

Performance of polygenic risk scores across different ancestry groups in the UK biobank data

Arina V. Nostaeva^{1,2}, Tatiana I. Shashkova¹, Lennart C. Karssen^{2,3}, Yurii S. Aulchenko^{1,2}

¹*Institute of Cytology and Genetics, Novosibirsk, Russia*, ²*PolyOmica, 's-Hertogenbosch, The Netherlands*

³*PolyKnomics, 's-Hertogenbosch, The Netherlands*

Presenter email: avnostaeva@gmail.com

ABSTRACT

Many studies have shown that Polygenic Risk Scores (PRS) are a powerful tool for prediction of the risk of future disease. For people of European descent, a large body of data is available to aid the development of PRS models, thanks to the active genetic research undertaken in these populations over the past decades. At the same time, resources to develop PRS models for people of non-European descent are limited. One potentially cost-effective approach to develop and improve PRS predictive models for underrepresented ancestry groups could be based on the transfer of models across populations and ethnic groups.

Here, we develop and validate genome-wide polygenic scores for body mass index (BMI) and breast cancer (BC) and analyse the portability of polygenic risk score models trained and calibrated in people of European ancestry to individuals of non-European descent.

We developed draft PRS models using the SBayesR method that utilises summary statistics from previously published genome-wide association studies (GWAS). To avoid “testimation” bias, we validated and calibrated the models in independent data sets not overlapping with the samples used for GWAS. These out-of-sample test datasets were subsamples of UK Biobank participants. We first used a random sample of individuals of white British born in England as a calibration data set. We then performed validation and estimated the predictive performance of the model in a series of datasets including random subsamples of white British born in England, Wales and Scotland, Ireland, and outside of the UK (“other Europeans”), as well as samples of individuals of Pakistani, Indian, African, Caribbean, and Chinese ancestry.

Our results demonstrate a low portability of PRS from European ancestry to non-European for both BMI and BC. For BMI, in case the model was not recalibrated, it performed reasonably well for different groups of European ancestry (the proportion of variance explained, R^2 , varies from 6% to 8% and does not deviate significantly from that observed in the calibration data set), but had a low to no predictive accuracy ($R^2 < 2\%$) for people of non-European descent. Recalibration allowed some prediction in people of non-European descent, but the proportion of variance explained (R^2 from 2% to 5%) was lower than in people of European descent (R^2 from 7% to 8%). A similar general picture was observed for BC, although the low number of cases in people of non-European descent make our conclusions less clear-cut.

We conclude that PRS models developed for European populations cannot be directly transferred to other populations. Minimally, the models, developed for European populations, require recalibration using data from the target population. The model performance drops with the increased genetic distance between the population where the model was developed (training and calibration) and the target population.