DOI: 10.1161/CIRCGEN.120.003128

Interactions between Enhanced Polygenic Risk Scores and Lifestyle for Cardiovascular Disease, Diabetes Mellitus and Lipid Levels

Running title: *Ye et al.; Interactions between Enhanced PRS and Lifestyle*

Yixuan Ye, MS[1]; Xi Chen, BS[2,3]; James Han, BS[2,3]; Wei Jiang, PhD[4]; Pradeep Natarajan, MD[5,6]; Hongyu Zhao, PhD[1,4]

[1]Program of Computational Biology and Bioinformatics, [2]Department of Statistics and Data Science, [3]Department of Molecular Biophysics and Biochemistry, Yale University; [4]Department of Biostatistics, Yale School of Public Health, New Haven, CT; [5]Center for Genomic Medicine and Cardiovascular Research Center, Massachusetts General Hospital, Boston; [6]Program in Medical and Population Genetics, Broad Institute of Harvard & Massachusetts Institute of Technology, Cambridge, MA

Correspondence:

Hongyu Zhao, PhD

Yale School of Public Health

P.O. Box 208034, 60 College Street

New Haven, CT 06520

Tel: 203.785.3613

Email: hongyu.zhao@yale.edu

Journal Subject Terms: Genetic, Association Studies; Lifestyle; Cardiovascular Disease; Diabetes, Type 2; Lipids and Cholesterol

DOI: 10.1161/CIRCGEN.120.003128

Abstract:

Background - Both lifestyle and genetic factors confer risk for cardiovascular diseases, type 2 diabetes (T2D), and dyslipidemia. However, the interactions between these two groups of risk factors were not comprehensively understood due to previous poor estimation of genetic risk. Here we set out to develop enhanced polygenic risk scores (PRS), and systematically investigate multiplicative and additive interactions between PRS and lifestyle for coronary artery disease, atrial fibrillation, T2D, total cholesterol, triglyceride, and LDL-cholesterol.

Methods - Our study included 276,096 unrelated white British participants from the UK Biobank. We investigated several PRS methods (P+T, LDpred, PRS-CS, and AnnoPred), and showed that AnnoPred achieved consistently improved prediction accuracy for all six diseases/traits. With enhanced PRS and combined lifestyle status categorized by smoking, body mass index, physical activity, and diet, we investigated both multiplicative and additive interactions between PRS and lifestyle using regression models.

Results - We observed that healthy lifestyle reduced disease incidence by similar multiplicative magnitude across different PRS groups. The absolute risk reduction (ARR) from lifestyle adherence was however significantly greater in individuals with higher PRS. Specifically, for T2D, the ARR from lifestyle adherence was 12.4% (95% CI, 10.0%-14.9%) in the top 1% PRS versus 2.8% (95% CI, 2.3%-3.3%) in the bottom PRS decile, leading to a ratio of more than 4.4. We also observed a significant interaction effect between PRS and lifestyle on triglyceride level.

Conclusions - By leveraging functional annotations, AnnoPred outperforms state-of-the-art methods on quantifying genetic risk through PRS. Our analyses based on enhanced PRS suggest that individuals with high genetic risk may derive similar relative but greater absolute benefit from lifestyle adherence.

Key words: genetics, association studies; lifestyle; cardiovascular disease; diabetes mellitus; lipids; polygenic risk score

DOI: 10.1161/CIRCGEN.120.003128

Nonstandard Abbreviations and Acronyms
cardiovascular diseases (CVD)
genome-wide association studies (GWAS)
type 2 diabetes (T2D)
polygenic risk scores (PRS)
coronary artery disease (CAD)
atrial fibrillation (AF)
UK Biobank (UKBB)
continuous shrinkage (CS)
total cholesterol (TC)
triglyceride (TG)
low-density lipoprotein cholesterol (LDL-C)
body mass index (BMI)
hazard ratios (HRs)
absolute risk reduction (ARR)
Atherosclerosis Risk in Communities study (ARIC)

Introduction

Poor lifestyle has long been known to confer risk for cardiovascular diseases (CVD)[1,2], which are the leading cause of morbidity and mortality in the world[3]. To improve cardiovascular health, the American College of Cardiology and the American Heart Association has recommended a healthy lifestyle with guidelines for diet, physical activity, obesity, and tobacco use[4].

Previous studies have also demonstrated that genetics plays an important role in CVD risk[5–7]. In recent years, large-scale genome-wide association studies (GWAS) have rapidly expanded our knowledge on the genetic variants that are associated with CVD and its key drivers including type 2 diabetes (T2D) and dyslipidemia[8–11]. Many methods have been subsequently proposed for aggregating information from GWAS results to quantify the genetic risk for an individual through polygenic risk scores (PRS)[12–17].

DOI: 10.1161/CIRCGEN.120.003128

As CVD risk is conferred by both lifestyle and genetic factors, several studies have sought to examine the interactions between these two groups of risk factors for CVD traits[18–20], and reached the conclusion that genetic factors (summarized through PRS) and lifestyle factors independently contribute to cardiovascular disorders and related diseases, including coronary artery disease (CAD), atrial fibrillation (AF), stroke, hypertension, and T2D[19,21]. However, the PRS used in previous studies were constructed either based on GWAS results from a limited number of samples or using less optimal risk prediction methods, providing a less accurate representation of an individual's genetic risk[12–15,18–21]. In addition, previous studies also coarsely stratified the study population by PRS into a limited number of risk groups without considering that the empirical risk of common diseases increased sharply in the extreme tails of the PRS distribution[22–25]. As summarized recently by Khera *et al*[25], larger GWASs and improved statistical methods could derive PRS with better prediction accuracy. Here, we set out to construct more predictive PRS for CAD, AF and T2D from the largest-to-date GWAS summary statistics, and then comprehensively investigate the extent to which the genetic predisposition of CAD, AF and T2D can alter the effect of lifestyle adherence on disease outcomes, especially for those with extremely high genetic risk. In addition, we also extend the investigation of potential interactions between PRS and lifestyle on blood lipid levels, which are crucial intermediate traits for CVD[6,26].

Methods

A full-length description of the methods is available as part of the Data Supplement (Methods, Table I-II, Figure I-II). Because of the sensitive nature of the individual-level data collected for this study, requests to access the dataset from qualified researchers trained in human subject

confidentiality protocols may be sent to UK Biobank (UKBB) at https://www.ukbiobank.ac.uk/register-apply. The summary-level data (e.g. PRS weights) are available from the corresponding author upon request. All analytical methods are available and reported[12–15]. The study was approved by the UKBB[27] and by the ethic committee of Yale University. All individuals have provided written informed consent.

Results

Population characteristics

Of the 502,618 participants aged 40-69 years in the UKBB, 226,522 were excluded according to the exclusion criteria stated in Methods in the Data Supplement. The remaining 276,096 individuals were divided into a training set of 92,928 individuals and a testing set of 183,168 individuals. In the training set, there were 4,746 CAD cases, 3,606 AF cases, and 4,639 T2D cases. A total of 16,719 individuals were excluded from lipid level analysis due to missingness or their taking cholesterol-lowering medication. The testing set for each disease was constructed by further excluding prevalent cases. This yielded 176,238 participants with 3,467 incident cases for CAD, 178,651 participants with 4,025 incident cases for AF, 178,138 participants with 4,659 incident cases for T2D, and 144,939 participants for the lipid testing set (Figure I in the Data Supplement). Baseline characteristics for the population are provided in Table 1.

Prediction performance of polygenic risk scores

Four PRS methods were considered in our study: P+T[12], LDPred[13], PRS-CS[14], and AnnoPred[15]. P+T is also the standard PRS method where the marginal effect sizes from GWAS summary statistics were directly used as weights, and the SNPs were selected after LD-clumping and *p*-value thresholding[12]. LDPred and PRS-CS are both Bayesian approaches that model the LD

DOI: 10.1161/CIRCGEN.120.003128

information extracted from a reference panel, where LDPred assumes an independent point-normal prior while PRS-CS assumes a continuous shrinkage (CS) prior on the SNP effect sizes[13,14]. AnnoPred is a Bayesian framework that further leverages functional annotations in quantifying genetic risk[15].

Our empirical results suggest that AnnoPred achieved the best predictive performance for all six traits in both the training and testing sets, with the AUC in the testing set being 0.643 (95% CI, 0.637-0.648) for CAD, 0.632 (95% CI, 0.625-0.639) for AF, and 0.645 (95% CI, 0.639-0.651) for T2D, respectively; and R^2 being 0.0751 (95% CI, 0.0728-0.0775) for total cholesterol (TC), 0.0744 (95% CI, 0.0720-0.0767) for triglyceride (TG), and 0.0705 (95% CI, 0.0682-0.0729) for low-density lipoprotein cholesterol (LDL-C), respectively (Figure 1, Table III in the Data Supplement). In particular, the optimal PRS involved 2,994,054 variants for CAD, 2,996,792 variants for AF, 2,996,760 variants for T2D, 1,198,743 variants for TC, 1,197,954 variants for TG, and 1,197,834 variants for LDL-C, respectively.

Associations of disease risk with PRS and combined and individual lifestyle factors

As shown in Figure 2, a risk gradient was clearly observed across the 10 genetic risk groups for each overall lifestyle status where individuals with high PRS were at higher risk of CAD, AF and T2D events than those with low PRS. This trend was especially visible for participants in the right tail of the PRS distribution, where the risk of developing disease increased sharply as PRS increased. For participants with an intermediate combined lifestyle status, the adjusted hazard ratios (HRs) were 4.23 (95% CI, 3.39-5.28) for CAD, 3.69 (95%, 2.96-4.60) for AF, and 3.67 (95%, 2.99-4.52) for T2D, when comparing the group with >99% PRS to the group with 40-60% PRS (Table IV-V in the Data Supplement).

DOI: 10.1161/CIRCGEN.120.003128

Within each of the 10 PRS-defined genetic risk groups, poor lifestyle was consistently associated with increased disease risk when compared to intermediate lifestyle, whereas healthy lifestyle was associated with decreased disease risk. Compared with the group of individuals with mid-range genetic risk (40-60% PRS) and intermediate combined lifestyle, the adjusted HRs for the group of individuals having the top 1% PRS while leading poor lifestyles increased to 5.23 (95% CI, 3.01-9.09) for CAD, 5.43 (95%, 3.34-8,82) for AF, and 6.67 (95%, 4.55-10.1) for T2D. On the contrary, for individuals with similarly high genetic risk (top 1% PRS) but healthy lifestyle, the corresponding risk decreased to 3.19 (95% CI, 1.64-6.19) for CAD, 2.15 (95%, 1.15-4.02) for AF, and 0.98 (95%, 0.41-2.36) for T2D, demonstrating the benefit of leading a healthy lifestyle even for this extremely high genetic risk group.

Relative disease risk reduction when leading a healthy lifestyle

As the three HR curves across the 10 genetic risk bins for the three lifestyle groups are almost parallel to each other (Figure 2), the UKBB data suggest that the effects of lifestyles may be independent of PRS for CAD, AF, and T2D at the log HR scale. We then performed formal statistical tests to investigate the multiplicative interactions between PRS and combined lifestyle on disease outcomes under the Cox proportional hazard regression model. Consistent with our impressions from the figures, no significant multiplicative interactions were identified after Bonferroni correction (Table VI-VII in the Data Supplement).

We further studied the joint effects of genetic risk and individual lifestyle factors as summarized in Figure III-V in the Data Supplement. At the individual lifestyle factor level, each lifestyle factor seems to still exert an effect independent of genetic risk on disease outcomes, albeit the magnitude of relative risk across each lifestyle factor varies among three diseases. Smoking and body mass index (BMI) were both important risk factors for CAD and AF, while

DOI: 10.1161/CIRCGEN.120.003128

diet and physical activity did not clearly influence risk. For T2D, BMI still played a prominent role, and the effect from physical activity could also be observed clearly, while no significant effects from smoking and diet were observed.

Absolute disease risk reduction when leading a healthy lifestyle

In contrast to the independent relationship between reduced HR and PRS, absolute risk reduction (ARR) from healthy lifestyle was greater in the group at higher PRS for CAD, AF and T2D (Figure 3, Table VIII in the Data Supplement). More specifically, after lifestyle modification, in the extremely high PRS group (PRS > 99%), AR of CAD reduced from 7.9% to 3.4% (ARR = 4.5%); while in the intermediate PRS group (40% < PRS < 60%), AR of CAD reduced from 2.8% to 0.9% (ARR = 1.9%); and in the low PRS group (PRS < 10%), AR of CAD reduced from 1.6% to 0.6% (ARR = 1.0%). For AF, after lifestyle modification, the AR reduced from 10.7% to 3.5% (ARR = 7.2%) in the extremely high PRS group (PRS > 99%), from 3.3% to 1.5% (ARR = 1.8%) in the intermediate PRS group (40% < PRS < 60%) and from 1.1% to 0.6% (ARR = 0.5%) in the low PRS group (PRS < 10%). Strikingly, there was an apparent sharp increase of ARR in the extreme right tail of AF PRS distribution, where the ARR was less than 3% even in group with PRS > 90% & PRS < 99%, but rose up to 7.2% in the group with PRS > 99%. Overall, the AR of T2D was larger than CAD and AF. The ARR from lifestyle adherence for T2D was also larger; after lifestyle modification, the AR reduced from 14.3% to 1.9% (ARR = 12.4%) in the extremely high PRS group (PRS > 99%), from 5.6% to 0.6% (ARR = 5.0%) in the intermediate PRS group (40% < PRS < 60%) and from 3.1% to 0.3% (ARR = 2.8%) in the low PRS group (PRS < 10%).

We also studied the ARR from individual lifestyle adherence within each PRS groups (Figure VI-VIII in the Data Supplement). For CAD, we could observe the same trend that ARR

was greater in the group at higher PRS from the modification of smoking behavior, BMI and physical activity. In the extremely high PRS group (PRS > 99%), changing the smoking status from poor to ideal alone could lead to an ARR of 5.2%. For AF, ARR from lifestyle adherence mainly came from the modification of BMI. Especially in the extremely high PRS group (PRS > 99%), ARR from BMI modification was as high as 6.7%. For T2D, the ARRs from all of the four individual lifestyle components increased with PRS. Specifically, BMI modification alone could lead to an ARR of 11.1% and smoking behavior modification alone could lead to an ARR of 6.3%.

Associations of lipid levels with PRS and combined and individual lifestyle factors

For all three blood lipid types, the mean levels increased across PRS quantiles for each combined lifestyle category, while healthy lifestyle was associated with decreased mean levels within each PRS group (Figure 3, Table IX in the Data Supplement).

According to the recommended guidelines set by the National Cholesterol Education Program (NCEP)[28,29], only the group with the lowest genetic risk (<5% PRS) in each lifestyle category had an average TC level at the desirable level (<5.172 mmol/L [<200 mg/dL]). Conversely, groups with high genetic risk (>90% PRS) in all lifestyle categories had an average TC level at high designation (≥6.206 mmol/L [≥240 mg/dL]). For groups with the same PRS grouping but different lifestyle categories, the average TC levels were mostly in the same NCEP designation, but healthy lifestyle tended to pull the average TC levels to a healthier designation. The protective effect of healthy lifestyle was most pronounced for TG. The normal designation for TG is <1.694 mmol/L [150 mg/dL]. While none of the PRS groups with poor lifestyle had average levels below this threshold, the average TG level was below this threshold for people with intermediate lifestyle and <40% PRS, whereas for people with healthy lifestyle, the average

DOI: 10.1161/CIRCGEN.120.003128

level was below the threshold for all the PRS groups except the one with the highest genetic risk, i.e. >99%. As for the high TG designation (≥2.258 mmol/L [≥200 mg/dL]), the average TG level was above this threshold for individuals with >60% PRS in poor lifestyle group versus >99% PRS in the intermediate lifestyle group and none of the PRS groups in the ideal lifestyle group. A similar result was also observed in the LDL-C analysis, with <5% PRS, <10% PRS, and <20% PRS in the poor, intermediate, and ideal lifestyle groups, respectively, were at the optimal or near-optimal levels of LDL-C (<3.362 mmol/L [<130 mg/dL]). Moreover, 5-60% PRS, 10-60% PRS, and 20-80% PRS in the respective groups were at borderline high levels (3.362-4.138 mmol/L [130-160 mg/dL]), and the remaining intervals of PRS had LDL-C levels designated as high (>4.138 mmol/L [>160 mg/dL]).

The results in Figure 4 suggest that lifestyle and genetic factors may exert independent effects on the blood lipid levels, as healthy lifestyle reduced the mean lipid levels similarly regardless of PRS grouping. We subsequently tested statistical interactions between PRS and combined lifestyle categories for all three types of lipid (Table X-XI in the Data Supplement). In the analyses based on the deciles of PRS, we observed no significant interactions after Bonferroni correction, as suggested by Figure 4. However, in the analyses based on continuous PRS, we observed a significant positive interaction between PRS and lifestyle for TG ($P = 0.0027$), suggesting that people at higher PRS could benefit more from lifestyle adherence.

We also studied the joint effects of genetic risk and individual lifestyle factors on lipid levels (Figures IX-XI in the Data Supplement). Individual lifestyle factors still exerted effects on lipid levels independent of genetic risk, with BMI as the prominent lifestyle factor for all three types of lipid and smoking playing an important role specifically for TG.

DOI: 10.1161/CIRCGEN.120.003128

Sex differences

We also analyzed the associations of diseases incidence and lipid levels with PRS and combined lifestyle stratified by sex as shown in Figures XII-XIV in the Data Supplement. The results were generally similar to the results from the combined analyses. However, given the same PRS quantile and combined lifestyle status, males tended to have a higher level of TG compared to females, and the larger ARRs among males suggested that males might have more absolute benefits from lifestyle adherence regarding to the prevention of CAD, AF and T2D.

Discussion

By using enhanced PRSs in this large-scale study of around 300,000 UKBB participants, no significant multiplicative interactions were found between genetic risk and lifestyle for CAD, AF, and T2D incidence under the Cox proportional hazard model. However, we observed a significant association between PRS grouping and ARR from lifestyle adherence for CAD, AF and T2D, where healthy lifestyle decreased the absolute risk much more significantly in individuals at high PRS-percentiles. We also found a significant positive interaction between PRS and lifestyle for TG.

In general, the impact of lifestyle factors and PRS on CAD, AF and T2D incidence observed in our study is in line with previous reports by Khera *et al*[18] and Said *et al*[19]. Both poorer lifestyle and higher genetic risk could lead to higher risks of developing CAD, AF and T2D. However, our study has extended the investigations of the previous studies in several ways.

First, we investigated the effect of genetic factor and its interaction with lifestyle based on more accurate quantification of genetic risks. More specifically, by leveraging functional annotations in genetic risk prediction, we developed enhanced PRS for CAD, AF, T2D, TC, TG,

and LDL-C using AnnoPred[15]. Applying to UKBB dataset, we showed that our enhanced PRS outperformed PRS developed by other state-of-the-art methods (P+T, LDpred, PRS-CS) for all six disease/traits with higher AUCs/R^2. It is worth noting that in both Khera's[18] and Said's[19] studies, they used restrictive PRS which could be regarded as a special case of P+T PRS. They generated the PRS based on empirical clumping and thresholding parameters without tuning; hence the performance of the PRS they used was expected to be even worse than the P+T PRS as we showed in Figure 1. Given the significant improvement of the prediction performance of AnnoPred PRS over P+T PRS, our investigation results based on the enhanced PRS would be more accurate and robust.

Second, we studied both multiplicative (HR) and additive (AR) interactions between PRS and lifestyle for CAD, AF, and T2D with a more refined population stratification. We observed that participants with higher PRS could get similar relative but greater absolute benefits by leading a healthy lifestyle. Previous work by Khera *et al*[18] also found that people with high PRS (> 80% PRS) could get larger ARR from lifestyle adherence compared to other PRS groups. However, based on their restricted PRS and coarse population stratification, they reported that of the 7814 participants in the Atherosclerosis Risk in Communities (ARIC) study, the ARR of CAD from healthy lifestyle was 2.7%, 2.5%, and 5.6% in low (< 20% PRS), intermediate (20%-80% PRS), and high PRS (> 80% PRS) groups, respectively; from which they could hardly conclude whether the large ARR only existed in the high PRS group, or there was a trend that higher PRS would lead to larger ARR. Our finer stratification allowed us to observe such a trend clearly as illustrated in Figures 2-3, where the HR lines were parallel to each other and ARRs increased along with PRS for each of the three diseases. These results also provided the insight that PRS was not only valuable to identify the high PRS group, but also informative for further

stratification among participants with low to intermediate PRS and among participants within the commonly categorized high PRS group (> 80% PRS).

In this study we also examined the characteristics of the extreme tails of the PRS distributions (PRS > 99%). We observed a sharply increased risk in this extremely high PRS group that deviates greatly from the rest of the population, which was consistent with previous reports[15,25]. Interestingly, the trend that higher PRS could lead to larger ARR from lifestyle adherence still held for this extremely high PRS group. And strikingly, based on our enhanced PRS, we were able to identify 1% (PRS > 99%) of the population that could get 4.5%, 7.2%, and 12.4% ARR from lifestyle adherence for CAD, AF, and T2D, respectively.

Besides, in addition to the traits analyzed by Khera *et al* and Said *et al*[18,19], we also considered intermediate traits, namely blood lipid levels (LDL-C, TG, TC). Although blood lipid levels have been implicated in CVD risk for some time[30], to our knowledge, our current study is the first to investigate the interactions between genetic risk and combined lifestyle on blood lipid levels through PRS. Consistent with previous studies[14,31], poorer lifestyle and higher PRS could lead to higher levels of LDL-C, TG, and TC. Among three types of lipids, being the only one of the diagnosis items of metabolic syndrome[32], TG appeared to be most responsive to lifestyle modification. This was also in line with the ACC/AHA Guideline where the main target of lifestyle therapies was metabolic syndrome[32]. And among the four individual lifestyle components, BMI played the most prominent role, as long recognized[33]. Additionally, we observed that people in all groups, either with higher or lower PRS, could benefit to an extent from lifestyle adherence in terms of lipids reduction. More specifically, we identified a positive significant interaction between continuous PRS and lifestyle for TG (Table IX in Data Supplement), suggesting people with higher PRS could benefit more from lifestyle modification

in terms of TG reduction. While the effect of lifestyle factors on TG is well-established and unsurprising[32], it is not expected that the effect is dependent on genetic factors and that this is not present for cholesterol (which is also influenced by genetic and lifestyle factors[32]). Furthermore, we also observed that the genetic burden on these lipid levels could not be completely overcome by lifestyle modification. For the group with extremely high PRS (> 99%), even with an ideal lifestyle, the mean levels of TC and LDL-C were still within an undesired range with high risk, and the mean level of TG was also at a broadline high designation. These findings suggest that lifestyle modification is likely to be adequate for people with low-middle PRS, especially for the management of TG; but more interventions (e.g. frequent surveillance, pharmaceutical interventions and more intense lifestyle interventions) are required for people with high PRS to manage LDL-C and TC. This also adds the justification that for high risk individuals, the strategy combining lifestyle and lipid lowering drug treatment since the start may be superior to the strategy stepping from lifestyle to drug therapy when the former fails[34,35].

We note several limitations of this study. First, although we have used enhanced PRSs with better prediction performance than other PRSs, the prediction capacities of PRSs are still moderate. The AUCs of using PRS alone in our study for three diseases ranged from 0.63 to 0.65, which were much lower than AUCs (0.75-0.80) of using comprehensive clinical prediction models as reported by previous studies[36–38]. And the lipids PRS could only explain < 10% variation of lipid levels, which were also relatively poor. Further development of risk prediction models incorporating other predictors could improve the performance and making these models relevant for clinical studies. Second, a causal relationship cannot be inferred between lifestyle and cardiovascular phenotypes given this study design[39], especially in the analysis of blood lipid, where the chronological order of lifestyle status and lipid levels was unknown. Third, since

large-scale GWAS summary statistics for stroke and heart failure independent of UKBB were unavailable, we only considered CAD and AF within the list of CVDs. Thus, future research is needed to investigate the interactions between PRS and lifestyle for all-cause CVD. Another limitation is that we used self-reported characteristics for lifestyle factors such as physical activity, smoking and diet status, which might be inaccurate and reduced the power of our study. In addition, although we analyzed the association of diseases/traits with lifestyle in PRS groups stratified by sex and found the results were generally similar to the combined analysis, PRS constructed based on sex-specific GWAS are required to further investigate possible sex-differences in the interactions between genetic risk and lifestyle factors[40–42]. Finally, the present analyses were performed only on individuals of white British descent, and the UKBB participants were reported to be possibly healthier than the general population[43], which together would decrease the generalizability of our results to other study populations.

In conclusion, genetic risk and combined lifestyle are independently associated with the risks of CAD, AF, and T2D with regard to the log scale of HR. However, individuals at high genetic risk could derive greater benefit from lifestyle adherence in terms of the management of lipid levels and diseases absolute risk reductions.

Acknowledgments: We conducted the research using the UK Biobank resource under an approved data request (ref: 29900). We sincerely thank many GWAS consortia for making their GWAS summary data publicly accessible (http://www.kp4cd.org/datasets/mi).

Sources of Funding: This study was supported in part by the Yale World Scholars Program sponsored by the China Scholarship Council, NIH grants R01 GM122078 and R01 GM134005, and National Science Foundation (NSF) grants DMS 1713120 and DMS 1902903.

Disclosures: None

DOI: 10.1161/CIRCGEN.120.003128

References:

1. Huxley RR, Misialek JR, Agarwal SK, Loehr LR, Soliman EZ, Chen LY, Alonso A. Physical activity, obesity, weight change, and risk of atrial fibrillation: the Atherosclerosis Risk in Communities study. *Circ Arrhythm Electrophysiol*. 2014;7:620–625.

2. Mozaffarian D. Dietary and Policy Priorities for Cardiovascular Disease, Diabetes, and Obesity: A Comprehensive Review. *Circulation*. 2016;133:187–225.

3. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Chang AR, Cheng S, Das SR, et al, American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association. *Circulation*. 2019;139:e56–e528.

4. Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, Himmelfarb CD, Khera A, Lloyd-Jones D, McEvoy JW, et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation*. 2019;140:e596–e646.

5. Kathiresan S, Srivastava D. Genetics of human cardiovascular disease. *Cell*. 2012;148:1242–1257.

6. Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat Rev Genet*. 2017;18:331–344.

7. Levin MG, Rader DJ. Polygenic Risk Scores and Coronary Artery Disease: Ready for Prime Time? *Circulation*. 2020;141:637–640.

8. Consortium TC, the CARDIoGRAMplusC4D Consortium. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 2015;47:1121–1130.

9. Scott RA, Scott LJ, Mägi R, Marullo L, Gaulton KJ, Kaakinen M, Pervjakova N, Pers TH, Johnson AD, Eicher JD, et al. An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes*. 2017;66:2888–2902.

10. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, Beckmann JS, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45:1274–1283.

11. Christophersen IE, Rienstra M, Roselli C, Yin X, Geelhoed B, Barnard J, Lin H, Arking DE, Smith AV, Albert CM, et al. Erratum: Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nat Genet*. 2017;49:1286.

12. 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;4:7.

13. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, Genovese G, Loh P-R, Bhatia G, Do R, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am J Hum Genet*. 2015;97:576–592.

14. Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun*. 2019;10:1776.

15. Hu Y, Lu Q, Powles R, Yao X, Yang C, Fang F, Xu X, Zhao H. Leveraging functional annotations in genetic risk prediction for human complex diseases. *PLoS Comput Biol*. 2017;13:e1005589.

16. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet*. 2013;9:e1003348.

17. Inouye M, Abraham G, Nelson CP, Wood AM, Sweeting MJ, Dudbridge F, Lai FY, Kaptoge S, Brozynska M, Wang T, et al, UK Biobank CardioMetabolic Consortium CHD Working Group. Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. *J Am Coll Cardiol*. 2018;72:1883–1893.

18. Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, Chasman DI, Baber U, Mehran R, Rader DJ, et al. Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. *N Engl J Med*. 2016;375:2349–2358.

19. Said MA, Verweij N, van der Harst P. Associations of Combined Genetic and Lifestyle Risks With Incident Cardiovascular Disease and Diabetes in the UK Biobank Study. *JAMA Cardiol*. 2018;3:693–702.

20. Hindy G, Wiberg F, Almgren P, Melander O, Orho-Melander M. Polygenic Risk Score for Coronary Heart Disease Modifies the Elevated Risk by Cigarette Smoking for Disease Incidence. *Circ Genom Precis Med*. 2018;11:e001856.

21. Merino J, Jablonski KA, Mercader JM, Kahn SE, Chen L, Harden M, Delahanty LM, Araneta MRG, Walford GA, Jacobs SBR, et al. Interaction Between Type 2 Diabetes Prevention Strategies and Genetic Determinants of Coronary Artery Disease on Cardiometabolic Risk Factors. *Diabetes*. 2020;69:112–120.

22. Khera AV, Chaffin M, Wade KH, Zahid S, Brancale J, Xia R, Distefano M, Senol-Cosar O, Haas ME, Bick A, et al. Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. *Cell*. 2019;177:587-596.e9.

23. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, Tyrer JP, Chen T-H, Wang Q, Bolla MK, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet*. 2019;104:21–34.

24. Natarajan P, Young R, Stitziel NO, Padmanabhan S, Baber U, Mehran R, Sartori S, Fuster V, Reilly DF, Butterworth A, et al. Polygenic Risk Score Identifies Subgroup With Higher Burden of Atherosclerosis and Greater Relative Benefit From Statin Therapy in the Primary Prevention Setting. *Circulation*. 2017;135:2091–2101.

25. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50:1219–1224.

26. Baigent C, Keech A, Kearney PM, Blackwell L, Buck G, Pollicino C, Kirby A, Sourjina T, Peto R, Collins R, et al. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. *Lancet*. 2005;366:1267–1278.

27. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12:e1001779.

28. Program NCE. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. *Circulation*. 2002;106:3143–3421.

29. Grundy SM, Cleeman JI, Merz CNB, Brewer HB Jr, Clark LT, Hunninghake DB, Pasternak RC, Smith SC Jr, Stone NJ, National Heart, Lung, and Blood Institute, et al. Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III guidelines. *Circulation*. 2004;110:227–239.

30. Shepherd J, Cobbe SM, Ford I, Isles CG, Lorimer AR, MacFarlane PW, McKillop JH, Packard CJ. Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. West of Scotland Coronary Prevention Study Group. *N Engl J Med*. 1995;333:1301–1307.

31. Mannu GS, Zaman MJS, Gupta A, Rehman HU, Myint PK. Evidence of lifestyle modification in the management of hypercholesterolemia. *Curr Cardiol Rev*. 2013;9:2–14.

32. Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, Braun LT, de Ferranti S, Faiella-Tommasino J, Forman DE, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: A report of the American college of cardiology/American heart association task force on clinical practice guidelines. *Circulation.* 2019;139:e1082-e1143.

33. Keys A. Effects of different dietary fats on plasma-lipid levels. *Lancet*. 1965;285:318–319.

34. Bonaccio M, Di Castelnuovo A, Costanzo S, Persichillo M, De Curtis A, Cerletti C, Donati MB, de Gaetano G, Iacoviello L, Moli-sani Study Investigators. Interaction between Mediterranean diet and statins on mortality risk in patients with cardiovascular disease: Findings from the Moli-sani Study. *Int J Cardiol*. 2019;276:248–254.

35. Stone NJ, Robinson JG, Lichtenstein AH, Bairey Merz CN, Blum CB, Eckel RH, Goldberg AC, Gordon D, Levy D, Lloyd-Jones DM, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation.* 2014;129(25 Suppl 2):S1-45.

36. Elliott J, Bodinier B, Bond TA, Chadeau-Hyam M, Evangelou E, Moons KGM, Dehghan A, Muller DC, Elliott P, Tzoulaki I. Predictive accuracy of a polygenic risk score-enhanced prediction model vs a clinical risk score for coronary artery disease. *JAMA*. 2020;323:636–645.

37. Alonso A, Krijthe BP, Aspelund T, Stepas KA, Pencina MJ, Moser CB, Sinner MF, Sotoodehnia N, Fontes JD, Janssens ACJW, et al. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. *J Am Heart Assoc*. 2013;2:e000102.

38. Wilkinson L, Yi N, Mehta T, Judd S, Garvey WT. Development and validation of a model for predicting incident type 2 diabetes using quantitative clinical data and a Bayesian logistic model: A nationwide cohort and modeling study. *PLoS Med*. 2020;17:e1003232.

39. Vandenbroucke JP, Broadbent A, Pearce N. Causality and causal inference in epidemiology: the need for a pluralistic approach. *Int J Epidemiol*. 2016;45:1776–1786.

40. Zeng Y, Nie C, Min J, Chen H, Liu X, Ye R, Chen Z, Bai C, Xie E, Yin Z, et al. Sex Differences in Genetic Associations With Longevity. *JAMA Netw Open*. 2018;1:e181670.

41. Winham SJ, de Andrade M, Miller VM. Genetics of cardiovascular disease: Importance of sex and ethnicity. *Atherosclerosis*. 2015;241:219–228.

42. Hartiala JA, Tang WHW, Wang Z, Crow AL, Stewart AFR, Roberts R, McPherson R, Erdmann J, Willenborg C, Hazen SL, et al. Genome-wide association study and targeted metabolomics identifies sex-specific association of CPS1 with coronary artery disease. *Nat Commun*. 2016;7:10558.

43. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, Collins R, Allen NE. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol*. 2017;186:1026–1034.

DOI: 10.1161/CIRCGEN.120.003128

Table 1: Baseline characteristics of the testing set

Characteristics	Testing Set (N = 183,168)
Age at recruitment, mean (SD), years	56.81 (8.00)
Number of females (%)	98,611 (53.84)
Years in education, mean (SD), years	15.08 (5.05)
Townsend deprivation index at recruitment, mean (SD)	-1.63 (2.90)
Annual household Income, No. (%), £	
<18,000	32,481 (17.73)
18,000 – 30,999	40,213 (21.95)
31,000 – 51,999	41,952 (22.90)
52,000 – 100,000	34,158 (18.65)
>100,000	9,223 (5.04)
Unknown	25,141 (13.73)
Smoking, No. (%)	
Ideal (Never)	101,012 (55.14)
Intermediate (Former)	65,762 (35.90)
Poor (Current)	16,394 (8.95)
Body mass index, No. (%)	
Mean (SD)	27.30 (4.71)
Ideal (< 25 kg/m^2 & ≥ 18.5 kg/m^2)	60,913 (33.26)
Intermediate (< 30 kg/m^2 & ≥ 25 kg/m^2)	77,966 (42.57)
Poor (≥ 30 kg/m^2)	42,802 (23.37)
Exclusion (missing or < 18.5 kg/m^2)	1,487 (0.81)
Physical activity, No. (%)	
Ideal (regular physical activity)	89,309 (48.76)
Intermediate (some physical activity)	53,613 (29.27)
Poor (limited physical activity)	40,246 (21.97)
Diet, No. (%)	
Ideal (adequate intake of >5 dietary components)	27,643 (15.09)
Poor (inadequate intake of >5 dietary components)	155,525 (84.91)
Combined lifestyles, No. (%)	
Ideal (≥ 3 lifestyle factors in Ideal status)	29,080 (15.88)
Intermediate (2 lifestyle factors in Ideal status)	138,410 (75.56)
Poor (≤ 1 lifestyle factors in Ideal status)	15,678 (8.56)

DOI: 10.1161/CIRCGEN.120.003128

Figure Legends:

Figure 1. Performance of polygenic risk scores (PRS) by different methods. Candidates PRS were generated using four PRS methods (P+T, LDpred, PRS-CS, and AnnoPred). Tuning parameters of each method were selected in the training set, and the predictive performance using the optimal tuning parameters was then assessed in the testing set. The prediction accuracy was measured by area under the receiver operator curve (AUC) and was provided with 95% CI.

Figure 2. Relative risk of coronary artery disease, atrial fibrillation, and type 2 diabetes stratified by the combination of genetic and lifestyle factors. We partitioned the testing set into 30 groups according to their PRS percentile (10 genetic risk bins) and lifestyle status (three lifestyle bins). The hazard ratios (HRs) were calculated by comparing each group to the group with 40%-60% PRS and intermediate lifestyle. All HRs were adjusted by age, sex and first four genetic principal components and were provided with their corresponding 95% CI. Y-axis was on log-scale.

Figure 3. Incident events of coronary artery disease, atrial fibrillation, and type 2 diabetes stratified by the combination of genetic and lifestyle factors. We partitioned the testing set into 21 groups according to their PRS percentile (7 genetic risk bins) and lifestyle status (three lifestyle bins). The absolute risk in each group was calculated as the incident rate of each disease in the group, and the absolute risk reduction (ARR) reflected the reduction of absolute risk when changing the lifestyle status from poor to ideal within the same PRS group.

DOI: 10.1161/CIRCGEN.120.003128

Figure 4. Lipid levels stratified by the combination of genetic and lifestyle factors. We partitioned the testing set into 30 groups according to their PRS percentile (10 genetic risk bins) and lifestyle status (three lifestyle bins). The mean level of lipid in each group was provided with its associated standard error (SD). Different background color indicated different designation according to the recommendation by the National Cholesterol Education Program (NCEP). Green, yellow and red indicated normal, border high, and high designation, respectively.