May 21, 2020

Inclusion of Variants Discovered from Diverse Populations Improves Polygenic Risk Score Transferability

Taylor B. Cavazos[1] and John S. Witte[1-3]

Affiliations:

[1] Biological and Medical Informatics, University of California San Francisco, San Francisco, CA

[2] Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA

[3] Institute for Human Genetics, University of California San Francisco, San Francisco, CA

Corresponding Author:

John S. Witte

University of California, San Francisco

1450 3rd Street, San Francisco, CA 94158

Phone: (415) 502-6882

Email: jwitte@ucsf.edu

ABSTRACT

The majority of polygenic risk scores (PRS) have been developed and optimized in individuals of European ancestry, and may have limited generalizability across other ancestral populations. Understanding aspects of PRS that contribute to this issue and determining solutions is complicated by disease-specific genetic architecture and limited knowledge of trans-ethnic sharing of causal variants and effect sizes. Motivated by these challenges, we undertook a simulation study to assess the relationship between ancestry and the potential bias in PRS developed in European populations. Our simulations show that the magnitude of this bias increases with increasing divergence from European ancestry, and this is attributed to population differences in linkage disequilibrium and allele frequencies of European discovered variants, likely as a result of genetic drift. Importantly, we find that including into the PRS variants discovered in African ancestry individuals has the potential to achieve unbiased estimates of genetic risk across global populations and admixed individuals. We confirm our simulation findings in an analysis of HbA1c in the UK Biobank. Given the demonstrated improvement in PRS prediction accuracy, recruiting larger diverse cohorts will be crucial—and potentially even necessary—for enabling accurate and equitable genetic risk prediction across populations.

INTRODUCTION

Increasing research into polygenic risk scores (PRS) for disease prediction highlights their clinical potential for informing screening, therapeutics, and lifestyle[1]. While their use enables risk prediction in individuals of European ancestry, PRS can have widely varying and much lower accuracy when applied to non-European populations[2–4]. Although the nature of this bias is not well understood, it can be attributed to the vast overrepresentation of European ancestry individuals in genome-wide association studies (GWAS), which is 4.5-fold higher than their percentage of the world population; conversely, there is underrepresentation of diverse populations such as individuals of African ancestry in GWAS, which is one fifth their percentage[3]. Potential explanations for the limited portability of European derived PRS across populations includes differences in population allele frequencies and linkage disequilibrium, the presence of population-specific causal variants or effects, or potential differences in gene-gene or gene-environment interactions[4]. Recent methods developed to improve PRS accuracy in non-Europeans have prioritized the use of European discovered variants and population specific weighting[5–7]. However, only small gains in accuracy are possible with limited sample sizes of non-European cohorts[4].

PRS have been applied and characterized within global populations, but there is limited understanding of PRS accuracy in recently admixed individuals and whether this varies with ancestry. Studies applying PRS in diverse populations[3,4,8] or exploring potential statistical approaches to improve accuracy in such populations[6,9] typically present performance metrics averaged across all admixed individuals. Only one study to date has suggested that PRS accuracy may be a function of genetic admixture (i.e., for height in the UK Biobank[5]). However, it is unknown if the relationship between PRS accuracy and ancestry exists in the absence of population-specific trait effects or what the best approach for applying PRS to admixed individuals will be when there exists adequately powered GWAS in non-European populations.

To help answer these questions, here we systematically and empirically explore the relationship between PRS performance and ancestry within African, European, and admixed populations through simulations. We highlight PRS building approaches that will achieve unbiased estimates across global populations and admixed individuals with future recruitment and representation of non-Europeans in GWAS. We also investigate reasons for loss of PRS accuracy, and attribute this to population differences in linkage disequilibrium (LD) tagging of causal variants by lead GWAS variants, as well as allele frequency biases due to genetic drift in Europeans. Finally, we confirm our simulation findings by application to data on HbA1c levels in individuals of European and individuals of African ancestry from the UK Biobank.

METHODS

Simulation of Population Genotypes

We used the coalescent model (msprime v.7.3[10]) to simulate European (CEU) and African (YRI) genotypes, based on HapMap populations, for chromosome 20 as described previously by Martin et al.[2] Genotypes were modeled after the demographic history of human expansion out of Africa[11], assuming a mutation rate of 2×10^{-8}. We simulated 200,000 Europeans and 200,000 Africans for each simulation trial, for a total of 50 independent simulations (20 million total individuals). We generated founders from an additional 1,000 Europeans and 1,000 Africans (10,000 total across the 50 simulations) to simulate 5,000 admixed individuals (250,000 total across the 50 simulations) with RFMIX v.2[12] assuming two-way admixture between Europeans and Africans with random mating and 8 generations of admixture.

True and GWAS Estimated Polygenic Risk Scores

We generated true genetic risk scores for all European, African, and admixed individuals within each simulation trial.[2] Briefly, m variants evenly spaced throughout the simulated genotypes were

selected to be causal and the effect sizes were drawn from a normal distribution with mean 0 and standard deviation h^2/m, where h^2 is the heritability; complete trans-ethnic sharing of effect sizes in Africans and Europeans was assumed. The true PRS was computed as the summation of all variant effects multiplied by their genotype for each individual. Finally, environmental 'noise' explaining the remainder of the phenotypic variation ($1-h^2$) was added to the genetic risk defining the total trait liability[2]. Cases were selected from the extreme tail of the liability distribution, assuming a 5% disease prevalence. An equal number of controls and 5,000 testing samples were randomly selected from the remainder of the distribution; all 5,000 admixed individuals were also used for testing.

The estimated PRS were constructed from GWAS of the simulated genotypes (modeled after chromosome 20) in Europeans and Africans, each with 10,000 cases and 10,000 controls. Odds ratios (ORs) were estimated for all variants with minor allele frequency (MAF) > 1% and statistical significance of association was assessed with a chi-squared test. For each population, variants were selected for inclusion into the estimated PRS by p-value thresholding ($p = 0.01$, $1x10^{-4}$, and $1x10^{-6}$) and clumping ($r^2 < 0.2$) in a 1 Mb window within the GWAS population, where r^2 is the squared Pearson correlation between pairs of variants. A fixed-effects meta-analysis was also performed to calculate the inverse-variance weighted average of the ORs in Africans and Europeans, and LD r^2 values for clumping used both datasets as the reference.

For each individual, an estimated PRS was calculated as the sum of the log(OR) (i.e., the PRS 'weights') multiplied by their genotype for all independent and significant variants at a given threshold. The PRS were constructed for testing samples with variants and weights each selected from European or African GWAS, or a fixed-effects meta of both combined. An additional local ancestry variant weighting approach was also explored for admixed individuals. Accuracy was measured by Pearson's correlation (*r*) between the true and estimated PRS within each

population, averaged across simulation trials. Ninety-five percent confidence intervals for r were calculated following a Fisher z-transformation for approximate normality, averaged across the 50 trials for each simulation scenario. The statistical significance of differences in accuracy between PRS approaches was assessed within each ancestry group with a z-test (also following Fisher transformation), taking the median p value across trials.

Local Ancestry Weighting PRS

In addition to genotypes of simulated admixed individuals, RFMIX[12] also outputs the local ancestry for every individual. Thus, we used this information to create a local ancestry weighted PRS. For every individual and PRS variant, the weight varied based on the individual's ancestry at that position. If at a given variant the local ancestry was European, then the weight from the European GWAS was used; similarly, if the local ancestry was of African descent then the weight from the African GWAS was used. If the local ancestry was heterogeneous, then the weight from the fixed-effects meta was used. In this way each individual has a PRS constructed from the same independent variants with personalized weights that are unique to the individual's local ancestry.

Application to Real Data

We obtained genome-wide summary statistics for HbA1c from a recent study that performed a within-ancestry fixed-effects meta-analysis in diabetes-free controls for 123,665 individuals of European ancestry and 7,564 individuals of African ancestry[13]. PRS were constructed from associated and independent GWAS variants within each population by p-value thresholding (p= {5×10^{-8}, 1×10^{-7}, 5×10^{-7}, 1×10^{-6}, 5×10^{-6}, 1×10^{-5}, 5×10^{-5}, 1×10^{-4}, 5×10^{-4}, 1×10^{-3}, 5×10^{-3}, 0.01, 0.05, 0.1, 0.5, 1}) and clumping (LD $r^2 < 0.2$) of variants within 250kb with PLINK[14]. Additionally a fixed-effects meta-analysis of the two populations was performed using METASOFT v2.0.1[15].

PRS performance was evaluated using genotype and phenotype data for 394,472 individuals of European ancestry and 5,886 individuals of African ancestry with HbA1c levels available from the UK Biobank, imputation and quality control previously described[16]. Ancestry was defined based on self-report; however, samples that did not fall within five standard deviations of the population mean for the first two principal components were excluded. For each individual, their PRS was computed as the weighted sum of the genotype estimates of effect on HbA1c from the above study, multiplied by the genotype at each variant. For each population-specific variant set, weights from either the European or African HbA1c summary statistics or the fixed-effects meta-analysis were used. A total of 96 polygenic risk scores were evaluated exploring the impact of ancestral population (two scenarios), p-value threshold (16 scenarios), and variant weighting (three scenarios). The proportion of variation explained by each PRS (partial-R^2) approach was assessed for UKB European-ancestry and African-ancestry individuals separately. The partial-R^2 was calculated from the difference in R^2 values following linear regression of HbA1c levels on age and sex, with and without the PRS also included.

RESULTS

Generalizability of European Derived Risk Scores to an Admixed Population

We constructed PRS from our simulated European datasets and applied them to independent simulated European, African, and admixed testing populations. On average, 1552 (range = [1134-1920]) variants were selected for inclusion into the PRS at p-value < 0.01 and LD r^2 < 0.2 (Table 1). The average accuracy, measured by the correlation (*r*) between the true and inferred genetic risk, was much higher when applying the PRS to Europeans (*r* = 0.77; 95% CI = [0.76, 0.78]) than to Africans (*r* = 0.45; 95% CI = [0.43, 0.48]; Figure 1). This is in agreement with decreased performance seen in real data when applying a European derived genetic risk score to an African population[2–4].

To understand the relationship between ancestry and PRS accuracy, admixed individuals were stratified by their proportion of genome-wide European (CEU) ancestry: high (100%>CEU>80%), intermediate (80%>CEU>20%), and low (20%>CEU>0%). PRS performance decreased linearly with decreasing European ancestry (Figure 1). Average accuracy (Pearson's correlation) for the high, intermediate, and low European ancestry groups was 0.73 (95% CI = [0.68, 0.77]), 0.61 (95% CI = [0.59, 0.63]), and 0.52 (95% CI = [0.45, 0.6]), respectively (Figure 1). In comparison to Europeans, the performance of the European derived PRS was significantly lower in individuals with intermediate (20% decrease, $p = 1.27x10^{-47}$), and low (32% decrease, $p = 6.48x10^{-16}$) European ancestry, and with African-only ancestry (41% decrease, $p = 8.00x10^{-155}$). There was no significant difference for individuals with high (5.3% decrease, $p = 0.09$) European ancestry. These trends remained consistent when varying the genetic architecture (Supplementary Figure 1), specifically decreasing the number of causal variants (m) and varying the trait heritability (h^2), as well as the p-value threshold used for variant selection (Supplementary Figure 2).

Population Specific Weighting of European Selected Variants

Using a well-powered GWAS from our simulated African cohort (10,000 cases and 10,000 controls), we aimed to explore the potential accuracy gains achieved from a PRS with European selected variants, but with population specific weighting of these variants. We applied three different weighting approaches to incorporate non-European effect sizes: (1) effect sizes from an African GWAS for all variants; (2) effect sixes from a fixed-effects meta-analysis of European and African GWAS for all variants, both having 10,000 cases and 10,000 controls; and (3) population specific weights based on the local ancestry for an individual at each variant in the PRS (Figure 2).

The most accurate PRS approach varied by the proportion of European ancestry. Populations with greater than 20% African ancestry benefited significantly from the inclusion of population

specific weights (Figure 2). Intermediate European ancestry benefitted most from using fixed-effects meta-analysis weighting instead of European weights (r = 0.64 vs. 0.61, p = 0.02). In contrast, variant weighting from an African GWAS instead of from European had higher accuracy in low European ancestry (r = 0.65 vs. 0.53, p = 0.009) and African-only (r = 0.64 vs. 0.45, p = 2.02×10^{-44}) populations. Individuals with high European ancestry had similar accuracy with weights from a fixed-effects meta-analysis as from European (r = 0.73 in both, p = 0.79), but decreased performance with the inclusion of weights from an African GWAS (r = 0.62 vs. 0.73, p = 0.01).

No clear benefits were observed for local ancestry informed weights compared to weights from a European or African GWAS or fixed-effects meta-analysis. Individuals with high, intermediate, and low European ancestry had similar accuracy using local ancestry informed weights as with the best weighting in each ancestry group: r = 0.72 vs. 0.73 (from fixed-effect or European weights; p = 0.86); r = 0.63 vs. 0.64 (from fixed-effect weights; p = 0.26); and r = 0.63 vs. 0.65 (from African weights; p = 0.50), respectively (Figure 2).

Performance of Non-European PRS Variant Selection and Weighting Approaches

In our simulations, population specific weighting of PRS SNPs discovered in Europeans improved PRS accuracy; however, the disparity between performance in Europeans versus Africans and admixed ancestry individuals remained large. We aimed to explore the potential improvements in PRS that could be gained by including variants discovered in large, adequately powered African ancestry cohorts. Following clumping and thresholding of significant variants using LD and summary statistics from the simulated African populations, an average of 5269 (range = [4462-6071]) variants were included in the PRS (Table 1) reflective of the greater genetic diversity and decreased LD compared to Europeans[17]. In contrast, when constructing a PRS using the same LD and p-value criteria applied to a fixed-effects meta-analysis of European and African ancestry,

an average of only 92 (range = [38-197]) variants were included in the PRS. This substantially smaller number was a result of few variants being statistically significant in both populations. Of the total number of variants included from the European GWAS, African GWAS, and fixed-effects meta, only 1.15%, 0.54%, and 15.0% on average were the exact causal variant from the simulation; an additional 3.72%, 5.34%, and 33.3% tagged at least one causal variant with $r^2 > 0.2$ (and were within ±1000 kb of that causal variant) in Europeans and 3.45%, 2.42%, and 28.1% in Africans (Table 1). The limited overlap between true causal and GWAS selected variants is a result of causal variants in our simulation arising from the total spectrum of allele frequencies, and therefore more likely to be rare, while GWAS is better powered to detect common variants in the study population from which they were identified[2]. These common variants may not adequately tag rare variants due to low correlation[18].

Overall, we constructed twelve PRS with variants selected from GWAS in Europeans or Africans or a fixed-effects meta of both (three scenarios) and weights from the same approaches plus an additional local ancestry specific weighting method (four scenarios) (Figure 2). For Europeans, the highest PRS accuracy was achieved with European selected variants and weights (r = 0.77; 95% CI = [0.76, 0.78]); however, a similar accuracy was observed for weights from a fixed-effects meta (r = 0.76; p = 0.53). For Africans, the highest PRS accuracy was with African selected variants and weights from a fixed-effects meta (r = 0.75; 95% CI = [0.73, 0.76]), similar performance was observed with African variants and weights (r = 0.74, p = 0.28). For admixed individuals, the highest performing PRS depended on the population ancestry percentage. In individuals with high European ancestry (>80%), the best PRS was with European selected variants and fixed-effects meta or European weights (r = 0.73; 95% CI = [0.68, 0.77]). For individuals with intermediate (20%-80%) or low (<20%) European ancestry, the most accurate PRS was from using African selected variants and weights from a fixed-effects meta-analysis (r = 0.68; 95% CI = [0.66, 0.69] and 0.71; 95% CI = [0.66, 0.76], respectively). Again, no benefit was

observed with the inclusion of local ancestry specific weights for any set of PRS variants. Using a more stringent p-value threshold and including fewer variants into the PRS did not result in a change of the best PRS variants and weights (Supplementary Figure 2).

Inclusion of Variants from Diverse Populations

We found that including in the PRS variants discovered in African GWAS with population specific weights results in less disparity in PRS accuracy across ancestries compared to European selected variants, confirming that GWAS in non-bottlenecked populations may yield a more unbiased set of disease variants[19]. For example, applying to Africans a PRS derived from GWAS variants discovered in Africans (with PRS weights from the African study) results in a 15.7% higher accuracy compared to using a PRS comprised of variants discovered in a European GWAS (also with African weights). In contrast, the gains in accuracy achieved by sourcing variants from ancestry-matched studies were much lower in Europeans. Compared to a PRS with variants from an African GWAS (with European weights), a PRS derived from a European GWAS (also with European weights) only gave a 3.9% higher accuracy. We also observed better generalization of PRS based on African selected variants across all admixed groups (Figure 2).

Based on simulations, the best PRS for admixed individuals with at least 20% African ancestry selects variants based on an African GWAS with variant weights from a fixed-effects meta-analysis. This assumes equal sized African and European cohorts (10,000 cases and 10,000 controls). When decreasing the number of African cases, we still see considerable improvements over using a European derived (European selected variants and weights) PRS, especially for low European ancestry (CEU < 20%) where even with 10-fold fewer African samples we see a 14.1% increase in PRS accuracy compared to the European derived risk score (Figure 3).

Allele Frequency and Linkage Disequilibrium of GWAS variants

We sought to understand what factors impacted PRS generalizability across the different variant selection approaches. GWAS performed in Europeans and Africans (for SNPs with MAF ≥ 0.01) were both more likely to identify significant variants that were more common in their own population than in the other. Approximately 60% of variants identified in Europeans had minor allele frequencies less than 1% in Africans and vice-versa; however, as expected, the smaller number of variants selected by a meta-analysis of the two populations tended to have more similar minor allele frequencies (Figure 4a). Although European and African GWAS were both better powered to detect variants at intermediate frequencies within the same study population, due to genetic drift GWAS in Europeans may be unable to capture derived risk variants that have remained in Africa, whereas GWAS in Africans are not subject to this bias[19].

We also examined LD tagging of causal variants by GWAS selected variants within our simulated European and African populations. This entailed computing the LD scores for every causal variant, where the LD score was the sum of the LD r^2 between that causal variant and every GWAS tag variant within $\pm$1000 kb. The LD scores calculated in Europeans and Africans were highly correlated (Pearson's $r > 0.7$) for the GWAS and fixed-effects meta selected variants. Variants selected from a fixed-effects meta had the highest LD score correlation between populations, as expected given that the variants reached significance in both populations and therefore were more common with similar LD patterns (Figure 4b). Since LD score correlation did not vary largely between simulations, we examined the raw LD scores for a single simulation in order to understand differences in LD score magnitude not captured by the Pearson's correlation.

We found that European selected variants had higher LD scores in Europeans compared to in Africans; however, variants selected from an African GWAS tagged causal variants in both populations more strongly (Figure 4c). This is unlikely to be due to the larger number of African

selected variants, as the results were unchanged following normalization of LD scores by size of the PRS (Supplemental Figure 3). Fixed-effects meta-analysis variants had similar LD score magnitudes. However, while a greater proportion of the fixed-effects meta selected variants were causal, fewer were strong tags for causal variants not directly identified, highlighting the potential need to allow for greater heterogeneity of effects for tag variants[20].

Application to Real Data

To corroborate our simulation findings, we undertook an analysis of 96 PRS developed for the prediction of HbA1c levels in 394,472 Europeans and 5,886 African-ancestry individuals from the UK Biobank. We tested variant selection strategies based on p-value thresholding and LD clumping of genome-wide summary statistics[13] computed in European (n = 123,665) or African (n = 7,564) cohorts and variant weights from the same approaches with an additional weighting from a trans-ethnic fixed-effects meta. Across the different p-value thresholds (Methods) we found a strong overlap (>98%) between PRS variants selected from independent European summary statistics and those available in UK Biobank Europeans (following imputation); however, overlap in UK Biobank Africans was much lower ranging from 60% to 77% (Supplementary Table 1). In contrast, variants selected from summary statistics based on an independent African ancestry population had strong overlap (>97%) across all p-value thresholds for both populations (Supplementary Table 1).

In UK Biobank Europeans, the best performing PRS (European variants, European weights, $p < 5x10^{-5}$) explained 2.5% of the phenotypic variance. In UK Biobank African ancestry individuals, with approximately 13.1% European ancestry[5], the best performing PRS (African variants, fixed-effect meta weights, $p < 0.5$) explained only 0.41% of the phenotypic variance. Although the proportion of variation explained by the PRS (partial-R^2) was consistently lower in UK Biobank Africans compared to Europeans, prediction was improved through the use of African selected

PRS variants (Supplementary Figure 4). This improvement was apparent at more inclusive p-value thresholds ($p > 0.05$), likely a reflection of the underpowered African GWAS with 16-fold fewer samples than the European GWAS[13]. Interestingly, we found that the proportion of variation explained in UK Biobank Europeans also increased when using African GWAS variants with increasing p-values, supporting our simulation finding that variants identified in African ancestry populations can be used for prediction in Europeans with limited bias.

DISCUSSION

Our work shows that incorporating variants selected from European GWAS into a PRS can result in less accurate and biased prediction in non-European and admixed populations. We demonstrate the anticipated improvements in PRS prediction accuracy that may be achieved with the inclusion of diverse individuals in GWAS, providing further evidence that supports the need to actively recruit non-European populations.

Our simulation finding that prediction accuracy of a European derived PRS varies with proportion of European ancestry in admixed African and European populations is consistent with a recent study of height where there was a 1.3% decrease for each 10% increase in European ancestry[5]. This decrease in prediction accuracy has been attributed to linkage disequilibrium and allele frequency differences, as well as differences in effect sizes across populations contributing to height[5]. Our work adds further insights into this reduction in PRS accuracy, showing that (1) it exists in the absence of trans-ancestry effect size differences, and (2) variants selected from an African population may not have these same biases. Although GWAS in African populations also identify variants with population allele frequency differences, they have more consistent LD tagging of causal variants across populations. These observations support the hypothesis that well-powered African GWAS will be able to more accurately capture disease associated loci that are more broadly applicable across populations, due to having undergone less genetic drift[19].

Current methods for improving PRS accuracy in diverse populations have prioritized the inclusion of variants from European GWAS, as these have higher statistical power to identify trait associated loci. For example, one such approach uses a two-component linear mixed model to allow for the incorporation of ethnic-specific weights[6]. Another method creates ancestry-specific partial PRS for each individual based on the local ancestry of variants selected from a European GWAS[7]. This approach was found to improve trait predictability, compared to a traditional PRS with population specific or European weights, in East Asians for BMI but not height[7]. In contrast, our simulation found that PRS accuracy was higher with African or fixed-effects meta weighting than with local ancestry in African populations. Note that unlike the previous method,[7] while our simulation is not impacted by incorrect local ancestry inference, we include weights from a fixed-effects meta and do not weight our combined PRS by proportion of overall ancestry. Our results suggest that true equality in performance will be difficult to obtain solely based on European-identified variants even with local ancestry-adjusted weights. Although local ancestry weighting may have greater benefits when complete trans-ethnic sharing is not assumed, we show that in the absence of population-specific factors, the optimal PRS approach involves using variants identified in an African population and population-specific weighting. Other approaches have focused on a mixture of PRS taking advantage of existing well-powered GWAS studies and supplementing with additional information that can be gained from a smaller study in the population of interest[9]. While this approach may offer relative improvement in PRS accuracy for non-Europeans compared to a European-derived PRS, our simulation suggests that the inclusion of significant tag variants discovered in Europeans may unnecessarily hinder predictive performance in non-Europeans.

An important novel finding of our work that the inclusion of variants from an African-ancestry population results in less disparity in PRS accuracy across other populations, illustrates the need

to recruit diverse populations in GWAS and make these data readily available. Large consortia such as H3Africa, PAGE, the Million Veterans Program, and All of Us are undertaking efforts to support this initiative. Based on our analysis of HbA1c in the UK Biobank, we find that improvement in PRS prediction accuracy is currently possible by incorporating findings from GWAS in African ancestry populations, albeit with lower power. In addition to smaller sample sizes, this potential improvement may be limited by ascertainment bias in what SNPs are included on genotyping arrays and poorer imputation in non-Europeans. GWAS arrays and their imputation have substantially higher coverage among Europeans, and this may result in decreased PRS portability of findings across traits; in such situations, whole genome sequencing in diverse populations may be needed in order to improve accuracy[21,22]. Our study and others support the immense genetic diversity that can be unlocked by studying underrepresented populations to both eliminate the disparity in genetics for prediction medicine and provide novel insights into disease biology for all populations[19,21,23].

Although our simulation study provides important insight into the future of PRS use, it has important limitations. First, while coalescent simulations allow for decreased computational burden, model assumptions may result in inaccurate long-range linkage disequilibrium especially for whole genome simulations[24]. However, given we only simulated chromosome 20, biases are expected to be modest[24]. In addition, our simulations assume random mating among admixed individuals and therefore do not reflect the more complex assortative mating that may be observed, which may impact the distribution of local ancestry tract lengths in our simulation and therefore hinder the improvement of PRS accuracy by local ancestry weighting[25]. Finally, we have only simulated individuals from Yoruba, a West African population, which may not be representative of the greater diversity in Sub Saharan Africa[26]. Future work must be done to ensure our findings can be extended to individuals from other regions of Africa.

Overall, our findings support the concern that while studies have demonstrated the potential clinical utility of PRS, adopting the current versions of these scores could contribute to inequality in healthcare[4]. Going forward, future studies should prioritize the inclusion of diverse participants and care must be taken with the interpretation of currently available risk scores. While statistical approaches may offer improvements in accuracy compared to current European-derived risk scores, in order to truly diminish the disparity and achieve PRS accuracies similar to in Europeans we must actively recruit and study diverse populations.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 1650113 and NIH grant CA201358. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This research has been conducted using the UK Biobank Resource under Application Number 14015. Furthermore, the authors thank Linda Kachuri for providing helpful feedback and discussion.

WEB RESOUCES

Simulation code: https://github.com/taylorcavazos/PRS_Admixture_Simulation

HBA1 summary statistics (Wheeler et. al.): https://www.magicinvestigators.org/downloads/

REFERENCES

1. Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of polygenic risk scores. *Nature Reviews Genetics* **19**, 581–590 (2018).
2. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).
3. Duncan, L. *et al.* Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun* **10**, 3328 (2019).
4. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics* **51**, 584 (2019).
5. Bitarello, B. D. & Mathieson, I. *Polygenic scores for height in admixed populations*. http://biorxiv.org/lookup/doi/10.1101/2020.04.08.030361 (2020) doi:10.1101/2020.04.08.030361.
6. Coram, M. A., Fang, H., Candille, S. I., Assimes, T. L. & Tang, H. Leveraging Multi-ethnic Evidence for Risk Assessment of Quantitative Traits in Minority Populations. *The American Journal of Human Genetics* **101**, 218–226 (2017).
7. Marnetto, D. *et al.* Ancestry deconvolution and partial polygenic score can improve susceptibility predictions in recently admixed individuals. *Nat Commun* **11**, 1628 (2020).
8. Grinde, K. E. *et al.* Generalizing polygenic risk scores from Europeans to Hispanics/Latinos. *Genetic Epidemiology* **0**,.
9. Márquez-Luna, C., Loh, P.-R. & Price, A. L. Multi-ethnic polygenic risk scores improve risk prediction in diverse populations. *Genet Epidemiol* **41**, 811–823 (2017).
10. Kelleher, J., Etheridge, A. M. & McVean, G. Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. *PLoS Comput Biol* **12**, e1004842 (2016).
11. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLoS Genet* **5**, e1000695 (2009).

12. Maples, B. K., Gravel, S., Kenny, E. E. & Bustamante, C. D. RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. *The American Journal of Human Genetics* **93**, 278–288 (2013).

13. Wheeler, E. *et al.* Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS Med* **14**, e1002383 (2017).

14. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **81**, 559–575 (2007).

15. Han, B. & Eskin, E. Random-Effects Model Aimed at Discovering Associations in Meta-Analysis of Genome-wide Association Studies. *The American Journal of Human Genetics* **88**, 586–598 (2011).

16. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

17. Campbell, M. C. & Tishkoff, S. A. African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. *Annu. Rev. Genom. Hum. Genet.* **9**, 403–433 (2008).

18. Li, B. & Leal, S. M. Methods for Detecting Associations with Rare Variants for Common Diseases: Application to Analysis of Sequence Data. *The American Journal of Human Genetics* **83**, 311–321 (2008).

19. Kim, M. S., Patel, K. P., Teng, A. K., Berens, A. J. & Lachance, J. Genetic disease risks can be misestimated across global populations. *Genome Biol* **19**, (2018).

20. Morris, A. P. Transethnic meta-analysis of genomewide association studies: Transethnic Meta-Analysis of GWAS. *Genet. Epidemiol.* **35**, 809–822 (2011).

21. Martin, A. R. *et al. Low-coverage sequencing cost-effectively detects known and novel variation in underrepresented populations*.

http://biorxiv.org/lookup/doi/10.1101/2020.04.27.064832 (2020) doi:10.1101/2020.04.27.064832.

22. Li, J. H., Mazur, C. A., Berisa, T. & Pickrell, J. K. *Low-pass sequencing increases the power of GWAS and decreases measurement error of polygenic risk scores compared to genotyping arrays*. http://biorxiv.org/lookup/doi/10.1101/2020.04.29.068452 (2020) doi:10.1101/2020.04.29.068452.

23. Bentley, A. R., Callier, S. L. & Rotimi, C. N. Evaluating the promise of inclusion of African ancestry populations in genomics. *npj Genom. Med.* **5**, 5 (2020).

24. Nelson, D. *et al.* Accounting for long-range correlations in genome-wide simulations of large cohorts. *PLoS Genet* **16**, e1008619 (2020).

25. Zaitlen, N. *et al.* The Effects of Migration and Assortative Mating on Admixture Linkage Disequilibrium. *Genetics* **205**, 375–383 (2017).

26. Gurdasani, D. *et al.* The African Genome Variation Project shapes medical genetics in Africa. *Nature* **517**, 327–332 (2015).

TABLES

Table 1. Summary of PRS Variants and Causal Tagging across Simulations

GWAS Population	Total # PRS Variants (p<0.01)	# Causal	# in LD with a Causal Variant			
European	1552 [1134-1920]	18 [10-26]	r^2> 0.8	r^2> 0.6	r^2> 0.4	r^2> 0.2
LD in Europeans			27 [16-40]	32 [22-44]	39 [25-55]	58 [38-80]
LD in Africans			20 [9-36]	25 [16-42]	34 [24-54]	53 [35-70]
African	5269 [4462-6071]	28 [18-40]	-	-	-	-
LD in Europeans			94 [67-122]	132 [95-171]	183 [123-238]	280 [202-364]
LD in Africans			37 [26-48]	48 [34-61]	67 [50-89]	127 [81-170]
Fixed-Effects Meta	92 [38-197]	12 [5-22]	-	-	-	-
LD in Europeans			15 [6-26]	17 [6-28]	21 [9-39]	29 [16-47]
LD in Africans			13 [6-21]	14 [6-25]	17 [9-29]	24 [10-43]

* The number of variants is reported as the average and range [low-high] across the 50 simulations.

Table 1 Legend: The set of PRS variants from each GWAS and the fixed-effects meta-analysis were selected by p-value thresholding ($p < 0.01$) and clumping ($r^2 < 0.2$) across the 50 simulations. Each PRS variant was compared to the causal set of variants ($m = 1000$) within each simulation to determine the direct overlap between the two sets and the LD r^2 between the PRS variant and every causal variant within a 1000 kb window. The total number of selected PRS variants that tag at least one causal variant at r^2 greater than 0.8, 0.6, 0.4, or 0.2 is listed in the table.

FIGURES

Figure 1. Accuracy of European Derived PRSs by Proportion of Total Ancestry

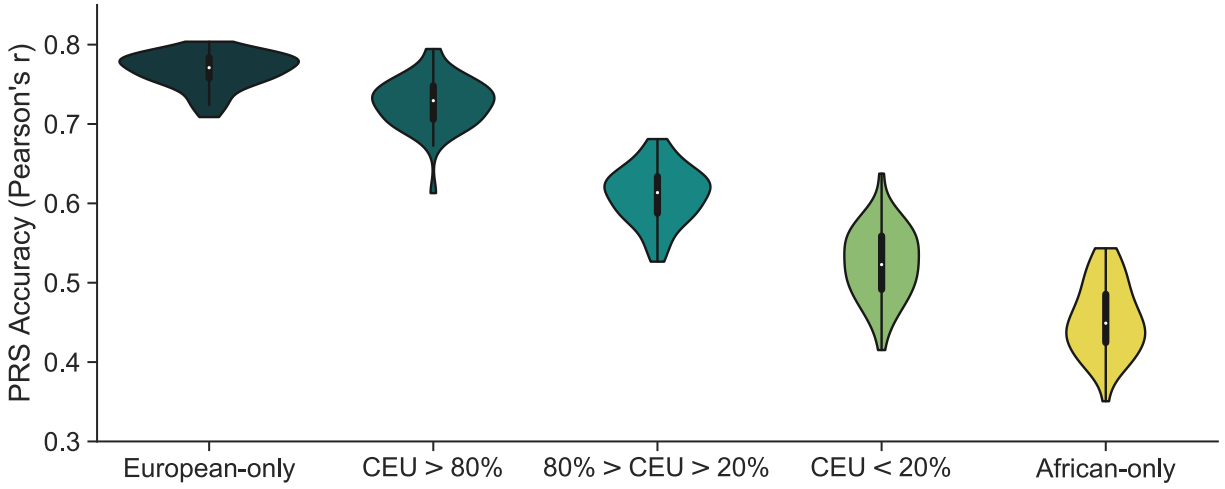

Figure 1 Legend: Accuracy of PRS, with variants and weights from a European GWAS, decreases linearly with increasing proportion of African ancestry. Variants and weights were extracted from a GWAS of 10000 European cases and 10000 European controls. PRS accuracy was computed as the Pearson's correlation between the true genetic risk and GWAS estimated risk score across 50 simulations in independent test populations of 5000 Europeans, 5000 Africans, and 5000 admixed individuals. Admixed individuals were grouped based on their proportion of genome-wide European ancestry. Simulations assume 1000 causal variants and a heritability of 0.5 to compute the true genetic risk. A p-value of 0.01 and LD r^2 cutoff of 0.2 was used to select variants for the estimated risk score.

Figure 2. PRS Construction Approaches and Performance in Admixed Individuals

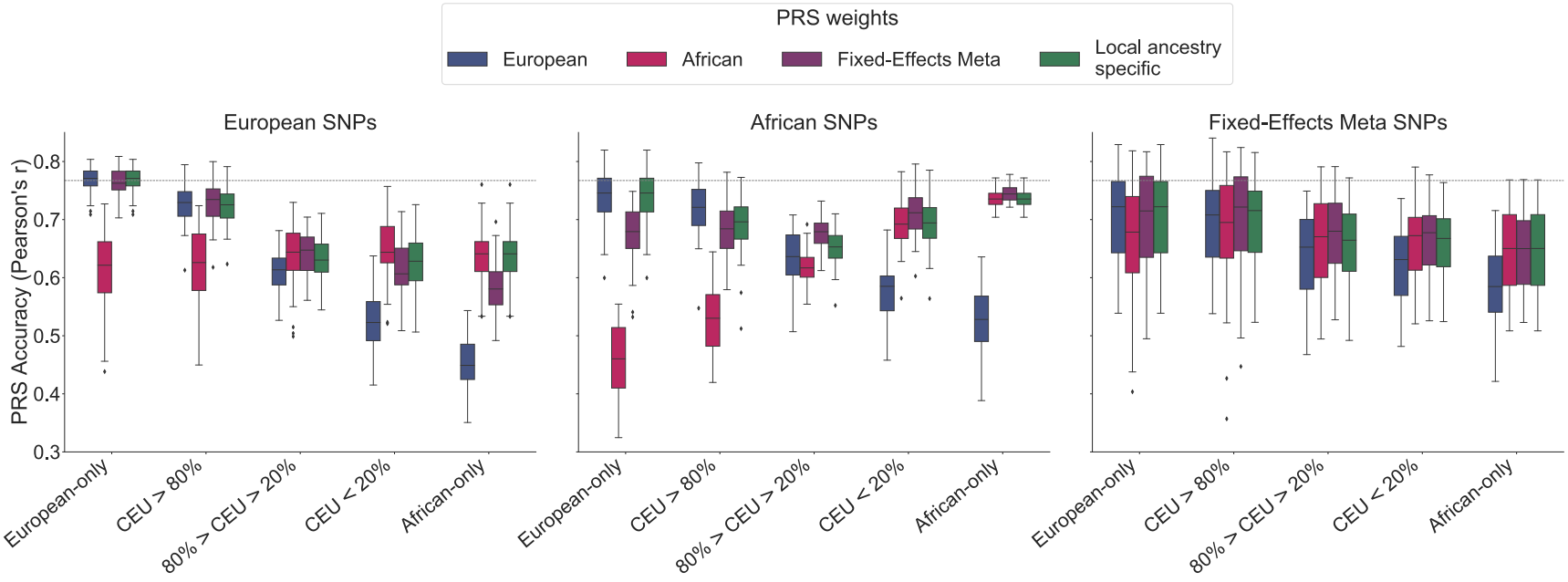

Figure 2 Legend: Using significant variants from an African GWAS with population-specific weights results in less disparity in PRS accuracy across populations. PRS were constructed using variants and weights selected from either a European or African population (10000 cases, 10000 controls each) or a fixed-effects meta-analysis of both. An additional local ancestry specific method was used for PRS weighting. Performance, measured as the Pearson's correlation between the true and GWAS estimated risk score, is shown across 50 simulations. Simulations assume 1000 causal variants and a heritability of 0.5 to compute the true genetic risk. A p-value of 0.01 and LD r^2 cutoff of 0.2 was used to select variants for the estimated risk scores.

Figure 3. Impact of African Sample Size on PRS Accuracy and Generalization

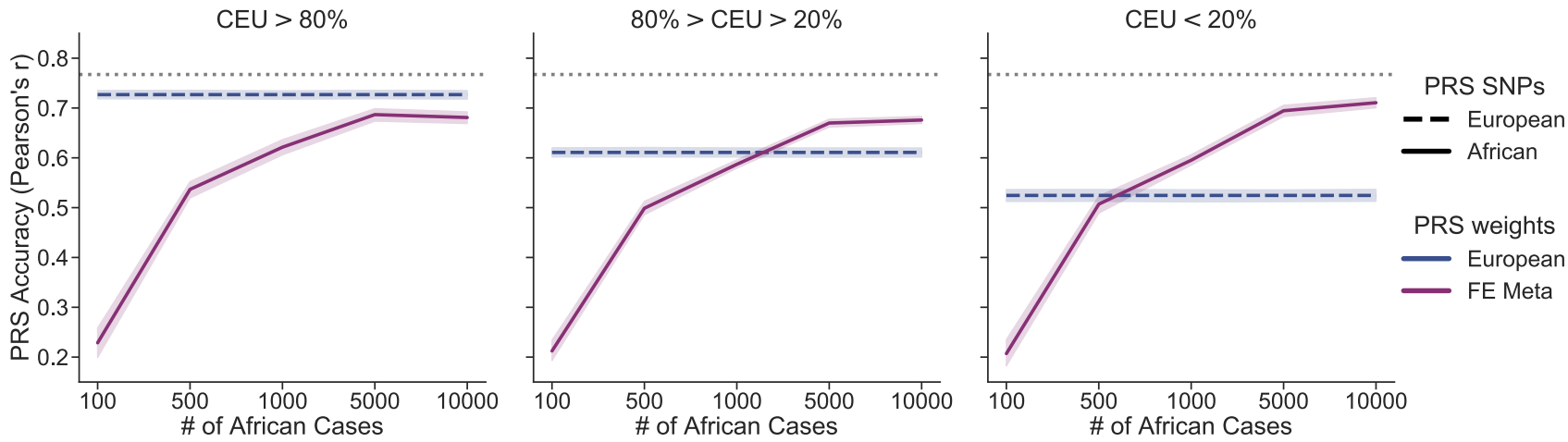

Figure 3 Legend: PRS accuracy in diverse populations can be improved by including data from an African GWAS with smaller sample sizes than in Europeans. The number of African samples used in the GWAS and subsequent PRS construction was decreased to reflect availability of diverse samples in real data. Analysis was conducted assuming 100, 500, 1000, 5000, and 10000 (matched size of European dataset) African cases. Average accuracy and the 95% confidence interval were reported across the 50 simulations for different variant selection and weighting approaches. Simulations assume 1000 causal variants and a heritability of 0.5 to compute the true genetic risk. A p-value of 0.01 and LD r^2 cutoff of 0.2 was used to select variants for the estimated risk score.

Figure 4. Allele Frequency Distribution of GWAS Selected Variants and LD Tagging of Causal Variants

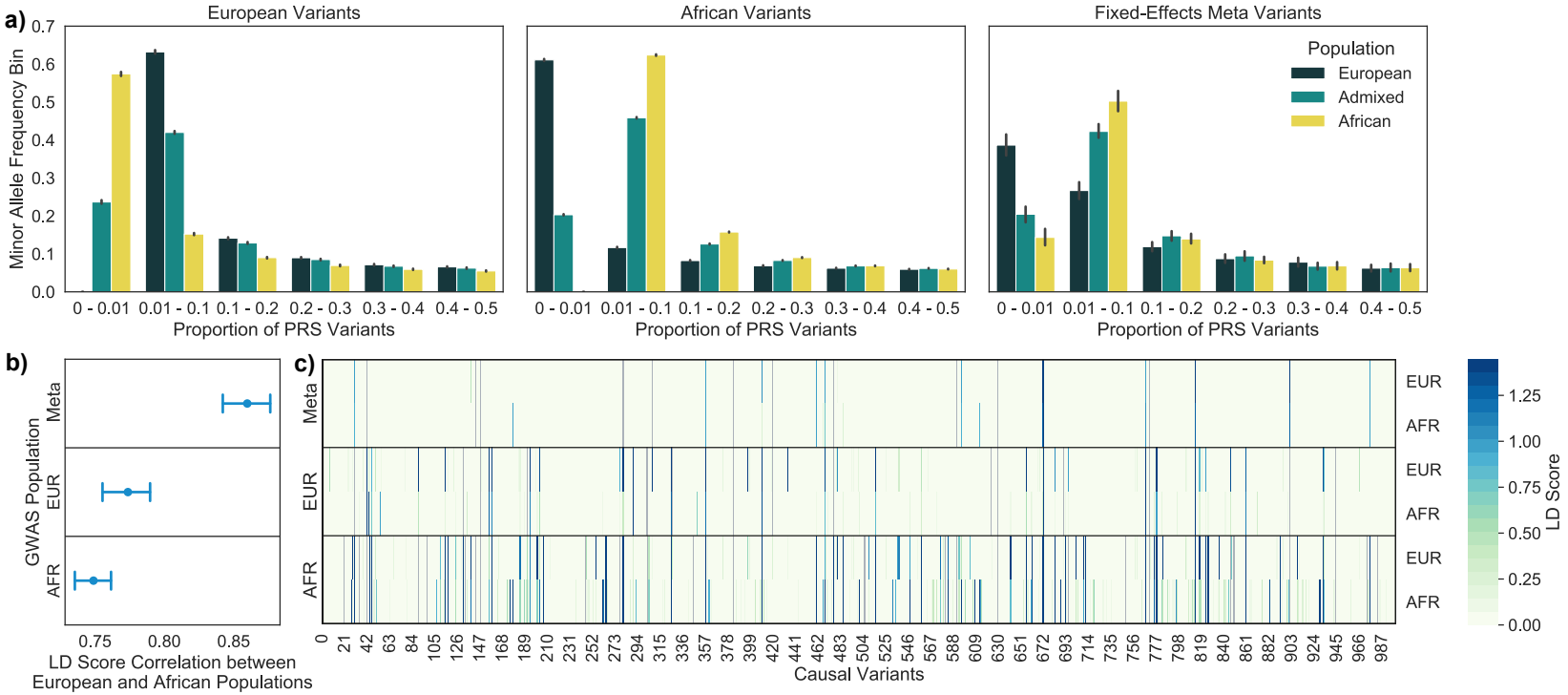

Figure 4 Legend: GWAS significant variants are more common in the study population from which they were discovered; however, African GWAS variants may result in better LD tagging across populations. Variants were selected from a European or African GWAS or a fixed-effects meta of both populations. 4a. GWAS variants were binned by their minor allele frequency estimated from the European, African, and admixed populations. The error bar represents the 95% CI across simulations. 4b. LD

scores were calculated for every causal variant by adding up the LD r^2 for each GWAS tag variant within ±1000 kb of the causal variant. LD scores calculated in a Europeans and Africans were compared by Pearson's correlation. The results were summarized across simulations as the average and 95% CI. 4c. Raw LD scores for each causal variant (m = 1000) calculated in a European or African population for one simulation. Each panel shows the approach used for variant selection. Causal variants directly discovered through the GWAS are colored in grey.