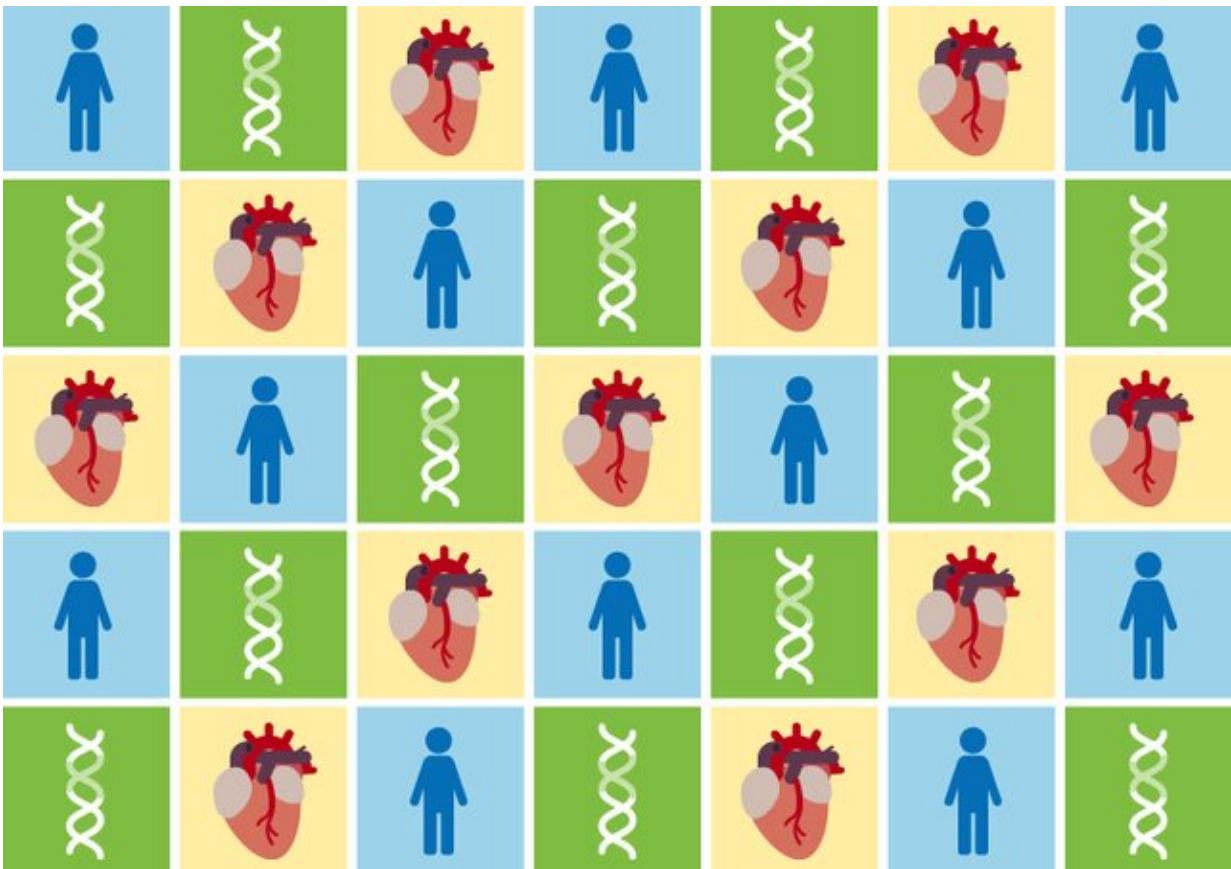


# More diverse datasets lead to better genetic risk prediction for heart disease

July 7 2023, by Leah Eisenstadt

---



Credit: Susanna Hamilton, Broad Communications

Over the past decade, researchers have been developing polygenic scores—calculations of a person's likelihood of getting a disease based

on the millions of small genetic differences across their genome. The accuracy of these scores has improved for some diseases and groups of people, but they continue to fall short for those of non-European ancestry, mainly because the genetic datasets used to calculate these scores have largely come from people of European ancestry.

A new approach from a team led by researchers in the Cardiovascular Disease Initiative at the Broad Institute of MIT and Harvard and at Massachusetts General Hospital (MGH) significantly improves the accuracy of genetic risk prediction of heart disease across all ancestries.

The scientists built a polygenic score using data from [genetic studies](#) involving more than 1 million people. To further improve the score, they also incorporated in their calculations genetic changes associated with 10 related traits such as blood pressure and body mass index. Their new score outperformed all existing scores in predicting risk for [coronary artery disease](#)—the leading cause of death worldwide—among participants of African, European, Hispanic, and South Asian ancestry.

The approach may one day allow clinicians to identify more high-risk individuals earlier in life and recommend interventions such as cholesterol-lowering medicines or lifestyle changes that have been shown to offset and even normalize high genetic risk. Published in *Nature Medicine*, the results suggest that the framework can be applied to improve genetic risk prediction for other traits and diseases, too.

"The ability to identify genetic risk early in life—technically possible even at birth—is powerful, because we don't have to wait for clinical factors like elevated cholesterol to arise," said co-senior author Amit V. Khera, who developed [polygenic scores](#) as a Merkin Institute Fellow at the Broad and is now vice president of genomic medicine at Verve Therapeutics and a cardiologist at Brigham and Women's Hospital.

"Using larger, more diverse datasets, our score can better identify individuals at high risk who would otherwise fly under the radar," said co-senior author Pradeep Natarajan, a Broad associate member who is director of preventive cardiology and the Paul & Phyllis Fireman Endowed Chair in Vascular Medicine at MGH, and an associate professor of medicine at Harvard Medical School.

## Prediction progress

To build the new score, the scientists gathered data on more than a million people, including nearly 270,000 individuals with coronary artery disease—far more than their previous 2018 study that analyzed only tens of thousands of individuals with the disease. With new, more diverse studies released in the past year, such as the U.S. Veterans Affairs Million Veteran Program, the team was also able to incorporate data on more people with African, Hispanic, and South Asian ancestry.

"The European-based scores developed in 2018 didn't work so well in predicting risk for people with other ancestries," said Aniruddh Patel, co-first author of the new study and a cardiologist and researcher in the Natarajan lab at MGH. "So the [scientific community](#) has been focusing on improving prediction across different ancestries."

In addition, Minxian (Wallace) Wang, a study co-first author and former computational biologist at the Broad, developed a pipeline to more precisely capture the influence of genetic variants with smaller impacts on heart disease risk, by prioritizing DNA changes known to influence both risk for heart disease and related traits, such as body mass index, smoking status, and blood pressure. Interestingly, about half of the score's predictive power came from studies of heart disease itself and the other half from studies of these other risk factors.

When applied to a separate dataset from individuals of diverse

ancestries, the new score, called GPSMult, identified more people at the highest and lowest risk of heart disease than all previous scores. For example, those in the lowest percentile of the score had a less than 1% chance of being diagnosed with heart disease by middle age, as compared to a 16% chance for those in the highest percentile. Strikingly, the team was able to identify 3% of unaffected individuals who—based on common DNA variation alone—have a risk for a future cardiac event such as a heart attack as high as people who've already been diagnosed with the disease.

The results demonstrate the benefits of combining traits and multi-ancestry data when calculating polygenic risk scores, and suggest that the approach could improve risk prediction for other illnesses. The scientists are continuing to refine their method by incorporating even larger, more diverse datasets, by employing new computational approaches that account for complex genome architecture, and by integrating clinical risk factors to improve the scores' relevance for physicians.

"What's exciting is that we still haven't reached the theoretical maximum for how good a genetic predictor for [heart disease](#) can be, so these tests will continue to get even better in coming years," said Khera. "We also have a lot of work to do to figure out how best to integrate these tests into clinical practice and ultimately make them standard of care."

**More information:** Aniruddh P. Patel et al, A multi-ancestry polygenic risk score improves risk prediction for coronary artery disease, *Nature Medicine* (2023). [DOI: 10.1038/s41591-023-02429-x](https://doi.org/10.1038/s41591-023-02429-x)

Provided by Broad Institute of MIT and Harvard

Citation: More diverse datasets lead to better genetic risk prediction for heart disease (2023, July

7) retrieved 29 April 2024 from

<https://medicalxpress.com/news/2023-07-diverse-datasets-genetic-heart-disease.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.