Quantifying factors that affect polygenic risk score performance across diverse ancestries and age groups for body mass index

Daniel Hui¹*, Brenda Xiao¹*, Ozan Dikilitas², Robert R. Freimuth³, Marguerite R. Irvin⁴, Gail P. Jarvik⁵, Leah Kottyan⁶, Ifitikhar Kullo⁷, Nita A. Limdi⁸, Cong Liu⁹, Yuan Luo¹⁰, Bahram Namjou¹¹, Megan J. Puckelwartz¹², Daniel Schaid¹³, Hemant Tiwari¹⁴, Wei-Qi Wei¹⁵, Shefali Verma¹⁶, Dokyoon Kim¹⁷, Marylyn D. Ritchie¹⁸**

Affiliations:
¹ Graduate Program in Genomics and Computational Biology, University of Pennsylvania, Philadelphia, PA, USA
² Department of Internal Medicine, Department of Cardiovascular Medicine, Clinician-Investigator Training Program, Mayo Clinic, Rochester MN
³ Department of Artificial Intelligence and Informatics, Mayo Clinic, Rochester, MN, USA
⁴ Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL, United States
⁵ Departments of Medicine and Genome Sciences, University of Washington, Seattle WA, USA
⁶ Center for Autoimmune Genomics and Etiology, Department of Pediatrics, University of Cincinnati, Cincinnati, OH, USA
⁷ Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN 55905, USA
⁸ Department of Neurology & Epidemiology, University of Alabama at Birmingham, Birmingham, AL, USA
⁹ Department of Biomedical Informatics, Columbia University, New York, NY, USA
¹⁰ Department of Preventive Medicine (Health and Biomedical Informatics), Northwestern University, Chicago, IL USA
¹¹ Department of Pediatrics, University of Cincinnati, Cincinnati, OH, USA
¹² Department of Pharmacology, Northwestern University, Chicago, IL USA
¹³ Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA.
¹⁴ Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL, United States
¹⁵ Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA
¹⁶ Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
¹⁷ Department of Biostatistics, Epidemiology and Informatics, Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
¹⁸ Department of Genetics, Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

* Equal contributions to the manuscript

**Correspondence to:
Marylyn D. Ritchie, PhD, FACMI
Perelman School of Medicine
University of Pennsylvania

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract
Polygenic risk scores (PRS) have led to enthusiasm for precision medicine. However, it is well documented that PRS do not generalize across groups differing in ancestry or sample characteristics e.g., age. Quantifying performance of PRS across different groups of study participants, using genome-wide association study (GWAS) summary statistics from multiple ancestry groups and sample sizes, and using different linkage disequilibrium (LD) reference panels may clarify factors limiting PRS transferability. To evaluate these factors in the PRS generation process, we generated body mass index (BMI) PRS (PRSBMI) in the Electronic Medical Records and Genomics network (N=75,661). Analyses were conducted in two ancestry groups (European and African) and three age ranges (adult, teenagers, and children). For PRSBMI calculations, we evaluated five LD reference panels and three GWAS summary statistics of varying sample size and ancestry. PRSBMI performance increased for both African and European ancestry individuals using cross-ancestry GWAS summary statistics compared to European-only summary statistics (6.3% and 3.7% relative R² increase, respectively, pAfrican=0.038, pEuropean=6.26x10⁻⁴). The effects of LD reference panels were more pronounced in African ancestry study datasets. PRSBMI performance degraded in children; R² was less than half of teenagers or adults. The effect of GWAS summary statistics sample size was small when modeled with the other factors. We also explored clinical comorbidities associated with the PRSBMI and identified associations with type 2 diabetes and coronary atherosclerosis. This study quantifies effects that ancestry, GWAS summary statistic sample size, and LD reference panel have on PRS performance, especially in cross-ancestry and age-specific analyses.
Introduction

Polygenic risk scores (PRS) provide individualized genetic estimates of a phenotype by aggregating genetic effects across hundreds or thousands of loci, typically from genome-wide association studies (GWAS). PRS are potentially a powerful source of increased prediction performance, even when combined with familial history (1,2). However, in recent years it has become increasingly apparent that performance of PRS is substantially reduced when the ancestry of the individuals in whom prediction is being done differs from the ancestry of the individuals from the GWAS used to generate SNP weights used for PRS construction. For instance, when using GWAS from European ancestry individuals, the prediction accuracy of polygenic scores in individuals of African or Hispanic/Latino ancestry have a relative performance of 25% and 65% compared to performance in European ancestry individuals (3). Additionally, evidence exists suggesting that for some traits, such as adiposity traits, this disparity may be further exacerbated by environmental, demographic, or social risk factors (including age, physical activity, smoking status, and alcohol use (4–7)). For example, differences in the genetic architecture of body mass index (BMI) has been shown to differ between age groups (8–11). Thus, the performance of PRS for BMI is also affected by the age of the individuals used in the GWAS and the test dataset (12). Broad-sense heritability estimates for BMI in adults range from 40%-90% when estimated in adults of different cohorts even of homogeneous ancestry (13) – even if heritability estimates are similar across populations, genetic architecture and enrichment for variants in different functional categories may still differ (14,15).

Several outstanding questions surrounding PRS, especially within the context of adiposity traits and BMI, warrant further investigation. For instance, when cross-ancestry summary statistics (i.e., those including individuals of multiple ancestry groups) are available, can they be used to improve prediction performance in individuals from one or more different ancestry groups. We need a more thorough evaluation of the potential prediction performance gain (or loss) in African ancestry individuals when cross-ancestry GWAS summary statistics are used to estimate the SNP weights. In addition, we need to improve our understanding of the impact of the composition of the linkage disequilibrium (LD) reference panel in combination with cross-ancestry GWAS summary statistics on the PRS prediction performance. For prediction of BMI specifically, how does prediction performance differ for individuals in different age groups, especially those who are not adults (i.e., less than age 18). Additionally, how much do these different variables impact the PRS performance when considered together; and how much do these differences affect associations with clinical comorbidities of obesity. Finally, the degree to which increased GWAS sample size increases prediction performance regardless of these other factors is important to determine.

We comprehensively investigated the influence of these factors on the performance of PRS using the Electronic Medical Records and Genomics (eMERGE) Network dataset. eMERGE is a prospective study that combines participants from multiple electronic health record (EHR) linked biobanks (16). In the present study, we included 75,661 individuals of diverse ancestry and age (14% African ancestry, 55% female, and 12% children age <13). These individuals were from the eMERGE III imputed array dataset (N=83,717) (dbGaP Study: phs001584.v2.p2), estimated European or African ancestry, and had BMI measurements available. For these analyses, we used published BMI GWAS summary statistics from the GIANT (Genetic Investigation of ANthropometric Traits) consortium, an international consortium that primarily studies anthropometric traits, which included participants (max N=339,224, mean N per variant=226,960) from European, African, and Asian ancestry groups.
We also used summary statistics from a European ancestry BMI GWAS (18) in the UK Biobank (UKBB) individuals (N=339,721), which was conducted using both the full sample size of the European ancestry UKBB, as well as after down sampling to the same number of individuals in the GIANT GWAS. This comparison allowed us to better evaluate whether it was the ancestry composition or the sample size of the dataset where the GWAS summary statistics were derived that affected the results of the PRS performance. We calculated PRS for BMI (PRSBMI) across 90 different combinations of analyses (described more in Methods) – 6 different groupings based on ancestry and age, five different LD reference panels (of varying ancestry and from three different cohorts), and the three mentioned sets of summary statistics. We then statistically compared the different sets of analyses to see what factors most influence the PRSBMI performance across various groupings of individuals based on ancestry and age. Lastly, we also tested the association of the best performing PRSBMI with common comorbidities across ancestry groups to identify the clinical relevance of the PRSBMI in phenotypes derived from an Electronic Health Record (EHR). Investigation of these variables elucidates our understanding of the factors that affect PRS performance and transferability across ancestries and populations, especially within the context of BMI.

**Methods**

**Overall study design**

The electronic Medical Records and Genomics (eMERGE) network dataset is a prospective study that combines participants from multiple electronic health record (EHR) linked biobanks. In this study, we included 75,661 individuals with available genetic and phenotypic data. The individuals in the eMERGE dataset include multiple ancestry groups and a large age distribution (14% genetically inferred African ancestry, 19% less than age 18, Figure 1). Briefly, we calculated PRSBMI for all individuals, for each combination where the following elements of the model varied: 1) LD panels that differed in ancestry, 2) GWAS summary statistics with variable ancestry composition, and 3) GWAS summary statistics for two different sample sizes. The details for each of these are provided more below. We then split the data by ancestry and age group, and statistically compared PRSBMI performance between all the different groups – in total, 90 sets of PRSBMI were compared. We first estimated the effect and significance of each variable (i.e., ancestry of GWAS summary statistics and test data, LD panel ancestry, size of GWAS summary statistics, and age of test individuals) on PRS performance. Next, we estimated how much each variable affects PRSBMI performance when all are modeled together, and finally we analyzed the potential clinical associations by testing the PRSBMI for association with common comorbid conditions from the EHR. For the primary results related to LD panel or ancestry of summary statistics and test data, we restricted analyses to adults as the other age groups were limited in sample size.

**Summary statistics to generate PRSBMI**

We obtained published GWAS summary statistics from the GIANT consortium (17) to use as one set of BMI GWAS summary statistics. Up to 322,154 adults of European ancestry, as well as an additional 17,072 adults of non-European ancestry (adults of African, East Asian, and South Asian ancestry), were included in the GIANT GWAS analysis.

For the second set of summary statistics, we performed a GWAS in the individuals of European ancestry from the UK Biobank (UKBB). Individuals were first filtered by low quality samples (sex mismatch between genetically inferred and self-reported, variant missingness
>5%), relatedness (no 2nd degree relatives or higher), and within the White British ancestry subset (with these individuals being defined by UKBB and selected based on self-reports and genetically determined ancestry) (18) – 377,921 individuals initially remained. Variants were filtered on imputation quality score (using the INFO metric (19)) > 0.30, and minor allele frequency >1% within this subset of individuals. In addition, we generated a second set of GWAS summary statistics from the UKBB, where we randomly down-sampled individuals to the sample size in the GIANT GWAS dataset (N=226,960). In each UKBB GWAS, data processing and modeling were performed similarly as in the GIANT GWAS – summary statistics were calculated using linear regression, with age, age$^2$, sex, and the first 5 genetic principal components (PCs) included as covariates. BMI, defined as weight in kilograms divided by squared height in meters, was first inverse-rank normal transformed.

After calculation of BMI GWAS summary statistics in each of the two datasets of UKBB individuals of European ancestry, we harmonized variants across all datasets used (UKBB, eMERGE, GIANT, and 1000 Genomes Phase 3). For the remainder of downstream analyses, we kept only those variants that were present in all datasets, and additionally excluded any strand-ambiguous SNPs (alleles A/T or C/G), and retained only biallelic variants; in total, 2,014,457 variants were retained for analyses.

**LD reference panels**

Five different LD reference panels were used for each set of PRS$_{BMI}$ calculations: 1) all of 1000 Genomes (1KGAll) (N=2,504), 2) 1000 Genomes European ancestry (1KGEUR) (N=503), 3) 5,000 randomly selected European ancestry individuals from the UK Biobank (UKBBEUR), 4) 5,000 randomly selected individuals from all of UK Biobank (UKBBAll), and 5) up to 5,000 randomly selected individuals from the test dataset for which PRS$_{BMI}$ were being calculated for (referred to as test data henceforward).

**Statistical methods**

**PRS software**

PRS$_{BMI}$ were calculated using pruning and thresholding method via PRSice v2.1.9 (20). Default parameters were used in all analyses (clumping performed in 250 kb windows using an R$^2$ of 0.1, p-value step size of 0.00005).

**Statistical comparisons**

Partial R$^2$ for PRS$_{BMI}$ was calculated by subtracting the R$^2$ using a model with only the covariates from the R$^2$ of the model using the covariates and the PRS$_{BMI}$ (the default option in PRSice). P-values for the differences between model performance from all different iterations were calculated by using the Wilcoxon rank-sum test between the distributions of the squared residuals of the different models – for comparisons between the same set of individuals, the paired Wilcoxon rank-sum test was used. When testing which of the five LD panels performed the best, we used a Bonferroni-corrected threshold of 0.05/10 = 0.005 (ten comparisons between five LD panels). When comparing the best performing PRS$_{BMI}$ across ancestries and summary statistics using their best LD panel, we used a Bonferroni threshold of 0.05/25 = 0.002 (25 comparisons between the five LD panels used).

**Proportion of variance explained by each individual variable**

We modeled all evaluated features together in the following linear regression model:
\[ R^2 \sim LD\ panel + N_{Sumstat} + \text{Age}_{Test} + \text{Ancestry}_{Sumstat} + \text{Ancestry}_{Test} + \text{Ancestry}_{Sumstat} \times \text{Ancestry}_{Test} \]

Where the Sumstat subscript is defined as a set of GWAS summary statistics, and the Test subscript is defined as a set of test individuals that PRS prediction is being assessed in. We quantified the variance in \( R^2 \) that could be explained by each of these different variables using type II sum of squares from ANOVA. The sum of squares of variables involving ancestry were summed together; an interaction term between summary statistics ancestry and test data ancestry was included to identify whether the ancestry of summary statistics and test data matched.

**Association of PRSBMI with comorbidities**
We selected the ten most frequent Phecodes (21) from the EHR data in the eMERGE dataset (which includes obesity as a positive control) to test their association with the PRSBMI (S Table 2). For each Phecode, individuals were classified as a case for the condition if there was at least one occurrence of the respective Phecode in their EHR record; individuals were classified as a control for that condition if there was no occurrence of the Phecode. This classification is a rule-of-one instance of a Phecode to define case status. For each eMERGE ancestry subgroup, we selected the best performing PRSBMI i.e., the PRSBMI with the highest \( R^2 \) and tested the association of the PRSBMI with the most prevalent conditions using a logistic regression model. PRSBMI was first mean-centered and standard deviation was set to 1. Sex, age, age\(^2\), and the first five genetic PCs were included as covariates.

**Data visualization**
The `ggplot2` R package was used for plotting, with the `geom_signif` package used to include significance bars. The association results were plotted using PheWAS-View (22).

**Results**

**Effect of LD panel**
For adults of African ancestry, when using the down-sampled UKBB GWAS summary statistics, using either cross-ancestry or African ancestry LD panels significantly improved PRSBMI performance compared to European ancestry LD panels (Figure 2). When using the UKBB summary statistics, the top PRSBMI \( R^2 \) was 0.0140 using the test data as LD panel, while the second-best performing LD panel (UKBB European) had an \( R^2 \) of 0.0109 (\( p = 4.94 \times 10^{-20} \)). When using the GIANT summary statistics, the top PRSBMI \( R^2 \) was 0.0149 using 1KGall as the reference panel. The PRSBMI calculated using the best European ancestry panel (1KG\text{EUR}) resulted in a \( R^2 \) of 0.141, but this difference between these two reference panels was not Bonferroni significant (\( p = 0.037 \)). However, the 1KG\text{all} LD panel performed significantly better than the two UKBB LD panels (UKBB\text{All}: \( R^2 = 0.134, \ p = 3.65 \times 10^{-5} \); UKBB European: \( R^2 = 0.0128, \ p = 3.65 \times 10^{-9} \)). The test data LD panel performed the second-best with an \( R^2 \) of 0.0142, and significantly outperformed the UKBB European LD panel (\( p = 4.78 \times 10^{-5} \)).

For adults of European ancestry, we observed more significant differences in performance when using the GIANT summary statistics compared to the down-sampled UKBB summary statistics. The 1KG\text{all} LD panel performed the best with a \( R^2 \) of 0.0612. It also significantly outperformed all other LD panels (1KG\text{EUR}: \( R^2 = 0.056, \ p = 5.54 \times 10^{-4} \); Test data: \( R^2 = 0.0564, \ p = 6.50 \times 10^{-67} \); UKBB\text{All}: \( R^2 = 0.0561, \ 8.09 \times 10^{-107} \); UKBB\text{EUR}: \( R^2 = 0.0561, \ p = 3.02 \times 10^{-77} \)). We note that this increase was larger when using the GIANT summary statistics but
was still present when using the UKBB summary statistics. When using the UKBB summary statistics, the choice of LD panel had a much smaller impact on prediction performance. While the 1KGAll LD panel performed the best, the difference in performance was much less significant between the next best performing LD panel ($R^2_{1KGAll} = 0.0590$, $R^2_{UKBBAll} = 0.0583$, $p = 3.48 \times 10^{-4}$). The difference between the best and worst performing scores – LD panel using 1KG all versus 1KG European – was also much less significant ($p= 1.15 \times 10^{-12}$). These results suggest that the choice of LD panel particularly matters when calculating PRSBMI using cross-ancestry GWAS, or for African ancestry individuals when the GWAS summary statistics are derived from European ancestry individuals.

However, we did observe a slight decrease in the impact of the choice of LD panel when using the full UKBB summary statistics for adults – again, the largest differences were observed in adults of African ancestry, but differences in performance across LD panels were not as significant. The test LD panel performed second best with the 1KGEUR LD panel performing best ($R^2_{Test} = 0.0197$, $R^2_{1KGEUR} = 0.0200$, $p = 0.18$). The 1KGAll LD panel was the worst performing LD panel with an $R^2$ of 0.185, and this difference between the 1KGEUR LD panel was significant after multiple hypothesis correction ($p = 5.08 \times 10^{-7}$).

**Effect of summary statistics and ancestry of test data**

As expected, the $R^2$ values of the PRSBMI were significantly higher when calculated for European ancestry adults than adults of African ancestry, even when using the cross-ancestry GIANT summary statistics (Figure 2). When using the GIANT summary statistics, the best performing PRSBMI in adults of European ancestry had an $R^2$ of 0.0612, which was significantly higher than the $R^2$ from the best performing PRSBMI in African ancestry adults ($R^2 = 0.0149$, $p < 4.9 \times 10^{-324}$).

In African ancestry adults, the $R^2$ when using the GIANT summary statistics was higher than the $R^2$ when using the down-sampled UKBB summary statistics with their respective best LD panel (GIANT (1KGAll LD panel): $R^2 = 0.0149$, UKBB (test data LD panel): $R^2 = 0.140$; $p = 0.038$). This difference was not statistically significant after multiple hypothesis correction. However, the GIANT summary statistics with the 1KGAll LD panel did significantly outperform the UKBB summary statistics with all other LD panels. When keeping the LD panel constant, the PRSBMI calculated using the GIANT summary statistics resulted in higher $R^2$ than using the UKBB summary statistics for all LD panels except for the test data LD panel, and this difference was statistically significant for the 1KGAll ($p = 1.55 \times 10^{-33}$), 1KGEUR ($p = 6.78 \times 10^{-18}$), and UKBBAI ($p = 1.28 \times 10^{-15}$) LD panels. Somewhat surprisingly, we observed higher $R^2$ values for European ancestry adults when using the cross-ancestry GIANT summary statistics versus the down-sampled European UKBB summary statistics ($R^2_{GIANT} = 0.0612$ versus $R^2_{UKBB} = 0.0590$), with this difference being statistically significant ($p = 6.26 \times 10^{-4}$); the best performing LD panel for both set of summary statistics was 1KGAll.

We also compared prediction performance in all individuals using the full (N=377,921) European UKBB GWAS versus the European UKBB GWAS down-sampled to GIANT’s sample size (N=226,960) (Figure 2, S Table 1). For consistency, UKBB European individuals were used for the European test ancestry comparisons, and for the African ancestry comparisons the test sets (i.e., African ancestry LD panels) were used as LD panels. Unanimously across test ancestry and age groups, we observed higher and statistically significant increases in $R^2$.

**Prediction performance across different age groups**
Across different ancestries and summary statistics, we broadly observed similar \( R^2 \) values for adults and teenagers, with substantially reduced performance in children (S Figure 1). \( R^2 \) values in children were consistently less than half of that in adults and teenagers, with differences in \( R^2 \) values for adults and teenagers being minimal (except in the case of African ancestry individuals using the GIANT summary statistics, with teenagers having more than double the \( R^2 \) of adults). Somewhat surprisingly, teenagers consistently had higher \( R^2 \) than adults across all analyses, although these differences were much less significant than those compared with children.

**Proportion of variance explained by each assessed factor**

While we observed significant differences due to ancestry, age, and number of individuals used to calculate summary statistics, we aimed to quantify the effect of these different variables on \( \text{PRS}_{\text{BMI}} \) performance when considered together (Table 1). We observed that 89.5\% of the variance in \( \text{PRS}_{\text{BMI}} \) \( R^2 \) could be explained using these variables, indicating that the majority of the effects of LD panel, ancestry, age, and sample size could be explained through linear relationships with \( \text{PRS}_{\text{BMI}} \) \( R^2 \). In the context of these comparisons, the ancestry of the summary statistics or test data accounts for 55.1\% of the variance explained in \( \text{PRS}_{\text{BMI}} \) \( R^2 \). Choice of LD panel and age of test individuals accounted for similar amounts of variance explained in \( \text{PRS}_{\text{BMI}} \) \( R^2 \) (16.5\% and 15.9\%, respectively), while the number of individuals used to calculate the GWAS summary statistics only accounted for 1.9\% of variance explained in \( \text{PRS}_{\text{BMI}} \) \( R^2 \). Per previous sections, while number of individuals used for summary statistics resulted in significant differences in \( \text{PRS}_{\text{BMI}} \) performance, its overall impact when modeled jointly with all the other factors in the context of these analyses seemed to be small.

**\( \text{PRS}_{\text{BMI}} \) association with comorbid traits**

To determine whether the \( \text{PRS}_{\text{BMI}} \) was associated with clinical comorbidities, we performed a Phenome-Wide Association Study (described more in Methods). Here, the \( \text{PRS}_{\text{BMI}} \) was tested for association with diagnosis codes (Phecodes) to evaluate whether the polygenic background for BMI associates with these clinical diagnoses. The \( \text{PRS}_{\text{BMI}} \) was significantly associated with several of the most frequent Phecodes in eMERGE, particularly in European adults (Figure 3a). As expected, obesity had the strongest association with \( \text{PRS}_{\text{BMI}} \) in all ancestry groups (\( p_{\text{EUR}} < 4.9 \times 10^{-324}; p_{\text{AFR}} = 5.17 \times 10^{-8} \)); this is a positive control. In European ancestry individuals, the best performing \( \text{PRS}_{\text{BMI}} \) was also significantly positively associated with type 2 diabetes (\( p_{\text{EUR}} = 1.04 \times 10^{-102} \)), essential hypertension (\( p_{\text{EUR}} = 7.12 \times 10^{-36} \)), coronary atherosclerosis (\( p_{\text{EUR}} = 3.61 \times 10^{-26} \)), hyperlipidemia (\( p_{\text{EUR}} = 4.38 \times 10^{-16} \)), depression (\( p_{\text{EUR}} = 1.95 \times 10^{-13} \)), hypercholesteremia (\( p_{\text{EUR}} = 3.64 \times 10^{-15} \)), asthma (\( p_{\text{EUR}} = 3.13 \times 10^{-13} \)), and diverticulosis (\( p_{\text{EUR}} = 0.0017 \)). These associations were less statistically significant in African ancestry individuals, potentially due to the lower sample size, and many associations were no longer significant after Bonferroni correction. Only type 2 diabetes (\( p_{\text{AFR}} = 1.2 \times 10^{-5} \)) and coronary atherosclerosis (\( p_{\text{AFR}} = 0.001 \)) were significantly associated with the \( \text{PRS}_{\text{BMI}} \) in African ancestry adults. We also looked at the prevalence of each condition per \( \text{PRS}_{\text{BMI}} \) quintile for the most significantly associated conditions (Figure 3b). The case prevalence generally increased in higher \( \text{PRS}_{\text{BMI}} \) quintile groups for conditions significantly associated with the \( \text{PRS}_{\text{BMI}} \), a trend matching the results we obtained from the association analysis. We performed similar analyses in teens and children but identified no statistically significant associations (results not shown). The much smaller sample sizes of the Phecodes in these age groups may have also contributed to the lack of statistically significant results – most of these diagnoses are adult-onset conditions.
Discussion

Using the eMERGE network data, a diverse dataset in terms of ancestry and age, we were able to quantify the effects that different variables (specifically LD panel, age, ancestry, and sample size of GWAS summary statistics) have on PRSBMI performance. We observed significant improvements in PRSBMI performance for African ancestry individuals using a cross-ancestry GWAS to generate the PRSBMI weights, compared to a European ancestry GWAS of the same size. Even for European ancestry individuals, we observed a significant increase in PRSBMI performance when using a cross-ancestry GWAS versus a solely European ancestry GWAS despite both GWAS being the same sample size. These results suggest that cross-ancestry GWAS summary statistics significantly improve PRSBMI performance not only for non-European ancestry individuals, but possibly for European ancestry individuals as well (or, at minimum, do not significantly degrade PRSBMI performance for European ancestry individuals). Furthermore, when all variables were considered in a multivariate model, we observed that the ancestry of the GWAS summary statistics and test data affects the PRSBMI performance the most. Despite the size of the GWAS having a significant effect on PRSBMI performance, its overall impact is small compared to the other variables we considered in this study, suggesting that for PRS prediction in non-European ancestry individuals, the impact of increasing the sample size of European ancestry GWAS will be minimal compared to performing GWAS in individuals of diverse ancestry. This also suggests that it would be advantageous to use GWAS summary statistics from a similar ancestry group (or cross-ancestry GWAS) even if the sample size is modest, rather than a larger sample size European ancestry GWAS. These results are consistent with previous reports that LD and allele frequency differences explain most of the loss of relative accuracy in PRS performance between European and African ancestry individuals (23). These results also highlight the utility and importance of increasing the number of non-European ancestry individuals used in GWAS, as their inclusion disproportionately increases PRS performance in non-Europeans and may even increase PRS performance for European ancestry individuals. We also found that the PRSBMI was positively associated with type 2 diabetes and coronary atherosclerosis in adults across ancestry groups, and we detected positive associations of the PRSBMI with many other common comorbidities, such as essential hypertension and hyperlipidemia, for European ancestry adults.

Somewhat unintuitively, African ancestry LD panels performed best for African ancestry individuals, regardless of whether European ancestry or cross-ancestry GWAS summary statistics were used. We observed minimal impact of the choice of LD panel when both test data and summary statistics were of European ancestry. These results suggest that as long as either the test data or GWAS summary statistics are of similar ancestry, or the test data and LD panel are of similar ancestry, the difference in PRS performance may be minimal as compared to if all the GWAS summary statistics, test data, and LD panel are of all the same ancestry.

While the findings in this study highlight many important strategies for performing PRS in different ancestry groups, there are limitations that should be addressed in future studies. First, inclusion of analyses that evaluate how different proportions of non-European ancestry individuals affect the prediction performance of PRS would be useful. The GIANT summary statistics we used in this study are only about 6% non-European ancestry. It may be useful to see how the PRS prediction performance changes in both European and non-European datasets as a function of the proportion of non-European ancestry samples included in the GWAS. Such analyses may be possible by combining African ancestry individuals from these different datasets. In addition, similar analyses to evaluate, in a more robust manner, the impact of the
overall size of the GWAS is greatly needed. These analyses will be possible once larger datasets
that include non-European ancestry cohorts are publicly available.

Overall, this study demonstrates the importance of expanding non-European ancestry
data resources for PRS, specifically in the generation of GWAS summary statistics and LD
reference panels. Failure to do so reduces the impact of PRS in diverse populations and increase
the potential for continued health disparities.

Description of supplemental data
Supplemental data include one figure and two tables.

Declaration of author competing interests
Authors have no competing interests to declare.

Data and code availability
Code supporting the current study are available from the corresponding author on request.

References
scores for biobank scale data by exploiting phenotypes from inferred relatives. Nat Commun.
2020 Jun 17;11(1):3074.

history of disease improves polygenic risk scores in diverse populations. Available from:
https://www.biorxiv.org/content/10.1101/2021.04.15.439975v1

3. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current

4. Rask-Andersen M, Karlsson T, Ek WE, Johansson Å. Gene-environment interaction study
for BMI reveals interactions between genetic factors and physical activity, alcohol

covariate interaction effects and the heritability of adult body mass index. Nat Genet. 2017

the overall contribution of gene-environment interaction for obesity-related traits. Nat Commun.

meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for


**Figures**

Figure 1. Flowchart of project. Max size of LD panel was 5,000 individuals. UK Biobank (UKBB) European GWAS summary statistics were downsampled to the mean sample size per variant of GIANT (N=226,960), full size of UKBB European was N=377,921. 1000 Genomes is abbreviated as 1KG.

Figure 2. PRS $R^2$ values across all runs in adults. Asterisks without bars indicate significantly different $R^2$ values between the other 4 LD panels used. Bars are present for significant differences between specific $R^2$ values.

Figure 3. a) Best $PRSBMI$ associations with top 9 most prevalent conditions overall in eMERGE adults b) Prevalence plots of significantly associated conditions in eMERGE adults by best performing PRS quintile

**Table 1. Proportion of variance in $R^2$ that can be explained by different variables using type II sum of squares from ANOVA.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Proportion of explained variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ancestry of summary statistics or test data</td>
<td>0.5510</td>
</tr>
<tr>
<td>Choice of LD reference panel</td>
<td>0.1650</td>
</tr>
<tr>
<td>Age of test individuals</td>
<td>0.1590</td>
</tr>
<tr>
<td>N individuals used to calculate summary statistics</td>
<td>0.0195</td>
</tr>
<tr>
<td>Residuals (unexplained variance)</td>
<td>0.1050</td>
</tr>
</tbody>
</table>