Polygenic risk, lifestyle, and cardiovascular mortality: a prospective population-based UK Biobank study

Jae-Seung Yun, M.D., Ph.D.^{1,2}, Sang-Hyuk Jung, B.E.^{1,3,4}, Manu Shivakumar, M.S.^{1,4}, Brenda Xiao, B.E.^{1,5}, Amit V. Khera, M.D., M.Sc.⁶, Woong-Yang Park, M.D., Ph.D.⁷, Hong-Hee Won, Ph.D.^{3,7*}, Dokyoon Kim, Ph.D.^{1,4*}

¹Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

²Division of Endocrinology and Metabolism, Department of Internal Medicine, St. Vincent's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

³Samsung Advanced Institute for Health Sciences and Technology (SAIHST), Sungkyunkwan University, Samsung Medical Center, Seoul, Republic of Korea

⁴Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA, USA

⁵Genomics and Computational Biology Graduate Group, University of Pennsylvania, Philadelphia, PA, USA

⁶Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

⁷Samsung Genome Institute, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

[§]Jae-Seung Yun and Sang-Hyuk Jung contributed equally as co-first authors.

^{*}Hong-Hee Won and Dokyoon Kim contributed equally as co-corresponding authors.

Correspondence to:

Hong-Hee Won

Samsung Advanced Institute for Health Sciences and Technology (SAIHST), Sungkyunkwan University, Samsung Medical Center, 81 Irwon-ro, Gangnam-gu, Seoul 06351, South Korea.

Phone: +82-2-2148-7566; Fax: +82-2-3410-0534; E-mail: wonhh@skku.edu

Dokyoon Kim

B304 Richards Building, 3700 Hamilton Walk, University of Pennsylvania, Philadelphia, PA 19104-6021

Phone: +1-215-573-5336; Fax: +1-215-573-3111;

E-mail: dokyoon.kim@pennmedicine.upenn.edu

Summary Box

What is already known on this topic

Polygenic risk scores quantify the inherited risk conferred by the cumulative impact of

common variants into a quantitative risk estimate.

Previous studies primarily targeted the ability of polygenic risk scores to predict a specific

disease, and only a few studies have investigated the association between genetic risk scores

and cardiovascular mortality.

The majority of previous analyses calculated polygenic risk scores from only a small number

of genetic variants or adjusted for only a few risk factors, and no studies have examined

whether the association of polygenic risk score with cardiovascular mortality differs by

lifestyle behavior.

What this study adds

Genetic risk and lifestyle are independent predictive factors for cardiovascular mortality, even

after adjustment for a comprehensive range of demographic and clinical factors.

A healthy lifestyle is associated with relative risk reduction for cardiovascular mortality

across all genetic risk categories, a finding that indicates the potential benefit of intensive

3

lifestyle modification in overcoming genetic risk for cardiovascular mortality.

Abstract

OBJECTIVE

To assess the prognostic ability of polygenic risk scores (PRSs) for coronary artery disease

(CAD) and type 2 diabetes mellitus (T2DM) for cardiovascular (CV) mortality, independent

of traditional risk factors, and further investigate the additive effect between lifestyle

behavior and PRS on CV mortality.

DESIGN

Prospective population-based cohort study.

SETTING

UK Biobank.

PARTICIPANTS

A total 377,909 unrelated participants of white British descent were included in the analyses

from the UK Biobank cohort.

MAIN OUTCOME MEASURES

Genome-wide PRSs were constructed using >6 million genetic variants. We stratified patients

into four PRS risk groups: low (0 to 19^{th} percentile), intermediate (20 to 79^{th} percentile), high

(80 to 98th percentile), and very high (99th percentile). We defined a favorable and

unfavorable lifestyle with four modifiable lifestyle components, including smoking, obesity,

physical activity, and diet. Cox proportional hazard models were used to analyze the

relationship between PRS and CV mortality with stratification by age, sex, disease status, and

4

lifestyle behavior.

RESULTS

Of 377,909 UK Biobank participants having European ancestry, 3,210 (0.8%) died due to CV disease during a median follow-up of 8.9 years. CV mortality risk was significantly associated with CAD PRS (low vs. very high genetic risk groups, CAD PRS hazard ratio [HR] 2.61 [2.02 to 3.36]) and T2DM PRS (HR 2.08 [1.58 to 2.73]), respectively. These relationships remained significant even after adjustment for a comprehensive range of demographic and clinical factors. In the very high genetic risk group, adherence to an unfavorable lifestyle was further associated with a substantially increased risk of CV mortality (favorable versus unfavorable lifestyle with very high genetic risk for CAD PRS, HR 8.31 [5.12 to 13.49]; T2DM PRS, HR 5.84 [3.39 to 10.04]). Across all genetic risk groups, 32.1% of CV mortality was attributable to lifestyle behavior (population attributable fraction [PAF] 32.1% [95% CI 28.8 to 35.3%]) and 14.1% was attributable to smoking (PAF 14.1% [95% CI 12.4 to 15.7%]). There was no evidence of significant interaction between PRSs and age, sex, or lifestyle behavior in predicting the risk of CV mortality.

CONCLUSION

PRSs for CAD or T2DM and lifestyle behaviors are independent predictive factors for future CV mortality in the white, middle-aged population. PRS-based risk assessment could be useful to identify individuals who need intensive behavioral or therapeutic interventions to reduce the risk of CV mortality.

Introduction

Common chronic diseases such as cardiovascular (CV) disease or diabetes represent a huge public health burden. Cardiovascular disease (CVD) is one of the leading causes of mortality worldwide; it has been estimated that CVD accounts for about 30% of all-cause mortality. It is well established that type 2 diabetes mellitus (T2DM) is also a major risk factor for mortality. Moreover, people with T2DM have a higher risk of CV mortality than people without diabetes, and risk of mortality continuously increases as glycemic levels increase, even if levels do not reach those needed for a diabetes diagnosis. Thus, early screening and prevention in individuals at risk of these diseases are important strategies for reducing CV morbidity and mortality.

Traditional risk factors for common chronic diseases do not typically manifest early in life, thus it is difficult to fully identify high-risk individuals. Polygenic risk scores (PRSs) comprised of single nucleotide polymorphisms (SNPs) offer a means for early screening of and preventive interventions in common chronic diseases, including coronary artery disease (CAD) and diabetes.⁴ Many studies have demonstrated the predictive ability of PRSs to identify those with higher genetic risk of incident disease. For instance, Khera et al. suggested that PRSs measuring the cumulative genetic burden of five common diseases are well-correlated with case status.⁵ However, most previous studies have assessed PRS in terms of predicting specific diseases; the utility of PRS in predicting mortality has not yet been evaluated comprehensively.⁵⁻⁷ A few studies have shown associations between PRS and disease-specific mortality but included only a small number of SNPs. ⁸⁻¹¹ Furthermore, a mortality event is a complicated outcome in which a number of potential confounding factors are involved, and the majority of previous analyses did not adjust for the comprehensive

range of demographic, lifestyle, and clinical factors that can affect the risk of mortality. It is

also uncertain if a healthy lifestyle can offset the genetic risk for CV mortality. Thus, it is

important to evaluate the association between CV mortality and a PRS for CAD and T2DM

calculated using millions of SNPs, to test this association for significance after more thorough

adjustments, and to examine any differences among individuals grouped by their age, sex,

disease status, or lifestyle behaviors.

The UK Biobank is a nationwide, prospective, population-based cohort of

participants aged 40 to 69 and provides a variety of genetic, phenotypic, and health-related

information. 12 We primarily aimed to investigate associations between the PRSs for CAD and

T2DM, lifestyle behaviors, and CV mortality risk in middle-aged UK Biobank participants of

European ancestry. We further investigated the association between PRS and CV mortality

according to age, sex, and disease status.

Methods

Study population

The UK Biobank recruited 502,505 participants aged 40 to 69 years between 2007 and

2010. 12 At baseline, participants provided their signed consent and completed a touchscreen

questionnaire, in-person interview, and physical assessment in one of 22 assessment centers

across the UK. The touchscreen questionnaires covered sociodemographic factors, lifestyle

behaviors, and health-related factors such as smoking status, alcohol frequency, eating habits,

and medical history. Anthropometric measurements, including height, weight, waist

circumference, and blood pressure, were measured by trained staff according to a standard

protocol. During the baseline assessment visit, blood samples were collected and processed

by standardized protocols.¹³ Only baseline assessments were used in the present study. The

UK Biobank was approved by the National Research Ethics Committee (June 17, 2011 [RES

reference 11/NW/0382]; extended on May 10, 2016 [RES reference 16/NW/0274]). This

research using the UK Biobank Resource was approved under Application Number 33002.

Genotyping and quality control

UK Biobank samples (version 3; March 2018) were genotyped using either the Affymetrix

UK BiLEVE Axiom array or the Affymetrix UK Biobank Axiom array; these include

>800,000 genotyped SNPs, 95% of which are shared between the two platforms. Imputation

via IMPUTE2 was carried out centrally by UK Biobank researchers using the merged 1000

Genomes Project panel and UK 10K panel. 4 We used genetic data from participants

identified as 'white-British' ancestry on the basis of both self-report and principal component

analysis of ancestry. We excluded individuals whose reported sex did not match with that

inferred from genetic data and individuals with second-degree or closer relatives also in the

Biobank. After exclusions, 377,909 individuals were eligible for the genetic analyses. After

imputation, variant-level quality control (QC) was carried out by filtering SNPs on the basis

of: (1) minor allele frequency <0.01 and (2) imputation quality score (INFO) <0.3. A total of

9,505,768 imputed autosomal SNPs passed the QC criteria.

Polygenic risk scores

To generate individual genetic risk scores, we derived PRSs based on the pre-calculated

weights for SNPs provided by Khera et al., which were determined using LDpred from large-

scale genome-wide association study (GWAS) summary statistics.⁵ PRSs for T2DM and

CAD were respectively constructed based on GWAS summary statistics from the Diabetes

Genetics Replication and Meta-analysis (DIAGRAM) consortium and the Coronary Artery

Disease Genome-wide Replication and Meta-analysis plus the Coronary Artery Disease

Genetics (CARDIOGRAMplusC4D) consortium. 15 16

Ascertainment of mortality outcomes

All participants provided their consent for follow-up linkage to their death registration and

health-related records. Date and underlying cause of death were obtained from death

certificates provided by the National Health Service Central Register (Scotland) and the

National Health Service Information Center (England and Wales). Mortality data for our UK

Biobank dataset were available until November 30, 2016 for centers in Scotland and until

January 31, 2018 for centers in England and Wales. Causes of death were classified using the

10th revision of the International Statistical Classification of Disease (ICD-10).

Ascertainment of variables

Information on smoking status, alcohol frequency, physical activity, eating habits, medical

history, and medication use were collected via a touchscreen questionnaire or in-person

interview during the baseline visit for the UK Biobank project. Height, weight, and waist

circumference were assessed by trained medical staff at the same time. Regular physical

activity was defined as participating in either moderate activity ≥5 days a week or vigorous

activity ≥ 3 days a week. We used four lifestyle factors: current smoking, obesity, physical

activity, and dietary pattern, as recommended by the strategic goals of the American Heart Association (AHA).^{7 17 18} Participants were categorized based on overall lifestyle scores into the following four risk subgroups: favorable (defined as having at least three healthy lifestyle factors), intermediate (having two healthy lifestyle factors), and unfavorable (having one or fewer healthy lifestyle factors). Additional definitions and details regarding lifestyle factors are provided in Supplementary Table 1.

Diagnosis of prevalent CAD or T2DM at baseline was based on self-report in the inperson interview at enrollment or on diagnostic and procedure codes in electronic health
records. Medical history of other major comorbidities was also collected concerning
dyslipidemia, hypertension, heart failure, ischemic stroke, chronic lung disease, chronic
kidney disease, and cancer. We used self-reported diagnoses, hospitalization records, and first
occurrence information to define prevalent comorbidities; these definitions are presented in
Supplementary Table 2. The Charlson comorbidity index (CCI), a composite comorbidity
score, was calculated based on ICD-10 codes and was used to assess the severity of any
underlying comorbidities.¹⁹ Estimated glomerular filtration rate (eGFR) was calculated using
the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation.²⁰

Statistical analysis

Continuous variables are reported as means with standard deviations, and categorical variables as frequencies and proportions. Baseline characteristics were compared between genetic risk groups using the chi-square test for categorical variables and ANOVA for continuous variables. The association between PRS status and mortality was investigated using Cox proportional hazards models. As in previous studies, we categorized study

participants on the basis of PRS into low, intermediate, or high groups ^{7 18} and further classified the top 1% of the PRS distribution as a very high risk group in light of the curve of cumulative incidence of prevalent disease over the PRS distribution (Supplementary Fig 1). Thus, participants were categorized into the following four risk subgroups: low (0 to 19th percentile), intermediate (20 to 79th percentile), high (80 to 98th percentile), and very high (99th percentile). Model 1 was adjusted for age, sex, genotyping array, and the first ten principal components (PCs) of ancestry. Model 2 was fully adjusted for the major variables associated with risk of CV mortality, such as age, sex, genotyping array, the first ten PCs, baseline blood pressure, lifestyle behavior, laboratory findings, baseline use of medications, CCI, and major comorbidities including hypertension, dyslipidemia, ischemic stroke, heart failure, cancer, chronic liver disease, chronic lung disease, and chronic kidney disease. We also calculated a P value for linear trends in regression based on the categorical risk groups. The population attributable fraction (PAF) was calculated to quantify the proportion of CV mortality in a population attributable to lifestyle behaviors. Stratified analyses were conducted using cutoffs and considered lifestyle behaviors, age at enrollment, sex, and the presence of CAD and T2DM at baseline. To investigate whether effect modification of the association between PRS and mortality occurred for lifestyle behaviors, age, sex, or disease status, we assessed multiplicative interactions between PRS and each of the stratification variables. In our survival analyses, individuals were censored according to the date of followup loss, the date of follow-up end (January 31, 2018 for England and Wales; November 30, 2016 for Scotland), or the date of death. Cases with missing data were excluded from modeling (Supplementary Table 3). Log minus log plots and Schoenfeld residuals were used to assess the proportional hazard assumption. All statistical tests were two-sided, and a P<0.05 was defined as statistically significant. All statistical analyses were performed using PLINK 1.9 and R (version 3.9.0).

Patient and public involvement

The development of the research question or outcome measures was not informed by patients'

priorities, experience, or preferences. No patients were involved directly in the setting of the

design or implementation of the study. There are no plans to involve patients in dissemination.

Results

Population characteristics

Baseline demographic and clinical characteristics of all 377,909 participants are presented in

Table 1; characteristics for each PRS group are presented in the Supplementary Table 4. Men

comprised 46.3% of participants, and the mean age at the study baseline was 56.5 years.

Overall, participants at higher genetic risk for CAD or T2DM tended to have higher blood

pressure, BMI, waist circumference, HbA1c, estimated glomerular filtration rate, CCI, and

more frequent use of aspirin, anti-hypertensive agents, or lipid-lowering agents than those at

low genetic risk.

Association of PRS with cardiovascular mortality

During the median follow-up of 8.9 years (IQR 8.3 to 9.5 years), 15,570 participants (4.1%)

died; of those deaths, 21.1% were related to CV. The overall mortality rate was 0.95 per 1000

person-years (95% CI 0.92 to 0.99). Compared with low genetic risk, higher genetic risk was

associated with a higher risk of CV mortality during follow-up (P for trend <0.001) (Figure 1, 12

Supplementary Table 5). In stepwise multivariable models, very high genetic risk for either CAD or T2DM remained significantly associated with a high risk of CV mortality, although the strength of the association was attenuated (Supplementary Table 5). In the fully-adjusted model, participants at very high genetic risk for CAD or T2DM had approximately two-fold increased risk of CV mortality relative to those with the corresponding low genetic risk. Associations between PRSs and mortality were attenuated when adjusting for PRS-specific disease status at baseline (Supplementary Table 5, model 4).

PRS and lifestyle behaviors for cardiovascular mortality

Compared to a favorable lifestyle, an unfavorable lifestyle was associated with an increased risk of CV mortality in all genetic risk groups for CAD and T2DM alike (Figure 2, Supplementary Table 6, 7, and Supplementary Fig 2). In the groups with high and very high genetic risk for CAD, adherence to an unfavorable lifestyle was respectively associated with 4.6- and 8.3-fold increased risk of CV mortality (P value <0.001). Similarly, in the high and very high genetic risk groups for T2DM, an unfavorable lifestyle was associated with 3.9- and 5.8-fold respective increased risk of CV mortality (P value <0.001). Adherence to a favorable lifestyle was associated with reduced risk of CV mortality across all genetic risk categories (unfavorable vs. favorable lifestyle group, HR 0.33 [95% CI 0.30 to 0.36]). However, participants at very high genetic risk with a favorable lifestyle still featured high mortality risk (hazard ratio [HR] 2.82 for CAD and 1.59 for T2DM). The population attributable fraction (PAF) of favorable lifestyle behaviors for CV mortality was 32.1% (95% CI 28.8 to 35.3), and the PAF of smoking was 14.1% (95% CI 12.4 to 15.7%). There were no significant interactions between PRSs and lifestyle behaviors in predicting risk of CV

mortality. Among the four lifestyle habits included in this analysis, smoking with very high

genetic risk for CAD and T2DM had the strongest effect on CV mortality (Supplementary Fig

3, 4).

PRS, age, sex, and disease status for cardiovascular mortality

We investigated the interactions of PRS, age, and sex in the context of mortality risks and

found that the association of mortality risk with PRS varied slightly by age and sex (Figure 3,

Supplementary Table 8, 9, and Supplementary Fig 5). Higher genetic risks of CAD and

T2DM were associated with CV mortality both in the young (<55 years) and the old (≥55

years) and were associated with increased risk of CV mortality in both sexes, with generally

higher risk in men. The CAD PRS was not associated with risk of CV mortality among

participants with CAD at baseline, even in those having very high genetic risk (HR 1.19 [95%

CI 0.63 to 2.24], P=0.60). By contrast, a robust association between very high genetic risk for

T2DM and mortality was observed among participants with CAD or T2DM at baseline (CAD

PRS, HR 2.31 [1.40 to 3.80]; T2DM PRS, HR 2.99 [1.59 to 5.62], P<0.001) (Figure 4,

Supplementary Table 10, and Supplementary Fig 6, 7).

Discussion

In this prospective nationwide cohort study, we comprehensively explored the prognostic

utility of PRS in predicting CV mortality. We found that the genetic risk of CAD and of

T2DM were independent predictive factors for CV mortality. Furthermore, participants at

very high genetic risk for CAD or T2DM, equivalent to 1% of the total population, had a 2.6-

and 2.1-fold increased risk of CV mortality, respectively, relative to those at low genetic risk. These associations were attenuated but still evident even after adjustment for a wide range of mortality-related clinical and lifestyle behavior factors. In all genetic risk groups, adherence to a favorable lifestyle was associated with a reduced risk of CV mortality. In groups with very high genetic risk for CAD or T2DM, CV mortality risk remained substantially high even with a favorable lifestyle.

Previous studies have identified an association between mortality and genetic risk scores incorporating CVD-related variants. Two studies reported that CV mortality is significantly associated with a PRS calculated with a small number of variants. 10 11 Recently, Meisner et al. described a significant association between CAD PRS and CAD-specific mortality using the UK Biobank dataset, ²¹ and Damask et al. likewise reported that those with high genetic risk for CAD had 50% higher risk of major adverse cardiac events.²² Fewer studies have investigated the association of genetic risk of diabetes with mortality. A previous longitudinal cohort study by Leong et al. reported a borderline significant association between genetic risk for T2DM and all-cause mortality. Others have suggested that genetic risk scores associated with hyperglycemia or diabetes could not predict mortality or age of death.²¹ In the context of CVD, diabetes and hyperglycemia are well-known major risk factors for CVD and related mortality,²³ but controversy exists regarding the effect of genetic variants associated with diabetes risk on CV mortality. Previous studies have suggested a shared genetic basis between T2DM and CAD with a complex bidirectional relationship, ²⁴ and several Mendelian randomization studies supported the hypothesis that genetic mechanisms linked with risk of T2DM may have causal roles in the etiology of CVD. 25 26 Our results expand on previous research by demonstrating the significant and robust association of genetic risk of T2DM as well as CAD with mortality using a large dataset comprised of UK Biobank participants. In the model fully adjusted for established clinical CV risk factors, the HRs of PRSs were attenuated, but both genetic scores remained significant independent predictors of future CV mortality. Our finding implies that although the traditional clinical risk factors included as covariates in this study influence CV mortality, there are multiple additional pathways related to PRS that affect CV risk and mortality, such as subclinical atherosclerosis, subclinical metabolic derangement, and clinical events related to CV mortality during follow-up periods.^{27 28}

In our study, we found that genetic risk and modifiable lifestyle habits were independently associated with risk of CV mortality. To the best of our knowledge, this is the first study to investigate the associations of lifestyle habits and genetic risk with CV mortality. Previous studies have mainly targeted the interaction between genetic risk and lifestyle habits for specific diseases. For example, Khera et al. suggested that both genetic factors and adherence to an unfavorable lifestyle conjointly contributed to an increased risk of CAD. Other studies reported that combined unfavorable lifestyle behaviors and genetic risk had an additive deleterious effect on the risk of developing CAD, stroke, and T2DM. The analyses in the current study showed that the risk of CV mortality was reduced by adherence to a favorable lifestyle, even in high genetic risk groups. These findings indicate the potential benefits of lifestyle modifications such as diet, weight control, regular physical activity, and smoking cessation in preventing CV mortality, regardless of genetic risk. Our analysis showed that over 30% of CV mortality might have been reduced if all participants would have adhered to favorable lifestyle behavior. In particular, patients could adjust their smoking habits, as current smoking was the most deleterious risk factor for CV mortality out of the

behaviors we analyzed. The effect of lifestyle intervention in individuals at high genetic risk for CV mortality should be verified in future studies.

Several studies have reported that the effect of genetic risk on disease onset or mortality varies according to age and sex. For example, Tada et al. reported that PRS risk for CVD was higher in younger individuals than in older ones¹¹, and Mars et al. suggested that higher genetic risk scores were associated with early onset of diseases including CAD and T2DM.⁶ An early onset of CVD in a high genetic risk group can be consistent with an increased risk of CV mortality in a young age group. Regarding risk and sex, while men have a higher rate of CVD incidence and mortality in the general population, the risks for CVD and its mortality have been shown to be greater in women with diabetes than in men with diabetes.^{29 30} However, we observed no significant interactions between PRS, age, and sex with regard to CV mortality. Our results indicated that the effects of PRSs on CV mortality were not stronger in women than in men, even among participants with diabetes.

This study also found that, among participants with T2DM at baseline, PRSs for T2DM and CAD were significantly associated with increased risk of CV mortality. Consistent with our findings, Cox et al. showed that a genetic risk score based on SNPs associated with CVD in patients with T2DM was associated with all-cause and CV mortality. Meanwhile, among individuals in our study with CAD at baseline, PRS for CAD was not associated with CV mortality. This finding is in line with previous results that PRSs determined using SNPs associated with CAD are not associated with recurrent CV events. Individuals with a medical history of CAD used more aggressive medications such as aspirin and lipid-lowering agents for secondary prevention; such post-event care efforts might reduce the impact of factors mediated by genetic risk for CAD on CV mortality. A previous post-hoc

analysis of clinical trials has shown that high genetic risk of CAD can be mitigated by statin treatment.³²

Several limitations of our study should be considered. First, our study could be affected by competing risk from other cause-specific mortalities, such as cancer. To account for the potential effect of competing risk of death from other causes, we adjusted for the presence at baseline of major comorbidities such as cancer, chronic lung disease, and chronic liver disease. Second, our definition of comorbidity included self-reported physician-made diagnosis, which may be incomplete and fail to include all true patients with comorbidities. However, self-reported disease status or history was obtained through a verbal interview with a trained nurse and was found to be the strongest predictor of all-cause mortality in men in a previous study.³³ Third, lifestyle behaviors were based on a single measurement at baseline and were nonrandomized. Fourth, the predictive power of PRS can be further refined by using summary statistics from larger GWASs, which may change the estimated risk of CV mortality in individuals at high genetic risk. Finally, the association between genetic risk and CV mortality was not validated in external cohorts nor in non-European populations. Further studies with other cohorts are warranted to verify the generalizability of our findings.

In summary, our study used a large, well-phenotyped, prospective cohort drawn from the UK Biobank to assess the association between PRSs and CV mortality. We found that PRSs for CAD and T2DM are independent predictive factors for future CV mortality for this white, middle-aged cohort. We were able to adjust for various demographic, clinical, and lifestyle factors to control for potential confounding factors for mortality. The associations remained robust after adjustment for a wide range of clinical variables, including traditional CV risk factors. PRS thus has prognostic clinical utility in identifying people at risk for CVD

and mortality. Furthermore, due to the considerable number of participants, we were also able

to perform analyses stratified by lifestyle risk. Our findings suggest that PRS-based risk

assessment could be useful among individuals with high genetic risk for CAD or T2DM who

need intensive behavioral or therapeutic interventions to prevent CV mortality. PRS can

stratify high risk groups associated with an increased risk of CV death as well as specific

diseases at an early age or early stage of disease. This stratification could be used for early

clinical support and lifelong intervention and thereby enable precision medicine. Further

studies are needed to verify the efficacy of intervention or cost-effectiveness of risk

assessment based on PRS.

Acknowledgments

This work was supported by a National Research Foundation of Korea (NRF) grant funded

by the Korea government (MSIT) (No. 2019R1A2C4070496) and by the National Institute of

General Medical Sciences (NIGMS) R01 GM138597.

Author Contributions

Jae-Seung Yun developed the study design, wrote the manuscript, and analyzed and

interpreted

data. Sang-Hyuk Jung analyzed data, interpreted data, and contributed to discussion. Manu

Shivakumar analyzed data and revised the manuscript. Brenda Xiao and Woong-Yang Park

revised the manuscript. Amit V. Khera, Hong-Hee Won, and Dokyoon Kim interpreted results,

19

contributed to discussion, and revised the manuscript.

Data sharing

The UK Biobank dataset was obtained from the UK Biobank (Application Number 33002), and a full list of the variables are available online. Data cannot be shared publicly due to the violation of patient privacy and the absence of informed consent for data sharing.

Declaration of Interests

Woong-Yang Park is employed by a commercial company, GENINUS.

References

- Roger VL, Go AS, Lloyd-Jones DM, et al. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. *Circulation* 2012;125(1):e2-e220. doi: 10.1161/CIR.0b013e31823ac046 [published Online First: 2011/12/20]
- 2. Haffner SM, Lehto S, Rönnemaa T, et al. Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. N Engl J Med 1998;339(4):229-34. doi: 10.1056/nejm199807233390404 [published Online First: 1998/07/23]
- Vistisen D, Witte DR, Brunner EJ, et al. Risk of Cardiovascular Disease and Death in Individuals With Prediabetes Defined by Different Criteria: The Whitehall II Study. Diabetes Care 2018;41(4):899-906. doi: 10.2337/dc17-2530 [published Online First: 2018/02/18]

- 4. Han X, Wei Y, Hu H, et al. Genetic Risk, a Healthy Lifestyle, and Type 2 Diabetes: the Dongfeng-Tongji Cohort Study. *J Clin Endocrinol Metab* 2020;105(4) doi: 10.1210/clinem/dgz325 [published Online First: 2020/01/05]
- Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* 2018;50(9):1219-24. doi: 10.1038/s41588-018-0183-z [published Online First: 2018/08/15]
- 6. Mars N, Koskela JT, Ripatti P, et al. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat Med* 2020;26(4):549-57. doi: 10.1038/s41591-020-0800-0 [published Online First: 2020/04/11]
- 7. 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(8):693-702. doi: 10.1001/jamacardio.2018.1717 [published Online First: 2018/06/30]
- 8. Cox AJ, Hsu FC, Ng MC, et al. Genetic risk score associations with cardiovascular disease and mortality in the Diabetes Heart Study. *Diabetes Care* 2014;37(4):1157-64. doi: 10.2337/dc13-1514 [published Online First: 2014/02/28]
- 9. Leong A, Porneala B, Dupuis J, et al. Type 2 Diabetes Genetic Predisposition, Obesity, and All-Cause Mortality Risk in the U.S.: A Multiethnic Analysis. *Diabetes Care* 2016;39(4):539-46. doi: 10.2337/dc15-2080 [published Online First: 2016/02/18]
- 10. Pereira A, Mendonca MI, Sousa AC, et al. Genetic risk score and cardiovascular mortality in a southern european population with coronary artery disease. *Int J Clin Pract* 2017;71(6) doi: 10.1111/ijcp.12956 [published Online First: 2017/05/16]

- 11. Tada H, Melander O, Louie JZ, et al. Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. *Eur Heart J* 2016;37(6):561-7. doi: 10.1093/eurheartj/ehv462 [published Online First: 2015/09/24]
- 12. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018;562(7726):203-09. doi: 10.1038/s41586-018-0579-z [published Online First: 2018/10/12]
- 13. Elliott P, Peakman TC. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int J Epidemiol* 2008;37(2):234-44. doi: 10.1093/ije/dym276 [published Online First: 2008/04/03]
- 14. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;5(6):e1000529. doi: 10.1371/journal.pgen.1000529 [published Online First: 2009/06/23]
- 15. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* 2015;47(10):1121-30. doi: 10.1038/ng.3396 [published Online First: 2015/09/08]
- 16. Scott RA, Scott LJ, Mägi R, et al. An Expanded Genome-Wide Association Study of Type
 2 Diabetes in Europeans. *Diabetes* 2017;66(11):2888-902. doi: 10.2337/db16-1253
 [published Online First: 2017/06/02]
- 17. Lloyd-Jones DM, Hong Y, Labarthe D, et al. Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic Impact Goal through 2020 and beyond. *Circulation* 2010;121(4):586-613. doi: 10.1161/circulationaha.109.192703 [published Online First: 2010/01/22]

- 18. Khera AV, Emdin CA, Drake I, et al. Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. N Engl J Med 2016;375(24):2349-58. doi: 10.1056/NEJMoa1605086 [published Online First: 2016/12/14]
- 19. Brusselaers N, Lagergren J. The Charlson Comorbidity Index in Registry-based Research.
 Methods Inf Med 2017;56(5):401-06. doi: 10.3414/me17-01-0051 [published Online
 First: 2017/01/01]
- Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med* 2009;150(9):604-12. doi: 10.7326/0003-4819-150-9-200905050-00006 [published Online First: 2009/05/06]
- 21. Meisner A, Kundu P, Zhang YD, et al. Combined Utility of 25 Disease and Risk Factor Polygenic Risk Scores for Stratifying Risk of All-Cause Mortality. Am J Hum Genet 2020;107(3):418-31. doi: 10.1016/j.ajhg.2020.07.002 [published Online First: 2020/08/08]
- 22. Damask A, Steg PG, Schwartz GG, et al. Patients With High Genome-Wide Polygenic Risk Scores for Coronary Artery Disease May Receive Greater Clinical Benefit From Alirocumab Treatment in the ODYSSEY OUTCOMES Trial. *Circulation* 2020;141(8):624-36. doi: 10.1161/circulationaha.119.044434 [published Online First: 2019/11/12]
- 23. Tancredi M, Rosengren A, Svensson AM, et al. Excess Mortality among Persons with Type 2 Diabetes. *N Engl J Med* 2015;373(18):1720-32. doi: 10.1056/NEJMoa1504347 [published Online First: 2015/10/29]
- 24. Zheng Q, Jiang J, Huo Y, et al. Genetic predisposition to type 2 diabetes is associated with severity of coronary artery disease in patients with acute coronary syndromes.
 Cardiovasc Diabetol 2019;18(1):131. doi: 10.1186/s12933-019-0930-1 [published]

Online First: 2019/10/09]

- 25. Ross S, Gerstein HC, Eikelboom J, et al. Mendelian randomization analysis supports the causal role of dysglycaemia and diabetes in the risk of coronary artery disease. *Eur Heart J* 2015;36(23):1454-62. doi: 10.1093/eurheartj/ehv083 [published Online First: 2015/04/01]
- 26. Li Y, Pan A, Wang DD, et al. Impact of Healthy Lifestyle Factors on Life Expectancies in the US Population. *Circulation* 2018;138(4):345-55. doi: 10.1161/circulationaha.117.032047 [published Online First: 2018/05/02]
- 27. Gan W, Bragg F, Walters RG, et al. Genetic Predisposition to Type 2 Diabetes and Risk of Subclinical Atherosclerosis and Cardiovascular Diseases Among 160,000 Chinese Adults. *Diabetes* 2019;68(11):2155-64. doi: 10.2337/db19-0224 [published Online First: 2019/08/11]
- 28. Meigs JB. The Genetic Epidemiology of Type 2 Diabetes: Opportunities for Health Translation. *Curr Diab Rep* 2019;19(8):62. doi: 10.1007/s11892-019-1173-y [published Online First: 2019/07/25]
- 29. Rutten-Jacobs LC, Larsson SC, Malik R, et al. Genetic risk, incident stroke, and the benefits of adhering to a healthy lifestyle: cohort study of 306□473 UK Biobank participants. *Bmj* 2018;363:k4168. doi: 10.1136/bmj.k4168 [published Online First: 2018/10/26]
- 30. Rao Kondapally Seshasai S, Kaptoge S, Thompson A, et al. Diabetes mellitus, fasting glucose, and risk of cause-specific death. N Engl J Med 2011;364(9):829-41. doi: 10.1056/NEJMoa1008862 [published Online First: 2011/03/04]
- 31. Labos C, Martinez SC, Leo Wang RH, et al. Utility of a genetic risk score to predict recurrent cardiovascular events 1 year after an acute coronary syndrome: A pooled

- analysis of the RISCA, PRAXY, and TRIUMPH cohorts. *Atherosclerosis* 2015;242(1):261-7. doi: 10.1016/j.atherosclerosis.2015.07.029 [published Online First: 2015/08/02]
- 32. Mega JL, Stitziel NO, Smith JG, et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. *Lancet* 2015;385(9984):2264-71. doi: 10.1016/s0140-6736(14)61730-x [published Online First: 2015/03/10]
- 33. Ganna A, Ingelsson E. 5 year mortality predictors in 498,103 UK Biobank participants: a prospective population-based study. *Lancet* 2015;386(9993):533-40. doi: 10.1016/s0140-6736(15)60175-1 [published Online First: 2015/06/08]

Table 1. Characteristics of participants at baseline

	Total (N=377,909)
Demographics & Physical measurement	
Age (year)	56.5 ± 7.9
Sex (men)	174,829 (46.3)
Systolic blood pressure (mmHg)	140.3 ± 19.7
Diastolic blood pressure (mmHg)	82.3 ± 10.7
BMI (kg/m²)	27.4 ± 4.8
Waist circumference (cm)	90.4 ± 13.5
Lifestyle behavior	
Favorable	196,450 (54.0)
Intermediate	122,695 (33.8)
Unfavorable	44,333 (12.2)
Laboratory findings	
HbA1c (%)	5.4 ± 0.6
Estimated GFR (mL/min/1.73m ²)	78.6 ± 14.2
Total cholesterol (mg/dL)	220.7 ± 44.2
Triglyceride (mg/dL)	155.7 ± 90.6
HDL cholesterol (mg/dL)	56.1 ± 14.8
LDL cholesterol (mg/dL)	137.9 ± 33.6
Baseline major comorbidity	
Coronary artery disease	21,678 (5.7)
Type 2 diabetes mellitus	15,034 (4.2)
Dyslipidemia	69,767 (18.5)
Hypertension	112,350 (29.7)
Myocardial infarction	9,472 (2.5)
Heart failure	2,626 (0.7)
Ischemic stroke	6,744 (1.8)
Chronic lung disease	55,410 (14.7)
Chronic kidney disease	5,766 (1.5)
Cancer	45,102 (11.9)
Charlson comorbidity index	0.57 ± 0.97

Data are n (%) or mean (SD). GFR, glomerular filtration rate

Figure Legends

Figure 1. Standardized cardiovascular mortality rates according to categories of polygenic

risk score for coronary artery disease (A) and type 2 diabetes mellitus (B)

Figure 2. Forest plot of cardiovascular mortality according to genetic risk and lifestyle risk.

Cox regression model were adjusted for age, sex, genotyping array, and first ten principal

components of ancestry.

P values are for testing the interaction between each genetic risk category and lifestyle

category.

Figure 3. Forest plot of cardiovascular mortality according to genetic risk, age, and sex.

Cox regression model were adjusted for age, sex, genotyping array, and first ten principal

components of ancestry.

P values are for testing the interaction between each genetic risk category, age, and sex.

Figure 4. Forest plot of cardiovascular mortality according to genetic risk and PRS-related

disease status.

PRS, polygenic risk score

Cox regression model were adjusted for age, sex, genotyping array, and first ten principal

components of ancestry.

P values are for testing the interaction between each genetic risk category and disease status.







