

NAME OF PRESENTING AUTHOR: Oliver Pain

EMAIL ADDRESS OF PRESENTING AUTHOR: oliver.pain@kcl.ac.uk

LOCATION OF PRESENTING AUTHOR: Europe

TIME ZONE OF PRESENTING AUTHOR: British Summer Time

TYPE OF SUBMISSION: Oral paper/ Poster

MEMBER STATUS: Non-member

ELIGIBLE FOR THOMPSON AWARD: No

ELIGIBLE FOR ROWEWARD: No

TITLE: Evaluation of Polygenic Prediction Methodology within a Reference-Standardized Framework

FULL AUTHOR LIST: Oliver Pain^{1,2}, Kylie P. Glanville¹, Saskia Hagenaars¹, Saskia Selzam¹, Anna E. Fürtjes¹, Helena Gaspar¹, Jonathan R. I. Coleman¹, Kaili Rimfeld¹, Gerome Breen^{1,2}, Robert Plomin¹, Lasse Folkersen², Cathryn M. Lewis^{1,2,3}.

AFFILIATIONS:

¹Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, SE5 8AF, United Kingdom

²NIHR Maudsley Biomedical Research Centre, South London and Maudsley NHS Trust, London, SE5 8AF, UK.

³Institute of Biological Psychiatry, Sankt Hans Hospital, Copenhagen, 4000 Roskilde, Denmark

⁴Department of Medical and Molecular Genetics, Faculty of Life Sciences and Medicine, King's College London, London, SE1 9RT, UK.

KEYWORDS: polygenic scores, reference-standardized, prediction, elastic net, GWAS

ABSTRACT:

Background: The predictive utility of polygenic scores is increasing but it is unclear which polygenic scoring method performs best. It is often advantageous to calculate polygenic scores using a reference-standardized framework, using a common set of variants and reference-estimated linkage disequilibrium and allele frequencies. This study evaluates the predictive utility of several polygenic scoring methods within a reference-standardized framework.

Methods: The following methods are evaluated using anthropometric, neuropsychiatric

and complex disease outcomes measured in UK Biobank and Twins Early Development Study (TEDS): p-value thresholding and clumping (pT+clump), SBLUP, lassosum, LDpred, PRSs and SBayesR. Strategies to identify optimal p-value threshold and shrinkage parameters are compared, including 10-fold cross validation, pseudovalidation (no validation sample), and multi-polygenic score elastic net models.

Results: Using 10-fold cross-validation to identify the most predictive p-value threshold or shrinkage parameter, lassosum, PRSs and LDpred provided optimal prediction (relative improvement of 14-17% over pT+clump). Using pseudovalidation to optimize the polygenic score, the best method was PRSs, with a relative improvement of >11% over other pseudovalidation methods (lassosum, SLBLUP, SBayesR, LDpred), and only 1% worse than the best polygenic score identified by 10-fold cross validation. Elastic net models containing polygenic scores based on a range of parameters consistently improve prediction over any single polygenic score.

Conclusion: Within a reference-standardized framework, the best polygenic prediction was achieved using lassosum, PRSs, modelling multiple polygenic scores that are derived using a range of parameters. As polygenic scores are widely applied in research studies, users should be aware of differences in prediction across methods.

GRANT SUPPORT: Funded by the MRC (MR/N015746/1), and the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London
