was taken. Only the last two readings were recorded and averaged for further analyses. Unit: mmHg |
| Diastolic blood pressure (DBP) | Measured simultaneously with SBP using a UA-779 digital sphygmomanometer. Unit: mmHg |
| Self-reported history of diabetes | Questionnaire Q1. Has a doctor ever told you that you had had diabetes? (yes/no) |
| Current daily smoker | Q2. How often do you smoke tobacco now? (0=do not smoke now, 1=only occasionally, 2=on most days, 3=daily or almost every day) The current daily smoker was defined as Q2=3 (yes/no) |
| Use of blood pressure-lowering treatment | Questionnaire Q3. Has a doctor ever told you that you had had hypertension? (yes/no) Q4. (if yes for Q3) Are you still on treatment for hypertension? (yes/no) Q5. Did you take any drugs to lower blood pressure in the last 2 days? (yes/no) Use of blood pressure-lowering treatment was defined as Q4=yes or Q5=yes (yes/no) |
| Waist circumference (WC) | WC was measured midway between the iliac crest and the lower rib margin at the end of normal expiration using a plastic flexible tape to the nearest 0.1 cm. |

Table S3. Quality control processes of PRS files from previous studies

| Index | PRS ID | Original PRS name | Development methods | First Author, Publication year | Original | Matched with CKB | Non-ambiguous | Non Ins/Del | Info \geq 0.8 in CKB | MAF \geq 1% in CKB | P _{HWE} \geq 1 \times 10 ⁻⁶ in CKB |
|-------|-----------|-------------------|---------------------|--------------------------------|-----------|------------------|---------------|-------------|------------------------|----------------------|--|
| 1 | PGS000012 | GRS49K | C+T | Abraham G, 2016 | 49,310 | 45,647 | 42,247 | 42,247 | 38,626 | 37,332 | 36,874 |
| 2 | PGS000013 | GPS_CAD | LDpred | Khera AV, 2018 | 6,630,150 | 5,532,782 | 5,532,782 | 5,532,782 | 4,874,459 | 4,609,648 | 4,547,075 |
| 3 | PGS000018 | metaGRS_CAD | metaGRS | Inouye M, 2018 | 1,745,179 | 1,088,507 | 1,087,387 | 1,087,387 | 833,239 | 800,771 | 783,616 |
| 4 | PGS000116 | CAD_EJ2020 | lassosum | Elliott J, 2020 | 40,079 | 29,626 | 26,570 | 26,570 | 22,884 | 21,519 | 21,293 |
| 5 | PGS000296 | GPS_CAD_SA | LDpred | Wang M, 2020 | 6,630,150 | 5,532,782 | 5,532,782 | 5,532,782 | 4,874,459 | 4,609,648 | 4,547,075 |
| 6 | PGS000329 | PRS_CHD | LDpred | Mars N, 2020 | 6,423,165 | 5,129,895 | 5,129,895 | 5,129,895 | 4,522,325 | 4,283,908 | 4,231,851 |
| 7 | PGS000337 | MetaPRS_CAD | C+T | Koyama S, 2020 | 75,028 | 75,026 | 64,471 | 64,471 | 61,616 | 61,251 | 59,951 |
| 8 | PGS001355 | CAD_AnnoPred_PRS | AnnoPred | Ye Y, 2021 | 2,994,055 | 2,853,286 | 2,853,286 | 2,853,286 | 2,672,372 | 2,540,206 | 2,524,097 |
| 9 | PGS001780 | CHD_PRSCS | PRS-CS | Tamlander M, 2022 | 1,090,048 | 1,052,151 | 1,052,151 | 1,052,151 | 981,507 | 953,733 | 942,050 |
| 10 | PGS002262 | metaPRS_CAD | metaGRS | Lu X, 2022 | 510 * | 508 | 451 | 451 | 445 | 434 | 426 |

Abbreviations: C+T, clumping & thresholding; CKB, China Kadoorie Biobank; HWE, Hardy-Weinberg Equilibrium; Info, imputation quality score; Ins/Del, insertion/deletion; MAF, minor allele frequency; PRS, polygenic risk score.

* We excluded 30 genetic variants whose weights were zero.

Table S4. Baseline characteristics of the training sets

| | Case | Control | P-values |
|--|---------------------|---------------------|----------|
| The training set for hard CAD | | | |
| Total participants | 3513 | 3513 | — |
| Array 1 | 2271 (64.6) | 2923 (83.2) | <0.001 |
| Urban | 1400 (39.9) | 1400 (39.9) | — |
| Women | 1496 (42.6) | 1496 (42.6) | — |
| Age, years | 63.1 (54.6-69.7) | 63.2 (54.6-69.7) | 0.727 |
| Primary school and below | 2173 (61.9) | 2262 (64.4) | 0.028 |
| Daily smoker | | | |
| Women | 93 (6.2) | 74 (4.9) | 0.130 |
| Men | 1142 (56.6) | 1093 (54.2) | 0.121 |
| Body mass index, kg/m ² | 23.5 (21.0-26.1) | 23.0 (20.5-25.7) | <0.001 |
| Waist circumference, cm | | | |
| Women | 82.3 (74.9-89.4) | 80.1 (73.3-87.5) | <0.001 |
| Men | 83.0 (75.2-90.6) | 80.9 (73.2-89.3) | <0.001 |
| Systolic blood pressure, mmHg | 142.0 (127.0-160.0) | 142.5 (127.0-161.0) | 0.560 |
| Diastolic blood pressure, mmHg | 80.0 (71.5-89.5) | 81.0 (72.5-90.5) | <0.001 |
| Use of blood pressure-lowering treatment, % | 974 (27.7) | 819 (23.3) | <0.001 |
| Hypertension | 2223 (63.3) | 2182 (62.1) | 0.312 |
| Diabetes | 553 (15.7) | 360 (10.2) | <0.001 |
| Family history of CVD | 814 (23.2) | 758 (21.6) | 0.109 |
| Self-report CAD | 329 (9.4) | 0 (0.0) | — |
| Follow-up to CAD event or censored date, years | 5.5 (3.2-7.8) | 8.2 (5.0-10.9) | <0.001 |
| The training set for soft CAD | | | |
| Total participants | 7142 | 7142 | — |
| Array 1 | 4991 (69.9) | 6013 (84.2) | <0.001 |
| Urban | 2746 (38.4) | 2746 (38.4) | — |
| Women | 3434 (48.1) | 3434 (48.1) | — |
| Age, years | 62.1 (54.2-68.7) | 62.1 (54.1-68.7) | 0.473 |
| Primary school and below | 4417 (61.8) | 4575 (64.1) | 0.006 |
| Daily smoker | | | |
| Women | 181 (5.3) | 176 (5.1) | 0.786 |
| Men | 2011 (54.2) | 1997 (53.9) | 0.744 |

| | Case | Control | P-values |
|--|---------------------|---------------------|-----------------|
| Body mass index, kg/m ² | 23.6 (21.1-26.3) | 22.9 (20.5-25.6) | <0.001 |
| Waist circumference, cm | | | |
| Women | 81.3 (74.6-88.8) | 79.8 (72.7-86.8) | <0.001 |
| Men | 82.8 (74.8-90.7) | 80.4 (73.0-88.6) | <0.001 |
| Systolic blood pressure, mmHg | 140.5 (125.5-158.5) | 140.5 (125.0-159.0) | 0.536 |
| Diastolic blood pressure, mmHg | 80.0 (72.0-89.0) | 80.5 (72.0-89.5) | 0.003 |
| Use of blood pressure-lowering treatment, % | 2010 (28.1) | 1582 (22.2) | <0.001 |
| Hypertension | 4353 (60.9) | 4194 (58.7) | 0.007 |
| Diabetes | 904 (12.7) | 686 (9.6) | <0.001 |
| Family history of CVD | 1702 (23.8) | 1584 (22.2) | 0.019 |
| Self-report CAD | 1101 (15.4) | 0 (0.0) | — |
| Follow-up to CAD event or censored date, years | 5.2 (2.7-7.8) | 9.7 (5.5-11.5) | <0.001 |

Abbreviations: CAD, coronary artery disease; CVD, cardiovascular disease.

Data are n (%) or median (25–75th percentile range) unless otherwise specified. The p-value comes from the Chi-square test or t-test. Hard CAD events included nonfatal myocardial infarction (I21–I23) and fatal CAD (I20–I25); soft CAD events included all fatal or nonfatal CAD (I20–I25).

Table S5. Baseline characteristics of the testing set

| Characteristics at baseline | PRS quintiles * | | | | | P-value for trend |
|---|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-------------------|
| | Q1 (1047.9~1067.6) | Q2 (1067.7~1071.2) | Q3 (1071.3~1074.2) | Q4 (1074.3~1077.8) | Q5 (1077.9~1099.5) | |
| Number of participants | 14,430 | 14,431 | 14,429 | 14,430 | 14,430 | |
| Women, % | 59.1 | 60.1 | 59.9 | 60.2 | 59.9 | 0.244 |
| Age, years | 51.9 | 51.8 | 51.6 | 51.6 | 51.6 | 0.007 |
| Urban, % | 46.1 | 46.7 | 47.3 | 47.2 | 47.9 | <0.001 |
| Array 1, % | 14.7 | 14.7 | 14.4 | 14.2 | 14.5 | 0.377 |
| Primary school and below, % | 48.5 | 49 | 49.1 | 49.7 | 49.6 | 0.005 |
| Daily smoker, % | | | | | | |
| Women | 2.1 | 1.9 | 2.3 | 2.1 | 2.3 | 0.117 |
| Men | 54.4 | 56.5 | 57.1 | 56.8 | 56.9 | <0.001 |
| Systolic blood pressure, mmHg | 128.7 | 130.1 | 130.8 | 131.7 | 132.7 | <0.001 |
| Diastolic blood pressure, mmHg | 76.8 | 77.5 | 77.7 | 78.2 | 78.7 | <0.001 |
| Use of blood pressure-lowering treatment, % | 9.4 | 9.9 | 11.2 | 11.2 | 13.2 | <0.001 |
| Body mass index, kg/m ² | 23.7 | 23.7 | 23.7 | 23.8 | 23.8 | <0.001 |
| Waist circumference, cm | | | | | | |
| Women | 78.4 | 78.5 | 78.6 | 78.6 | 78.9 | <0.001 |
| Men | 81.9 | 81.8 | 82 | 82 | 81.9 | 0.657 |
| Diabetes, % | 4.9 | 5.6 | 5.4 | 5.6 | 6.4 | <0.001 |
| Family history of CVD, % | 18.8 | 20.1 | 19.9 | 20.3 | 21.5 | <0.001 |

Abbreviations: CAD, coronary artery disease; CVD, cardiovascular disease; PRS, polygenic risk score.

Baseline characteristics were presented as mean or percentage. Age, sex, urban areas, and array were not adjusted. Others were adjusted for age, sex, and study regions as appropriate.

* The PRS displayed here was the optimal PRS for hard CAD (PGS000337, see **Table 1** for details).

Table S6. Adjusted HRs for CAD associated with the optimal PRS for hard CAD in different models

| Items | Total | PRS quintiles* | | | | | HR per SD | P-value for trend |
|---------------------|---------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|---------------------|-------------------|
| | | Q1 (1047.9~1067.6) | Q2 (1067.7~1071.2) | Q3 (1071.3~1074.2) | Q4 (1074.3~1077.8) | Q5 (1077.9~1099.5) | | |
| No. of participants | 72,150 | 14,430 | 14,431 | 14,429 | 14,430 | 14,430 | | |
| Hard CAD | | | | | | | | |
| No. of PYs | 805,881 | 161,520 | 161,754 | 161,220 | 160,952 | 160,435 | | |
| No. of cases | 1,214 | 197 | 216 | 227 | 257 | 317 | | |
| Cases/PYs (1/1000) | 1.51 | 1.22 | 1.34 | 1.41 | 1.6 | 1.98 | | |
| Model 1 | | Reference | 1.15 (0.95 to 1.39) | 1.26 (1.04 to 1.53) | 1.47 (1.22 to 1.77) | 1.87 (1.57 to 2.24) | 1.26 (1.19 to 1.33) | <0.001 |
| Model 2 | | Reference | 1.14 (0.94 to 1.38) | 1.25 (1.03 to 1.52) | 1.47 (1.22 to 1.78) | 1.87 (1.56 to 2.23) | 1.26 (1.19 to 1.33) | <0.001 |
| Model 3 | | Reference | 1.14 (0.94 to 1.38) | 1.24 (1.03 to 1.50) | 1.40 (1.16 to 1.69) | 1.79 (1.50 to 2.14) | 1.23 (1.17 to 1.31) | <0.001 |
| Soft CAD | | | | | | | | |
| No. of PYs | 777,571 | 156,347 | 155,947 | 155,771 | 155,334 | 154,172 | | |
| No. of cases | 7,201 | 1,332 | 1,407 | 1,425 | 1,433 | 1,604 | | |
| Cases/ PYs (1/1000) | 9.26 | 8.52 | 9.02 | 9.15 | 9.23 | 10.4 | | |
| Model 1 | | Reference | 1.10 (1.02 to 1.19) | 1.15 (1.07 to 1.24) | 1.17 (1.08 to 1.26) | 1.36 (1.27 to 1.47) | 1.11 (1.09 to 1.14) | <0.001 |
| Model 2 | | Reference | 1.10 (1.02 to 1.19) | 1.15 (1.07 to 1.24) | 1.17 (1.08 to 1.26) | 1.36 (1.27 to 1.47) | 1.11 (1.09 to 1.14) | <0.001 |
| Model 3 | | Reference | 1.10 (1.02 to 1.18) | 1.14 (1.06 to 1.23) | 1.14 (1.06 to 1.23) | 1.33 (1.24 to 1.44) | 1.10 (1.08 to 1.13) | <0.001 |

Abbreviations: CAD, coronary artery disease; PRS, polygenic risk score; PY, person year. SD, standard deviation.

Hard CAD events included nonfatal myocardial infarction (I21–I23) and fatal CAD (I20–I25); soft CAD events included all fatal or nonfatal CAD (I20–I25).

All models were stratified by sex and ten study regions.

Model 1: no adjustment for covariates;

Model 2: adjusted for the top 10 principal components of ancestry and array versions;

Model 3: Model 2 + further adjusted for level of education (no formal school, primary school, middle school, high school, college or university, or higher), current daily smoker (yes/no), systolic blood pressure (mmHg), diabetes (yes/no), and waist circumference (cm).

* The PRS displayed here was the optimal PRS for hard CAD (PGS000337, see **Table 1** for details).

Table S7. Reclassification based on the continuous NRI and relative IDI

| | continuous NRI | relative IDI *, % |
|-----------|-----------------------|-------------------|
| Hard CAD | | |
| Women | | |
| cases | 0.108 (0.017, 0.200) | — |
| non-cases | 0.103 (0.094, 0.113) | — |
| total | 0.212 (0.119, 0.305) | 3.9 (0.3, 7.4) |
| Men | | |
| cases | 0.090 (0.005, 0.174) | — |
| non-cases | 0.103 (0.092, 0.115) | — |
| total | 0.193 (0.108, 0.278) | 4.6 (1.9, 7.2) |
| Soft CAD | | |
| Women | | |
| cases | 0.026 (-0.007, 0.059) | — |
| non-cases | 0.042 (0.031, 0.052) | — |
| total | 0.068 (0.034, 0.101) | 0.9 (0.4, 1.4) |
| Men | | |
| cases | 0.026 (-0.015, 0.067) | — |
| non-cases | 0.052 (0.040, 0.064) | — |
| total | 0.077 (0.034, 0.121) | 1.7 (0.9, 2.6) |

Abbreviations: IDI, integrated discrimination improvement; NRI, net reclassification improvement; CAD, coronary artery disease. The PRS reported here is the optimal PRS for hard CAD (PGS000337, see **Table 1** for details). Hard CAD events included nonfatal myocardial infarction (I21–I23) and fatal CAD (I20–I25); soft CAD events included all fatal or nonfatal CAD (I20–I25). Numbers in the brackets represent the 95% confidence intervals, which were calculated by 100 bootstrap replications using the BCa method in Stata.

* When calculating relative IDI, cases were defined as participants who developed CAD within 10 years of follow-up; non-cases were defined as those who were followed up for more than 10 years, including participants who developed CAD after 10 years.

Supplementary References

1. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics* 2010;26:2867-2873. doi: 10.1093/bioinformatics/btq559.
2. 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. doi: 10.1038/s41588-021-00783-5.
3. Vilhjalmsdottir BJ, Yang J, Finucane HK, Gusev A, Lindstrom S, Ripke S, *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am J Hum Genet* 2015;97:576-592. doi: 10.1016/j.ajhg.2015.09.001.
4. Wang M, Menon R, Mishra S, Patel AP, Chaffin M, Tanneer D, *et al.* Validation of a Genome-Wide Polygenic Score for Coronary Artery Disease in South Asians. *J Am Coll Cardiol* 2020;76:703-714. doi: 10.1016/j.jacc.2020.06.024.