# Improving polygenic risk prediction in non-European populations: reweighting SNPs by fixation index

**Shu Mei Teo**[1,2], Woei-Yuh Saw[1], Mike Inouye[1,2]

*[1]Baker Heart and Diabetes Institute, Melbourne; [2]Public Health and Primary Care, University of Cambridge*

## Introduction

Polygenic risk scores (PRS) combine the totality of genomewide disease-associated signals. They can identify individuals at high genetic predisposition to complex human diseases. However, current PRS are biased toward prediction in European populations, and do not transfer well to other populations. This could be due to different SNP allele frequencies and linkage disequilibrium.

## Hypothesis

SNPs with little population differentiation (low $F_{ST}$) between the discovery (GWAS) and target populations should show less attenuation of trait-SNP effect sizes and hence more transferable.

## Materials and methods

Calculate per-SNP fixation index ($F_{ST,}$ Weir and Cockerham)[1] for European vs. African and European vs. South Asian in the 1000 Genomes defined "super populations"[2].

Calculate PRS for individuals in the UK Biobank[3] by summing risk alleles weighted by effect sizes derived from respective CAD and T2D* GWAS[4,5] for SNPs with $F_{ST}$ < lim for
lim = {0.05, 0.1, 0.15, 0.2, 0.25}.

Evaluate PRS performance for
British white (N = 407,632, Self-reported British + PCA refinement),
South Asians (N = 5,668, Self-reported Indian, Pakistani and Bangladeshi + PCA-refinement),
Blacks (N = 5,082, Self-reported Caribbean, African and Any other black background + PCA refinement).

*summary statistics with UKB excluded

## Results

1. For T2D, the best PRS (in red) for Blacks and South Asians resulted in a 7.9% and 1.4% relative improvement in OR over the full score, respectively (**Fig. 1a and b**).
2. For CAD, the best PRS (in blue) for Blacks and South Asians resulted in a 6.9% and 0.7% relative improvement in HR over the full score, respectively (**Fig. 1c and d**).
3. For both CAD and T2D, the $F_{ST}$ filtered PRS did not offer any improvement in the British white population.

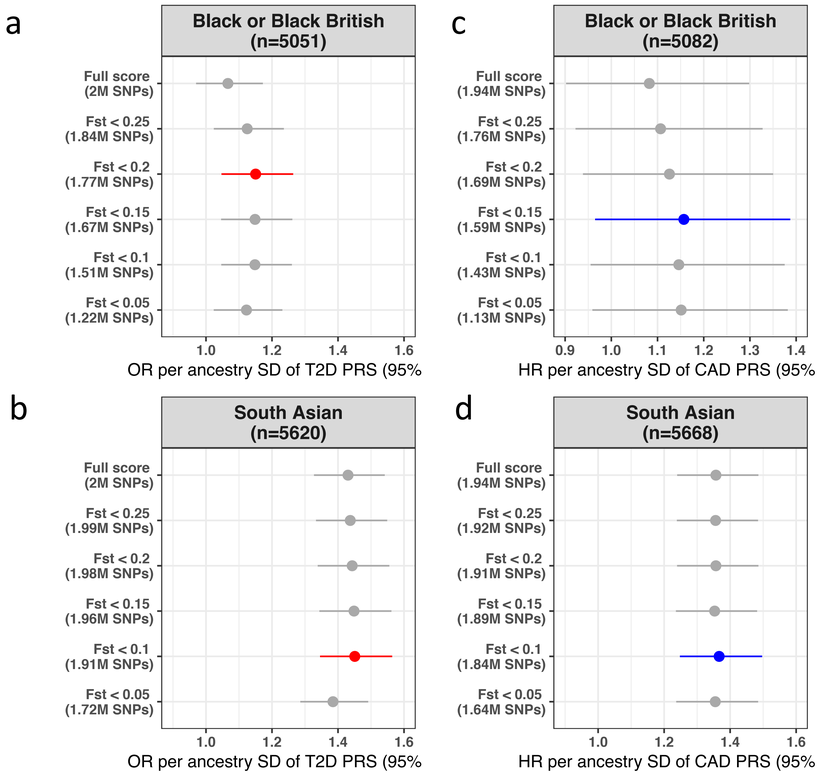


**Fig. 1**: (**a-b**) Odds ratio (OR, x-axis) for self-reported doctor diagnosed diabetes estimated using logistic regression adjusted for age, gender and genotyping array.
(**c-d**) Hazard ratio (HR, x-axis) for CAD (defined as fatal or nonfatal myocardial infarction cases, percutaneous transluminal coronary angioplasty, or coronary artery bypass grafting) estimated using gender-stratified Cox proportional-hazards regression, adjusted for genotyping array.

## Conclusions

PRS derived from SNPs with low $F_{ST}$ improved prediction for CAD and T2D in non-European populations.

This suggests the highly differentiated SNPs added noise for prediction in non-European populations.

Marginal effect sizes (estimated in European populations) for these SNPs may be inaccurate for other populations due to LD differences.

## Future directions

- Using effects sizes from trans-ethnic or population-specific GWAS for high-$F_{ST}$ SNPs.
- Meta-PRS containing

low $F_{ST}$ SNPs weighted by large European GWAS

+

high $F_{ST}$ SNPs weighted by population-specific GWAS

## References

[1] Weir BS, Cockerham CC. Evolution. 1984;38(6):1358-70.
[2] Genomes Project C, et al. Nature. 2015;526(7571):68-74
[3] Bycroft C et al. Nature. 2018;562(7726):203-9.
[4] Nikpay M et al. Nat Genet. 2015;47(10):1121-30.
[5] Mahajan A et al. Nat Genet. 2018;50(11):1505-13.