SUPPLEMENTAL MATERIAL

SUPPLEMENTAL METHODS

Study populations

The CARDIA study recruited 5,115 healthy Black and White participants aged 18-30 years from four communities in 1985-1986 to study the natural history of cardiovascular disease beginning in young adulthood.[22,23] Genotype and phenotype data were available for a subset of 1,436 White participants through the dbGaP resource (phs000285). The Framingham Offspring Cohort study recruited 5,124 adult children and their spouses of participants of the original Framingham study in 1971-1975.[38] Genotype and phenotype data were available for a subset of 1,416 participants aged 18 to 35 at their first visit (phs000007). Use of de-identified data made available through dbGaP was approved by the Institutional Review Board of Vanderbilt University Medical Center.

Genetic data

For each data set, single nucleotide polymorphism (SNP) genotype data were acquired using the Affymetrix 500K (Framingham) or Affymetrix 6.0 SNP (CARDIA) array platforms. Data sets were validated against the 1000G reference using the HRC-1000G-check tool v4.2.5 (http://www.well.ox.ac.uk/~wrayner/tools/). One of each pair of related individuals (pi-hat>0.2) was removed in CARDIA. SNPs with >2% missingness, deviating from Hardy-Weinberg proportions ($p<10^{-6}$) or with allele frequencies not matching the 1000G reference standard for Europeans were removed prior to imputation. Genotype data were pre-phased using EAGLE (https://data.broadinstitute.org/alkesgroup/Eagle/) and imputed using the 10/2014 release of the 1,000 Genomes cosmopolitan reference haplotypes. Principal components (PCs) were generated among the European ancestry subpopulations using the SNPRelate package[39] and used to adjust analyses to ensure that associations were not due to ancestral differences between populations.

Polygenic risk score

These analyses used a CHD PRS developed by Khera et al using the LDpred[40] software package in conjunction with summary statistics from the CARDIOGRAMplusC4D CHD GWAS.[2] The PRS comprises 6,630,149 SNPs, and weightings were downloaded from http://www.broadcvdi.org/informational/data. For each participant, the PRS value was computed by summing the product of the allele weighting and the allele dosage across all SNPs. Post-imputation SNPs that comprised the PRS and that did not meet post-imputation quality standards (Hardy-Weinberg deviations and imputation r-square<0.3) were identified. A summary of these SNPs is presented in **Table 1**.

Table 1. Post-imputation SNP characteristics and PRS performance.

	CARDIA	Framingham
Imputed SNPs	6,163,735	6,630,138
HWE<10^{-10}	157,198	373,784
R-square<0.3	121	303,068
Correlation between PRS and PRS excluding low quality SNPs[1]	0.998	0.991
CAC Association (PRS using all SNPs)	1.54 (1.27 - 1.87)	2.11 (1.57 - 2.83)
CAC Association (PRS excluding low quality SNPs)	1.52 (1.25 - 1.84)	2.05 (1.53 - 2.75)

Footnotes:

1. The Pearson correlation coefficient between the PRS that includes all SNPs and the PRS that excludes low quality SNPs.

PRSs were computed that included and excluded low quality SNPs. The Pearson's correlation coefficient between the two PRS was >0.99 (**Table 1**). The association statistics between the PRS and a CAC Agatston score>0 (CARDIA) or CAC>300 (Framingham) were similar for both PRSs (**Table 1**). The point estimates were nominally higher when all SNPs were included in the PRS. For this reason, the PRS including all SNPs was used in these analyses.

Baseline phenotypes

Clinical variables including age, sex, smoking status, SBP, diastolic blood pressure, anti-hypertensive medication use, body mass index, waist circumference (CARDIA only), total cholesterol, triglycerides, LDL-C, high density lipoprotein cholesterol (HDL-C) and glucose levels were ascertained from the visit 1 exams (CARDIA: 1985-1986; Framingham: 1971-5). In CARDIA, current smokers were participants who smoked >5 cigarettes/day for a duration >3 months. No CARDIA participants were taking lipid-lowering medications at Visit 1.

Coronary artery calcium scoring

Participants underwent two computed tomography scans from the root of the aorta to the apex of the heart at year 15 (2000-2001) in CARDIA or year 30 (2002-2005) in Framingham.[41,42] From these, Agatston scores, adjusted using a standard calcium phantom scanned underneath each participant, were computed for the four major arteries. The CAC Agatston score is the average of two scans.[10] Binary classifiers for Agatston scores above thresholds of >0, >10 and >20 were constructed in CARDIA and >100, >200 and >300 were constructed in Framingham. The primary analyses examined the >20 (CARDIA) and >300 (Framingham) thresholds, though results are presented for other thresholds for comparison.

Analysis
Participants with missing CAC data, or covariate data were excluded from the analyses.

In each cohort, the PRS was set to have a standard deviation of 1, so the units of all reported associations are the change per standard deviation change in the PRS. When risk factors are noted to be included in the models, these were: baseline waist circumference [or BMI in Framingham], total cholesterol, HDL-C, log-transformed triglycerides, SBP and smoking status.

All models were adjusted for age, sex and 5 PCs. In the Framingham cohort, models included a family structure variable as a random effect to adjust for relatedness among individuals.

Associations between baseline characteristics and the PRS were tested using either linear regression (continuous traits) or logistic regression (binary traits) models. Associations with a nominal p-value are highlighted.

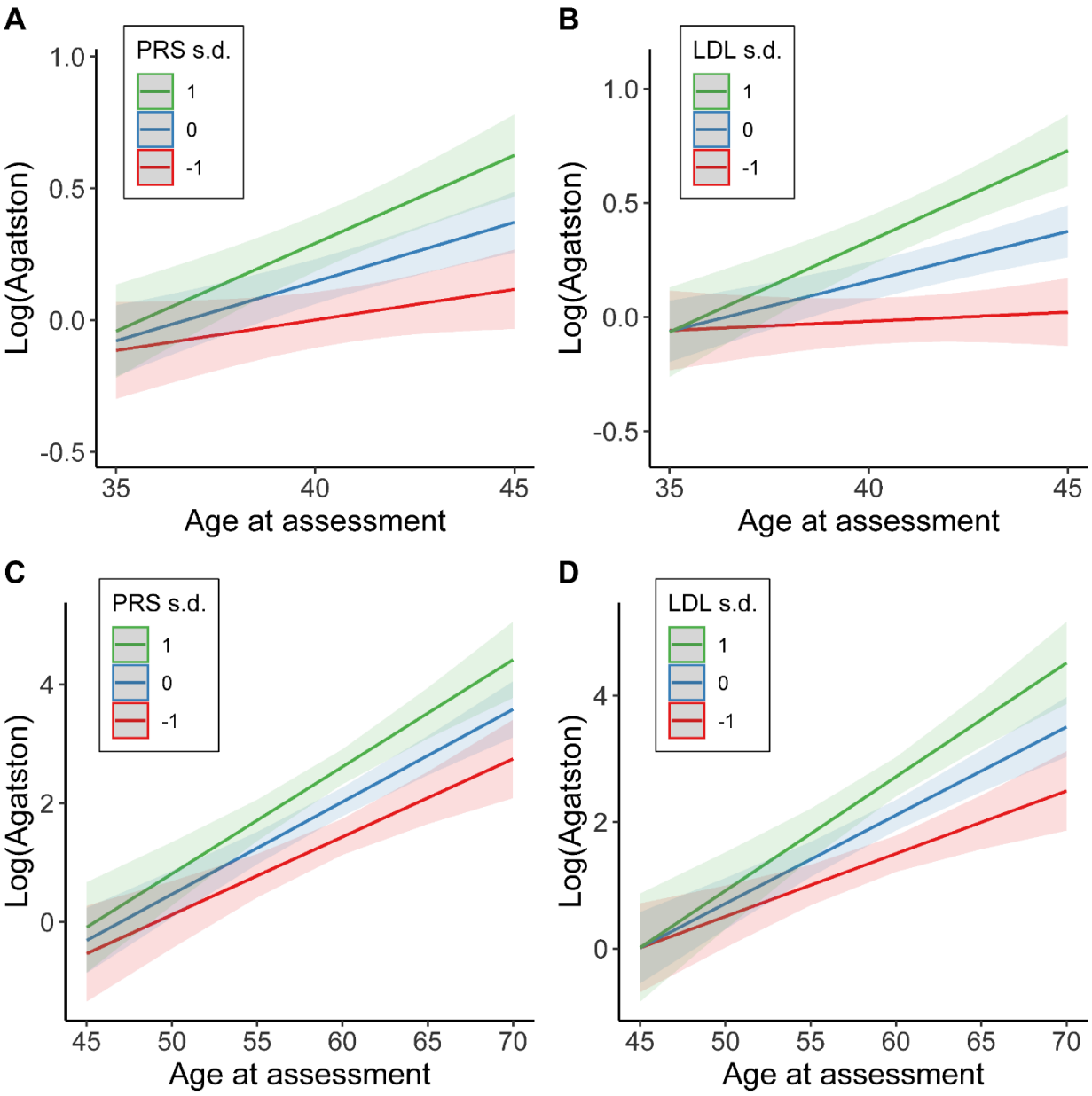
To examine the association between continuous CAC scores and the PRS, linear regression models with and without additional adjustment for risk factors, were used to estimate changes in log-transformed scores. Because CAC has a zero-inflated distribution, log(Agatston score +1) was used as the dependent variable. To examine whether there was an interaction between the PRS and duration of exposure to genetic risk, a linear regression model that included an interaction term between the PRS and age of assessment of CAC was also evaluated. An interaction term $p<0.05$ was considered significant.

Logistic regression was used to test the association between having a CAC score greater than a specific threshold and the PRS, with and without additional adjustment for risk factors. C-statistics were used to evaluate model discrimination. For Framingham, c-statistics were based on fixed effect estimates. Bootstrapping was used to estimate 95% confidence intervals for C-statistic estimates and differences between C-statistic estimates due to the effects of additional covariates added to the model. A confidence interval that did not cross zero was considered statistically significant. Data were visualized using Receiver Operating Characteristic curves. Goodness-of-fit, including slope estimates from goodness-of-fit plots, was assessed and visualized using the rms package (val.prob package).

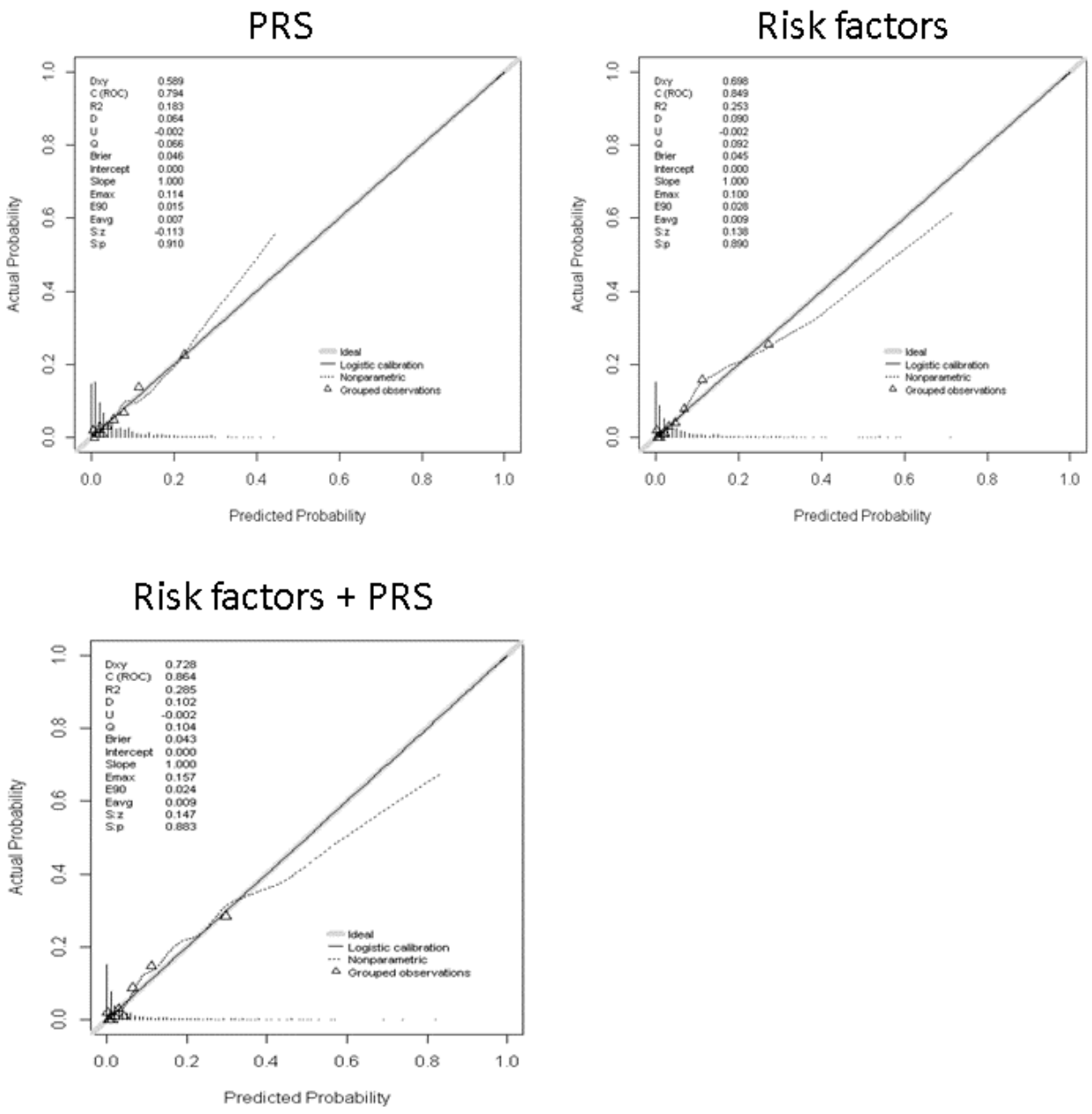
The integrated discrimination improvement (IDI) was used to measure the extent that addition of a covariate to the predictive model increased risk estimates among participants with CAC>0 and decreased risk in those without CAC. A higher value is indicative of improved discrimination. Bootstrapping was used to computed 95% confidence intervals.

The performance of the PRS as a binary risk classifier was evaluated using select thresholds between the top 5th and 20th percentiles of the distribution to identify subjects at elevated polygenic risk.

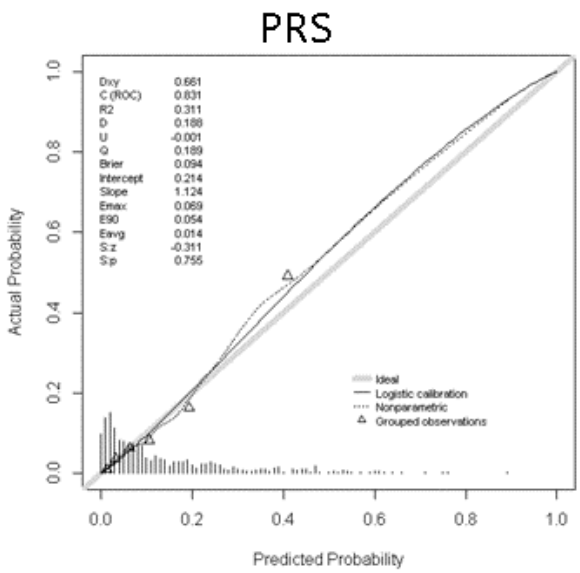
All analyses used R v4.0.0.

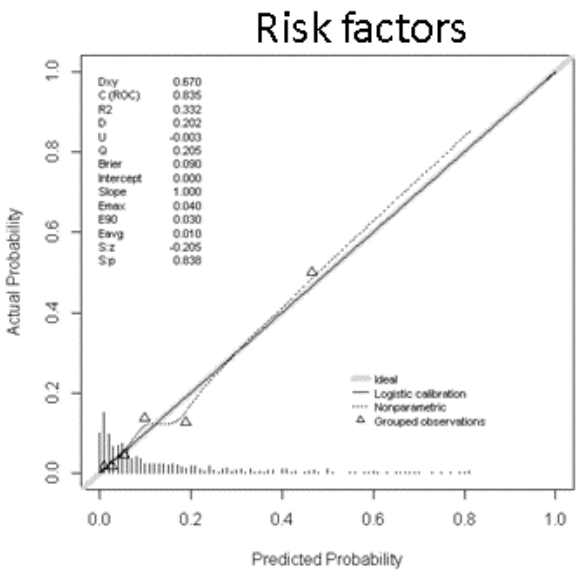


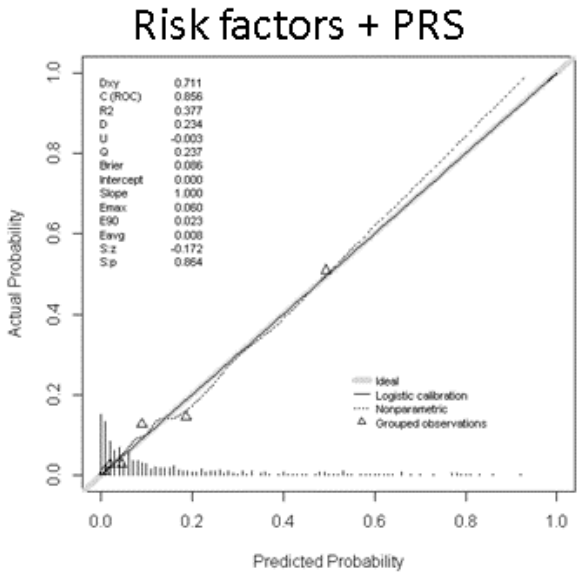
Supplemental Figure I: Predicted log(Agatston score +1) levels from a linear regression model that included an interaction term between either the PRS or LDL-C and the participant's age at CAC. Shown are predicted levels for participants at the mean (0) or +/-1 standard deviations from the mean of the PRS or LDL-C distribution in A&B) CARDIA and C&D) Framingham.



Supplemental Figure II: Goodness-of-fit plots for logistic regression models in CARDIA looking at associations with CAC>20. Logistic regression models included: the 1) CHD PRS; 2) modifiable risk factors; or 3) modifiable risk factors and the PRS. All models were adjusted sex, age and 5 PCs. The solid line shows the goodness-of-fit curve for a logistic regression model.







Supplemental Figure III: Goodness-of-fit plots for logistic regression models in the Framingham population looking at associations with CAC>300. Logistic regression models included: the 1) CHD PRS; 2) modifiable risk factors; or 3) modifiable risk factors and the PRS. All models were adjusted sex, age and 5 PCs. The solid line shows the goodness-of-fit curve for a logistic regression model.

Supplemental table I: Overview of exclusions and population selection for CARDIA and Framingham cohorts.

CARDIA

Category	Excluded	Remaining
Genotyped individuals	n/a	1,441
Related Individuals	7	1,434
Failed genotype imputation	2	1,432
Missing coronary calcium data	286	1,146
Missing baseline covariates	14	1,132

Framingham Offspring Cohort

Category	Excluded	Remaining
Visit 1 age 18-35	n/a	1,441
Missing coronary calcium data	726	690
Failed genotype imputation	2	688
Missing baseline covariates	25	663

Supplemental table II: Characteristics of the CARDIA cohort at visit 6 (Year 15). Either a paired T-test or McNemar's test were used to examine for significant changes between Visit 1 and Visit 6.

Characteristic	Visit 1	Visit 6	p- value
Male (%)	545 (48.1%)	-	-
Age (years)	25.6 (3.3)	40.7 (3.3)	<1E-300
Systolic blood pressure (mmHg)	108.8 (11)	109.9 (12.7)	8.3E-04
Diastolic blood pressure (mmHg)	68 (9.1)	72.3 (10.3)	8.5E-41
Body mass index (kg/m^2)	23.6 (3.8)	27.1 (5.5)	1.3E-178
Waist Circumference (cm)	77.2 (10.3)	87.5 (14.2)	2.1E-214
Total cholesterol (mg/dl)	176.2 (31.7)	186.6 (33.9)	4.8E-30
Triglycerides (mg/dl)	65 (47 - 90)	89 (64 - 132)	9.3E-85
HDL cholesterol (mg/dl)	51.7 (12.9)	49.7 (14.6)	1.0E+00
LDL cholesterol (mg/dl)	108.8 (29.3)	114.6 (30.9)	1.4E-12
Glucose (mg/dl)	82 (77 - 87)	90 (85 - 98)	4.1E-135
Current smoker	298 (26.3%)	198 (17.5%)	4.8E-14
Type 2 diabetes	4 (0.4%)	27 (2.4%)	1.6E-06

Supplemental table III: Comparison of subjects included and excluded from the CARDIA cohort analyses.

Characteristic	Included	Excluded
Total	1132	300
Male (%)	545 (48.1%)	132 (44%)
Age (years)	25.6 (3.3)	25 (3.4)*
Systolic blood pressure (mmHg)	108.8 (11)	110 (10.9)*
Diastolic blood pressure (mmHg)	68 (9.1)	68.4 (9.3)
Body mass index (kg/m^2)	23.6 (3.8)	24 (4.8)
Waist Circumference (cm)	77.2 (10.3)	77.7 (12.6)
Total cholesterol (mg/dl)	176.2 (31.7)	177.3 (32)
Triglycerides (mg/dl)	65 (47 - 90)	73 (53 - 98)*
HDL cholesterol (mg/dl)	51.7 (12.9)	51.3 (13.7)
LDL cholesterol (mg/dl)	108.8 (29.3)	108.4 (30.3)
Glucose (mg/dl)	82 (77 - 87)	83 (78 - 89)
Anti-hypertensive medications	9 (0.8%)	2 (0.7%)
Diabetes medications	2 (0.2%)	1 (0.3%)
Mother MI Prior to Age 60	24 (2.1%)	8 (2.7%)
Father MI Prior to Age 60	122 (10.8%)	41 (13.7%)
Current smoker	298 (26.3%)	78 (26%)
Type 2 diabetes	4 (0.4%)	1 (0.3%)

*****Denotes a significant difference ($p<0.05$) between the included and excluded subjects.

Supplemental table IV: Comparison of subjects included and excluded from the Framingham cohort analyses. There were no significant differences in the baseline characteristic.

Characteristic	Included subjects	Excluded subjects
Total	663	753
Male (%)	308 (46.5%)	333 (44.2%)
Age (years)	27.8 (4.7)	28.3 (4.5)
Systolic blood pressure (mmHg)	116 (11.8)	117 (12.1)
Diastolic blood pressure (mmHg)	75.4 (8.6)	74.8 (9.1)
Body mass index (kg/m^2)	24.31 (3.9)	24.7 (4.3)
Total cholesterol (mg/dl)	181.2 (33.0)	182.7 (34.7)
Triglycerides (mg/dl)	215 (151-328)	213 (143 - 356)
HDL cholesterol (mg/dl)	50.4 (13.2)	50.2 (13.7)
LDL cholesterol (mg/dl)	114.1 (30.4)	114.7 (31.1)
Glucose (mg/dl)	98 (92-104)	99 (93 - 105)
Anti-hypertensive medications	3 (0.45)	5 (0.7)
Mother MI Prior to Age 55	17 (2.9%)	20 (2.7%)
Father MI Prior to Age 55	76 (12.7%)	95 (12.6%)
Risk Factors		
Current smoker	280 (42.2%)	351 (46.6%)
Type 2 diabetes	3 (0.5%)	0 (0%)

Supplemental table V: Overview of findings associated with selected CAC Agatston score cut-offs in CARDIA. The base model is adjusted for age, sex and 5 PCs.

	CAC Threshold	>0	>10	>20
	Cases	140	78	61
PRS Beta Coefficient	**Unadjusted**	1.54 (1.27 - 2.87)	1.62 (1.27, 2.09)	1.87 (1.41 - 2.50)
	Adjusted	1.49 (1.22 - 1.82)	1.53 (1.19 - 2.00)	1.74 (1.29 - 2.36)
Likelihood ratio test	**Add PRS to Base**	7.1E-06	8.1E-05	8.9E-06
	Add risk factors to Base	1.4E-14	2.1E-10	7.6E-09
	Add PRS to risk factors	9.2E-05	1.0E-03	2.0E-04
	Add risk factors to PRS	1.4E-13	1.9E-09	1.1E-07
C-statistics	**Base**	0.701 (0.643, 0.740)	0.738 (0.672, 0.786)	0.764 (0.692, 0.816)
	Base + PRS	0.728 (0.672, 0.768)	0.759 (0.694, 0.804)	0.794 (0.728, 0.840)
	Base + risk factors	0.793 (0.745, 0.825)	0.828 (0.774, 0.862)	0.849 (0.789, 0.889)
	Base + risk factors + PRS	0.806 (0.761, 0.839)	0.840 (0.786, 0.877)	0.864 (0.807, 0.904)
C-statistic change	**Add PRS to Base**	0.026 (0.004, 0.049)	0.020 (-0.005, 0.042)	0.030 (0.007, 0.054)
	Add risk factors to Base	0.092 (0.051, 0.131)	0.089 (0.044, 0.133)	0.085 (0.036, 0.131)
	Add PRS to risk factors	0.013 (0.001, 0.027)	0.013 (0.001, 0.026)	0.015 (0.004, 0.028)
	Add risk factors to PRS	0.078 (0.047, 0.113)	0.081 (0.040, 0.124)	0.070 (0.033, 0.109)
IDI	**Add PRS to risk factors**	0.020 (0.000, 0.039)	0.016 (-0.009, 0.0373)	0.027 (-0.006, 0.054)
	Add Risk factors to PRS	0.079 (0.028, 0.113)	0.070 (0.0001, 0.106)	0.070 (-0.005, 0.111)

Supplemental table VI: Overview of findings associated with selected CAC Agatston score cut-offs in the Framingham Offspring cohort.

	CAC Threshold	CAC>100	CAC>200	CAC>300
	Cases	178	122	93
PRS Beta Coefficient	**Unadjusted**	1.79 (1.49, 2.16)	2.01 (1.66, 2.42)	2.11 (1.57, 2.83)
	Adjusted	1.63 (1.31, 2.02)	1.76 (1.35, 2.30)	1.90 (1.42, 2.54)
Likelihood ratio test	**Add PRS to Base**	4.49E-08	3.64E-08	2.11E-08
	Add risk factors to Base	3.76E-11	1.09E-12	1.01E-11
	Add PRS to risk factors	7.33E-06	9.07E-06	5.31E-06
	Add risk factors to PRS	3.62E-09	1.60E-10	1.46E-09
C-statistics	**Base**	0.736 (0.690, 0.775)	0.752 (0.700, 0.792)	0.759 (0.702, 0.808)
	Base + PRS	0.771 (0.725, 0.805)	0.794 (0.743, 0.836)	0.804 (0.751, 0.845)
	Base + risk factors	0.794 (0.746, 0.822)	0.829 (0.781, 0.861)	0.835 (0.781, 0.871)
	Base + risk factors + PRS	0.811 (0.764, 0.840)	0.846 (0.801, 0.875)	0.855 (0.805, 0.887)
C-statistic change	**Add PRS to Base**	0.035 (0.008, 0.058)	0.041 (0.012, 0.071)	0.045 (0.010, 0.076)
	Add risk factors to Base	0.058 (0.026, 0.081)	0.077 (0.036, 0.110)	0.076 (0.031, 0.11)
	Add PRS to risk factors	0.017 (-0.0001, 0.033)	0.017 (-0.001, 0.034)	0.020 (0.001, 0.039)
	Add risk factors to PRS	0.040 (0.015, 0.061)	0.052 (0.019, 0.079)	0.051 (0.017, 0.079)
IDI	**Add PRS to risk factors**	0.031 (0.0008, 0.057)	0.030 (-0.002, 0.059)	0.039 (0.0005, 0.072)
	Add Risk factors to PRS	0.072 (0.024, 0.103)	0.100 (0.036, 0.142)	0.111 (0.036, 0.158)