

**Article:**

Tristan J Hayeck, George B Busby, Sung Chun, Anna C F Lewis, Megan C Roberts, and Bjarni J Vilhjálmsón.

Polygenic Risk Scores: Genomes to Risk Prediction

Clin Chem 2023; 69(6): 551-7. <https://doi.org/10.1093/clinchem/hvac099>

Guest: Tristan J. Hayeck from the Division of Genomic Diagnostics within the Department of Pathology at the Children's Hospital of Philadelphia and University of Pennsylvania.

Bob Barrett:

This is a podcast from *Clinical Chemistry*, a production of the American Association for Clinical Chemistry. I'm Bob Barrett. Risk prediction, or the estimation of an individual's chances of developing a given medical condition, can help determine frequency of screening, encourage lifestyle changes, or guide treatment decisions. In many cases, genetic information is a key determinant of disease risk and increased testing availability has enabled personalization of risk estimates on the basis of genetic factors.

While estimating risk is relatively simple in conditions defined by an alteration in a single gene that directly causes disease, it becomes harder in conditions that involve hundreds or thousands of small genetic changes. To address the latter scenario, researchers have turned to polygenic risk scores, complicated mathematical models that combine multiple determinants of disease risk into a single predictive score. While polygenic risk scores have shown promise, some questions must be resolved in order to broaden their positive impact.

For instance, risk scores have largely been developed using data from individuals of European descent. How accurately do they predict risk in patients of other genetic backgrounds? How feasible is it to apply population risk estimates to guide care decisions for an individual patient? A Q&A article appearing in the June 2023 issue of *Clinical Chemistry* tackles these questions, and today, we are excited to talk with the article's moderator. Tristan J. Hayeck is an Assistant Professor in the Division of Genomic Diagnostics within the Department of Pathology at the Children's Hospital of Philadelphia and University of Pennsylvania. His clinical and research interests are in improving statistical genetics methods and guidelines for translational medicine.

So, Dr. Hayeck, let me play devil's advocate to start here. Let's say that I'm trying to get a polygenic risk score for an individual. Does it really matter if the score is based on information from a different population than the patient I'm looking at?

Tristan Hayeck: Yeah. I mean, that's a good question and something that's come up multiple times in clinical settings. So, in short, the answer is yes, it can matter a lot. It's been reported there can be large discrepancies in prediction accuracy when we're looking at the discovery population that was used to estimate the genetic effects and then applying that to individuals coming from a different population to get their polygenic risk score. So, let's say you have risk estimates generated from a large European population, and you're looking at type 2 diabetes. Well, that may not be as effective at predicting the genetic risk in an African-American population.

So, genetic or heritable contribution to disease is part of a system or the way I sometimes, and others in the field like to think about it is, it's a complicated network of effects where the genetic contribution is just part of it. So, across different populations, we see different genetic patterns, say, differences in allele frequencies and patterns of linkage disequilibrium. So strong linkage disequilibrium is when you see two alleles together more frequently than would be expected randomly. This tends to be stronger with individuals within the same population because their linkage disequilibrium, their LD patterns, reflect generations of evolutionary history.

Sometimes, differences in effects can be due to interactions across variants or interactions with different environmental exposures. So, getting back to that question, although not universally, there can be low transferability across genetic ancestries in polygenic risk assessment.

Bob Barrett: Right, so that brings up another question. There is a lot that goes into developing polygenic risk scores, including things like genome-wide association studies, or GWAS. What are factors that we should be thinking about addressing even before someone sits down to calculate a polygenic risk score for an individual?

Tristan Hayeck: Yeah. A major thing that sort of all of the experts touch on in the manuscript is the importance of representation, from more and diverse populations, to really get the most effective risk prediction. So, most genome-wide association studies include samples that are predominantly of European descent. From an equity standpoint, studies really need to be planned, executed, and include more diverse groups. In addition, on top of equity, it's also just very important to have diverse samples in terms of effectiveness. So, less studied populations may actually offer lower-hanging fruit in terms of genetic discovery.

So, the idea is, if the prevalence of a given disease of interest is higher in the non-European population, this in turn means

the truly causal variants are potentially at a much higher frequency in the understudied population.

So, in the extreme case, the causal variants may be rare or not even observed in the European population. So, there's no way to even detect them, whatsoever, if you use a study that's purely based off a European population. So, the more information we have, the better we can mechanistically understand the different contributions to disease.

So, then, another factor that we need to be thinking about prospectively in studies is sort of the technology perspective, thinking about the platform that the data are collected on. So, GWAS studies tend to come from SNP array data historically, and that favors common variants. Well, depending on the genetic architecture of the disease, this may make less sense and you need to be looking for more rare variants. Then additionally, to have the most effective models, what you want is you want to have all the different possible risk factors, so including other things like electronic health records or other omics data. Then where possible, trying to get good continuous measurements. So, running longitudinal studies to better understand the interplay of both genetic and other environmental factors over time is going to help get you sort of the best risk prediction.

Bob Barrett: So clearly, there is a lot of nuance here on top of the fact that these are complicated mathematical models. So, how do you think experts should be explaining polygenic risk scores to patients?

Tristan Hayeck: Yeah. I believe sometimes, we sort of need to help patients make a mental shift in interpreting genetic tests or genetic data. It's not uncommon for patients to think of genetic risk factors is being deterministic, and Dr. Lewis talks about this in the manuscript. I was listening to an NPR podcast a while ago and they were talking about using genetic information to better understand disease risk, and the narrator, likely in a somewhat cheeky manner, posed, 'if we have sequence data for all individuals, are we going to be able to predict when everyone is going to die?'

Well, I mean, like we already talked about, genetic factors are important but they're only one piece of the puzzle, and there are potentially a lot of other factors that play a part in disease. So, polygenic risk prediction is not a diagnosis. I think Dr. Busby makes a good comparison when saying it's like if you see high LDL in a patient. That's going to help you assess the risk of cardiovascular disease, or you know that somebody is a smoker, so they're more likely to have higher blood pressure. All these things help us better assess the risk.

So, what we need on top of this is also trained health professionals--physicians, nurses, genetic counselors--sitting down and trying to help patients understand this and understand the different nuances for their specific cases, you know, really, as best they can.

Bob Barrett: So, in terms of reporting standards and guidelines for polygenic risk prediction, what safeguards should be in place with effective evaluation of polygenic risk models and tools?

Tristan Hayeck: Yeah. I mean, this gets back to some of the same points as before. First demonstrating accurate risk prediction in diverse populations. So, training and validation in representative populations so that you know that what is being implemented is going to be effective on the group being studied, or striving to get as close as possible to that. So, disease and genetic ancestry-specific estimates. And then how do we assess, what metrics are we using? Without getting too technical, are we in a setting where we care about touching every possible variant?

Other times, the best utility for a polygenic risk prediction is looking at the extremes. Say, we want to figure out which individuals are in the top ten percentage of being at high disease risk. In the literature and as method developers, we may sometimes reasonably focus on things like predictive accuracy or strength of correlation, but we also need to be thinking about how the models are going to be used in application. Then, no matter the type of study, a lot of the time, in order to protect the privacy of the patients, which is very important, not all the genetic data are shared or only summary level data is available. So, this often makes it either difficult or impossible to assess the predictive utility.

So, this goes back to your earlier question. We need to be planning ahead to have some data for sub-cohort analysis including different, and additional, summary statistics looking at things like genetic ancestry but also other demographic factors and other risk factors too, to both fit and assess our models.

Then lastly, but still very important, we should also be paying attention to, and making sure that, these models aren't overfitting, because if they're overfitting, then there really isn't going to be any reproducibility.

Bob Barrett: Well, finally, Doctor, where do you see polygenic risk scores playing a role in the future?

Tristan Hayeck: Yeah. There are several areas where PRS could play a role, from screening, treatment choices, and diagnosis. Dr. Vilhjálmsón talks about this. You know, thinking about polygenic risk scores from the perspective of screening. It

makes sense to try and leverage polygenic risk scores when there are different preventative measures that can help the patient. For example, if we have individuals that we know are at high risk, it can be recommended to take certain medications, or they can make different lifestyle changes like changing their diets or exercising. Then even simpler things like if we identify the high risk patients, well then, we should just have increased monitoring of them.

Another area is treatment choices. So, in some settings, polygenic risk scores have been shown to correlate with drug efficacy. You know, we don't expect all drugs to be effective for all people in all settings. So, polygenic risk scores are a way to potentially better inform treatment decisions and PRS certainly has the potential to be part of the promise of precision medicine. Then related is diagnosis refinement. Let's say you're looking at clinical differentiation, so you have two different potential health outcomes that you're considering with very similar symptoms, but the treatment course may change. So, if you have a more accurate diagnosis, say, an individual is type 1 versus type 2 diabetic, that can potentially influence how you prescribe, or if you prescribe, insulin.

So, PRS has the potential to improve both the short- and long-term treatment and really, a major factor is we need to have more diverse studies to improve polygenic risk prediction, which in term has the potential to provide real precision medicine solutions. You know what? It's not one pill for one person. It's using information from many studies and many diverse populations in order to better inform what we do at an individual level, integrating genetic data but also integrating all the other different risk factors. I think that's where a lot of the real promise for polygenic risk prediction is going to be.

Bob Barrett:

That was Dr. Tristan Hayeck from the Children's Hospital of Philadelphia. He served as moderator for a Q&A article on polygenic risk scores in the June 2023 issue of *Clinical Chemistry* and he's been our guest in this podcast on that topic. I'm Bob Barrett. Thanks for listening.