The American Journal of Human Genetics, Volume *107*

Supplemental Data

Combined Utility of 25 Disease and Risk Factor Polygenic Risk Scores for Stratifying Risk of All-Cause Mortality

Allison Meisner, Prosenjit Kundu, Yan Dora Zhang, Lauren V. Lan, Sungwon Kim, Disha Ghandwani, Parichoy Pal Choudhury, Sonja I. Berndt, Neal D. Freedman, Montserrat Garcia-Closas, and Nilanjan Chatterjee

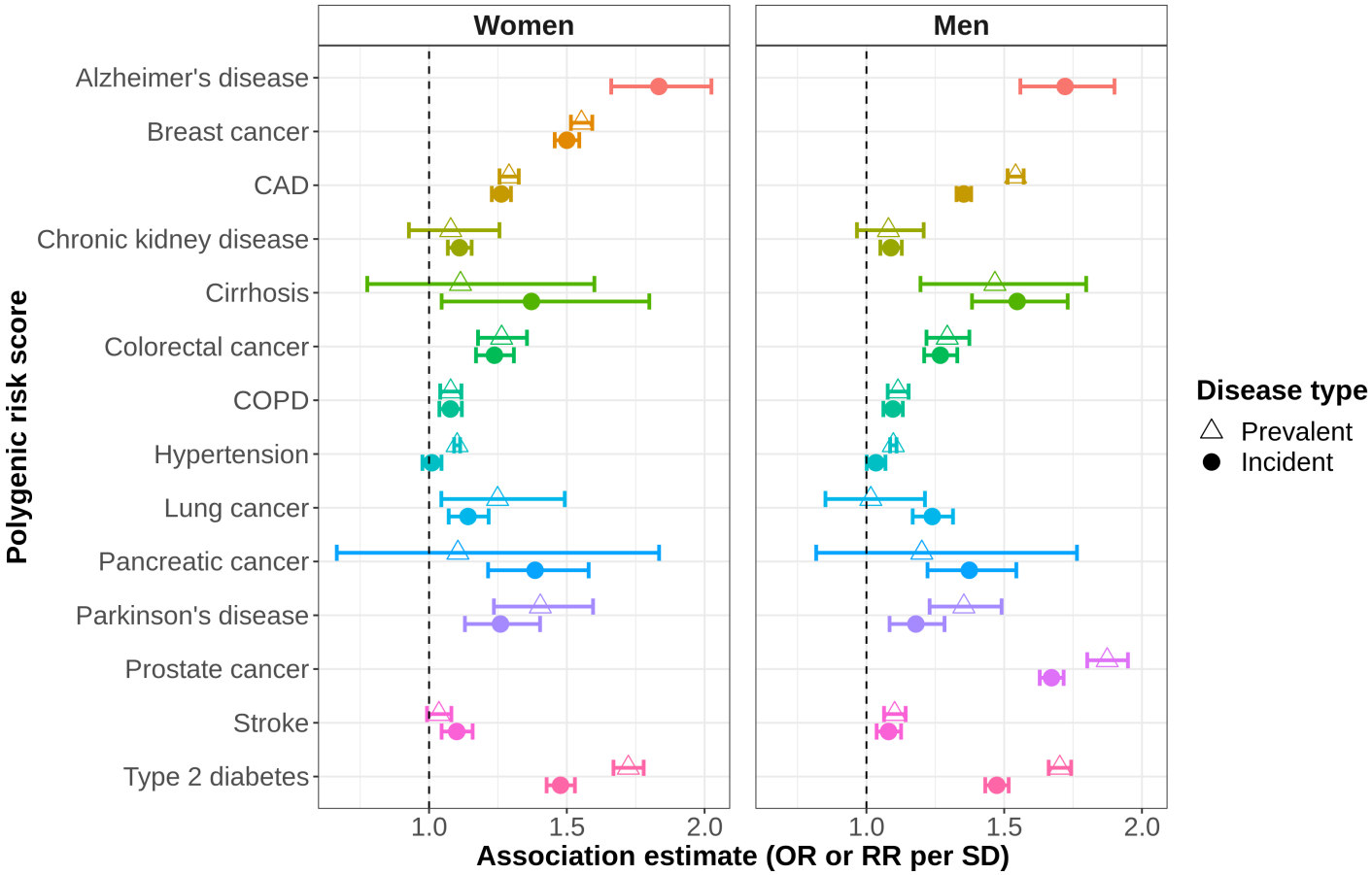

Figure S1. The estimated association between each disease PRS and prevalent and incident disease. The results are presented for women (left panel) and men (right panel) separately. For prevalent disease (open triangles in the plot), sex-specific logistic regression models were fit in the full cohort. For incident disease (closed circles in the plot), sex-specific modified Poisson regression models with robust standard error estimates were fit to the full cohort, excluding individuals with the disease at baseline (prevalent cases). All models included adjustment for age at entry. The estimates are presented as the estimated OR or RR per standard deviation of the PRS. The horizontal lines indicate 95% confidence intervals. As the number of prevalent cases of Alzheimer's disease was quite low for both men and women, these estimates are not presented. CAD: coronary artery disease; COPD: chronic obstructive pulmonary disease; OR: odds ratio; RR: relative risk; SD: standard deviation; PRS: polygenic risk score.

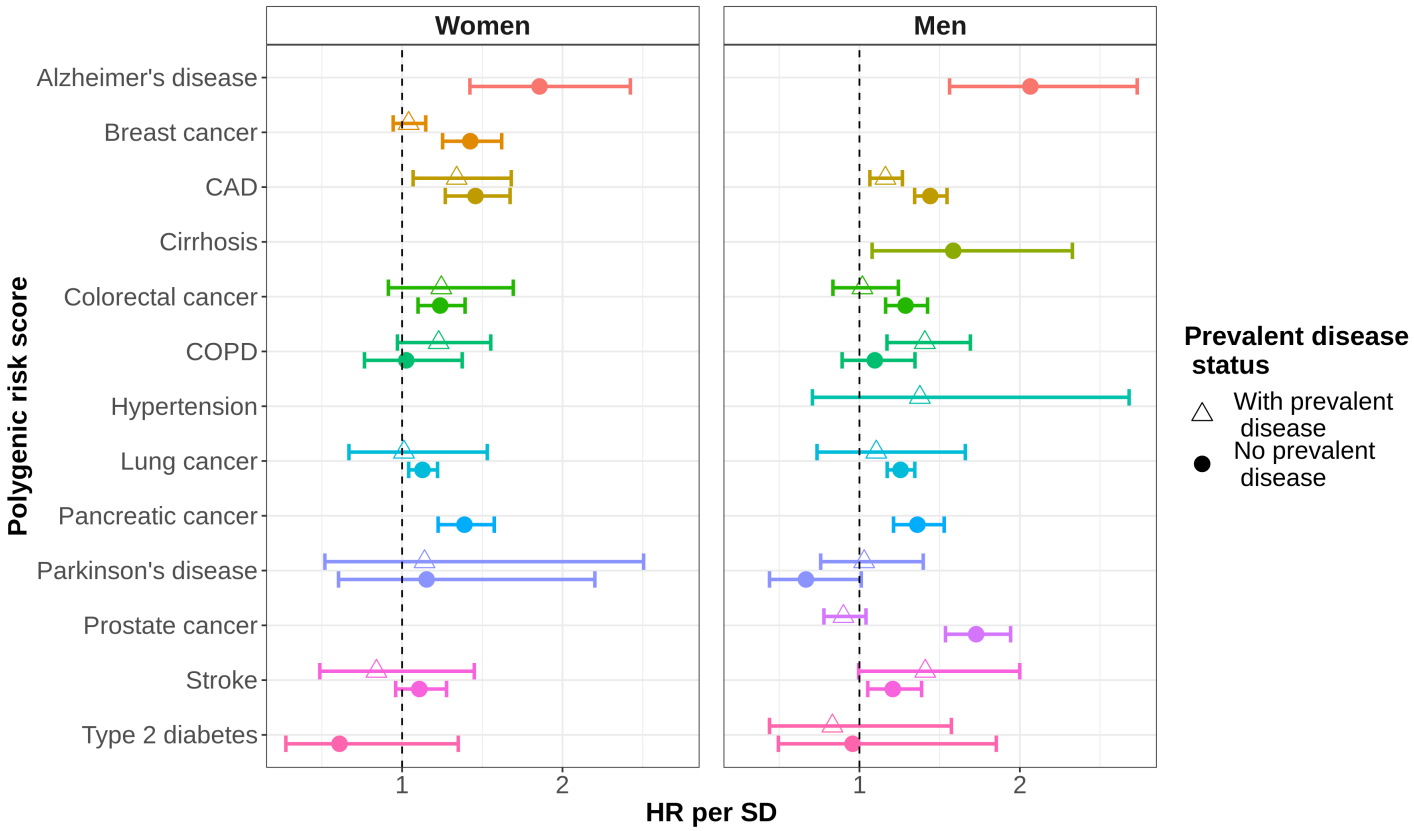

Figure S2. Cause-specific mortality results, stratified by the presence of disease at study baseline. For each disease, we used the data from the full cohort to evaluate the association between the disease PRS and mortality from the disease based on sex-specific Cox proportional hazards models of age at death in individuals with the disease at baseline (open triangles in the plot) and in individuals without the disease at baseline (closed circles in the plot). Deaths from other causes were treated as censoring events. Some causes did not have enough observations or deaths to yield stable estimates (< 30 observations or < 6 deaths); in these cases, estimates are not provided. Each PRS was standardized to have unit variance so the estimates correspond to the HR per SD of the PRS. The horizontal lines indicate 95% confidence intervals. CAD: coronary artery disease; COPD: chronic obstructive pulmonary disease; HR: hazard ratio; SD: standard deviation; PRS: polygenic risk score.

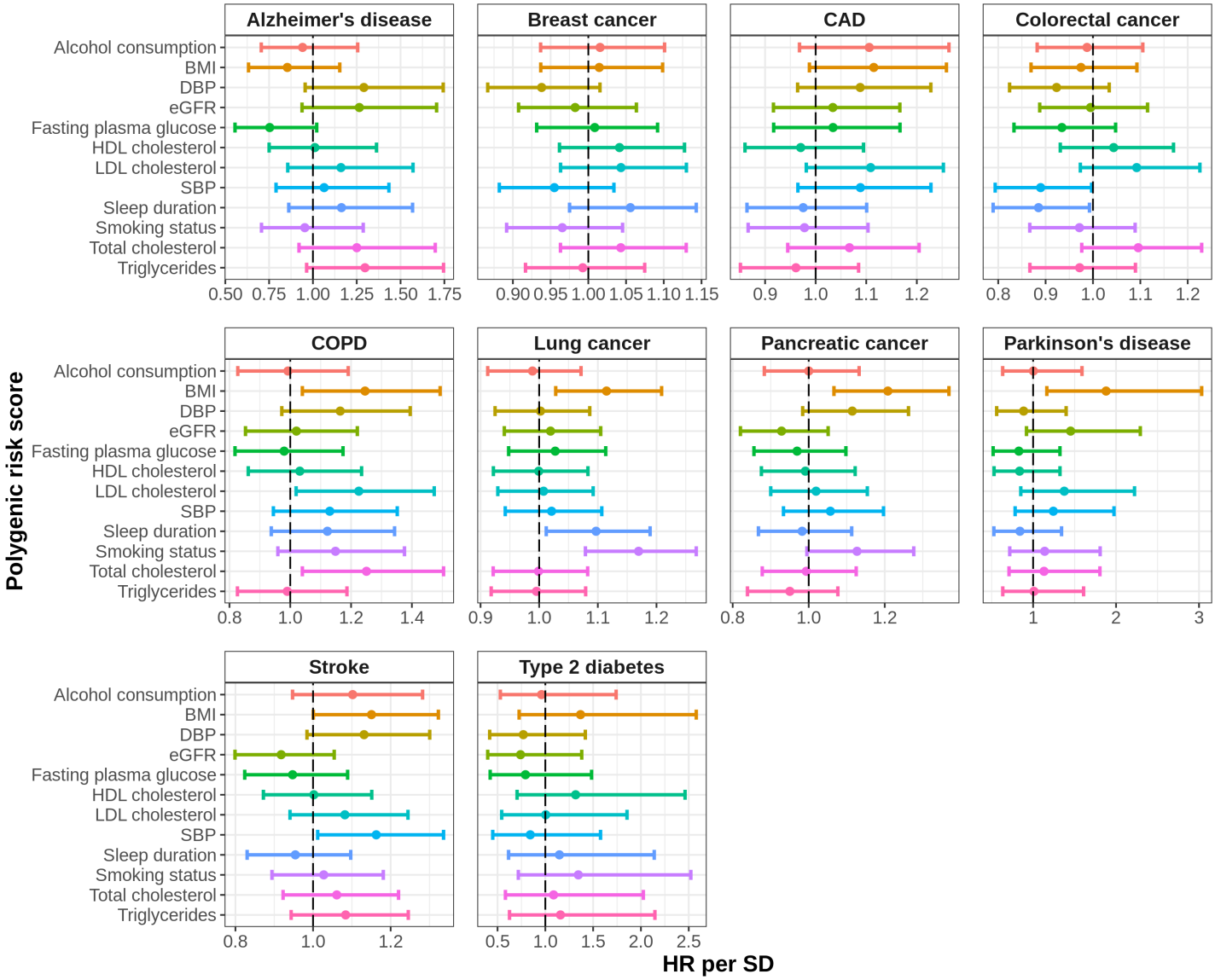

Figure S3. The estimated association between each mortality risk factor PRS and mortality due to each of the top causes of death among women. For each disease, we evaluated the association between each of the risk factor PRS and mortality from the disease based on Cox proportional hazards models of age at death in women in the full cohort. Deaths from other causes were treated as censoring events. Some causes did not have enough deaths to yield stable estimates (< 6 deaths); in these cases, estimates are not provided. Each PRS was standardized to have unit variance so the estimates correspond to the HR per SD of the PRS. The horizontal lines indicate 95% confidence intervals. BMI: body mass index; CAD: coronary artery disease; COPD: chronic obstructive pulmonary disease; DBP: diastolic blood pressure; eGFR: estimated glomerular filtration rate; HDL: high-density lipoprotein; LDL: low-density lipoprotein; SBP: systolic blood pressure; HR: hazard ratio; SD: standard deviation; PRS: polygenic risk score.

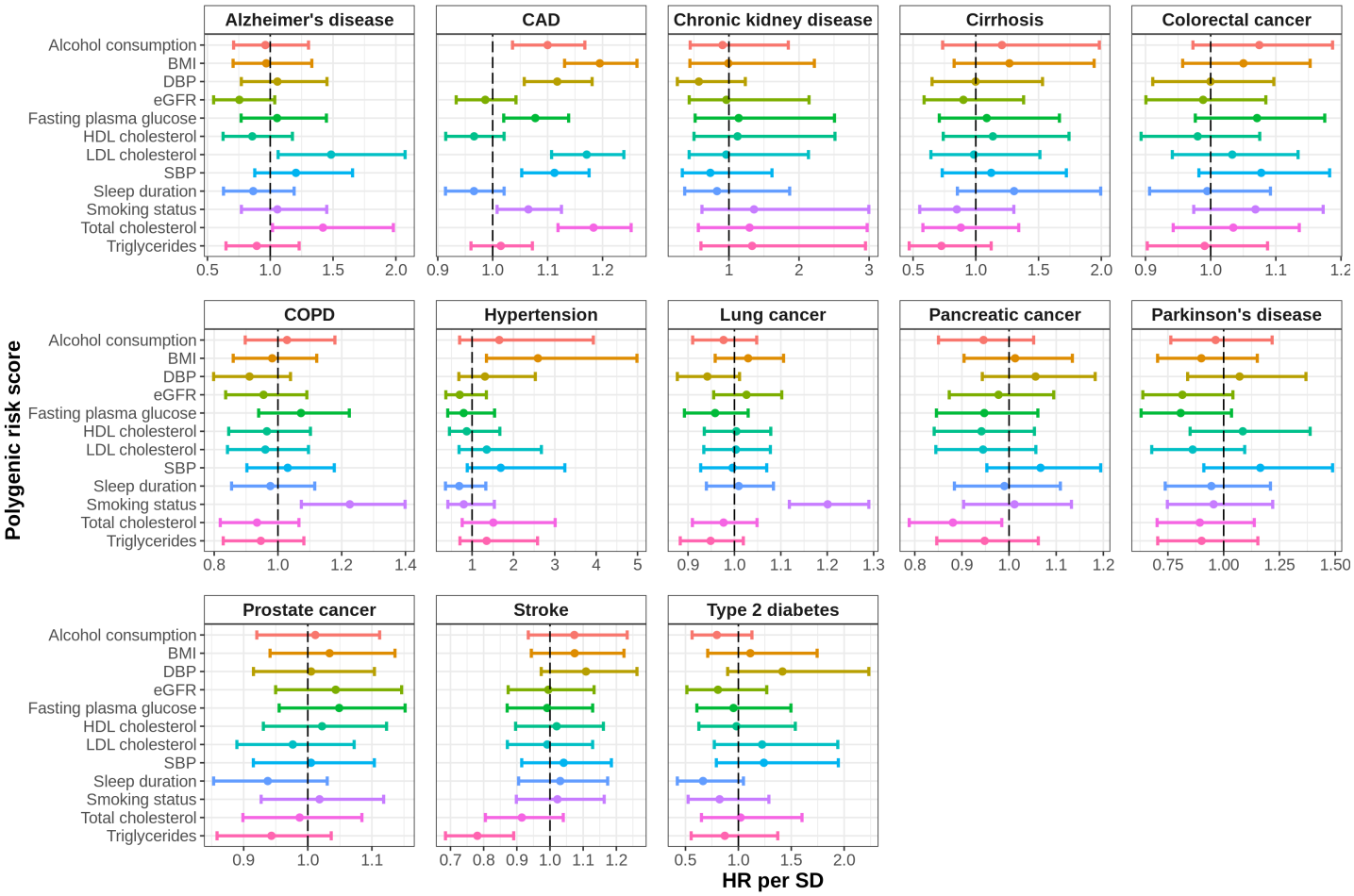

Figure S4. The estimated association between each mortality risk factor PRS and mortality due to each of the top causes of death among men. For each disease, we evaluated the association between each of the risk factor PRS and mortality from the disease based on Cox proportional hazards models of age at death in men in the full cohort. Deaths from other causes were treated as censoring events. Each PRS was standardized to have unit variance so the estimates correspond to the HR per SD of the PRS. The horizontal lines indicate 95% confidence intervals. BMI: body mass index; CAD: coronary artery disease; COPD: chronic obstructive pulmonary disease; DBP: diastolic blood pressure; eGFR: estimated glomerular filtration rate; HDL: high-density lipoprotein; LDL: low-density lipoprotein; SBP: systolic blood pressure; HR: hazard ratio; SD: standard deviation; PRS: polygenic risk score.

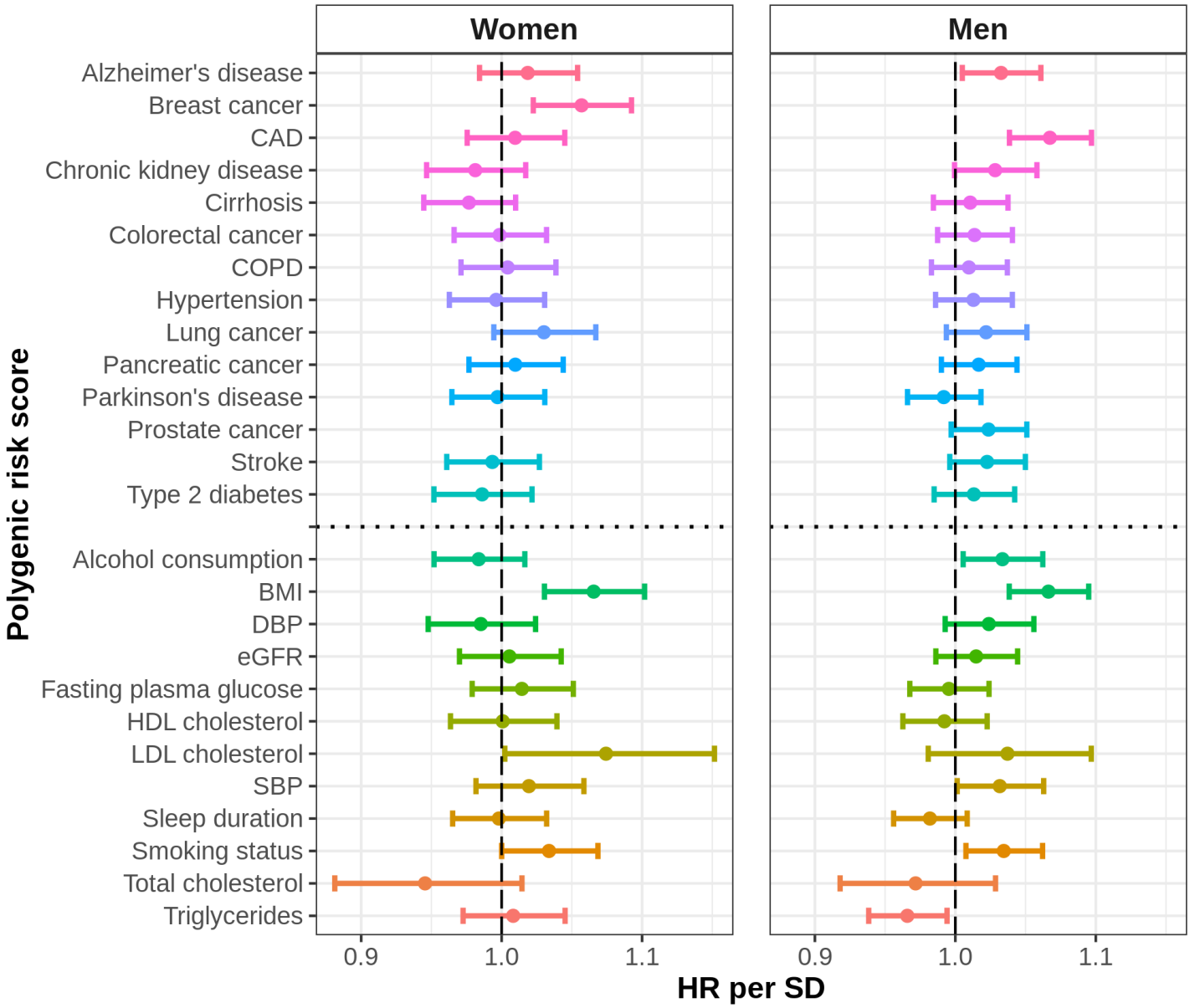

Figure S5. Association of trait-specific PRS with all-cause mortality in the training data based on models with all 25 PRS. The estimates are based on sex-specific Cox proportional hazards models of age at death with all 25 PRS, fit in the training data. These association estimates were used to weight each PRS to form the cPRS. Each PRS was standardized to have unit variance so the estimates correspond to the HR per SD of the PRS. The horizontal lines indicate 95% confidence intervals. BMI: body mass index; CAD: coronary artery disease; COPD: chronic obstructive pulmonary disease; DBP: diastolic blood pressure; eGFR: estimated glomerular filtration rate; HDL: high-density lipoprotein; LDL: low-density lipoprotein; SBP: systolic blood pressure; HR: hazard ratio; SD: standard deviation; PRS: polygenic risk score.

Table S1. ICD-10 codes for the top causes of death. The top causes of death ("CDC Definition") based on the CDC WONDER database and the corresponding specific cause of death included in the analysis ("Our Definition") are both presented. Ranking in the US based on data for 2017 from CDC WONDER for non-Hispanic whites aged 40 and over; ranking in the UK based on data for 2017 from the Office of National Statistics for individuals aged 40 and over.

CDC Definition				Our Definition	
Ranking in US (UK)		Cause	ICD-10 codes	Cause	ICD-10 codes
Women	Men				
1 (2)	1 (2)	Diseases of heart	I00-I09, I11, I13, I20-I51	CAD	I20-I25
2 (1)	2 (1)	Malignant neoplasms	C00-C97	Pancreatic	C25
				Colorectal	C18-C20
				Breast	C50
				Lung	C33-C34
				Prostate	C61
3 (4)	3 (3)	Chronic lower respiratory diseases	J40-J47	Chronic obstructive pulmonary disease	J41-J44
4 (5)	5 (5)	Alzheimer's disease	G30	Alzheimer's disease	G30
5 (3)	4 (4)	Cerebrovascular diseases	I60-I69	Stroke	I60, I61, I63, I64
6 (6)	6 (9)	Diabetes mellitus	E10-E14	Type 2 diabetes	E11
7 (10)	8 (10)	Nephritis, nephrotic syndrome and nephrosis	N00-N07, N17-N19, N25-N27	Chronic kidney disease	N18
8 (11)	10 (11)	Essential hypertension and hypertensive renal disease	I10, I12, I15	Hypertension	I10
9 (7)	7 (6)	Chronic liver disease and cirrhosis	K70, K73-K74	Alcoholic liver cirrhosis	K70.3
10 (8)	9 (7)	Parkinson's disease	G20-G21	Parkinson's disease	G20

CAD: Coronary artery disease; CDC: Centers for Disease Control; ICD: International Classification of Diseases; WONDER: Wide-ranging ONline Data for Epidemiologic Research.

Table S2. Methods for identifying prevalent and incident cases of each disease included in the analysis.

Cause of death	ICD Codes		Prevalent Definition	Incident Definition
	ICD9	ICD10		
Coronary artery disease	410-414	I20-I25	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date prior to the date of baseline assessment (b) Self-report: self-reported CAD at baseline	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date after the date of baseline assessment (b) Mortality data: one of the ICD10 codes listed as a primary or secondary cause of death
Pancreatic cancer	157	C25	Cancer registry: one of the ICD (9/10) codes in the cancer registry with an initial date prior to the date of baseline assessment	Cancer registry: one of the ICD (9/10) codes in the cancer registry with an initial date after date of baseline assessment
Colorectal cancer	153, 154.0, 154.1, 154.8	C18-C20		
Breast cancer	174	C50		
Lung cancer	162	C33-C34		
Prostate cancer	185	C61		
COPD	491, 492, 496	J41-J44	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date prior to the date of baseline assessment (b) Self-report: self-reported COPD, emphysema, or chronic bronchitis at baseline	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date after the date of baseline assessment (b) Mortality data: one of the ICD10 codes listed as a primary or secondary cause of death
Alzheimer's disease	331.0	G30 and F00	(a) HES: one of the ICD (9/10) codes in the primary or any secondary position with an initial code date is prior to the date of baseline assessment.	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date after the date of baseline assessment (b) Mortality data: one of the ICD10 codes listed as a primary or secondary cause of death
Stroke	430, 431, 434, 436	I60, I61, I63, I64	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date prior to the date of baseline assessment (b) Self-report: self-reported stroke at baseline	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date after the date of baseline assessment (b) Mortality data: one of the ICD10 codes listed as a primary or secondary cause of death
Type 2 diabetes	Defined based on algorithms in Eastwood et al. (1)			
Chronic kidney disease	585	N18	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date prior to the date of baseline assessment	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date after the date of baseline assessment (b) Mortality data: one of the ICD10 codes listed as a primary or secondary cause of death
Hypertension	401	I10	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date prior to the date of baseline assessment (b) Self-report: (i) self-reported essential hypertension or "any hypertension" but not "gestational hypertension/pre-eclampsia" at	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date after the date of baseline assessment (b) Mortality data: one of the ICD10 codes listed as a primary or secondary cause of death

			baseline or (ii) hypertension medication usage at baseline (c) SBP/DBP measures: systolic blood pressure ≥140 mmHg, or diastolic blood pressure ≥90 mmHg at baseline	
Alcoholic liver cirrhosis	571.2	K70.3	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date prior to the date of baseline assessment	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date after the date of baseline assessment (b) Mortality data: one of the ICD10 codes listed as a primary or secondary cause of death
Parkinson's disease	332.0	G20	(a) HES: ICD (9/10) codes in HES in the primary or secondary position with an initial code date prior to the date of baseline assessment (b) Self-report: self-reported Parkinson's disease at baseline	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date after the date of baseline assessment (b) Mortality data: one of the ICD10 codes listed as a primary or secondary cause of death

COPD: chronic obstructive pulmonary disease; HES: hospital episode statistics data; ICD: International Classification of Diseases.

Table S3. Conversion of self-reported alcohol intake to grams of alcohol per day. To compute grams of alcohol per day: (1) for each source of alcohol, multiply by the given factor and divide by 7 (if input is weekly intake) or 30 (if input is monthly intake) to get units/day; (2) multiply units/day by 8 to obtain grams/day; (3) sum grams/day intake of each source of alcohol to get total grams of alcohol per day.

Source	**Factor**
Red wine intake	1.5
Champagne/white wine	1.5
Beer/cider	2.5
Spirits	1
Fortified wine	1
Other alcoholic drinks	1.5

Table S4. The number of SNPs included in each PRS after removing SNPs in linkage disequilibrium via clumping.

Trait	**# SNPs**
Alcohol consumption	58
Alzheimer's disease	31
BMI	1,458
Breast cancer	153
CAD	207
Chronic kidney disease	4
Cirrhosis	2
Colorectal cancer	34
COPD	20
DBP	352
eGFR	31
Fasting blood glucose	24
HDL cholesterol	223
Hypertension	7
LDL cholesterol	195
Lung cancer	17
Pancreatic cancer	18
Parkinson's disease	44
Prostate cancer	123
SBP	390
Sleep duration	95
Smoking status	127
Stroke	79
Total cholesterol	240
Triglycerides	138
Type 2 diabetes	175
Total number of unique SNPs	*3,941*

BMI: body mass index; CAD: coronary artery disease; COPD: chronic obstructive pulmonary disease; DBP: diastolic blood pressure; eGFR: estimated glomerular filtration rate; HDL: high-density lipoprotein; LDL: low-density lipoprotein; SBP: systolic blood pressure; SNP: single nucleotide polymorphism; PRS: polygenic risk score.

Table S5. Summary statistics for the full cohort. Individuals who were related, were not of British ancestry, or had withdrawn their consent to participate were removed.

	Women	Men
Deaths by cause (n)		
Alzheimer's disease	43	38
with prevalent disease	0	1
without prevalent disease	43	37
Breast cancer	609	0
with prevalent disease	384	0
without prevalent disease	225	0
CAD	264	1,267
with prevalent disease	68	486
without prevalent disease	196	781
Chronic kidney disease	2	6
with prevalent disease	1	1
without prevalent disease	1	5
Cirrhosis	4	21
with prevalent disease	0	2
without prevalent disease	4	19
Colorectal cancer	295	445
with prevalent disease	35	94
without prevalent disease	260	351
COPD	119	218
with prevalent disease	74	126
without prevalent disease	45	92
Hypertension	4	9
with prevalent disease	4	9
without prevalent disease	0	0
Lung cancer	592	753
with prevalent disease	26	31
without prevalent disease	566	722
Pancreatic cancer	249	301
with prevalent disease	4	12
without prevalent disease	245	289
Parkinson's disease	18	64
with prevalent disease	9	41
without prevalent disease	9	23
Prostate cancer	0	436
with prevalent disease	0	183
without prevalent disease	0	253
Stroke	199	229
with prevalent disease	13	32
without prevalent disease	186	197
Type 2 diabetes	10	19
with prevalent disease	4	10
without prevalent disease	6	9
Prevalent disease (n)		
Alzheimer's disease	4	7
Breast cancer	6,323	0
CAD	5,445	12,530
Chronic kidney disease	170	311
Cirrhosis	27	70
Colorectal cancer	736	1,009
COPD	3,115	3,450
Hypertension	85,464	95,002
Lung cancer	107	122
Pancreatic cancer	15	26

Parkinson's disease	230	405
Prostate cancer	0	2,382
Stroke	2,126	3,092
Type 2 diabetes	4,072	7,576
Incident disease (n)		
Alzheimer's disease	314	345
Breast cancer	4,082	0
CAD	4,966	9,070
Chronic kidney disease	2,512	2,912
Cirrhosis	48	235
Colorectal cancer	1,036	1,437
COPD	2,740	3,576
Hypertension	3,154	3,371
Lung cancer	790	907
Pancreatic cancer	230	266
Parkinson's disease	299	493
Prostate cancer	0	4,542
Stroke	1,491	2,140
Type 2 diabetes	3,080	4,392
Mortality risk factors (mean (SD))		
Alcohol consumption (grams/day)	13.55 (12.34)	27.19 (23.39)
BMI (kg/m^2)	27.03 (5.14)	27.82 (4.21)
DBP (mmHg)	80.63 (9.93)	84.14 (9.99)
eGFR (mL/min/1.73 m^2)	85.57 (16.23)	87.61 (16.63)
Blood glucose (mmol/L)	5.07 (1.04)	5.18 (1.37)
HDL cholesterol (mmol/L)	1.60 (0.38)	1.28 (0.31)
LDL cholesterol (mmol/L)	3.64 (0.87)	3.49 (0.86)
SBP (mmHg)	135.60 (19.21)	141.30 (17.44)
Sleep duration (hours/day)	7.19 (1.10)	7.15 (1.07)
Smoking status (# ever smokers (%))	73,159 (40.5%)	79,226 (50.9%)
Total cholesterol (mmol/L)	5.90 (1.13)	5.50 (1.13)
Triglycerides (mmol/L)	1.56 (0.86)	1.98 (1.14)

CAD: coronary artery disease; COPD: chronic obstructive pulmonary disease; BMI: body mass index; DBP: diastolic blood pressure; eGFR: estimated glomerular filtration rate; HDL: high-density lipoprotein; LDL: low-density lipoprotein; SBP: systolic blood pressure; SD: standard deviation

Table S6. The estimated association between each mortality risk factor PRS and the risk factor measured at study baseline in women and men. Estimates are based on sex-specific linear regression models with robust standard error estimates, with the exception of smoking status, which was modeled using sex-specific logistic regression models. All models included adjustment for age at entry. Estimates are reported per standard deviation of the PRS.

Mortality risk factor	Women	Men
Alcohol consumption (grams/day)	0.90 (0.83, 0.96)	1.90 (1.79, 2.02)
BMI (kg/m^2)	1.46 (1.44, 1.49)	1.26 (1.24, 1.28)
DBP (mm Hg)	1.90 (1.85, 1.95)	1.63 (1.58, 1.68)
eGFR (mL/min/1.73 m^2)	2.54 (2.47, 2.61)	2.36 (2.28, 2.44)
Blood glucose (mmol/L)	0.065 (0.060, 0.070)	0.077 (0.069, 0.084)
HDL cholesterol (mmol/L)	0.118 (0.117, 0.120)	0.095 (0.094, 0.097)
LDL cholesterol (mmol/L)	0.234 (0.230, 0.238)	0.194 (0.190, 0.198)
SBP (mm Hg)	3.82 (3.74, 3.90)	3.06 (2.98, 3.14)
Sleep duration (hour)	0.092 (0.087, 0.097)	0.082 (0.077, 0.087)
Smoking status (odds ratio for ever smoking)	1.20 (1.19, 1.22)	1.22 (1.21, 1.23)
Total cholesterol (mmol/L)	0.300 (0.295, 0.305)	0.257 (0.251, 0.262)
Triglycerides (mmol/L)	0.187 (0.183, 0.191)	0.269 (0.263, 0.275)

BMI: body mass index; DBP: diastolic blood pressure; eGFR: estimated glomerular filtration rate; HDL: high-density lipoprotein; LDL: low-density lipoprotein; SBP: systolic blood pressure; PRS: polygenic risk score.

Table S7. The results of the analysis of all-cause mortality and the cPRS fitted in the training data and evaluated in the healthy subset of the test data. The cPRS were evaluated by fitting sex-specific Cox proportional hazards models of the association between age at death from all causes and the cPRS in the healthy subset of the test data. The healthy subset of the test data was defined as the test data with individuals with any of the diseases included as a top cause of death at baseline (prevalent cases). Both the continuous cPRS and categorical cPRS were modeled. The estimated HRs and CIs were converted to estimated years of life lost.

	Women	Men
Population in test data: N (deaths)		
Total population	29,379 (444)	18,249 (531)
Top 5% of cPRS	1,371 (27)	588 (21)
Middle 20% of cPRS	5,843 (80)	3,680 (107)
Bottom 5% of cPRS	1,647 (21)	1,145 (28)
Summary statistics for test data		
Age at entry (years; mean (SD))	54.4 (7.9)	54.3 (8.2)
Follow-up (years; mean (SD))	8.9 (0.9)	8.8 (1.1)
cPRS: HR (95% CI)		
Per SD of cPRS	1.07 (0.98, 1.18)	1.15 (1.06, 1.26)
Top 5% vs. middle 20% of cPRS	1.46 (0.94, 2.25)	1.28 (0.80, 2.04)
Bottom 5% vs. middle 20% of cPRS	0.89 (0.55, 1.44)	0.78 (0.51, 1.18)
cPRS: years of life lost (95% CI)		
Per SD of cPRS	0.71 (-0.21, 1.63)	1.43 (0.56, 2.31)
Top 5% vs. middle 20% of cPRS	3.75 (-0.61, 8.12)	2.45 (-2.23, 7.13)
Bottom 5% vs. middle 20% of cPRS	-1.16 (-5.97, 3.64)	-2.50 (-6.67, 1.66)

HR: hazard ratio; CI: confidence interval; cPRS: composite PRS; PRS: polygenic risk score; SD: standard deviation.

Table S8. Comparison of more sophisticated PRS with PRS based on GWAS Catalog. We compared the performance of three approaches for constructing PRS: (a) using results from the GWAS Catalog, (b) LD clumping and thresholding (C+T), and (c) LDpred. For the binary traits, we present the AUC for incident disease, while for the continuous risk factors, we present the R^2 for the measurement at baseline. See the Supplemental Methods for more details.

Trait (Reference)[a,b]	Best AUC or R^2 from C+T/LDpred	GWAS Catalog AUC or R^2
Alzheimer's disease (2)	0.67	0.66
Breast cancer (3)	0.64	0.62
Chronic kidney disease (4)	0.54	0.53
Coronary artery disease (5)	0.57	0.58
Prostate cancer (6)	0.63	0.65
Stroke (7)	0.55	0.52[c]
Type 2 diabetes (8)	0.61	0.61
BMI (9)	0.063	0.083
HDL cholesterol (10)	0.004	0.079
LDL cholesterol (10)	0.015	0.061
Total cholesterol (10)	0.009	0.060
Triglycerides (10)	< 0.001	0.048
Fasting plasma glucose (11)	0.006	0.003
eGFR (12)	0.027	0.022

[a]Summary statistics for several of the disease traits were not available.
[b]Available summary statistics for several of the mortality risk factors were based on data from the UK Biobank, precluding evaluation of the performance of the PRS in our data.
[c]PRS based on recently published stroke PRS, not results from the GWAS Catalog (13).
AUC: area under the receiver operating characteristic curve; BMI: body mass index; C+T: clumping and thresholding; COPD: chronic obstructive pulmonary disease; DBP: diastolic blood pressure; eGFR: estimated glomerular filtration rate; GWAS: genome-wide association study; HDL: high-density lipoprotein; LD: linkage disequilibrium; LDL: low-density lipoprotein; PRS: polygenic risk score; SBP: systolic blood pressure; SNP: single nucleotide polymorphism.

SUPPLEMENTAL METHODS

Constructing Alternative PRS

In addition to the PRS based on genome-wide significant results from the GWAS Catalog, which were the focus of the paper, we considered PRS constructed by two alternative approaches: linkage disequilibrium (LD) clumping and thresholding (C+T) (14) and LDpred (15). Briefly, C+T involves first clumping variants so as to remove correlated SNPs. This is accomplished by selecting the most significant variant and removing nearby variants that are correlated beyond some r^2. Next, the clumped variants are thresholded, i.e., those with p-values larger than some p are removed. LDpred is a Bayesian method that leverages information on LD structure to infer the posterior mean effect size of each SNP. This method utilizes a prior effect size distribution which requires specification of the proportion of causal SNPs, q. Thus, C+T involves selecting two tuning parameters, r^2 and p, while LDpred requires selecting one tuning parameter, q.

We tuned the parameters for C+T and LDpred using data from the UK Biobank. In particular, we considered $r^2 = (0.1, 0.2, 0.4, 0.6, 0.8)$, $p = (5 \times 10^{-8}, 5 \times 10^{-6}, 5 \times 10^{-4}, 0.05, 1)$, and $q = (0, 0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 1)$. For the binary disease traits, we chose the tuning parameters based on the AUC for prevalent disease. For the continuous risk factors, we chose the tuning parameters based on the R^2 for the baseline risk factor measurements.

Evaluating Alternative PRS

For each trait, we compared the three PRS (GWAS Catalog, C+T, and LDpred) in terms of their ability to predict the trait. Specifically, for the binary disease traits, we evaluated the AUC for incident disease in the UK Biobank. For the continuous risk factors, we evaluated the R^2 for the baseline risk factor measurements in the UK Biobank. We compared the best AUC/R^2 among the PRS constructed by C+T and LDpred to the AUC/R^2 for the GWAS Catalog PRS for the same trait.

Mortality Analysis with Alternative PRS

The performance of the GWAS Catalog PRS in terms of the AUC/R^2 was generally quite good (relative to the PRS built using C+T or LDpred), though we found meaningful gains for the LDpred PRS for breast cancer and stroke (Table S8). Thus, we repeated the main composite PRS mortality analysis, using the LDpred PRS for breast cancer and stroke (with $q = 0.1$ and 0.03, respectively) in place of the GWAS Catalog versions of these PRS. In particular, we constructed a new composite PRS in the training data using the LDpred PRS for breast cancer and stroke and the GWAS Catalog PRS for the other traits. All other aspects of the construction of this composite PRS were as described for the main composite PRS mortality analysis. We then evaluated the association between the new composite PRS and all-cause mortality by estimating the hazard ratio per standard deviation of the composite PRS in the test data, adjusting for the first ten principal components.

SUPPLEMENTAL REFERENCES

1. Eastwood S V., Mathur R, Atkinson M, Brophy S, Sudlow C, Flaig R, De Lusignan S, Allen N, Chaturvedi N. Algorithms for the capture and adjudication of prevalent and incident diabetes in UK Biobank. PLoS One. 2016;11(9).
2. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, Jun G, DeStefano AL, Bis JC, Beecham GW, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 2013;45(12):1452–8.
3. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, Lemaçon A, Soucy P, Glubb D, Rostamianfar A, et al. Association analysis identifies 65 new breast cancer risk loci. Nature. 2017;551(7678):92–4.
4. Pattaro C, Teumer A, Gorski M, Chu AY, Li M, Mijatovic V, Garnaas M, Tin A, Sorice R, Li Y, et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. Nat Commun. 2016;7(1):1–19.
5. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, Saleheen D, Kyriakou T, Nelson CP, CHopewell J, et al. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015;47(10):1121–30.
6. Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, Dadaev T, Leongamornlert D, Anokian E, Cieza-Borrella C, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet. 2018;50(7):928–36.
7. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, Rutten-Jacobs L, Giese AK, Van Der Laan SW, Gretarsdottir S, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. Nat Genet. 2018;50(4):524–37.
8. 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(11):2888–902.
9. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015;518(7538):197–206.
10. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013;45(11):1274–85.
11. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, MäGi R, Strawbridge RJ, Rehnberg E, Gustafsson S, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet. 2012;44(9):991–1005.
12. Li M, Li Y, Weeks O, Mijatovic V, Teumer A, Huffman JE, Tromp G, Fuchsberger C, Gorski M, Lyytikäinen LP, et al. SOS2 and ACP1 loci identified through large-scale exome chip analysis regulate kidney development and function. J Am Soc Nephrol. 2017;28(3):981–94.
13. Rutten-Jacobs LCA, Larsson SC, Malik R, Rannikmäe K, Sudlow CL, Dichgans M, Markus HS, Traylor M. Genetic risk, incident stroke, and the benefits of adhering to a healthy lifestyle: Cohort study of 306 473 UK Biobank participants. BMJ. 2018;363.
14. Privé F, Vilhjálmsson BJ, Aschard H, Blum MGB. Making the most of clumping and thresholding for polygenic scores. Am J Hum Genet. 2019;105(6):1213–21.
15. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, Genovese G, Loh PR, Bhatia G, Do R, et al. Modeling linkage disequilibrium increases accuracy of

polygenic risk scores. Am J Hum Genet. 2015;97(4):576–92.