

Dissecting Cross-Population Performance of Polygenic Scores Through Standard and Sibling GWAS

1. Background

Polygenic scores (PGS) are widely applied in predicting complex traits and disease risks at the individual level, but their performance at the population level, especially in cross-population comparisons, remains limited. Differences in mean PGS across populations may reflect true genetic effects, but may also be due to allele frequency structure, linkage disequilibrium differences, or GWAS-related biases.

Most PGS are constructed using standard GWAS, whose effect estimates may be confounded by population stratification and shared environmental factors. In contrast, sibling-based GWAS reduce such confounding by comparing siblings within families, providing a cleaner reference for genetic effects. Comparing PGS derived from standard and sibling GWAS therefore offers a useful framework for dissecting cross-population PGS performance.

This study compares population-mean PGS derived from standard and sibling GWAS across diverse populations and examine their relationships with population-level phenotypic means to assess the performance and limitations of PGS in cross-population settings.

2. Methods & Data

2.1 GWAS Summary Statistics

This study used GWAS summary statistics published by Howe et al. (2022) as the base data for PGS construction. The dataset includes results from both standard GWAS and sibling-based (within-family) GWAS for multiple complex traits.

Quality control was applied to the original summary statistics prior to analysis. The analysis was restricted to autosomal chromosomes (1–22). Traits with SNP-based heritability (h_{SNP}^2) greater than 0.05 were retained. Duplicate SNP records were removed based on genomic position. Ambiguous SNPs with A/T or C/G alleles were excluded to avoid potential strand alignment errors.

2.2 Genotype and Phenotype Data

Genotype data were obtained from the 1000 Genomes Project Phase 3 (2013 release), consisting of whole-genome sequencing data for 2,504 individuals from 26 populations. These populations were grouped into five super-populations: African (AFR), European (EUR), East Asian (EAS), South Asian (SAS), and Admixed American (AMR).

Population-level phenotype data were obtained from the NCD Risk Factor Collaboration (NCD-RISC) Global Database. Phenotype values were extracted for the

year 2013 to match the release year of the 1000 Genomes Project. For height, mean height estimates for individuals aged 18–19 years were used. For body mass index (BMI), obesity prevalence was used as the phenotype measure. Population-level data for systolic blood pressure (SBP) and high-density lipoprotein cholesterol (HDL) were obtained from the same database.

For downstream population-level analyses, each 1000 Genomes population was further mapped to a corresponding country or region based on sampling information provided by the project. Population codes, super-population assignments, and country/region mappings are summarized in Table 1. This mapping was used to align genotype-derived population-level polygenic scores with the corresponding country-level phenotype data.

Table 1. 1000 Genomes population–country mapping

Population Code	Super-Population	Country / Region
BEB	SAS	Bangladesh
ACB	AFR	Barbados
CEU	EUR	United States of America
CDX	EAS	China
CHB	CHN	EAS
CHS	EAS	China
CLM	AMR	Colombia
FIN	EUR	Finland
GWD	AFR	Gambia
GIH	IND	SAS
ITU		SAS
TSI	EUR	Italy
JPT	EAS	Japan
LWK	AFR	Kenya
MXL	AMR	Mexico
ESN	ESN	AFR
YRI		AFR
PJL	SAS	Pakistan
PEL	AMR	Peru
PUR	AMR	Puerto Rico
MSL	AFR	Sierra Leone
IBS	EUR	Spain
STU	SAS	Sri Lanka
GBR	EUR	United Kingdom
KHV	EAS	Vietnam

2.3 SNP Selection and Allele Frequency

GWAS SNPs were first clumped to obtain approximately independent variants, and SNPs passing a p-value threshold of 1×10^{-4} were retained for downstream analyses. Effect allele frequencies were then calculated for each population using the target genotype data.

Random SNP sets were generated by sampling from the QC'ed autosomal SNP pool such that their EUR allele-frequency distribution and SNP count matched those of the clumped GWAS SNPs. Cross-population allele-frequency histograms were computed for these random SNP sets and used as a demographic baseline for comparison with the GWAS SNP patterns.

2.4 Polygenic Score Construction

PGS were constructed using GWAS summary statistics from standard GWAS and sibling-based GWAS. PGS were calculated at the individual level using PLINK v1.9 as the weighted sum of effect allele dosages across included variants, with weights given by the corresponding GWAS effect size estimates.

For individual i , the PGS is calculated as

$$\text{PGS}_i = \sum_{j=1}^M \beta_j \times G_{ij}$$

where M is the number of SNPs included in the score, β_j is the effect size estimate of SNP j obtained from GWAS summary statistics, and G_{ij} is the genotype dosage of the effect allele for individual i (coded as 0, 1, or 2).

PGS were computed separately for standard and sibling-based GWAS results and aggregated to the population level by calculating mean PGS values for each population.

2.5 Statistical Analysis

Population-level mean PGS were standardized across all populations using z-score normalization prior to analysis. Associations between population-level PGS derived from standard GWAS and sibling-based GWAS were assessed using linear regression analyses and correlation. Analyses were performed separately for males and females.

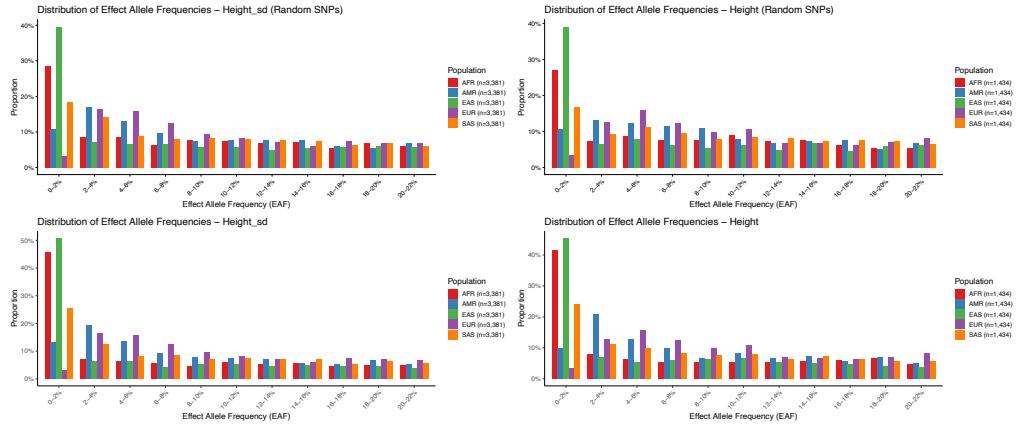
3. Results

3.1 Allele Frequency Distributions of Standard and Sibling GWAS SNPs and Random SNPs

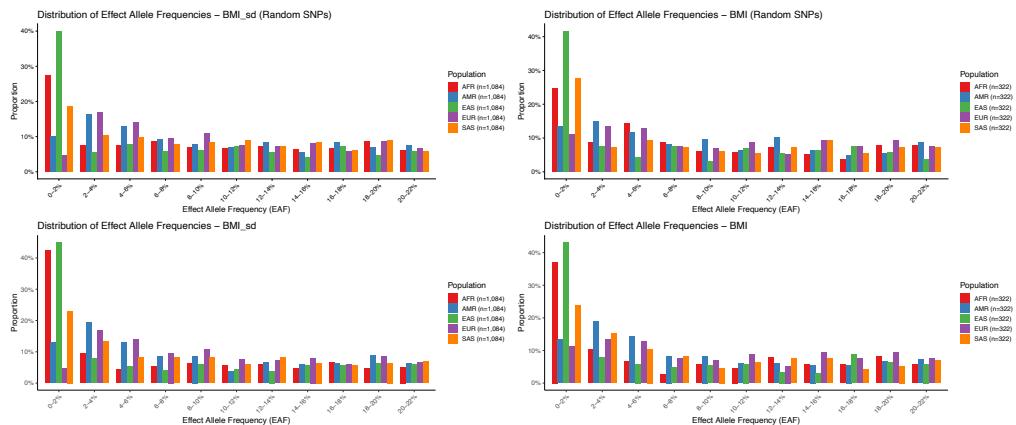
Across traits, allele frequency distributions were examined for SNPs derived from standard and sibling-based GWAS, together with frequency-matched random SNPs. For each trait, distributions were compared across five super-populations to assess

population-level differences in effect allele frequency patterns. Random SNP sets with matched allele frequency distributions served as a demographic baseline for comparison. In the figures, labels with the suffix “_sd” denote SNPs derived from standard GWAS.

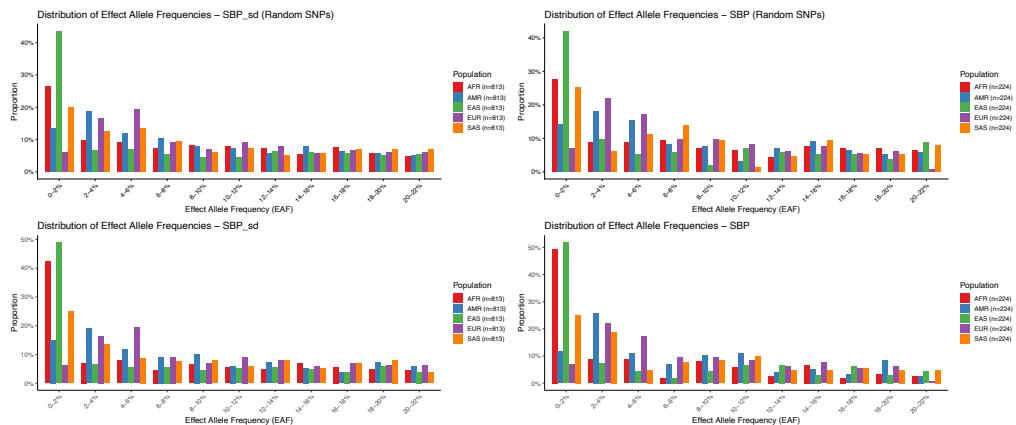
3.1.1 Height



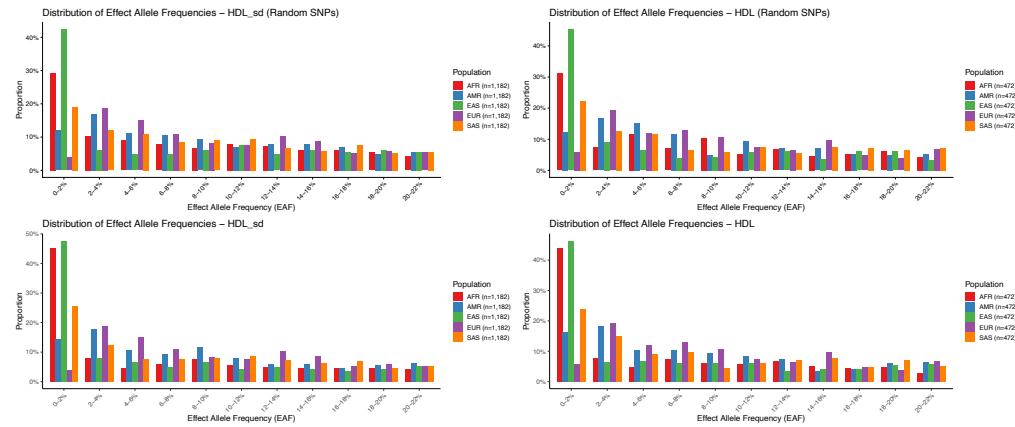
3.1.2 BMI



3.1.3 SBP



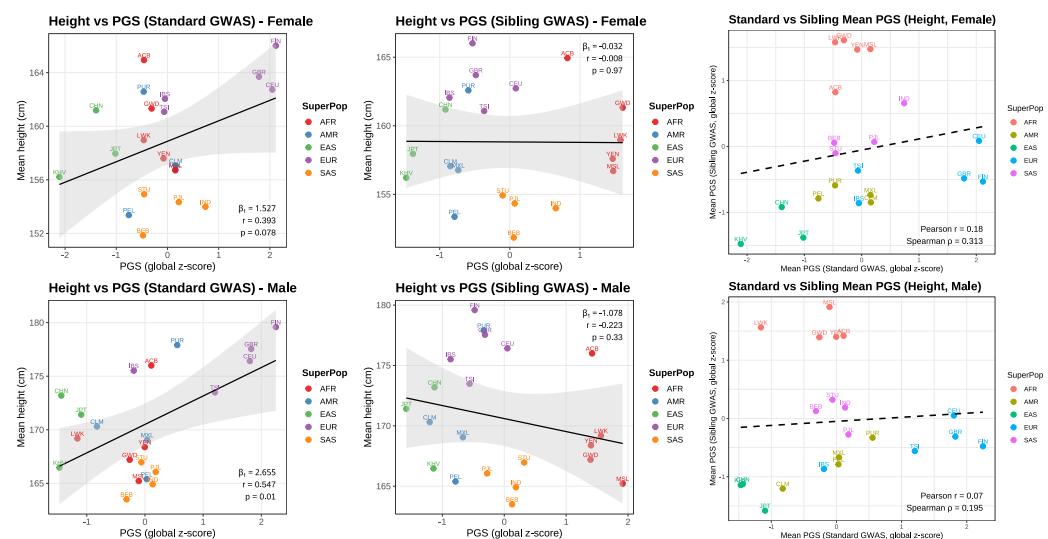
3.1.4 HDL



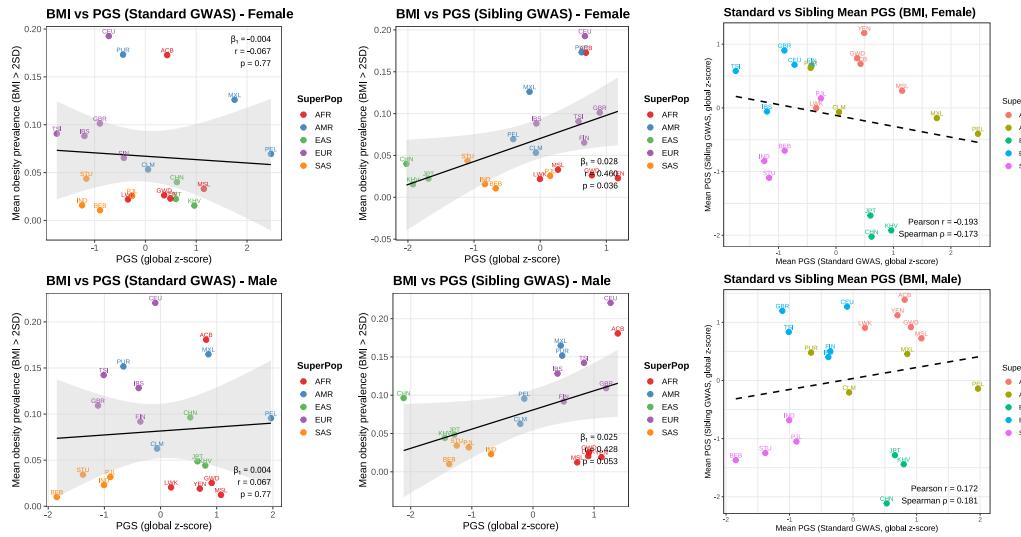
3.2 Population-Level Associations Between Traits and PGS from Standard and Sibling GWAS

Across traits, population-level relationships between phenotypic measures and corresponding PGS derived from standard and sibling-based GWAS were examined. For each trait, associations between population-level phenotypes and PGS were assessed separately for females and males across five super-populations. In addition, population-mean PGS derived from standard and sibling-based GWAS were directly compared to evaluate their concordance across populations. In the figures, each point represents a population, with PGS values standardized to global z-scores, and points colored by super-population.

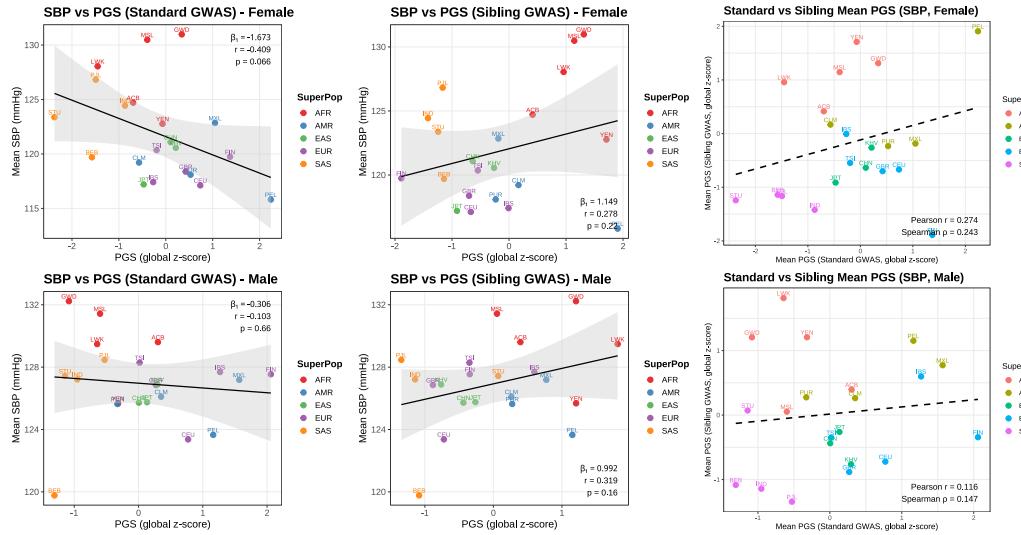
3.2.1 Height



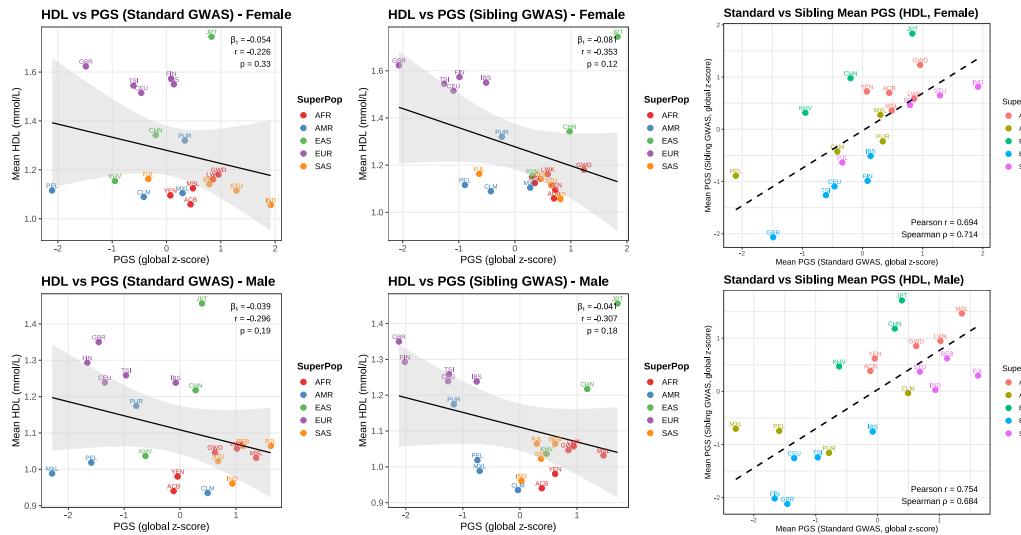
3.2.2 BMI



3.2.3 SBP



3.2.4 HDL



4. Discussion

This study compared PGS derived from standard and sibling-based GWAS across populations and used frequency-matched random SNPs to investigate the sources of population-level PGS differences.

Across traits, SNPs identified by GWAS showed consistently lower allele frequencies in non-European populations, particularly in AFR and EAS, compared with random SNPs matched on European allele frequencies. This pattern was observed for both standard and sibling-based GWAS, indicating that it is not driven solely by environmental confounding or population stratification. Instead, it reflects the way GWAS identifies variants: SNPs that are detected tend to have sufficiently high frequencies and effects in the discovery population, while the same variants may be rarer in other populations.

Associations between population-level PGS and phenotypes varied across traits. For height, standard GWAS PGS showed a clearer relationship with population-level phenotypes, whereas such associations were weak or absent for BMI, SBP, and HDL. This suggests that cross-population PGS performance depends not only on genetic effects but also on whether consistent phenotypic differences exist across populations.

Although correlations between population-mean PGS derived from standard and sibling-based GWAS were limited, both approaches showed highly similar allele frequency patterns across populations. This indicates that cross-population PGS differences are driven largely by allele frequency structure and GWAS ascertainment rather than by family-level or environmental confounding.

Overall, these results highlight that population-level differences in PGS reflect a combination of allele frequency differences, GWAS SNP selection, and trait-specific phenotypic variation, rather than direct differences in genetic effects alone.