International Journal of Social Science and Economic Research

ISSN: 2455-8834

Volume:09, Issue:10 "October 2024"

Integration of Polygenic Risk Scores and Artificial Intelligence into Healthcare

Shannon Victor

Independent Researcher, USA

DOI: 10.46609/IJSSER.2024.v09i10.053 URL: https://doi.org/10.46609/IJSSER.2024.v09i10.053

Received: 4 October 2024 / Accepted: 26 October 2024 / Published: 30 October 2024

ABSTRACT

The introduction of artificial intelligence technology has been groundbreaking in increasing positive outcomes and simplifying processes for numerous industries. In recent years, the healthcare field has followed suit by leveraging advancements to allow for predictive analysis. The integration of artificial intelligence into scientific techniques has transformed the way we identify the population's genetic predispositions to diseases. Polygenic risk scores (PRS) are a calculation of an individual's risk for a certain genetic disease. However, since most complex diseases are also affected by environmental and lifestyle factors that these tests don't take into consideration, utilizing artificial intelligence in addition to this provides a much more accurate approach. This will aid in identifying individuals at higher risk, usually because of family history, in order to provide a comprehensive diagnosis and treatment plan. While this combination holds significant promise in advancing personalized medicine, it's important to be aware of the possible racial and socioeconomic underrepresentation present in the current development of these technologies.

Introduction

Family history is a well-established risk factor for numerous diseases, including cardiovascular disease, diabetes, and various forms of cancer (Chacko et al., 2020). Gaining an understanding of an individual's genetic predisposition can enable early intervention and personalized treatment strategies, potentially reducing the frequency and severity of these conditions.

Polygenic risk scores have emerged as a novel tool in quantifying genetic risk by summing the effects of multiple genetic variants. They are limited in their current form because they do not account for environmental and lifestyle factors, which are critical in the development of many diseases. This gap reduces the overall accuracy of PRS-based risk assessments, as genetic predisposition alone does not fully explain an individual's likelihood of developing a condition.

ISSN: 2455-8834

Volume:09, Issue:10 "October 2024"

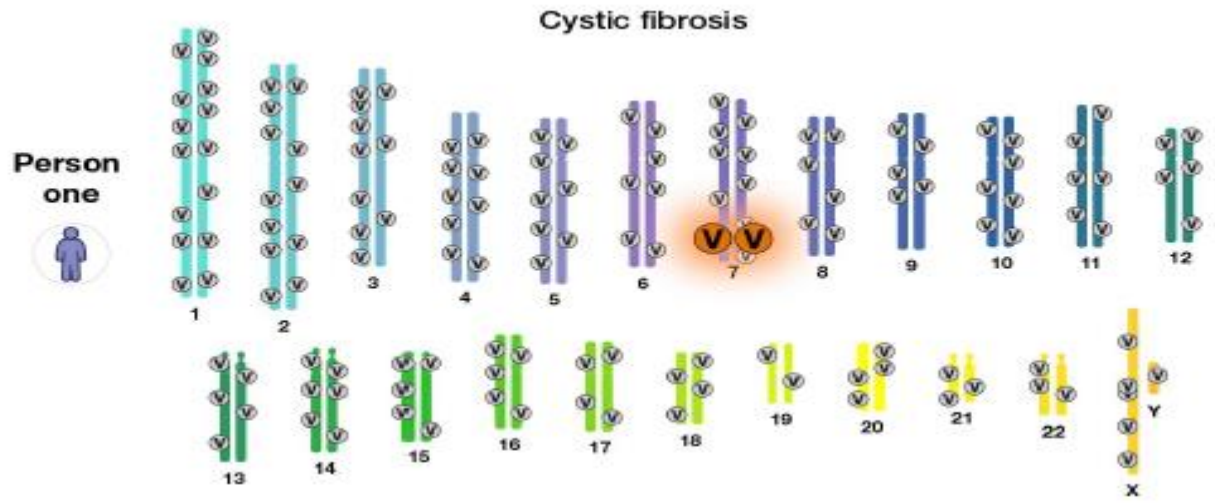
Further, the developmental nature of this procedure causes PRS to be strictly experimental and not implemented in clinical settings.

AI offers a solution to this gap by utilizing electronic medical records that include not only physical health information but also external factors like environmental exposures, diet, exercise habits, and socioeconomic conditions. AI's ability to process large datasets, recognize patterns, and integrate diverse sources of information makes it an ideal partner. Using the provided information, it will calculate which patients are at high risk based on the result of feeding patient health information to a series of programs. Then, these profiles will be evaluated with the results of their individual PSR scores. This integration of genetic and non-genetic data through AI has the potential to revolutionize how we assess and address disease risk.

In the constantly evolving healthcare field, understanding the applications, benefits, and potential pitfalls of predictive analytics becomes increasingly important. By constructing a path to combine the results of predictive score analysis and artificial intelligence, these upcoming technologies can be applied in clinical settings to transform care of hereditary diseases.

Polygenic Risk Score

A polygenic risk score (PRS) is a statistical calculation that uses genomic information alone to estimate an individual's likelihood of developing a particular medical condition. By analyzing the presence or absence of multiple genomic variants, PRS assesses a person's genetic risk for hereditary diseases like coronary artery disease, breast cancer, or diabetes.



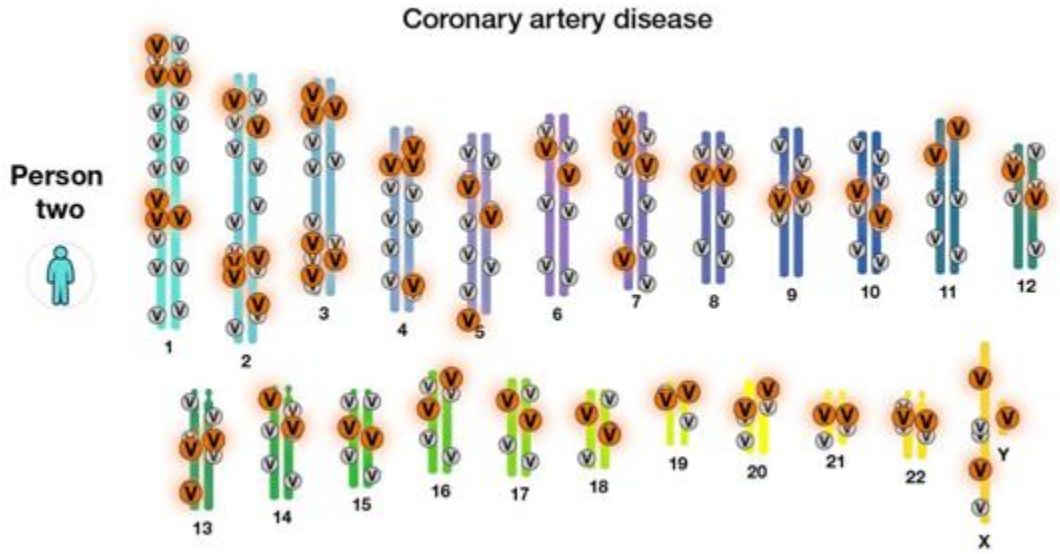
Each small “v” represents a genomic variant that is present in an individual’s genome but is notassociated with cystic fibrosis. Each larger “V” represents a CFTR gene mutation.

ISSN: 2455-8834

Volume:09, Issue:10 "October 2024"

Many inherited diseases can be traced to variants in a single gene. For example, Cystic fibrosis, a progressive genetic disease that causes long-term lung infections, is caused by variants in the cystic fibrosis transmembrane conductance regulator (CFTR) gene on chromosome 7 (Polygenic Risk Scores in Complex Disease Research, n.d.).

Complex, or polygenic, diseases occur as a result of many genomic variants, paired with environmental influences such as diet, sleep, stress and smoking. Medical practices for identifying risk for these diseases are more limited, with one of the leading techniques being polygenic risk scores.



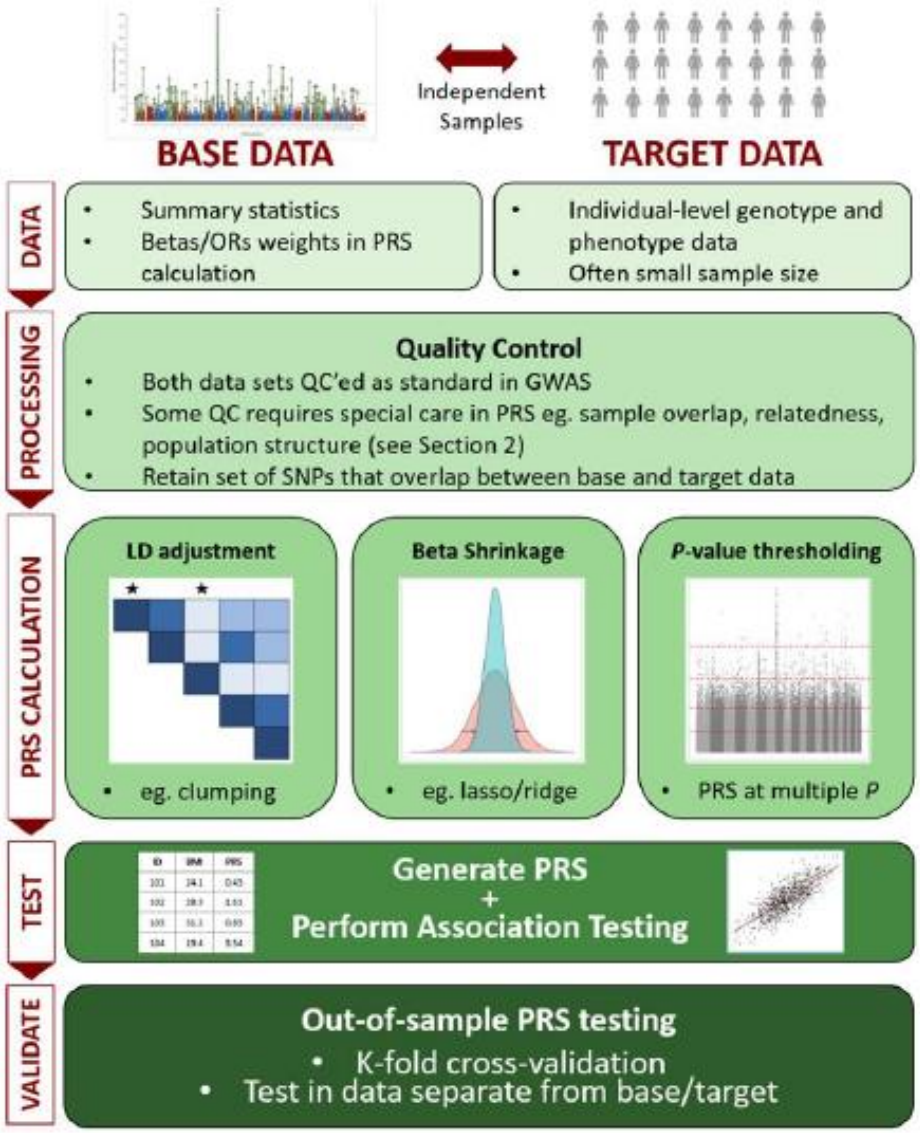
Each red "v" represents variants in an individual's genome that are associated with coronary artery disease. Each smaller gray "v" is a variant that is also present in the person's genome but is not implicated indisease.

Researchers identify genomic variants associated with complex diseases by comparing the genomes of individuals with and without those diseases. There are around 4 to 5 million such variants in an individual's genome, some of which increase or decrease disease risk, while others may have no effect. The risks are calculated by summing the number of risk alleles carried by an individual at specific genetic variants, typically single nucleotide polymorphisms (SNPs), weighted by the effect size estimate of the most powerful GWAS on the phenotype (Choi et al., 2020). The data from Genome-Wide Association Studies (GWAS) provides summary statistics like betas (measures the strength of associations) and P-values (indicates the likelihood that the associations are real) (Uffelmann et al., 2021). These statistics represent how different genetic variants are related to traits across the entire genome.

ISSN: 2455-8834

Volume:09, Issue:10 "October 2024"

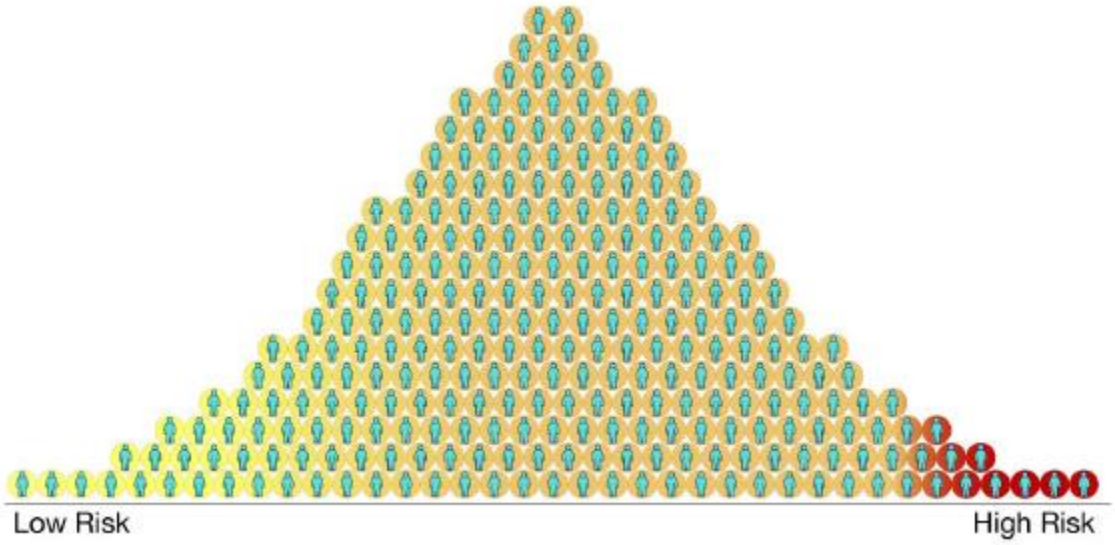
When the P-value shows the data is significant and issues like strand mismatches, incorrect effect allele, or ambiguous SNPs are eliminated after a thorough quality check, scientists can proceed to calculate the PRS. As GWAS sample sizes continue to get larger, errors in the data will be minimized and the process will be streamlined. Then, methods like shrinkage are used to fine-tune the estimates, either by adjusting all SNP effect sizes or using a threshold to include only SNPs with significant effects (Lewis & Vassos, 2020). It is important to note that PRS results can be skewed by differences in population structure between the base and target samples. Factors like location, age, and socioeconomic status can cause PRS to either inflate or deflate predictions, leading to potential inaccuracies (Mostafavi et al., 2020).



ISSN: 2455-8834

Volume:09, Issue:10 "October 2024"

Each polygenic risk score can be put on a bell curve distribution. Most people will find their scores to be in the middle, indicating average risk for developing a disease. Others may find themselves on the extreme ends, putting them at either low or high risk. While these scores indicate a person's risk for conditions compared to others, it does not provide a timeline or guarantee of disease progression (National Human Genome Research Institute, 2020). For example, two people with high PRS for cancer—one who is a young guy and the other an old smoker—would have vastly different lifetime risks. PRS shows



correlations, not causation, so its interpretation must be cautious. However, people who score in the top percentiles of genetic risk may require future discussions with their physicians for more frequent screening or lifestyle changes.

Artificial Intelligence

Polygenic risk scores are calculated solely based on genetics. However, as mentioned before, environmental and lifestyle factors also play a large part in complex diseases.

For instance, children with a genetic variant called MET, which plays a role in brain development, are at an increased risk of developing autism when exposed to high levels of air pollution (Volk et al., 2014). However, this variant does not elevate autism risk in the 75% of the population exposed to lower pollution levels, suggesting that autism may arise from an interaction between genetic and environmental factors. Additionally, an international study, including NIEHS scientists, found that children with similar variations in the TLR4 gene

ISSN: 2455-8834

Volume:09, Issue:10 "October 2024"

exposed to certain environmental triggers were more likely to develop RSV bronchiolitis, a life-threatening respiratory disease (Gene and Environment Interaction, 2023).

Social determinants of health (SDOH) are recognized internationally as non-medical reasons that influence health outcomes. Numerous studies suggest that these variables account for between 30-55% of health outcomes (WHO, 2024).

Below are some social determinants of health recognized internationally (Social Determinants of Health, n.d) :

- Economic stability
- Education access and quality
- Health care access and quality
- Neighborhood and work environment
- Social and community context

Lifestyle factors such as diet, exercise, and substance use have also been linked to increased risk for a variety of diseases. Members of a medical care team typically question and document if any

ISSN: 2455-8834

Volume:09, Issue:10 "October 2024"

of these determinants or lifestyle choices can hinder medical outcomes for a patient. Along with other notes about the patients, this information is recorded in electronic medical records (EMR) and hospital databases. These records contain years of detailed patient information which tend to be skimmed through by doctors. By implementing an artificial intelligence program to scan and analyze longitudinal information derived from EMRs, the possibility of dismissing key external information for a patient's health will be greatly reduced.

When coupled with the polygenic risk scores, this technology has the potential to accurately pinpoint populations at risk by analyzing genetic and external factors to provide timely interventions. Say, for example, a patient with a family history of diabetes comes in for a medical check-up. Firstly, we would test their sample to calculate their genetic risk factor for the said disease. If the results show they're not at risk, general precautions should still be discussed with their physicians but they would be otherwise considered safe. However, if their genetics are considered high risk, AI can highlight their past/current symptoms and lifestyle to accurately understand if they do have the condition. Early detection would lead to further testing, lifestyle changes, and treatment which would mitigate the effects of the disease.

Limitations

While this technology holds significant promise in advancing the healthcare field, there are also critical limitations to consider.

Current PRS methods rely on an individual's genetic ancestry being similar to the large GWAS study from which reference effect sizes are taken for PRS calculation and may require access to an ancestry-matched genotype-level reference panel. Such studies are currently only widely available in European ancestries and the data varies by race (Lewis & Vassos, 2020). Because of this, the majority of the studies discussed in this paper so far have only been conducted on European participants, raising questions about the generalizability of the data for other regions. As it becomes an increasingly popular technique, there's a dire need for more research in order to derive the data for making polygenic risk scores useful for other populations. This historic lack of diversity in genomic studies is also a concern for any future breakthroughs regarding this topic due to its impacts being limited to certain races.

Similarly, when incorporating automated mechanisms into a complex field, the chances of misdiagnosis increase. Predictive AI models that are incomplete or unrepresentative can exacerbate bias, leading to misdiagnosis. For instance, overdiagnosis can occur when a single abnormal test result is accepted as diagnostic. Many factors can contribute to abnormal results in patients. Even something as simple as a common cold can present symptoms that align with symptoms of a hereditary disease. This tendency narrows definitions of normalcy or defines

ISSN: 2455-8834

Volume:09, Issue:10 "October 2024"

screening outcomes in ways that favor false-positive results. Not only could this lead to unnecessary interventions, it could also evoke stress and fear for the patients. Conversely, AI models may under-diagnose certain groups by excluding patients with missing data due to limited access to care or by ignoring groups (typically minorities) for whom data are not recorded. If population subgroups do not frequently attend physicians and are subsequently underdiagnosed, their characteristics (e.g., age, race, sex) may be incorrectly interpreted as possessing lower risk for the specific diseases being tracked (Hernandez, 2016). This exemplifies a broader issue in which predictive AI often ignores underlying parameters which are not recorded.

Future Directions

Future directions for combining polygenic risk scores with artificial intelligence must prioritize addressing ethical and regulatory challenges. Since patient information is easily accessible and used in multiple programs, it is essential to ensure that individuals' information isn't exposed in case of a security breach or hack (source). One way this can be done is through anonymizing data as it is transferred from electronic medical records to software programs, minimizing risks of re-identification.

The medical field contains a lot of bias, which oftentimes is hard to quantify and detect. There are 2 main types of bias when involving artificial intelligence in healthcare. Statistical bias can cause an algorithm to produce an output that differs from the true estimate. Social bias, by contrast, refers to inequities that may result in some populations having access to this technology while others do not (Norori et al., 2021). To mitigate these challenges, it's vital that more diverse groups are involved in the testing and development of this technology. Once it is approved for clinical usage, these technologies should be accessible in multiple regions for free as part of preventative care.

While mistakes are inevitable when newly implementing this process, clear accountability chains can increase trust in the system. Identifying who is responsible for errors- whether it's healthcare providers, AI developers, or other stakeholders- can allow for future improvements. Mistakes should also be immediately communicated to the patient, along with the implications of the error.

Finally, it's crucial to remember genetic research and artificial intelligence are both highly experimental fields. While this article proposes the utilization of PRS combined with AI based on current understandings, there is a constant development of new knowledge and studies. It's vital that this proposed process has the ability to adapt to the latest breakthroughs in order to stay current and accurate.

ISSN: 2455-8834

Volume:09, Issue:10 "October 2024"

References

Chacko, M., Sarma, P. S., Harikrishnan, S., Zachariah, G., & Jeemon, P. (2020). Family history of cardiovascular disease and risk of premature coronary heart disease: A matched case-control study. *Wellcome Open Research*, *5*(5), 70. https://doi.org/10.12688/wellcomeopenres.15829.2

CDC. (2024, January 17). *Social Determinants of Health (SDOH)*. CDC. https://www.cdc.gov/about/priorities/why-is-addressing-sdoh-important.html

Choi, S. W., Mak, T. S.-H., & O'Reilly, P. F. (2020). Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols*, *15*(9), 2759–2772. https://doi.org/10.1038/s41596-020-0353-1

Gene and Environment Interaction. (2018). National Institute of Environmental Health Sciences. https://www.niehs.nih.gov/health/topics/science/gene-env#:~:text=Introduction

Hernandez, L. M. (2016). Genetics and Health. Nih.gov; National Academies Press (US). https://www.ncbi.nlm.nih.gov/books/NBK19932

Krisela Steyn, & Albertino Damasceno. (2009). *Lifestyle and Related Risk Factors for Chronic Diseases*. Nih.gov; The International Bank for Reconstruction and Development / The World Bank. https://www.ncbi.nlm.nih.gov/books/NBK2290/

Lewis, C. M., & Vassos, E. (2020). Polygenic risk scores: from research tools to clinical instruments. *Genome Medicine*, *12*(1). https://doi.org/10.1186/s13073-020-00742-5

Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J. K., & Przeworski, M. (2020). Variable prediction accuracy of polygenic scores within an ancestry group. ELife, 9. https://doi.org/10.7554/elife.48376

Norori, N., Hu, Q., Aellen, F. M., Faraci, F. D., & Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. Patterns, 2(10), 100347. https://doi.org/10.1016/j.patter.2021.100347

National Human Genome Research Institute. (2020, August 11). *Polygenic Risk Scores*. Genome.gov. https://www.genome.gov/Health/Genomics-and-Medicine/Polygenic-risk-scores

Polygenic Risk Score (PRS). (n.d.). Genome.gov. https://www.genome.gov/genetics-glossary/Polygenic-Risk-Score

ISSN: 2455-8834

Volume:09, Issue:10 "October 2024"

Polygenic Risk Scores in Complex Disease Research. (2020). Illumina.com. https://www.illumina.com/areas-of-interest/complex-disease-genomics/polygenic-risk-scores.html#:~:text=Polygenic%20risk%20scores%20

Social determinants of health. (n.d.). Www.who.int.

https://www.who.int/health-topics/social-determinants-of-health#tab=tab_

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. Nature Reviews Methods Primers, 1(1). https://doi.org/10.1038/s43586-021-00056-9

Volk, H. E., Kerin, T., Lurmann, F., Hertz-Picciotto, I., McConnell, R., & Campbell, D. B. (2014). Autism Spectrum Disorder. Epidemiology, 25(1), 44–47. https://doi.org/10.1097/ede.0000000000000030

What are complex or multifactorial disorders?: MedlinePlus Genetics. (2021).

Medlineplus.gov. https://medlineplus.gov/genetics/understanding/mutationsanddisorders/complexdisorders/#:~:text=Common%20health%20 problems%20such%20as

WHO. (2024). Social Determinants of Health. World Health Organization. https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1

