Risk factors affecting polygenic score performance across diverse cohorts

Daniel Hui[1], Scott Dudek[1], Krzysztof Kiryluk[2], Theresa L. Walunas[3], Iftikhar J. Kullo[4], Wei-Qi Wei[5], Hemant K. Tiwari[6], Josh F. Peterson[5], Wendy K. Chung[7], Brittney Davis[8], Atlas Khan[2], Leah Kottyan[9], Nita A. Limdi[8], Qiping Feng[10], Megan J. Puckelwartz[11], Chunhua Weng[12], Johanna L. Smith[4], Elizabeth W. Karlson[13], Regeneron Genetics Center[14], Gail P. Jarvik[15], Marylyn D. Ritchie[1]

[1]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.
[2]Division of Nephrology, Department of Medicine, Columbia University, NY, New York.
[3]Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA.
[4]Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN.
[5]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN.
[6]Department of Pediatrics, University of Alabama at Birmingham, Birmingham, AL.
[7]Departments of Pediatrics and Medicine, Columbia University Irving Medical Center, Columbia University, New York, NY.
[8]Department of Neurology, School of Medicine, University of Alabama at Birmingham, Birmingham, AL.
[9]The Center for Autoimmune Genomics and Etiology, Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH.
[10]Division of Clinical Pharmacology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA.
[11]Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL.
[12]Department of Biomedical Informatics, Vagelos College of Physicians & Surgeons, Columbia University, New York, NY.
[13]Division of Rheumatology, Inflammation, and Immunity, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.
[14]Regeneron Pharmaceuticals Inc., Tarrytown, NY.
[15]Departments of Medicine (Medical Genetics) and Genome Sciences, University of Washington Medical Center, Seattle, WA.

Correspondence to:
Marylyn D. Ritchie, PhD
marylyn@pennmedicine.upenn.edu

Abstract

Apart from ancestry, personal or environmental covariates may contribute to differences in polygenic score (PGS) performance. We analyzed effects of covariate stratification and interaction on body mass index (BMI) PGS (PGS_{BMI}) across four cohorts of European (N=491,111) and African (N=21,612) ancestry. Stratifying on binary covariates and quintiles for continuous covariates, 18/62 covariates had significant and replicable R^2 differences among strata. Covariates with the largest differences included age, sex, blood lipids, physical activity, and alcohol consumption, with R^2 being nearly double between best and worst performing quintiles for certain covariates. 28 covariates had significant PGS_{BMI}-covariate interaction effects, modifying PGS_{BMI} effects by nearly 20% per standard deviation change. We observed overlap with covariates that had significant R^2 differences between strata and interaction effects – across all covariates, their main effects on BMI were correlated with maximum R^2 differences and interaction effects (0.56 and 0.58, respectively), suggesting high-PGS-score individuals have highest R^2 and PGS effect increases. Given significant and replicable evidence for context-specific PGS_{BMI} performance and effects, we investigated ways to increase model performance taking into account non-linear effects. Machine learning models (neural networks) increased relative model R^2 (mean 23%) across datasets. Finally, creating PGS_{BMI} directly from GxAge GWAS effects increased relative R^2 by 7.8%. These results demonstrate that certain covariates, especially those most associated with BMI, significantly affect both PGS_{BMI} performance and effects across diverse cohorts and ancestries, and we provide avenues to improve model performance that consider these effects.

Introduction

Polygenic scores (PGS) provide individualized genetic predictors of a phenotype by aggregating genetic effects across hundreds or thousands of loci, typically from genome-wide association studies (GWAS). In recent years it has become increasingly apparent that the transferability of PGS performance across different cohorts is poor (1). Most analyses to date have focused on ancestry differences as the main driver of this lack of transferability (2–4). However, a growing body of evidence has demonstrated that PGS performance and effect estimates are influenced by differences in certain environmental (classically termed "gene-environment" effects or interactions) or personal-level covariates – different phenotypes seem to be differently affected by these covariates, with most evidence being for adiposity traits such as body mass index (BMI) (5–14).

There are several gaps in current knowledge about these covariate-specific effects. Many analyses have assessed only a handful of these covariates due to the myriad of choices possible in typical large-scale biobanks. Little investigation has been done to systematically understand why certain covariates affect PGS performance, with such knowledge being useful to reduce the potential search for variables that impart context-specific effects. Furthermore, most studies investigating PGS-covariate interactions have been in European ancestry individuals; unfortunately, comparing differences in PGS performance and prediction while controlling for differences in ancestry versus differences in context has not been assessed in previous studies. Moreover, covariate-specific effects are notorious for having poor replicability in human genetics studies, and previous PGS-covariate interaction studies have been predominantly performed in the UK Biobank (UKBB) (15), where the majority of individuals are aged 40-69 (i.e., excluding young adults), are overall healthier than those from other e.g., hospital-based cohorts, and are predominantly European ancestry. Additionally, PGS performance is often assessed using linear models and in isolation of clinical covariates, which in practice would often be available. Machine learning models can have increased performance over linear models and are capable of modeling complex relationships and interactions between variables, which may serve to increase predictive performance, especially given evidence for PGS-covariate specific effects. Finally, given evidence for context-specific effects, it should be possible to directly incorporate SNP-covariate interaction effects from a GWAS directly to improve prediction performance, instead of relying on post-hoc interactions from a typical PGS calculated from main GWAS effects.

In this study we aimed to address the previously mentioned concerns. Using genetic data with linked-phenotypic information from electronic health records, we estimated the effects of covariate stratification and interaction on performance and effect estimates of PGS for BMI (PGS_{BMI}) – Figure 1 is a flowchart of the project. These analyses were done across four datasets: UK Biobank (UKBB), Penn Medicine BioBank (PMBB) (15), Electronic Medical Records and Genomics (eMERGE) network dataset (16), and Genetic Epidemiology Research on Adult Health and Aging (GERA). These datasets include participants from two ancestry groups (N=491,111 European ancestry (EUR), N=21,612 African ancestry (AFR)), and 62 covariates (25 present in multiple datasets) representing laboratory, survey, and biometric data types typically associated with cardiometabolic health and adiposity (including blood lipids, socioeconomic status, blood pressure, diet, physical activity, alcohol intake, smoking, and lung function, among others). We calculated the difference in PGS_{BMI} R^2 among individuals stratified by different quintiles of these variables (different groups for binary variables) and assessed whether a variable's association with BMI could explain its effect on PGS_{BMI} performance across

139 groups. We also assessed the significance of PGS_{BMI}-covariate interaction terms, and their
140 increases to model R^2 over models only using main effects, as well as correlation of main effect,
141 interaction effect, and R^2 differences. Replication and comparisons across all datasets and
142 ancestries were conducted. We then used machine learning models and evaluated their increase
143 in performance over linear models, even those that included regularization and interaction terms.
144 Finally, we conducted BMI GWAS using GxAge interaction terms, created PGS_{BMI} directly
145 using these effects, and assessed performance gains over using typical main effect GWAS. This
146 study addresses a plethora of open issues considering performance and effects of PGS on
147 individuals from diverse backgrounds.
148
149 **Results**
150 Effect of covariate stratification on PGS_{BMI} performance
151 We assessed 62 covariates for R^2 differences (25 present, or suitable proxies, in multiple datasets)
152 after stratifying on binary covariates and quintiles for continuous covariates. With UKBB EUR
153 as discovery (N=376,729), 18 covariates had significant differences (Bonferroni $p<.05/62$) in R^2
154 among groups (Figure 2a), including age, sex, alcohol consumption, different physical activity
155 measurements, Townsend deprivation index, different dietary measurements, lipids, blood
156 pressure, and A1c, with 40 covariates having suggestive ($p<.05$) evidence of R^2 differences.
157 From an original PGS_{BMI} R^2 of .076, R^2 increased to .094, .093, and .088 for those in the bottom
158 physical activity, alcohol intake, and high-density lipoprotein (HDL) cholesterol quintiles, and
159 decreased to .067, .57, and .049 for those in the top quintile (71%, 61%, and 56% relative R^2),
160 respectively, comparable to differences due to ancestry (1). We note that the differences in R^2
161 due to alcohol intake and HDL were larger than those of any physical activity phenotype, despite
162 physical activity having one of the oldest and most replicable evidence of interaction with
163 genetic effects of BMI (33,34). Despite considerable published evidence suggesting covariate-
164 specific genetic effects between BMI and smoking behaviors (6,8), we were only able to find
165 suggestive evidence for R^2 differences when stratifying individuals across several smoking
166 phenotypes (minimum p=0.016, for smoking pack years). R^2 differences due to educational
167 attainment were also only suggestive (p=0.015), of which published evidence is conflicting (35–
168 37).
169 We replicated these analyses in three additional large-scale cohorts of European and
170 African ancestry individuals (Figure 2b, Supplemental Table 3), as well as in African ancestry
171 UKBB individuals. For covariates with significant performance differences in the discovery
172 analysis, we were able to replicate significant ($p<.05$) R^2 differences for age, HDL cholesterol,
173 alcohol intake frequency, physical activity, and HbA1c, despite much smaller sample sizes. We
174 again observed largely insignificant differences across cohorts and ancestries when stratifying
175 due to different smoking phenotypes and educational attainment. Combining across cohorts, for
176 each covariate and ancestry we regressed R^2 values across groupings on covariate values in a
177 weighted linear regression, weighting by sample size. Slope of a weighted linear regression
178 across cohorts was directionally inconsistent between ancestries for the same covariate
179 (triglyceride levels, HbA1c, diastolic blood pressure, and sex), although larger sample sizes may
180 be needed to confirm these differences.
181 We noted several observations related to age-specific effects on PGS_{BMI}. First, in the
182 weighted linear regression of all R^2 across ancestries, expected R^2 for African ancestry
183 individuals overcomes that of European ancestry individuals among individuals within bottom
184 and top age quintiles observed in these data. For instance, the predicted R^2 of .048 for 80 year-

old European ancestry individuals would be lower than that of African ancestry individuals aged 24.7 and lower, indicating that certain differences in covariates can overcome differences in PGS_{BMI} performance due to ancestry (with age being one of the most common variables used for assessing and determining health). Second, we obtained these results despite the average age of GWAS individuals being 57.8, which should increase PGS_{BMI} R^2 of individuals closest to this age (13). This result suggests that in the case of these analyses, PGS performance due to decreased heritability with age cannot be reconciled by having GWAS individuals of similar age being used to create the PGS_{BMI}. Finally, we observed that PGS_{BMI} R^2 increases as age decreases, consistent with published evidence suggesting that the heritability of BMI decreases with age (38,39).

PGS-covariate interaction effects

Next, we estimated difference in PGS effects due to interactions with covariates, by modeling interaction terms between PGS_{BMI} for each covariate (Methods). We implemented a correction for shared heritability between covariates of interest and outcome (which can inflate test statistics (28)) to better measure the environmental component of each covariate, and show that this correction successfully reduces significance of interaction estimates (Supplemental Figure 1). Again, using UKBB EUR as discovery, we observed 28 covariates with significant (Bonferroni $p<.05/62$) PGS-covariate interactions (Table 1), with 38 having suggestive ($p<.05$) evidence (Supplemental Table 4). We observed the largest effect of PGS-covariate interaction with alcohol drinking frequency (20.0% decrease in PGS effect per 1 standard deviation (SD) increase, $p=2.62\times10^{-55}$), with large effects for different physical activity measures (9.4%-12.5% decrease/SD, minimum $p=3.11\times10^{-66}$), HDL cholesterol (15.3% decrease/SD, $p=1.71\times10^{-96}$) and total cholesterol (12.7% decrease/SD, $p=1.64\times10^{-71}$). We observed significant interactions with diastolic blood pressure (10.8% increase/SD, $p=6.06\times10^{-60}$), but interactions with systolic blood pressure were much smaller (1.17% increase/SD, $p=4.41\times10^{-3}$). Significant interactions with HbA1c (4.63% increase/SD, $p=5.37\times10^{-14}$) and type 2 diabetes (27.2% PGS effect increase in cases, $p=1.83\times10^{-7}$) were also observed. Other significant PGS-covariate interactions included lung function, age, sex, and LDL cholesterol – various dietary measurements also had significant interactions, albeit with smaller effects than other significant covariates. We were able to find significant interaction effects for smoking pack years (4.78% increase/SD, $p=3.68\times10^{-7}$), but other smoking phenotypes had insignificant interactions after multiple testing correction (minimum $p=2.7\times10^{-3}$) – interactions with educational attainment were also insignificant ($p=4.54\times10^{-2}$).

We replicated these analyses across ancestries and the other cohorts in this study (Figure 3, Supplemental Table 4). For age and sex, which were available for all cohorts, interactions were significant ($p<.05$) and directionally consistent across all cohorts and ancestries (except for GERA AFR both effects were reversed, albeit with small sample size (N=1,789)). We were able to test interactions with alcohol intake frequency and physical activity in GERA, and we were able to replicate significant and directionally consistent association. We observed poor replication for LDL cholesterol, HbA1c, and smoking pack years, with insignificant and directionally inconsistent interaction effects across cohorts. Educational attainment was available in GERA, and interactions were once again insignificant. We observed significant and directionally consistent interaction effects for TG in eMERGE EUR and PMBB EUR, while the effect was inconsistent in UKBB EUR despite much larger sample size.

However, despite significance of interaction terms, increases in model R^2 when including PGS-covariate interaction terms were small. For instance, the maximum increase among all covariates in UKBB EUR was only .0024 from a base R^2 of .1049 (2.1% relative increase), for alcohol drink frequency per week. Across all cohorts and ancestries, the maximum increase in R^2 was only .0058 from a base R^2 of 0.09454 (6.1% relative increase), when adding a PGS-age interaction term for eMERGE EUR ($p=5.40x10^{-46}$) – this was also the largest relative increase among models with significant interaction terms. This result suggests that, while interaction effects can significantly modify the effect of PGS_{BMI}, their overall impact on model performance is relatively small, despite the observation that stratifying by covariates produces large differences in R^2.

Correlations between R^2 differences, interaction effects, and main effects

We next investigated the relationship between interaction effects, maximum R^2 differences across quintiles, and main effects of covariates on BMI. We first estimated main effects of each covariate on BMI (Methods, Supplemental Table 5), and then calculated the correlation weighted by sample size between main effects, maximum PGS_{BMI} R^2 across quintiles, and PGS-covariate interaction effects (Figure 4) across all cohorts and ancestries – GERA data were excluded from these analyses due to slightly different phenotype definition (Supplemental Table 6), as were binary variables. Interaction effects and maximum R^2 differences had a .80 correlation ($p=2.1x10^{-27}$), indicating that variables with the largest interaction effects also had the largest effect on PGS_{BMI} performance across quintiles, and that covariates that increase PGS_{BMI} effect also have the largest effect on PGS_{BMI} performance i.e., individuals most at risk for obesity will have both disproportionately larger PGS_{BMI} effect and R^2. Main effects and maximum R^2 differences had a .56 correlation ($p=1.3x10^{-11}$), while main effects and interaction effects had a .58 correlation ($7.6x10^{-12}$) again suggesting that PGS_{BMI} are more predictive in individuals with higher values of BMI-associated covariates, although predictive to a lower degree than estimating the interaction effects themselves directly. However, this result demonstrates that covariates that influence both PGS_{BMI} effect and performance can be predicted just using main effects of covariates, which are often known for certain phenotypes and easier to calculate, as genetic data and PGS construction would not be required.

Effects of machine learning approaches on predictive performance

Given evidence of PGS-covariate dependence, we aimed to assess increases in R^2 when using machine learning models (neural networks), which can inherently model interactions and other nonlinearities, over linear models even with interaction terms. We first included age and sex as the only covariates (along with genotype PCs and PGS_{BMI}), as age and sex were present in all datasets and had significant and replicable evidence for PGS-dependence across our analyses. Three models were assessed – L1-regularized (i.e., LASSO) linear regression without any interaction terms, LASSO including a PGS-age and PGS-sex interaction term, and neural networks (without interaction terms). When comparing neural networks to LASSO with interaction terms, the relative 10-fold cross-validated R^2 increased up to 67% (mean 23%) across cohorts and ancestries (Figure 5, Supplemental Table 7). The inclusion of interaction terms increased cross-validated R^2 up to 12% (mean 3.9%) when comparing LASSO including interaction terms to LASSO with main effects only.

We then modeled all available covariates for each cohort and did similar comparisons. Cross-validated R^2 increased up to 17.6% (mean 9.5%) when using neural networks versus

LASSO with interaction terms, and up to 7.0% (mean 2.0%) when comparing LASSO with interaction terms to LASSO with main effects only. Increases in model performance using neural networks were smaller in UKBB, perhaps due to the age range being smaller than other cohorts (all participants aged 39-73). This result suggests that additional variance explained through non-linear effects with age and sex are explained by other variables present in the remainder of the datasets. Our findings suggest that machine learning methods can improve model R^2 that include PGS_{BMI} as variables beyond including interaction terms in linear models, even when variable selection is performed using LASSO.

<u>Calculating PGS directly from GxAge GWAS effects</u>

Previous studies (13) have created PGS using GWAS stratified by different personal-level covariates, but for practical purposes this leads to a large loss of power as the full size of the GWAS is not utilized for each strata and continuous variables are forced into bins. We developed a novel strategy where PGS are instead created from a full-size GWAS that includes SNP-covariate interaction terms (Methods). We focused on age interactions, given their large and replicable effects based on our results – similar to a previous study, we conducted these analyses in the European UKBB. We used a 60% random split of study individuals to conduct three sets of PGS using GWAS of the following designs: main effects only, main effects also with a SNP-age interaction term, and main effects but stratified into quartiles by age. 20% of the remaining data were used for training and the final 20% were held-out as a test set. The best performing PGS created from SNP-age interaction terms (PGS_{GxAge}) increased test R^2 to 0.0771 from 0.0715 (7.8% relative increase) compared to the best performing main effect PGS (Figure 6, Supplemental Table 9) – age-stratified PGS had much lower performance than both other strategies (unsurprising given reasons previously mentioned). Including a PGS_{GxAge}-age interaction term only marginally increased R^2 (.0001 increase), with similar increases for the other two sets of PGS, further demonstrating that post-hoc modeling of interactions cannot reconcile performance gained through directly incorporating interaction effects from the original GWAS. The strategy of creating PGS directly from full-sized SNP-covariate interactions is potentially quite useful as it increases PGS performance without the need for additional data – there are almost certainly a variety of points of improvement (Discussion), but we consider their investigation outside the scope of this study.

Discussion

Across four large-scale cohorts of diverse ancestry, we uncover replicable effects of covariates on both performance and effects of PGS_{BMI}. When stratifying by quintiles of different covariates, we find certain covariates that have significant and replicable evidence for differences in PGS_{BMI} R^2, with R^2 being nearly double between top and bottom performing quintiles for covariates with the largest differences. When testing covariates for PGS-covariate interaction effects, we also find covariates with significant interaction effects, where, for the largest effect covariates, each standard deviation change affects PGS_{BMI} effect by nearly 20%. Across analyses, we find age and sex had the most replicable interaction effects, with levels of serum cholesterol, physical activity, and alcohol consumption having the largest effects across cohorts. We also observed that interaction effects and R^2 differences were strongly correlated, with main effects also being correlated with interaction effects and R^2 differences, suggesting that covariates with the largest interaction effects also contribute to the largest R^2 differences, with simple main effects also being predictive of expected differences in R^2 and interaction effects. While in this study we

322 stratified by quintile, similar trends of performance differences would be apparent in smaller
323 percentiles. This observation is relevant to current research and clinical use of PGS, as
324 individuals above a percentile cutoff are designated high risk (40). These analyses imply that
325 individuals most at-risk for obesity should have both disproportionately higher predicted BMI
326 and increased BMI prediction performance compared to the general population – more work is
327 required to understand why, and if similar trends can be observed in other traits relevant to
328 human health. We also employ machine learning methods for prediction of BMI with models
329 that include PGS_{BMI} and demonstrate that these methods outperform regularized linear regression
330 models that include interaction effects. Finally, we employ a novel strategy that directly
331 incorporates SNP interaction effects into PGS construction, and demonstrate that this strategy
332 improves PGS performance when modeling SNP-age interactions compared to PGS created only
333 from main effects.
334 Future work may include replicating these analyses across additional traits, and further
335 exploring machine learning and deep learning methods on other phenotypes to determine if this
336 trend of inclusion of PGS along with covariate interaction effects outperforms linear models for
337 risk prediction. Additionally, the method of including a PGS for the covariate to better measure
338 its environmental effect is potentially worth exploring further, which should improve in the
339 future as PGS performance continues to increase. A slight limitation of this method in our study
340 is that for the UKBB analyses the GWAS used for PGS construction were also from UKBB thus
341 not out-of-sample, although many of the covariates only have GWAS available through UKBB
342 individuals. Furthermore, a variety of improvements are likely possible when creating PGS
343 directly from SNP-covariate interaction terms. First, we used the same SNPs that were selected
344 by pruning and thresholding based on their main effect p-values, but selection of SNPs based on
345 their interaction p-values should also be possible and would likely improve performance.
346 Additionally, performance of pruning and thresholding-based strategies have largely been
347 overtaken by methods that first adjust all SNP effects for LD and don't require exclusion of
348 SNPs, and a method that could do a similar adjustment for interaction effects would likely
349 outperform most current methods for traits with significant context-specific effects. Finally,
350 incorporation of additional SNP-covariate interactions (e.g., SNP-gender) would also likely
351 further improve prediction performance, although any SNP selection/adjustment procedures may
352 further complicated by additional interaction terms.
353 We propose several ideas on why BMI-associated covariates have larger interaction
354 effects and impact on R^2 for PGS_{BMI}. Age may be a proxy for accumulated gene-environment
355 interactions as younger individuals have less exposure to environmental influences on weight
356 compared to older individuals; therefore, one may expect that in younger individuals their
357 phenotype could be better explained by genetics compared to environment. Secondly, PGS may
358 more readily explain more extreme phenotypes because they incorporate large effect variants
359 associated with very high weight or height (41). For BMI, the distribution is skewed toward the
360 right i.e., effects in trait-increasing variants may have a larger potential to affect trait variation
361 compared to trait-reducing variants. This explanation would likely be better suited to traits with
362 left-tailed distributions and is not fully satisfactory as first log-transforming or rank-normal
363 transforming the phenotype, as was done in this study, may invalidate this explanation.
364 PGS is a promising technique to stratify individuals for their risk of common, complex
365 disease. In order to achieve accurate predictions as well as promote equity, further research is
366 required regarding PGS methods and assessments. This research provides firm evidence on the
367 context-specific nature of PGS and the incorporation of covariate interaction effects for

predicting BMI and, apart from promoting equitable use of PGS across ancestries and cohorts, demonstrates ways to improve PGS performance incorporating context-specific effects.

Methods

Study datasets

Individual inclusion criteria and sample sizes per cohort are described below – additional information is available in Supplemental Table 1.

UKBB

Individual-level quality-control and filtering have been described elsewhere (17) for European ancestry individuals. Briefly, individuals were split by ancestry according to both genetically inferred ancestry and self-reported ethnicity (15). Individuals with low genotyping quality and sex mismatch were removed, only unrelated individuals (pi-hat < .250) were retained, and variants were filtered at INFO > .30 and minor allele frequency (MAF) > .01. For African ancestry, individuals were first selected based on self-reported ethnicity "Black or Black British", "Caribbean", "African", or "Any other Black background". Individuals that were low quality i.e., "Outliers for heterozygosity or missing rate", and that were Caucasian from "Genetic ethnic grouping" were removed. Of these individuals, those that were +/- 6 standard deviations from the mean of the first 5 genetic principal components provided by UKBB were excluded. Finally, only unrelated individuals were retained up to the second degree using plink2 (18) "-king-cutoff .125". After QC and consideration of phenotype, a total of 7,046 individuals in the UKBB AFR data who also had BMI available were used for downstream analyses. In total, 383,775 individuals were used for analysis (N_{EUR}=376,729, N_{AFR}=7,046).

eMERGE

Ancestry and relatedness inference have been described elsewhere (15). Individuals were split into European and African ancestry cohorts, and related individuals were removed (pi-hat > .250) such that all were unrelated. 35,064 individuals (N_{EUR}=31,961, N_{AFR}=3,103) were used for analysis.

GERA

Ancestry inference has been described elsewhere (19), and study individuals were divided into European and African ancestry cohorts. Related individuals were removed using plink2 "-king-cutoff .125". 57,838 individuals (N_{EUR}=56,049, N_{AFR}=1,789) were used for analysis.

PMBB

Ancestry inference and relatedness inference have been described elsewhere (20). Individuals were split into European and African ancestry cohorts, and related individuals were removed at pi-hat > .250. 36,046 individuals (N_{EUR}=26,372 , N_{AFR}=9,674) were used for analysis.

Choice of covariates

A total of 62 covariates were included in the analyses, 25 of which were present (or similar proxies) in multiple datasets. These covariates were selected based on relevance to cardiometabolic health and obesity, and previous evidence of context-specific effects with BMI (5,6,8,21–23). For UKBB, phenotype values were used from the collection that was closest to recruitment, and for PMBB the median values were used – for GERA and eMERGE only one

value was available. Additional details on covariate constructions, transformations, filtering, and cohort availability are in Supplemental Table 2.

PGS construction

PGS for BMI (PGS_{BMI}) were constructed using PRS-CSx (24), using GWAS summary statistics for individuals of European (25), African (26), and East Asian (27) ancestry that were out-of-sample of study participants. A set of 1.29 million HapMap3 SNPs provided by PRS-CSx was used for PGS calculation, which are generally well-imputed and variable frequency across global populations. Default settings for PRS-CSx (downloaded November 22, 2021) were used, which have been demonstrated to perform well for highly polygenic traits such as BMI (list of parameters in Supplemental Table 8). The final PGS_{BMI} per ancestry and cohort was calculated by regressing log(BMI) on the PGS_{BMI} per ancestry without covariates – the combined, predicted value was used as a single PGS_{BMI} in downstream analyses.

For GERA, BMI was not transformed, as it was already binned into a categorical variable with 5 levels (18≤, 19-25, 26-29, 30-39, >40). Additionally, for GERA the uncombined ancestry-specific PGS_{BMI} were used in the final models, as it had higher R^2 than using the combined PGS_{BMI} (data not shown).

PGS_{BMI} performance after covariate stratification

Analyses were performed separately for each cohort and ancestry. For each covariate, individuals were binned by binary covariates or quintiles for continuous covariates. Incremental PGS_{BMI} R^2 was calculated by taking the difference in R^2 between:

$\log(BMI) \sim PGS_{BMI} + Age + Sex + PCs_{1\text{-}5}$
$\log(BMI) \sim Age + Sex + PCs_{1\text{-}5}$

We performed 5,000 bootstrap replications to obtain a bootstrapped distribution of R^2. P-values for differences in R^2 were calculated between groups by calculating the proportion of overlap between two normal distributions of the R^2 value using the standard deviations of the bootstrap distributions. Again for GERA, BMI was not transformed.

PGS_{BMI} interaction modelling

Evidence for interaction with each covariate with the PGS_{BMI} was evaluated using linear regression. It has been reported that the inclusion of covariates that share heritability with the outcome can inflate test statistic estimates (28–30). To assuage these concerns, we introduced a novel correction, where we first calculated a PGS for the covariate ($PGS_{Covariate}$) and included it in the model, as well as an interaction term between PGS_{BMI} and $PGS_{Covariate}$. The $PGS_{Covariate}$ terms were calculated using the European ancestry Neale Lab summary statistics (URLs) and PRS-CS (31). To standardize effect sizes across analyses, PGS_{BMI} and Covariate were first converted to mean zero and standard deviation of 1 (binary covariates were not standardized). We demonstrate inclusion of $PGS_{Covariate}$ terms successfully reduced significance of the PGS_{BMI}*Covariate term (Supplemental Figure 1). The final model used to evaluate PGS_{BMI} and Covariate interactions was:

$\log(BMI) \sim PGS_{BMI}*Covariate + PGS_{BMI} + Covariate + PGS_{Covariate} + PGS_{BMI}*PGS_{Covariate} + Age + Sex + PCs_{1\text{-}5}$

We report the effect estimates of the PGS_{BMI}*Covariate term, and differences in model R^2 with and without the PGS_{BMI}*Covariate term. Again for GERA, BMI was not transformed.

Correlation between R^2 differences, interaction effects, and main effects

We estimated main effects of each covariate on BMI with the following model:

$\log(BMI) \sim Covariate + Age + Sex + PCs_{1\text{-}5}$

Note that we ran new models with main effects only, instead of using the main effect from the interaction models (as the main effects in the interaction models depend on the interaction terms, and main effects used to create interaction terms are sensitive to centering of variables despite the scale invariance of linear regression itself (32)). We then estimated the correlation between main effects, interaction effects, and maximum R^2 differences across all cohorts and ancestries weighting by sample size, analyzing quantitative and binary variables separately.

Machine learning models

UKBB EUR and GERA EUR models were restricted to 30,000 random individuals, for computational reasons – BMI distributions did not differ from the full-sized datasets (Kolmogorov-Smirnov p-value of 0.29 and 0.57, respectively). PGS_{BMI} and top five genetic principal components were included as features in all models. Two sets of models were evaluated for each cohort and ancestry: including age and sex as features, and including all available covariates in each cohort as features. 'Ever Smoker' status was used in favor of 'Never' vs 'Current smoking' status (if present), as individuals with 'Never' vs 'Current' status are a subset of those with 'Ever Smoker' status. UKBB AFR with all covariates was excluded due to small sample size (N=53).

Neural networks were used as the model of choice, given their inherent ability model interactions and other nonlinear dependencies. Prior to modeling, all features were scaled to be between 0 and 1. We used average 10-fold cross validation R^2 to evaluate model performance. Separate models were trained using untransformed and log(BMI). L1-regularized linear regression was used with 18 values of lambda (1.0, 5.0×10^{-1}, 2.0×10^{-1}, 1.0×10^{-1}, 5.0×10^{-2}, 2.0×10^{-2}, …, 1.0×10^{-5}, 5.0×10^{-6}, 2.0×10^{-6}). Models were trained without inclusion of interaction terms (which neural networks can implicitly model), using 1,000 iterations of random search with the following hyperparameter ranges: size of hidden layers [10, 200], learning rate [.01, .0001], type of learning rate [constant, inverse scaling], power t [.4, .6], momentum [.80, 1.0], batch size [32, 256], and number of hidden layers [1, 2].

GxAge PGS_{BMI} creation and assessment

Analyses were conducted in the European UKBB (N=376,629), as was done in a study on a similar topic (13). Three sets of analyses were performed, using GWAS conducted in a 60% random split of individuals using the following models (BMI was rank-normal transformed before GWAS):

1) $BMI \sim SNP + Age + Sex + PCs_{1\text{-}5}$

2) $BMI \sim SNP + Age*SNP + Age + Sex + PCs_{1\text{-}5}$

3) Using the model in 1) but stratified into quartiles by age. BMI was rank-normal transformed within each quartile.

Using each set of GWAS, PGS was first calculated in a 20% randomly selected training set of the dataset using pruning and thresholding using 10 p-value thresholds (.50, .10, …, $5.0\text{x}10^{-5}$, $1.0\text{x}10^{-5}$) and remaining settings as default in Plink 1.9. For 2), GxAge PGS_{BMI} were calculated using SNPs clumped by their main effect p-values from 1), and additionally incorporating the GxAge interaction effects per SNP. In other words, instead of typical PGS construction as:

$$PGS_i = \beta_1 k_1 + \beta_2 k_2 + \ldots + \beta_n k_n$$

For an individual *i*'s PGS calculation, with main SNP effect β, and *n* SNPs, PGS incorporating GxAge effects (PGS_{GxAge}) was calculated as:

$$PGS_{GxAge,i} = \beta_1 k_1 + \beta_{GxAge,1} k_1 Age_i + \beta_2 k_2 + \beta_{GxAge,2} k_2 Age_i \ldots + \beta_n k_n + \beta_{GxAge,n} k_n Age_i$$

Where β_{GxAge} is the GxAge effect for each SNP *n* and Age_i is the age for individual *i*.

For each of the three analyses, the parameters and models resulting in the best performing PGS (highest incremental R^2, using same main effect covariates as in the three GWAS) from the training set were evaluated in the remaining 20% of the study individuals. For 3), models were first trained within each quartile separately. To maintain sense of scale across quartiles (after rank-normal transformation), R^2 between all predicted values and true values was calculated together.

URLs

Neale Lab UKBB summary statistics: http://www.nealelab.is/uk-biobank

Data Availability statement

UK Biobank data was accessed under project #32133. eMERGE data is available at dbGaP in phs001584.v2.p2. GERA data is available at dbGaP in phs000674.v3.p3. Summary statistics for PMBB are available from authors upon request.

Regeneron Genetics Center Banner Author List and Contribution Statements

RGC Management and Leadership Team

Goncalo Abecasis, PhD , Aris Baras, M.D. , Michael Cantor, M.D. , Giovanni Coppola, M.D. , Andrew Deubler , Aris Economides, Ph.D. , Luca A. Lotta, M.D., Ph.D. , John D. Overton, Ph.D. , Jeffrey G. Reid, Ph.D. , Katherine Siminovitch, M.D. , Alan Shuldiner, M.D.

Sequencing and Lab Operations

Christina Beechert , Caitlin Forsythe, M.S. , Erin D. Fuller , Zhenhua Gu, M.S. , Michael Lattari , Alexander Lopez, M.S. , John D. Overton, Ph.D. , Maria Sotiropoulos Padilla, M.S. , Manasi Pradhan, M.S. , Kia Manoochehri, B.S. , Thomas D. Schleicher, M.S. , Louis Widom , Sarah E. Wolf, M.S. , Ricardo H. Ulloa, B.S.

Clinical Informatics

Amelia Averitt, Ph.D. , Nilanjana Banerjee, Ph.D. , Michael Cantor, M.D. , Dadong Li, Ph.D. , Sameer Malhotra, M.D. , Deepika Sharma, MHI , Jeffrey Staples , Ph.D.

Genome Informatics

Xiaodong Bai, Ph.D. , Suganthi Balasubramanian, Ph.D. , Suying Bao, Ph.D. , Boris Boutkov, Ph.D. , Siying Chen, Ph.D. , Gisu Eom, B.S. , Lukas Habegger, Ph.D. , Alicia Hawes, B.S. , Shareef Khalid , Olga Krasheninina, M.S. , Rouel Lanche, B.S. , Adam J. Mansfield, B.A. , Evan K. Maxwell, Ph.D. , George Mitra, B.A. , Mona Nafde, M.S. , Sean O'Keeffe, Ph.D. , Max Orelus, B.B.A. , Razvan Panea, Ph.D. , Tommy Polanco, B.A. , Ayesha Rasool, M.S. , Jeffrey G. Reid, Ph.D. , William Salerno, Ph.D. , Jeffrey C. Staples, Ph.D. , Kathie Sun, Ph.D.

Analytical Genomics and Data Science

Goncalo Abecasis, D.Phil. , Joshua Backman, Ph.D. , Amy Damask, Ph.D. , Lee Dobbyn, Ph.D. , Manuel Allen Revez Ferreira, Ph.D. , Arkopravo Ghosh, M.S. , Christopher Gillies, Ph.D. , Lauren Gurski, B.S. , Eric Jorgenson, Ph.D. , Hyun Min Kang, Ph.D. , Michael Kessler, Ph.D. , Jack Kosmicki, Ph.D. , Alexander Li , Ph.D. , Nan Lin, Ph.D. , Daren Liu, M.S. , Adam Locke, Ph.D. , Jonathan Marchini, Ph.D. , Anthony Marcketta, M.S. , Joelle Mbatchou, Ph.D. , Arden Moscati, Ph.D. , Charles Paulding, Ph.D. , Carlo Sidore, Ph.D. , Eli Stahl, Ph.D. , Kyoko Watanabe, Ph.D. , Bin Ye, Ph.D. , Blair Zhang, Ph.D. , Andrey Ziyatdinov, Ph.D.

Therapeutic Area Genetics

Ariane Ayer, B.S. , Aysegul Guvenek, Ph.D. , George Hindy, Ph.D. , Giovanni Coppola, M.D. , Jan Freudenberg, M.D. , Jonas Bovijn M.D. , Katherine Siminovitch, M.D. , Kavita Praveen, Ph.D. , Luca A. Lotta, M.D. , Manav Kapoor, Ph.D. , Mary Haas, Ph.D. , Moeen Riaz , Ph.D. , Niek Verweij, Ph.D. , Olukayode Sosina, Ph.D. , Parsa Akbari, Ph.D. , Priyanka Nakka, Ph.D. , Sahar Gelfman, Ph.D. , Sujit Gokhale, B.E. , Tanima De, Ph.D. , Veera Rajagopal, Ph.D. , Alan Shuldiner, M.D. , Bin Ye, Ph.D. , Gannie Tzoneva, Ph.D. , Juan Rodriguez-Flores, Ph.D.

Research Program Management & Strategic Initiatives

Esteban Chen, M.S. , Marcus B. Jones, Ph.D. , Michelle G. LeBlanc, Ph.D. , Jason Mighty, Ph.D., Lyndon J. Mitnaul, Ph.D. , Nirupama Nishtala, Ph.D. , Nadia Rana, Ph.D. , Jaimee Hernandez

Acknowledgements

We acknowledge David Crosslin for helping clean the eMERGE data.

Funding

M.D.R., S.D., and D.H. are funded by AI077505. W.K.C is funded by grant NIDDK DK52431. J.F.P. is funded by grant U01 HG011166. C.W. is funded by U01 HG008680. This phase of the eMERGE Network was initiated and funded by the NHGRI through the following grants: U01HG008657 (Group Health Cooperative/University of Washington); U01HG008685 (Brigham and Women's Hospital); U01HG008672 (Vanderbilt University Medical Center); U01HG008666 (Cincinnati Children's Hospital Medical Center); U01HG006379 (Mayo Clinic); U01HG008679 (Geisinger Clinic); U01HG008680 (Columbia University Health Sciences); U01HG008684 (Children's Hospital of Philadelphia); U01HG008673 (Northwestern University); U01HG008701 (Vanderbilt University Medical Center serving as the Coordinating

599 Center); U01HG008676 (Partners Healthcare/Broad Institute); U01HG008664 (Baylor College
600 of Medicine); and U54MD007593 (Meharry Medical College).
601
602 **References**

603 1. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current
604 polygenic risk scores may exacerbate health disparities. Nat Genet. 2019 Apr;51(4):584–91.

605 2. Wang Y, Guo J, Ni G, Yang J, Visscher PM, Yengo L. Theoretical and empirical
606 quantification of the accuracy of polygenic scores in ancestry divergent populations. Nat
607 Commun. 2020 Jul 31;11(1):3865.

608 3. Galinsky KJ, Reshef YA, Finucane HK, Loh PR, Zaitlen N, Patterson NJ, et al. Estimating
609 cross-population genetic correlations of causal effect sizes. Genet Epidemiol. 2019
610 Mar;43(2):180–8.

611 4. Shi H, Gazal S, Kanai M, Koch EM, Schoech AP, Siewert KM, et al. Population-specific
612 causal disease effect sizes in functionally important regions impacted by selection. Nat
613 Commun. 2021 Feb 17;12(1):1098.

614 5. Rask-Andersen M, Karlsson T, Ek WE, Johansson Å. Gene-environment interaction study
615 for BMI reveals interactions between genetic factors and physical activity, alcohol
616 consumption and socioeconomic status. PLoS Genet. 2017 Sep;13(9):e1006977.

617 6. Robinson MR, English G, Moser G, Lloyd-Jones LR, Triplett MA, Zhu Z, et al. Genotype-
618 covariate interaction effects and the heritability of adult body mass index. Nat Genet. 2017
619 Aug;49(8):1174–81.

620 7. Sulc J, Mounier N, Günther F, Winkler T, Wood AR, Frayling TM, et al. Quantification of
621 the overall contribution of gene-environment interaction for obesity-related traits. Nat
622 Commun. 2020 Mar 13;11(1):1385.

623 8. Justice AE, Winkler TW, Feitosa MF, Graff M, Fisher VA, Young K, et al. Genome-wide
624 meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for
625 obesity traits. Nat Commun. 2017 Apr 26;8:14977.

626 9. Helgeland Ø, Vaudel M, Juliusson PB, Lingaas Holmen O, Juodakis J, Bacelis J, et al.
627 Genome-wide association study reveals dynamic role of genetic variation in infant and early
628 childhood growth. Nat Commun. 2019 Oct 1;10(1):4448.

629 10. Vogelezang S, Bradfield JP, Ahluwalia TS, Curtin JA, Lakka TA, Grarup N, et al. Novel loci
630 for childhood body mass index and shared heritability with adult cardiometabolic traits.
631 PLoS Genet. 2020 Oct;16(10):e1008718.

632 11. Couto Alves A, De Silva NMG, Karhunen V, Sovio U, Das S, Taal HR, et al. GWAS on
633 longitudinal growth traits reveals different genetic factors influencing infant, child, and adult
634 BMI. Sci Adv. 2019 Sep;5(9):eaaw3095.

12. Choh AC, Lee M, Kent JW, Diego VP, Johnson W, Curran JE, et al. Gene-by-age effects on BMI from birth to adulthood: the Fels Longitudinal Study. Obes Silver Spring Md. 2014 Mar;22(3):875–81.

13. Mostafavi H, Harpak A, Agarwal I, Conley D, Pritchard JK, Przeworski M. Variable prediction accuracy of polygenic scores within an ancestry group. eLife. 2020 Jan 30;9:e48376.

14. Elks CE, den Hoed M, Zhao JH, Sharp SJ, Wareham NJ, Loos RJF, et al. Variability in the heritability of body mass index: a systematic review and meta-regression. Front Endocrinol. 2012;3:29.

15. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018 Oct;562(7726):203–9.

16. Stanaway IB, Hall TO, Rosenthal EA, Palmer M, Naranbhai V, Knevel R, et al. The eMERGE genotype set of 83,717 subjects imputed to ~40 million variants genome wide and association with the herpes zoster medical record phenotype. Genet Epidemiol. 2019 Feb;43(1):63–81.

17. Zhang X, Lucas AM, Veturi Y, Drivas TG, Bone WP, Verma A, et al. Large-scale genomic analyses reveal insights into pleiotropy across circulatory system diseases and nervous system disorders. Nat Commun. 2022 Jun 14;13(1):3428.

18. 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.

19. Banda Y, Kvale MN, Hoffmann TJ, Hesselson SE, Ranatunga D, Tang H, et al. Characterizing Race/Ethnicity and Genetic Ancestry for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. Genetics. 2015 Aug;200(4):1285–95.

20. Penn Medicine BioBank [Internet]. Available from: https://pmbb.med.upenn.edu/

21. Tyrrell J, Wood AR, Ames RM, Yaghootkar H, Beaumont RN, Jones SE, et al. Gene-obesogenic environment interactions in the UK Biobank study. Int J Epidemiol. 2017 Apr 1;46(2):559–75.

22. Young AI, Wauthier F, Donnelly P. Multiple novel gene-by-environment interactions modify the effect of FTO variants on body mass index. Nat Commun. 2016 Sep 6;7:12724.

23. Winkler TW, Justice AE, Graff M, Barata L, Feitosa MF, Chu S, et al. The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. PLoS Genet. 2015 Oct;11(10):e1005378.

24. Ruan Y, Lin YF, Feng YCA, Chen CY, Lam M, Guo Z, et al. Improving polygenic prediction in ancestrally diverse populations. Nat Genet. 2022 May;54(5):573–80.

670 25. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body
671 mass index yield new insights for obesity biology. Nature. 2015 Feb 12;518(7538):197–206.

672 26. Ng MCY, Graff M, Lu Y, Justice AE, Mudgal P, Liu CT, et al. Discovery and fine-mapping
673 of adiposity loci using high density imputation of genome-wide association studies in
674 individuals of African ancestry: African Ancestry Anthropometry Genetics Consortium.
675 PLoS Genet. 2017 Apr;13(4):e1006719.

676 27. Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, et al. A cross-
677 population atlas of genetic associations for 220 human phenotypes. Nat Genet. 2021
678 Oct;53(10):1415–24.

679 28. Aschard H, Vilhjálmsson BJ, Joshi AD, Price AL, Kraft P. Adjusting for heritable covariates
680 can bias effect estimates in genome-wide association studies. Am J Hum Genet. 2015 Feb
681 5;96(2):329–39.

682 29. Kerin M, Marchini J. Inferring Gene-by-Environment Interactions with a Bayesian Whole-
683 Genome Regression Model. Am J Hum Genet. 2020 Oct 1;107(4):698–713.

684 30. Vanderweele TJ, Ko YA, Mukherjee B. Environmental confounding in gene-environment
685 interaction studies. Am J Epidemiol. 2013 Jul 1;178(1):144–52.

686 31. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression
687 and continuous shrinkage priors. Nat Commun. 2019 Apr 16;10(1):1776.

688 32. Afshartous D, Preston RA. Key Results of Interaction Models with Centering. J Stat Educ.
689 2011 Nov;19(3):1.

690 33. Kilpeläinen TO, Qi L, Brage S, Sharp SJ, Sonestedt E, Demerath E, et al. Physical activity
691 attenuates the influence of FTO variants on obesity risk: a meta-analysis of 218,166 adults
692 and 19,268 children. PLoS Med. 2011 Nov;8(11):e1001116.

693 34. Rampersaud E, Mitchell BD, Pollin TI, Fu M, Shen H, O'Connell JR, et al. Physical activity
694 and the association of common FTO gene variants with body mass index and obesity. Arch
695 Intern Med. 2008 Sep 8;168(16):1791–7.

696 35. Amin V, Dunn P, Spector T. Does education attenuate the genetic risk of obesity? Evidence
697 from U.K. Twins. Econ Hum Biol. 2018 Sep;31:200–8.

698 36. Li Y, Cai T, Wang H, Guo G. Achieved educational attainment, inherited genetic
699 endowment for education, and obesity. Biodemography Soc Biol. 2021 Jun;66(2):132–44.

700 37. Frank M, Dragano N, Arendt M, Forstner AJ, Nöthen MM, Moebus S, et al. A genetic sum
701 score of risk alleles associated with body mass index interacts with socioeconomic position
702 in the Heinz Nixdorf Recall Study. PloS One. 2019;14(8):e0221252.

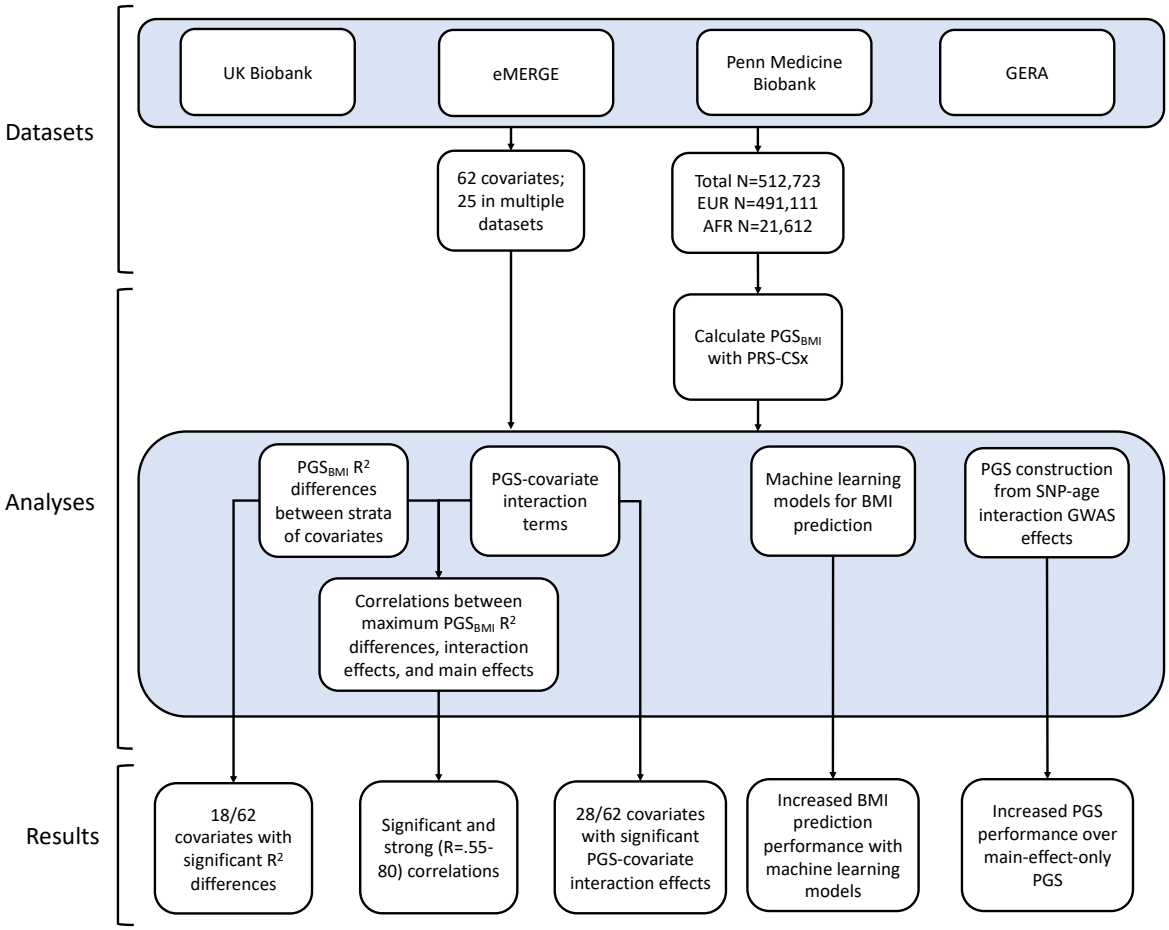
703 38. Ge T, Chen CY, Neale BM, Sabuncu MR, Smoller JW. Phenome-wide heritability analysis
704 of the UK Biobank. PLoS Genet. 2017 Apr;13(4):e1006711.

39. Min J, Chiu DT, Wang Y. Variation in the heritability of body mass index based on diverse twin studies: a systematic review. Obes Rev Off J Int Assoc Study Obes. 2013 Nov;14(11):871–82.

40. Linder JE, Allworth A, Bland ST, Caraballo PJ, Chisholm RL, Clayton EW, et al. Returning integrated genomic risk and clinical recommendations: The eMERGE study. Genet Med Off J Am Coll Med Genet. 2023 Jan 6;25(4):100006.

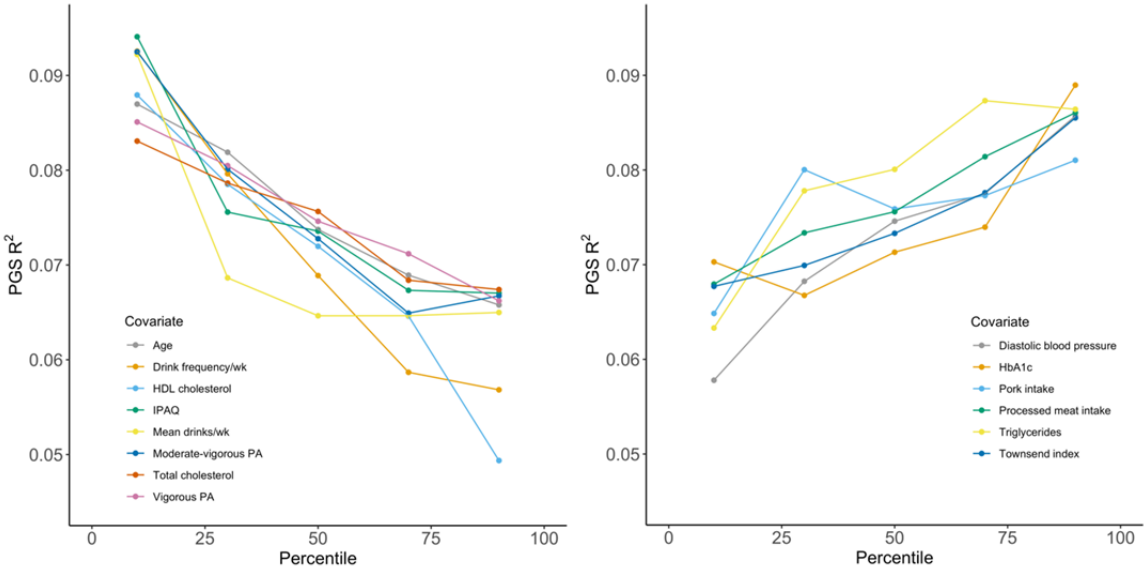
41. Robinson PN, Arteaga-Solis E, Baldock C, Collod-Béroud G, Booms P, De Paepe A, et al. The molecular genetics of Marfan syndrome and related disorders. J Med Genet. 2006 Oct;43(10):769–87.

749 **Figures**



750
751 Figure 1. A flowchart of the project.
752
753
754
755
756
757
758
759
760
761
762

a)



b)

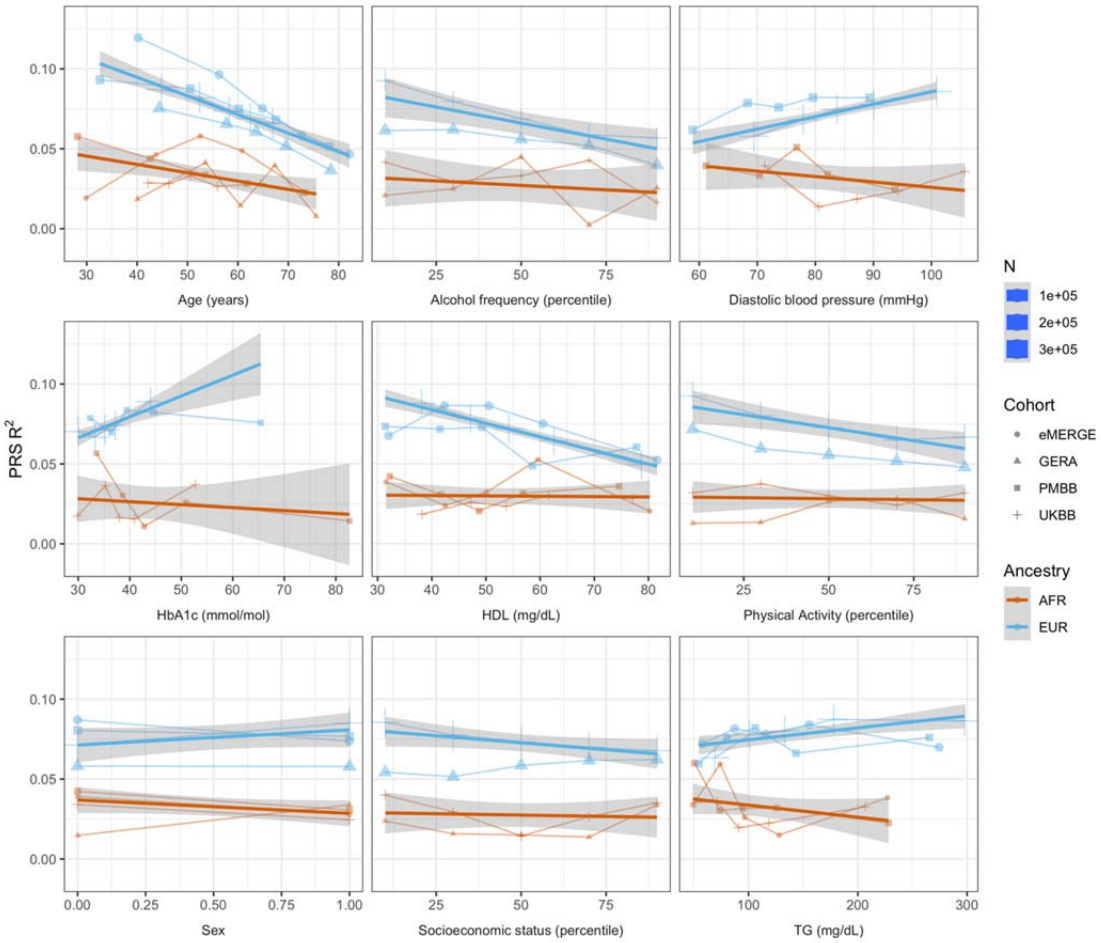


Figure 2. PGS R^2 stratified by quintiles for quantitative variables and by binary variables. a) Continuous covariates with significant ($p < 7.7x10^{-4}$) R^2 differences across quintiles in UKBB

770 EUR. Pork and processed meat consumption per week were excluded from this plot in favor of
771 pork and processed meat intake. b) Covariates with significant differences that were available in
772 multiple cohorts, note actual values instead of percentiles plotted on x-axis (except percentiles
773 were used for physical activity, alcohol intake frequency, and socioeconomic status, which had
774 slightly differing phenotype definitions across cohorts). Townsend index and income were used
775 as variables for socioeconomic status UKBB and GERA, respectively. Note that the sign for
776 Townsend index was reversed, since increasing Townsend index is lower socioeconomic status,
777 while increasing income is higher socioeconomic status. Abbreviations: physical activity (PA),
778 International Physical Activity Questionnaire IPAQ).
779
780
781
782
783
784
785
786

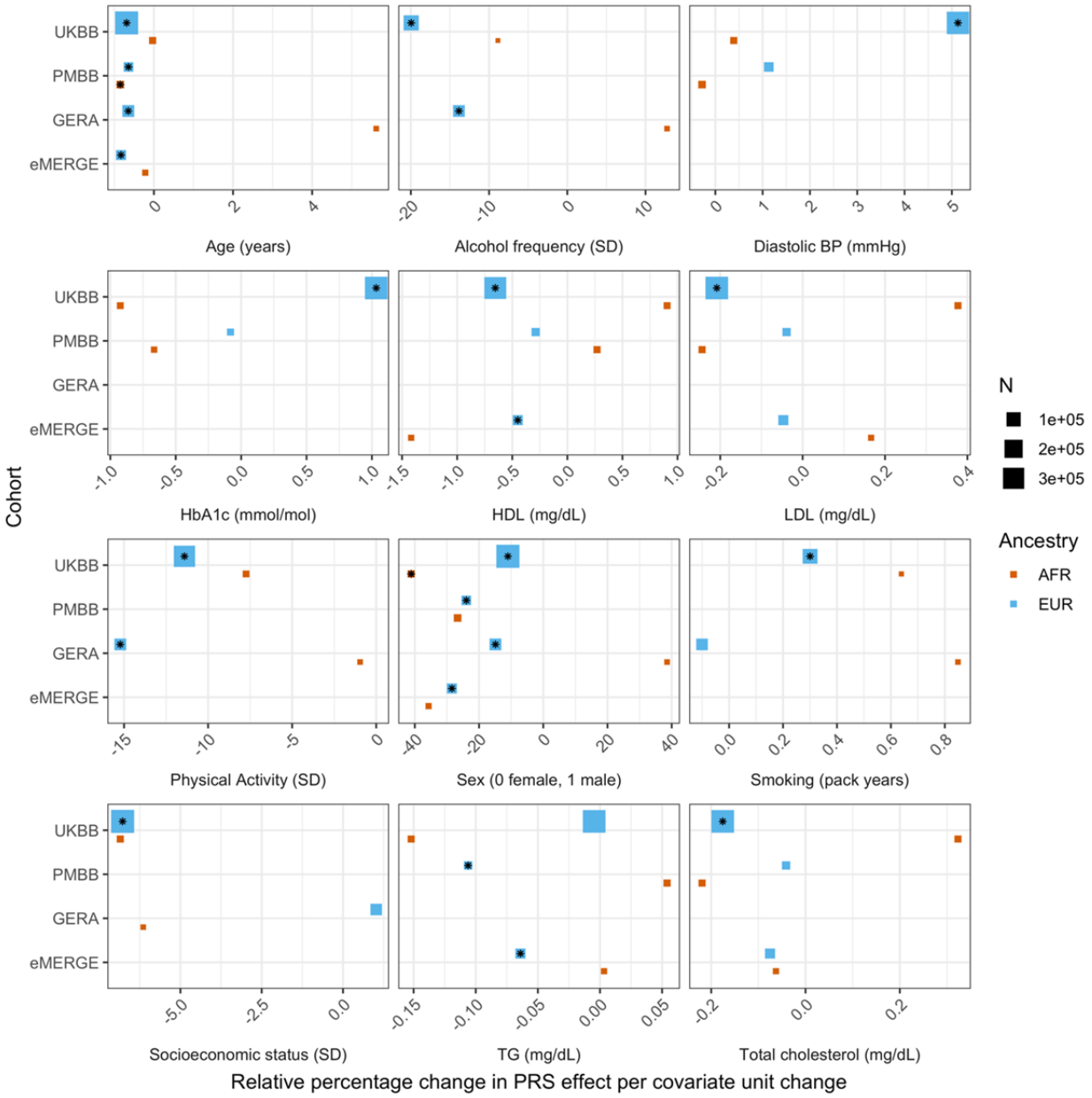


Figure 3. Relative percentage changes in PGS effect per unit change in covariate, for covariates that significantly changed PGS effect (i.e., significant interaction beta at Bonferroni $p < 7.7x10^{-4}$ – denoted by asterisks) and were present in multiple cohorts and ancestries. Same covariate groupings and transformations were performed as with Figure 1.

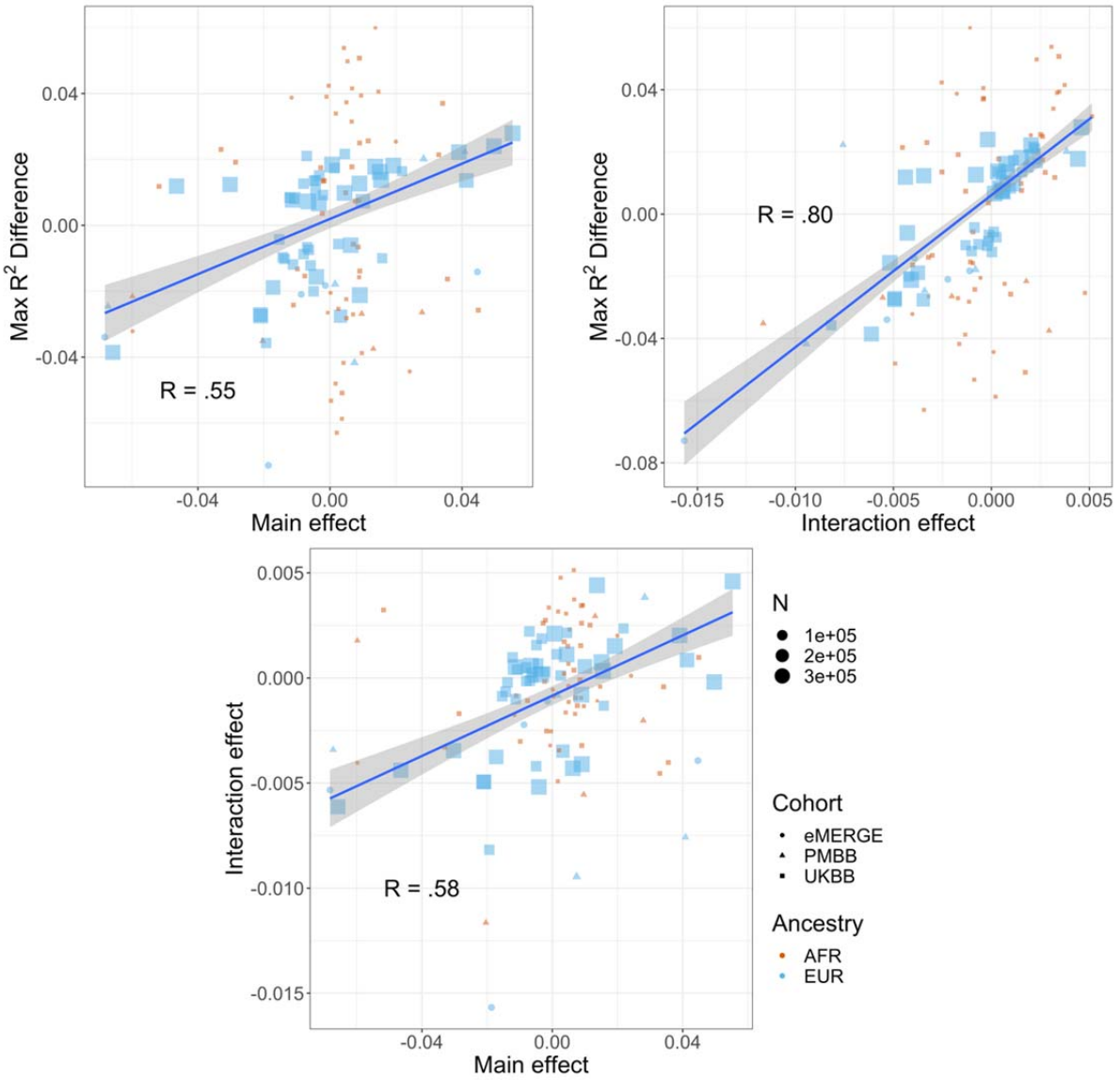


Figure 4. Pearson correlations weighted by sample size between maximum R^2 differences across strata, main effects of covariate on log(BMI), and PGS-covariate interaction effects on log(BMI). Main effect units are in standard deviations, interaction effect units are in PGS standard deviations multiplied by covariate standard deviations. Only continuous variables are plotted and modeled. GERA was excluded due to slightly different phenotype definitions.

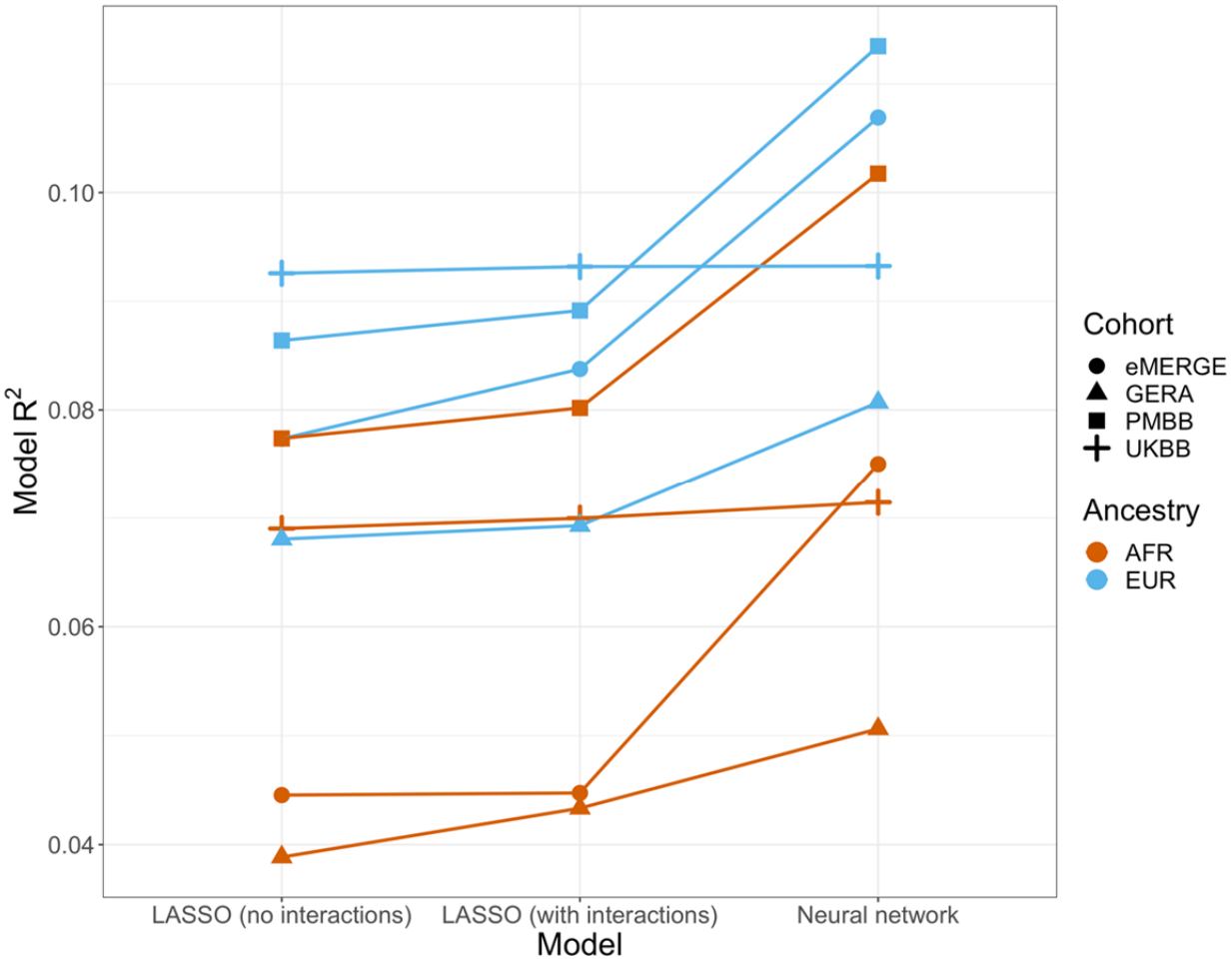


Figure 5. Model R^2 across cohorts and ancestries using age and gender as covariates (along with PGS_{BMI} and PCs 1-5). Across all cohorts and ancestries, LASSO with PGS-age and PGS-gender interaction terms had better average 10-fold cross-validation R^2 than LASSO without interaction terms, while neural networks outperformed LASSO models.

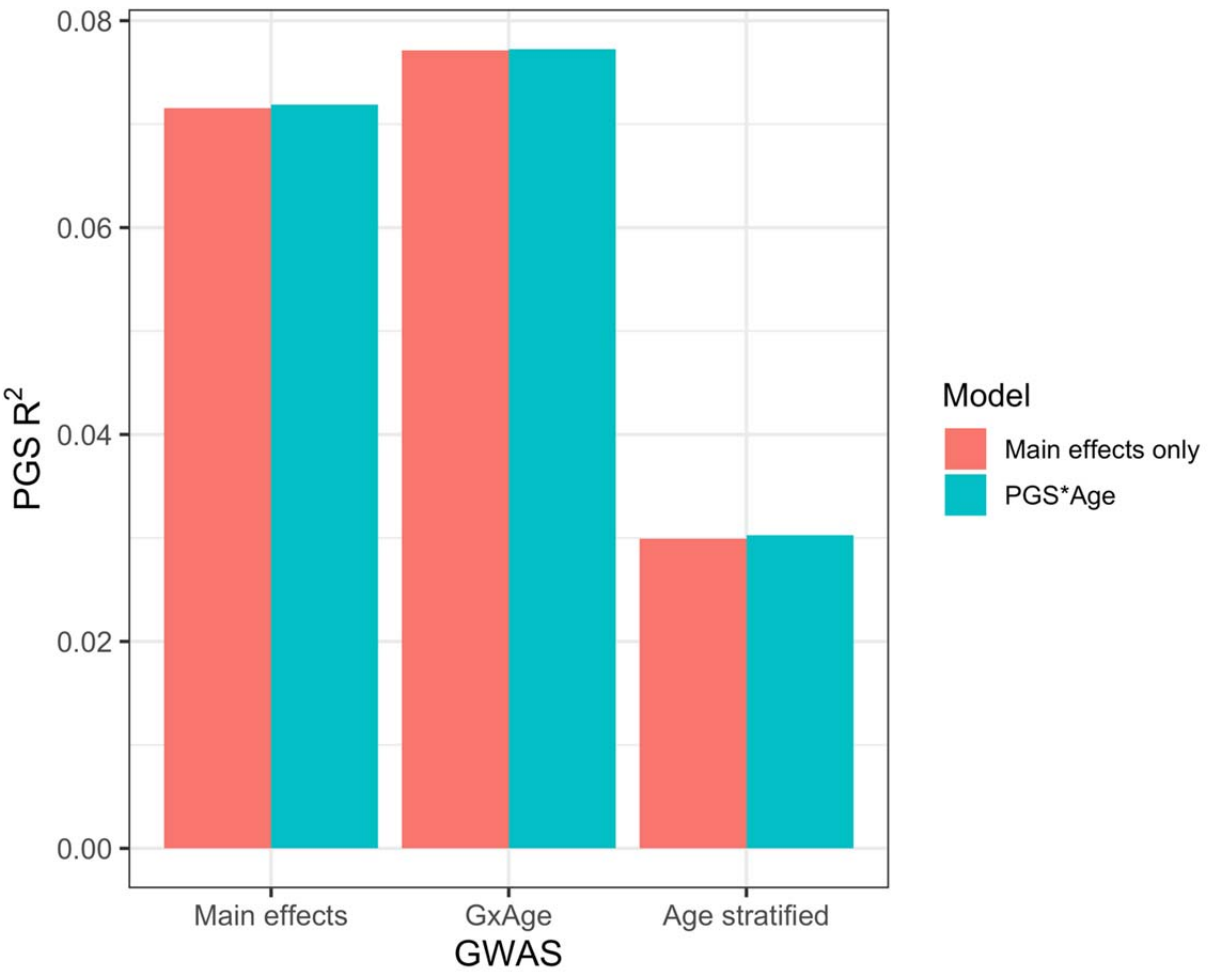
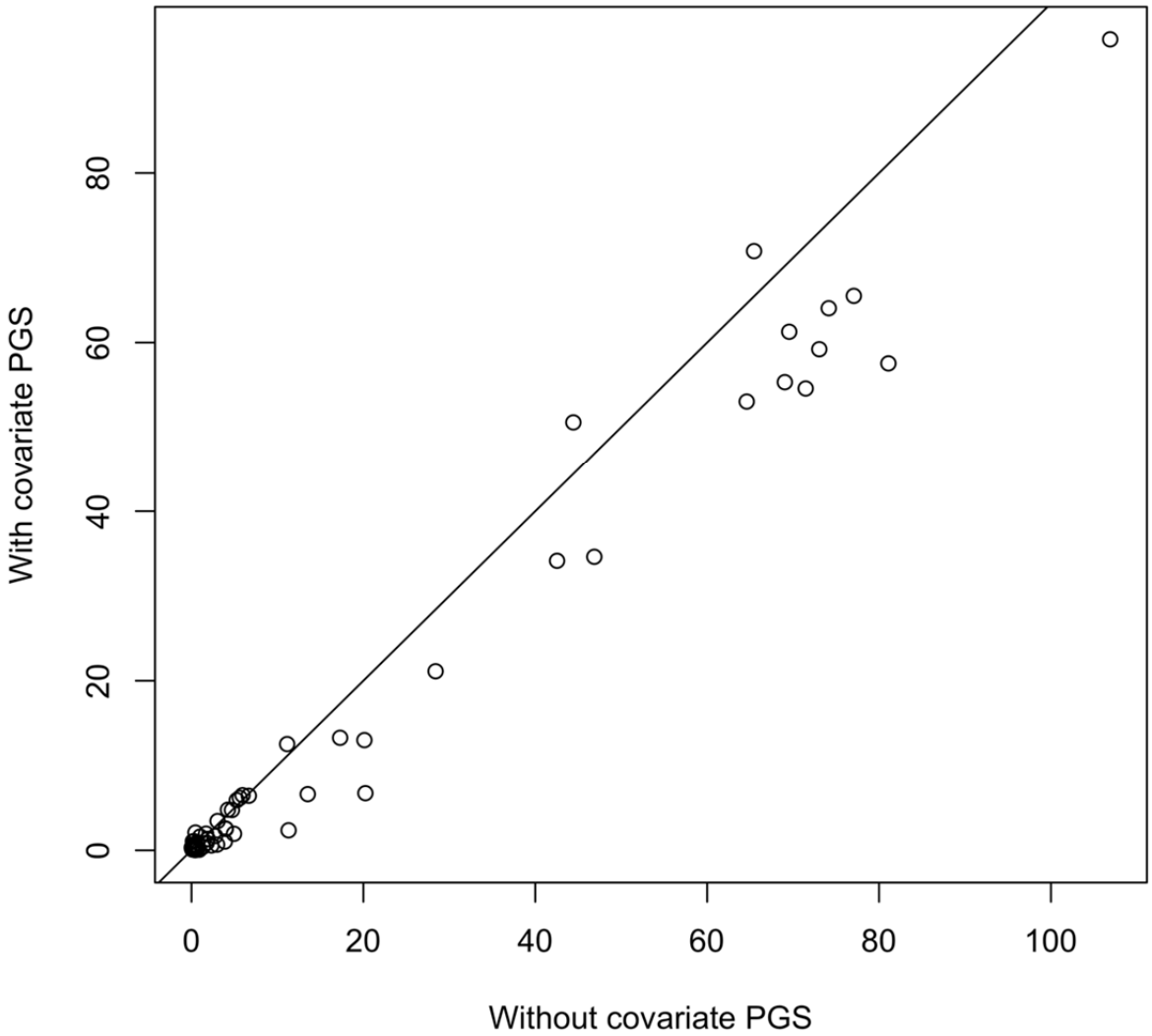


Figure 6. PGS R^2 based on three sets of GWAS. "Main effects" were from a typical main effect GWAS, "GxAge" effects were from a GWAS with a SNP-age interaction term, and "Age stratified" GWAS had main effects only but were conducted in four age quartiles. PGS R^2 was evaluated using two models: one with main effects only, and one with an additional PGS*Age interaction term.

858 Supplemental Figure:



859
860 S Figure 1. PGS-covariate interaction term p-values in UKBB EUR, with and without including
861 the covariate PGS in the model – the mean $-\log_{10}(p)$ is reduced from 18.0899 to 14.97072 with
862 their inclusions. Note age and sex PGS were not calculated, and their interaction p-values are
863 excluded from this figure.
864
865
866
867
868
869
870
871

872 **Table**

Variable type	Covariate	% change in β_{PGS} per covariate unit change	Interaction P	R^2 increase with interaction term	N
Continuous	HDL cholesterol	-15.29	1.71×10^{-96}	0.0012	328719
	Total cholesterol	-12.70	1.64×10^{-71}	0.00082	359221
	IPAQ	-12.50	3.11×10^{-66}	0.001	304951
	Moderate-vigorous PA	-11.41	8.92×10^{-65}	0.001	304951
	Diastolic BP	10.84	6.06×10^{-60}	0.0007	352804
	Townsend Index	6.78	2.86×10^{-58}	0.00089	376283
	Age	-9.02	3.60×10^{-57}	0.00061	376729
	FVC	-9.66	4.69×10^{-56}	0.0008	343467
	Drink frequency/wk	-19.96	2.62×10^{-55}	0.0024	122281
	LDL cholesterol	-9.86	2.63×10^{-51}	0.00058	358556
	N days vigorous PA/wk	-9.37	2.42×10^{-35}	0.0007	299963
	FEV1	-7.38	7.15×10^{-35}	0.0005	343544
	Mean alcohol consumption	-7.38	7.65×10^{-22}	0.00113	126756
	HbA1c	4.63	5.37×10^{-14}	0.0002	358798
	Mean drinks/wk	-7.66	1.01×10^{-13}	0.0008	112204
	Water intake	4.60	2.97×10^{-13}	0.00014	347472
	Processed meat intake	3.70	2.38×10^{-7}	0.0002	376205
	Starch mean	5.51	3.15×10^{-7}	0.00018	128346
	Smoking pack years	4.78	3.68×10^{-7}	0.0002	114135
	Protein mean	4.82	6.52×10^{-7}	0.00018	128181
	Saturated fat mean	4.92	1.23×10^{-6}	0.00017	127899
	Fat mean	4.40	1.64×10^{-5}	0.00013	128092
	Saturated fat grams/wk	2.46	1.79×10^{-5}	4.00E-05	364629
	Retinol mean	3.77	3.54×10^{-4}	9.00E-05	126029
Binary	IPAQ	-12.68	5.30×10^{-62}	0.0009	304951
	Vigorous PA/wk	-20.55	9.07×10^{-54}	0.0009	304951
	Sex	-11.02	1.41×10^{-24}	0.00025	376729
	Diabetes	27.19	1.83×10^{-7}	0.0004	375903

873

874 Table 1. Model descriptive statistics on 28 of 62 covariates, which have significant ($p<.05/62$)
875 PGS-covariate interaction terms, in UKBB EUR. The third column is the percentage change in
876 PGS effect per unit change (standard deviations for continuous variables, binary variables
877 encoded as 0 or 1) in covariate. The fifth column is the increase in model R^2 with a PGS-
878 covariate interaction term versus a main effects only model. Abbreviations: blood pressure (BP),
879 physical activity (PA), forced vital capacity (FVC), forced expiratory volume in 1-second
880 (FEV1), International Physical Activity Questionnaire (IPAQ).