Improving interpretability of polygenic scores using only summary statistics

Oliver Pain[1,2], Alexandra C. Gillett[1], Jehannine C. Austin[3], Lasse Folkersen[4], Cathryn M. Lewis[1,2,5].

[1]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK
[2]NIHR Maudsley Biomedical Research Centre, South London and Maudsley NHS Trust, London, UK
[3]Departments of Psychiatry and Medical Genetics, University of British Columbia, Vancouver, Canada
[4]Institute of Biological Psychiatry, Sankt Hans Hospital, Copenhagen, Denmark
[5]Department of Medical and Molecular Genetics, Faculty of Life Sciences and Medicine, King's College London, UK

Correspondence: oliver.pain@kcl.ac.uk

Introduction

- Genetic variation leads to substantial individual differences in health and disease[1]
- Polygenic scores (PGS) capture a part of the genetic liability/propensity for a given outcome and can inform personalised medicine[2]
- To accurately interpret the meaning of a PGS, it must be converted to the absolute scale
  - E.g. given a PGS for schizophrenia in the 90th percentile of the general population, what is the probability of developing schizophrenia?
- We have developed an approach for converting PGS to the absolute scale, using only summary statistics, for binary and normally distributed outcomes:
  - Binary: Population prevalence and PGS area-under-the-ROC curve (AUC).
  - Normally distributed: Trait mean and standard deviation, and PGS variance explained (R^2).
- Given PGS AUC/R^2 is often unknown, we explore approaches estimating the AUC/R^2 of PGS using genome-wide association study (GWAS) summary statistics only.

Methods

Conversion to absolute scale

- PGS can be modelled as a mixture of two normal distributions for binary outcomes, and single normal distribution for normally distributed outcomes
- We estimate outcome values within each PGS quantile:
  - Binary outcomes: Probability of being a case
  - Normally distributed outcomes: Trait mean and standard deviation

Estimation of PGS AUC/R^2

- The lassosum pseudovalidation method[3] estimates the correlation (R) between PGS and outcome. This correlation can then be converted into an AUC or R^2

Validation

- 11 outcomes in UK Biobank[4] using PGS derived using the DBSLMM method[5]
- Comparison of observed and estimated values on the absolute scale, using either the observed AUC/R^2 in UK Biobank, or lassosum estimated AUC/R^2

Results

Using observed PGS AUC/R^2

- Observed and estimated values on the absolute scale were highly concordant (Figure 1)
- The concordance was reduced for body mass index (BMI) due to skewness of BMI in UK Biobank

Estimation of PGS AUC/R^2

- Lassosum estimates of AUC/R^2 were highly concordant with observed values for 8/11 outcomes
- Results were discordant for the three autoimmune disorders, leading to discordance between estimated and observed absolute values at the extremes of the PGS distribution

Discussion

- Our summary-statistic based approach for converting PGS to the absolute scale provides highly accurate results when the PGS AUC/R^2 is accurately specified
- Lassosum pseudovalidation provides an accurate estimates of PGS AUC/R^2 in most scenarios, though specifying the observed PGS AUC/R^2 is more reliable
- We created online implementations of the absolute scale conversions for both binary and normally distributed outcomes to help individuals interpret PGS (Figures 3-4)
- Further information available at https://opain.github.io/GenoPred/

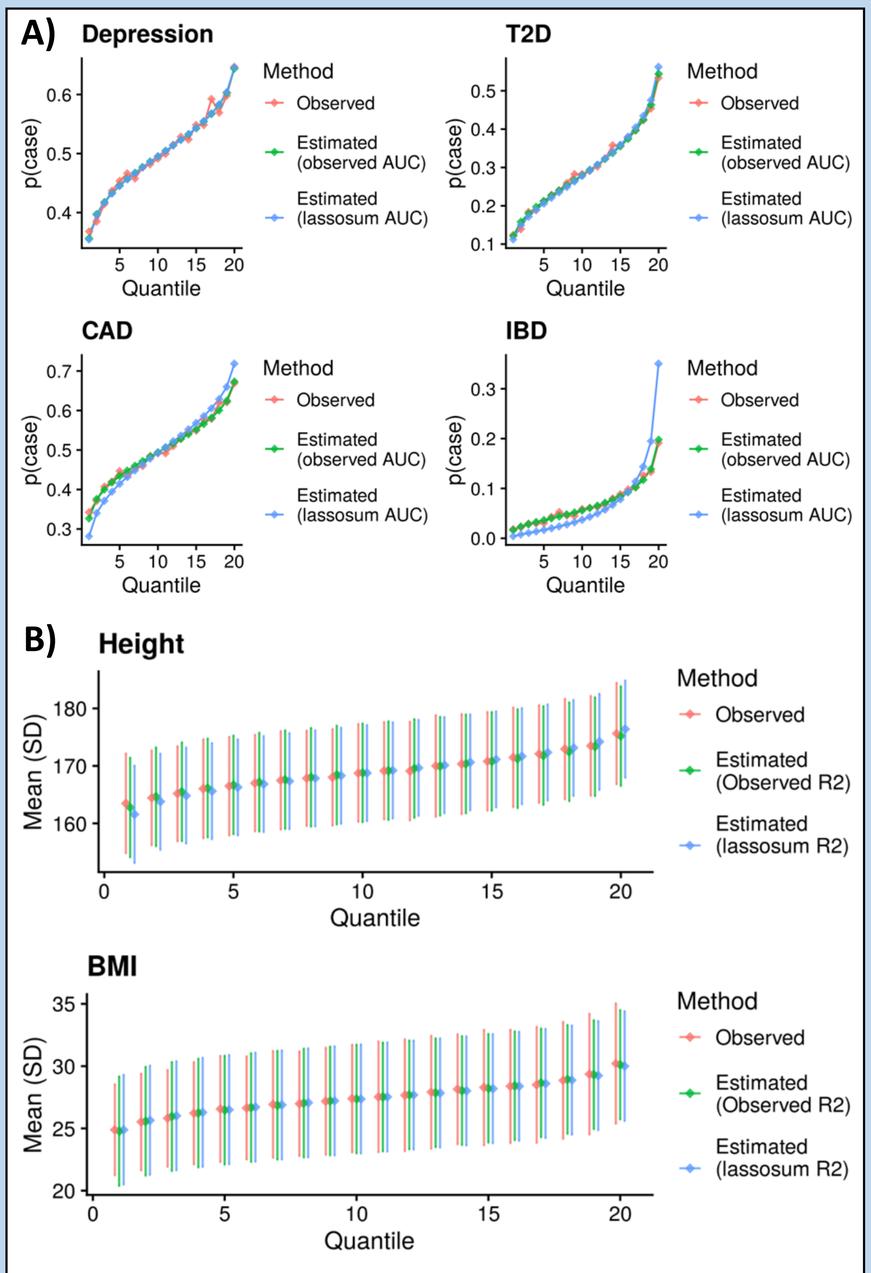

Figure 1. Comparison of observed and estimated values on the absolute scale across PGS quantiles. **A)** Compares the probability of being a case. **B)** Compares the mean and standard deviation of the trait. Estimated values are based on either the observed PGS AUC/R^2 or lassosum estimated AUC/R^2. Showing results for selected outcomes.

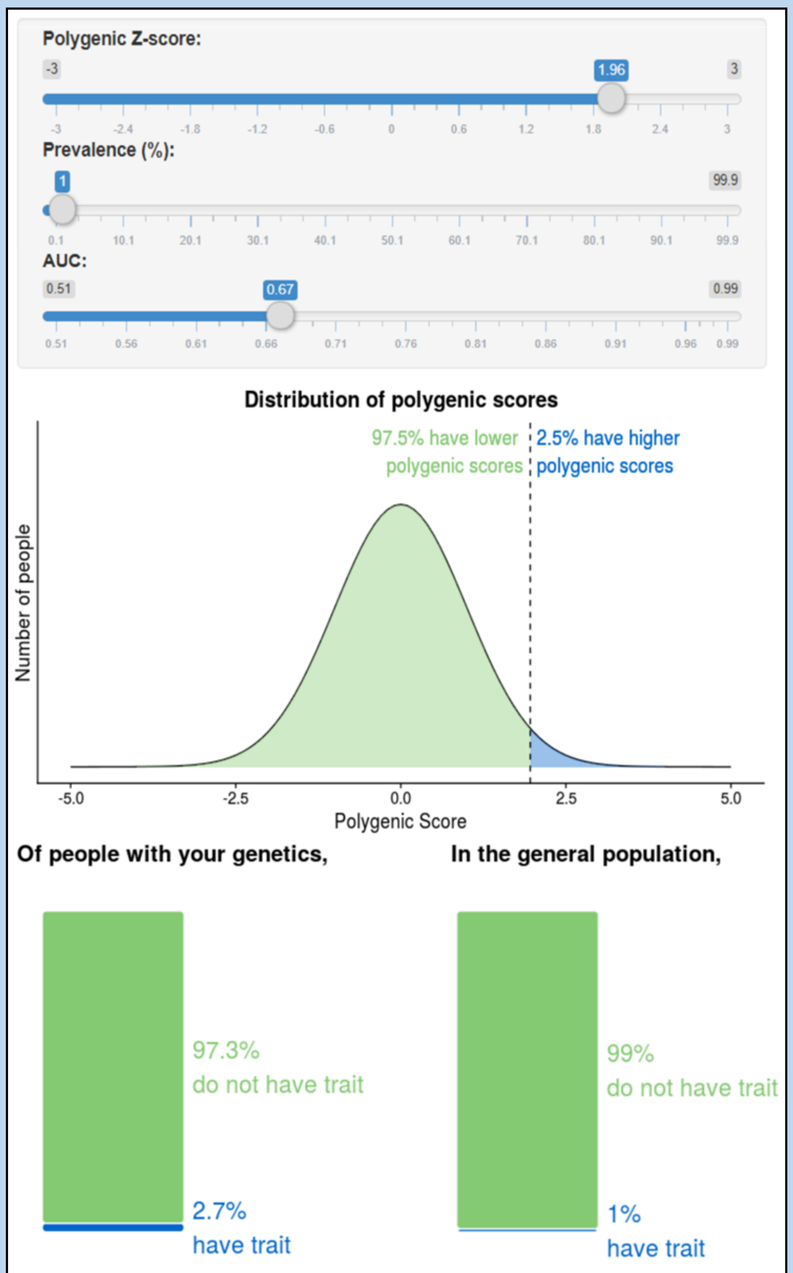

Figure 2. Webtool for converting PGS for binary outcomes onto the absolute scale based on the population prevalence of the outcome and the PGS AUC.

Link: https://opain.shinyapps.io/risk_visualiser/

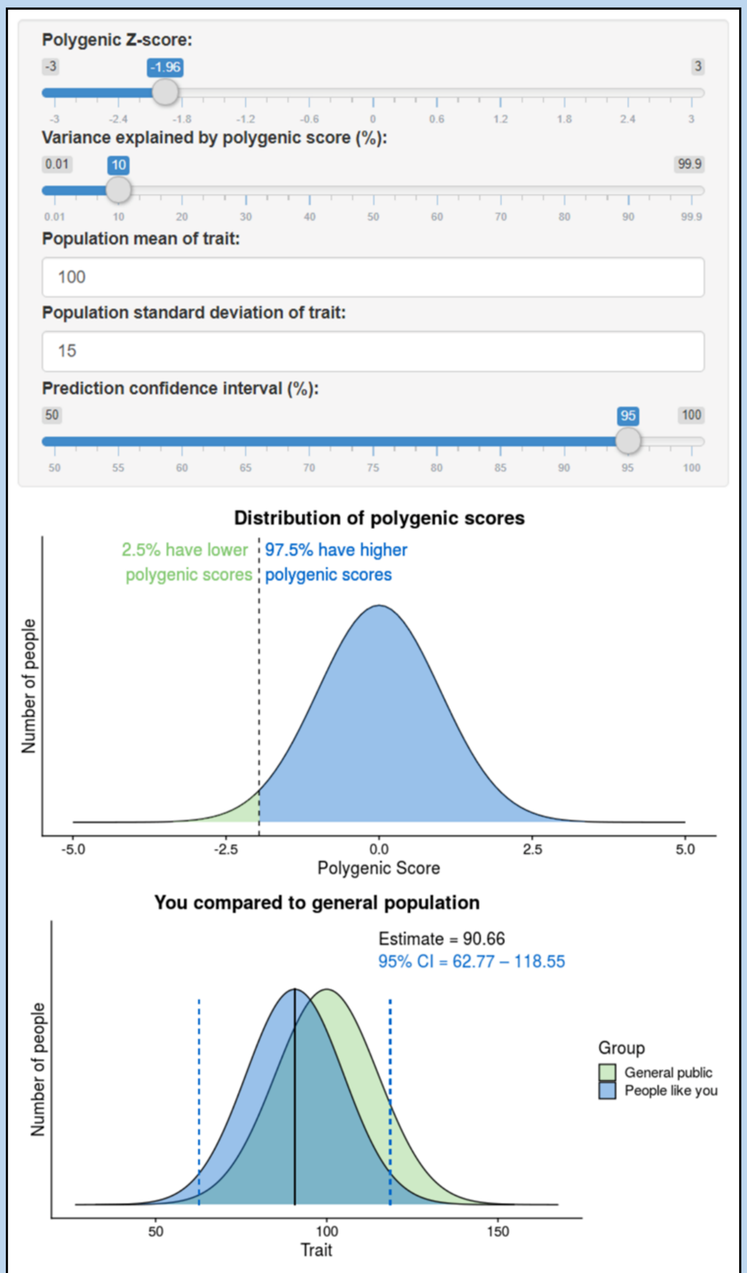

Figure 3. Webtool for converting PGS for normally distributed outcomes onto the absolute scale based on the population trait mean and standard deviation, and the PGS R^2.

Link: https://opain.shinyapps.io/trait_visualiser/

References

1. Polderman, T. J. C. et al. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, *47*(7), 702–709.

2. Lewis, C. M., & Vassos, E. (2020). Polygenic risk scores: from research tools to clinical instruments. Genome Medicine, 12, 1–11.

3. Mak, T. S. H. et al. (2017). Polygenic scores via penalized regression on summary statistics. Genetic Epidemiology, 41(6), 469–480.

4. Bycroft, C. et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature, 562(7726), 203–209.

5. Yang, S., & Zhou, X. (2020). Accurate and scalable construction of polygenic scores in large biobank data sets. The American Journal of Human Genetics, 106(5), 679–693.