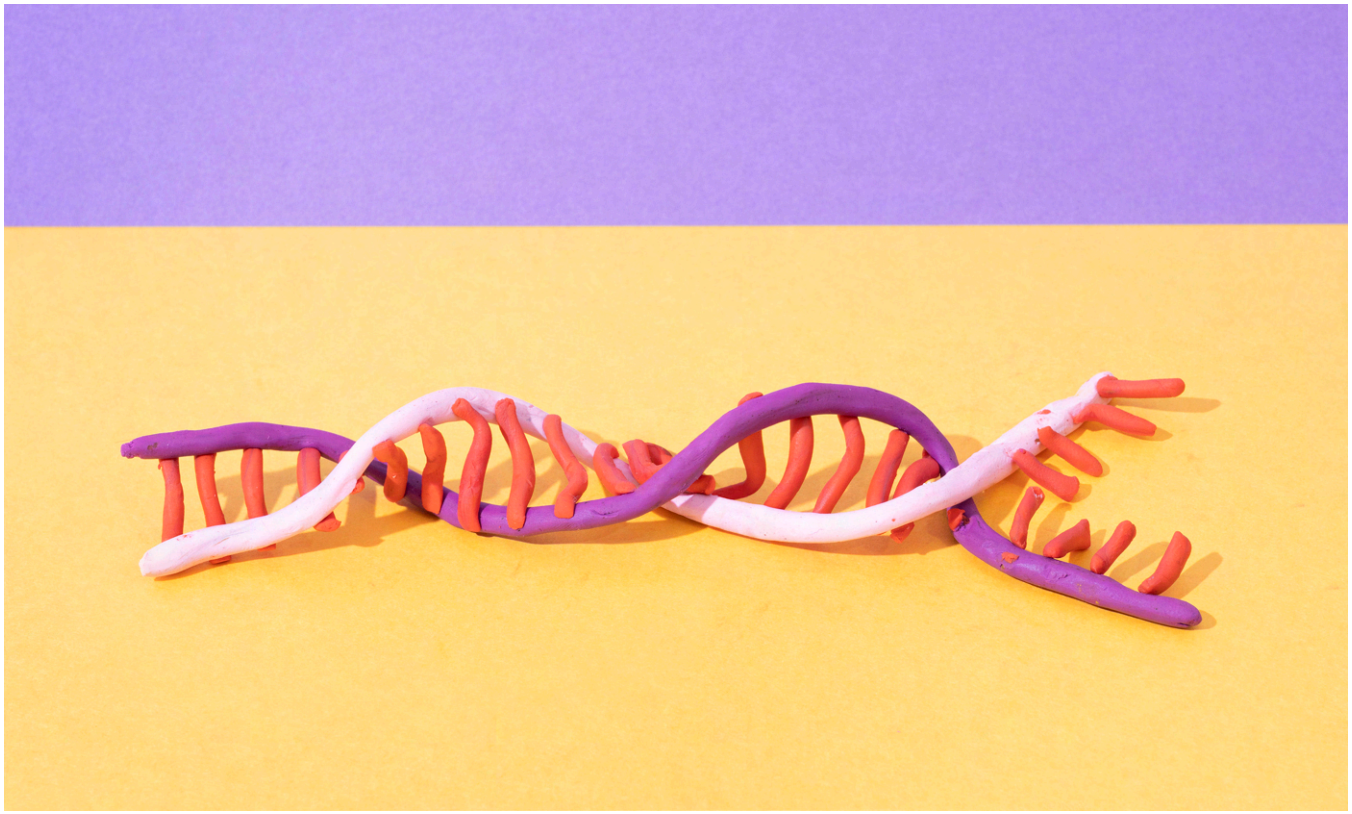


Clay illustration by Lily Offit; Photographed by Ben Denzer

The Danger of Genetic Risk Scores for Worsening Race-Based Disparities in Healthcare

Sohail Zahid, MD, PhD[1]

[1]Harvard Medical School, Boston, MA 02115
Correspondence should be addressed to S.Z. (s.zahid@gmail.com)

Over the last twenty years, there has been an explosion of new technological advancements in sequencing the human genome. It originally cost $2.7 billion dollars to sequence the first human genome, but now, human sequencing costs have dropped to below $1000 per person.[1] During this period, there also have been many advancements in our understanding of the genetics underpinning common, non-communicable disorders such as diabetes, coronary artery disease, and depression.[2] We now know that these common diseases often have a genetic contribution from thousands of genetic variants throughout the genome, each of which confer a small effect on disease risk. The cumulative sum of these genetic variants (known as polygenic risk scores) have been shown to have a significant effect on disease risk, similar to the effects of rare Mendelian disorders.[3]

In the last ten years, research groups and genetic companies have been creating personalized polygenic risk scores with the goal of predicting an individual's susceptibility to chronic diseases. These scores have shown promise in predicting the onset and severity of future diseases before the emergence

of symptoms or traditional risk factors.[4] With these scores, clinicians have the potential to more effectively target intervention (i.e. reducing LDL cholesterol with statins in patients with high polygenic risk for coronary artery disease[5]) and prevention (i.e. early mammogram screening for patients with high polygenic risk for breast cancer).[6]

A significant concern about personalized polygenic risk scores is that they have been developed, optimized, and validated in white individuals with European ancestry. As a result, personalized genetic risk scores have not been as validated in non-white demographics.[7] In fact, these risk scores have poorer performance in non-white groups and may misrepresent their genetic risk for diseases.[7] African descent populations, which have the most health disparities worldwide, are expected to benefit the least from personalized genetic risk score assessments.[7]

In this paper, I will discuss the reasons why there are differences in clinical polygenic risk score efficacy between different demographics, the potential danger in worsening race-based disparities in healthcare, and methods to improve parity of polygenic risk prediction.

Differences in Clinical Efficacy of Polygenic Risk Scores Between Demographics

In *Low Representation of Non-White Groups in Genetic Databases*

There are several steps needed to calculate an individual's personalized genetic risk score for a specific disease.[8] First, you need an ancestry-specific genome-wide association study (GWAS) that identifies the disease risk for each genetic locus (or single nucleotide polymorphism [SNP]). Then you would need an unbiased reference population of genomes for that same ancestry. To calculate an individual's polygenic risk score, one sums the cumulative effect of SNPs calculated from GWAS and compares this number to the population reference. This score is often referred to as a percentile (i.e. in the top fifth percentile of genetic risk).

The main reason why polygenic risk scores perform better in white groups is that this demographic is the best represented in genetic databases. Approximately 80% of all individuals in GWAS databases are whites of European descent.[9] Since 2010, the recruitment of white Europeans in genetic databases has skyrocketed whereas the fraction of non-white groups have remain relatively stagnant (Figure 1).[9] This stagnation is caused by several different factors ranging from lagging diversity in the scientific community, limited engagement with volunteer participants from minority backgrounds, preference of researchers to study European ancestry cohorts, challenges in publication, and analytic challenges.[10]

As a result, for white individuals with European ancestry, there have been a greater number of genetic loci identified as significantly associated with disease, greater accuracy in calculating the genome-wide polygenic risk score, and a more reliable population reference cohort for comparison.[11] Conversely, for the same reasons, there have been fewer discoveries of genetic variants significantly associated with disease in African and East Asian populations.[11] The predictive accuracy of polygenic risk scores is also significantly worse in East Asian and African populations compared to Europeans (Figure 2).[7]

Complicated Gene-Environment Interactions

Most common non-communicable diseases are influenced by both genes and environment. Environmental pressures have led to selective selection of specific genes, such as the sickle cell trait to protect against malaria. However, it is currently unknown what extent environmental pressures played in the up or down regulation of genetic variants at the genome-wide level. This is an important consideration because GWAS calculations and polygenic risk score estimates assume that all causal genetic loci have the same effect across populations.[8]

Certain traits such as height and weight are heavily influenced by genetics, but also significantly affected by access to food and resources.[12] This makes it difficult to disentangle which part of someone's trait is caused by the environment and which is caused by genetics.[12] The complexity of cultural factors with biology can also play a huge role in disease susceptibility. For example, the differences in alcohol use disorder between Asians and Europeans is partially explained by genetic risk scores, but also

have a significant contribution from availability to alcohol and differences in alcohol metabolism.[13]

Certain genetic data banks such as the UK Biobank addressed this challenge of gene-environment interactions by recruiting many healthy people (around 500,000) from Britain and cataloguing many different types of sociodemographic information such as income and housing.[14] However, white Europeans comprise most of this genetic data bank and minority groups are poorly represented.[14]

Poor Generalizability in Different Ethnic Groups

One possibility is to use white European cohorts as reference populations or polygenic risk calculations for everyone, regardless of demographic. However, polygenic risk scores have been shown to have poor generalizability across different populations. Martin et al. evaluated the performance of polygenic risk scores derived from white European GWAS data across 17 anthropometric and blood-panel traits and found that the prediction accuracy is far worse for non-European white groups.[7] The prediction drop-off was 1.6 fold in Hispanics, 2.0 fold in East Asians, and 4.5 fold in Africans.[7]

Danger in Worsening Race-Based Healthcare Disparities

Limited Access to Genetic Risk Scores and Healthcare Guidance

Commercial deployment of polygenic risk scores in their current state has the potential to worsen race-based healthcare disparities. The demographic with the greatest access to genetic risk scores are white individuals. Carroll et al. studied the consumer behavior of approximately 57,000 people at Kaiser Parmenente and found that white individuals were more likely to receive direct-to-consumer genetic testing, clinician-ordered testing, and research-related testing.[15] Among those who received a genetic test and received a notification about a potential genetic abnormality, Hispanic, Black, and Asian individuals were less likely to speak to a medical professional for healthcare guidance and/or understand the meaning of the abnormal results.[16]

Different Benefits from Genetic Risk Score Prediction

At present, clinical application of polygenic risk scores are less useful in minority populations such as those from African descent. Martin et al. studied the polygenic risk prediction in African-descent individuals and found that the prediction accuracy is barely above random chance.[7] Conversely, the prediction accuracy is much stronger for white individuals with British ancestry (Figure 3).[7] Current clinical use of polygenic risk scores is problematic because they will only meaningfully benefit white individuals. African descent populations, which already experience the most disparities in healthcare worldwide, will marginally benefit, if at all.

Methods to Improve Parity of Polygenic Risk Score Prediction

Increase Diversity and Representation in Genetic Databases

There are a couple solutions to address this concern of disparities in polygenic risk score prediction. The most obvious solution is to increase the number of underrepresented groups in genetic databases. Some initiatives with this goal include the All of Us Research Program and the Population Architecture using Genomics and Epidemiology Consortium.[17] The All of Us Research Program, for example, is an NIH sponsored genetic database recruiting over 1,000,000 people with goals to represent a diversity of races, ethnicities, sexes, genders and sexual orientations. Participants who volunteer in this program will also get information about their health and have full access to their own information and records.[17] This latter point is especially important given the pernicious history of medical exploitation of minority groups. A large diverse data bank can help identify more genetic variants associated with disease, uncover new disease biology, and improve genetic risk prediction.

Other countries are also spurring new initiatives to increase the number of underrepresented individuals in genetic data banks. For example, the Human Hereditary and Health in Africa Initiative has invested more than $200 million for genomics research in Africa.[18] Similarly, China and Japan are creating large national biobanks in order to build more accurate ancestry-derived genome wide association studies for various diseases.

Increase Analytical Tools to Improve Accuracy in Different Populations

In addition to recruitment of a diversity of individuals in genetic databases, there needs to be greater effort in methods development to improve polygenic risk score predictions for different ancestry groups.[10] Several different research groups have started to tackle this challenge.[10] Grinde et al. used ancestry-derived meta-analyses to weight genetic loci to improve polygenic prediction accuracy.[19] Similarly Marquez-Luna et al. developed a new tool called MultiPred that creates multi-ethnic polygenic risk scores derived from European and non-European ancestry groups.[20] It is critical that these novel analytic tools become open sourced to facilitate widespread adoption.

FURTHER CONSIDERATIONS

There is still much debate over whether mandatory or case-by-case mental health referrals are better safeguards for patients requesting PAD medication, and it is important to conduct further investigation to answer this question. Additionally, the efficacy of the various evidence-based psychotherapies for terminally ill patients should be further explored in the context of PAD, with special attention given to their effectiveness on different cultures and groups of terminally ill patients, including those with terminal illnesses outside of cancer. [(26),(27)]

CONCLUSION

Polygenic risk scores have tremendous potential in improving diagnostic prediction of disease and guiding treatment management. Because genetic studies have predominated in white European groups in the last twenty years, polygenic risk score prediction is well-characterized and optimized for these groups. Unfortunately, non-white individuals will benefit the least from polygenic risk scores if they were clinically deployed today. It is essential that we increase diversity in genetic databases, and we develop new tools to improve polygenic prediction in underrepresented groups. Otherwise, we will worsen race-based disparities in healthcare.

REFERENCES

1. Dewey FE, Grove ME, Pan C, et al. Clinical interpretation and implications of whole-genome sequencing. JAMA. 2014;311(10):1035-1045.
2. MacArthur J, Bowler E, Cerezo M, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 2017;45(D1):D896-D901.
3. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018;50(9):1219-1224.
4. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. Nat Rev Genet. 2018;19(9):581-590.
5. Natarajan P, Young R, Stitziel NO, et al. Polygenic Risk Score Identifies Subgroup With Higher Burden of Atherosclerosis and Greater Relative Benefit From Statin Therapy in the Primary Prevention Setting. Circulation. 2017;135(22):2091-2101.
6. Mavaddat N, Pharoah PD, Michailidou K, et al. Prediction of breast cancer risk based on profiling with common genetic vari-ants. J Natl Cancer Inst. 2015;107(5).
7. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacer-bate health disparities. Nat Genet. 2019;51(4):584-591.
8. Choi SW, Mak TS, O'Reilly PF. Tutorial: a guide to per-forming polygenic risk score analyses. Nat Protoc. 2020;15(9):2759-2772.
9. Mills MC, Rahal C. A scientometric review of genome-wide association studies. Commun Biol. 2019;2:9.
10. Peterson RE, Kuchenbaecker K, Walters RK, et al. Ge-nome-wide Association Studies in Ancestrally Diverse Popula-tions: Opportunities, Methods, Pitfalls, and Recommendations. Cell. 2019;179(3):589-603.
11. Duncan L, Shen H, Gelaye B, et al. Analysis of polygenic risk score usage and performance in diverse human populations. Nat Commun. 2019;10(1):3328.
12. Sohail M, Maier RM, Ganna A, et al. Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. Elife. 2019;8.
13. Li D, Zhao H, Gelernter J. Strong protective effect of the aldehyde dehydrogenase gene (ALDH2) 504lys (*2) allele against alcoholism and alcohol-induced medical diseases in Asians. Hum Genet. 2012;131(5):725-737.
14. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12(3):e1001779.
16. Carroll NM, Blum-Barnett E, Madrid SD, et al. Demo-graphic differences in the utilization of clinical and direct-to-consumer genetic testing. J Genet Couns. 2020;29(4):634-643.
17. The "All of Us" Research Program. New England Journal of Medicine. 2019;381(7):668-676.
18. Martin AR, Teferra S, Moller M, Hoal EG, Daly MJ. The critical needs and challenges for genetic architecture studies in Africa. Curr Opin Genet Dev. 2018;53:113-120.
19. Grinde KE, Qi Q, Thornton TA, et al. Generalizing poly-genic risk scores from Europeans to Hispanics/Latinos. Genet Epi-demiol. 2019;43(1):50-62.
20. Marquez-Luna C, Loh PR, South Asian Type 2 Diabetes C, Consortium STD, Price AL. Multiethnic polygenic risk scores improve risk prediction in diverse populations. Genet Epidemiol. 2017;41(8):811-82

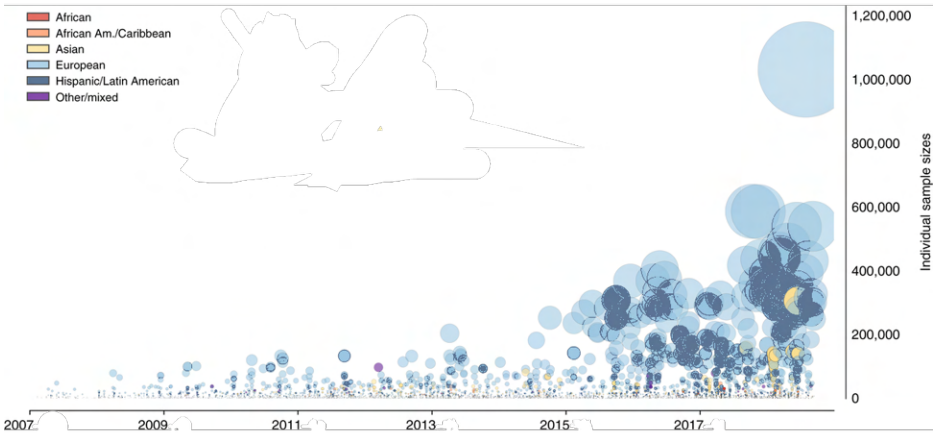


Figure 1. Non-white individuals are poorly represented in genetic databases. Inclusion of non-white groups have stagnated in the last five years. Adapted from Mills et al.[9]

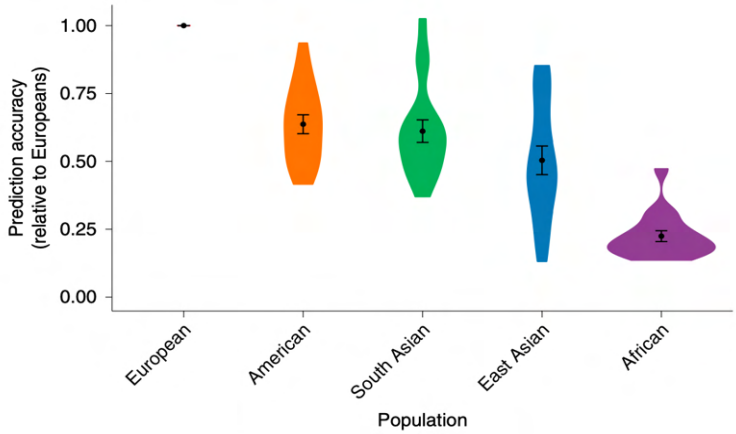


Figure 2. The predictive accuracy of polygenic risk scores across 17 anthropomorphic and blood panel traits is worse in non-white demographics. Adapted from Martin et al.[7]

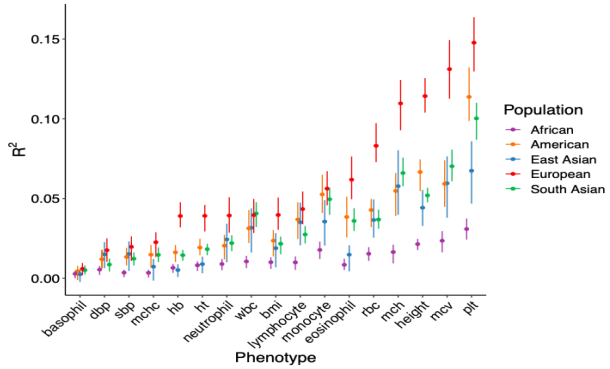


Figure 3. The only demographic that will meaningfully benefit from accurate polygenic risk score prediction are white individuals with European ancestry. The predictive accuracy for those from African ancestry are barely better than random chance. Adapted from Martin et al.[7]