Understanding Society Polygenic Scores, *2019*

USER GUIDE

Version 1, *September 2022*

Authors: Anna Dearman, Yanchun Bao and Meena Kumari

Institute for Social and Economic Research
University of Essex
Colchester
Essex

Contents

1. Introduction

A polygenic score (PGS) is a continuous variable that reflects an individual's propensity towards a given trait, as calculated by comparing their genetic profile to the summary statistics generated from a genome-wide association study (GWAS) of the trait. Traits are wide-ranging and can include disease status, behaviours, and blood levels of biomolecules, among many others. In the case of disease status, polygenic scores are sometimes termed "polygenic risk scores" (PRSs) as they reflect the risk, or liability, of having the disease.

For a given trait, the GWAS summary statistics comprise a long list of single-nucleotide polymorphisms (SNPs), or points on the genome which have two possible variants (alleles), which have been tested for association with the trait of interest. For each SNP, one of the alleles, often referred to as the "effect allele", has an effect size and p-value reported in the summary statistics. An individual has 0, 1 or 2 copies of the effect allele per SNP. PGSs are sums - usually weighted by effect size - of the effect alleles that an individual possesses. A biological glossary can be found on the _Understanding Society_ biomarker topic website.

2. Methods

For each PGS that is made available in _Understanding_ Society, a corresponding README file and a separate list of all SNPs included in the calculation will also be made available. The README contains a reference to the GWAS paper and details of the methods used to generate the PGS in _Understanding Society_. PGS calculation methods vary in terms of the software used, the exclusion or inclusion of SNPs based on different clumping methods and p-value thresholds, weighted vs unweighted approaches, etc.

3. Discussion and Recommendations

PGSs generated in _Understanding Society_ can be used as explanatory variables in a range of analyses, including as an instrumental variable (IV) in Mendelian Randomization (MR) studies to investigate the causal effect of the PGS trait with health/social outcomes – for examples using the PGS of body mass index (BMI) please see [1, 2]. MR entails a special set of assumptions; for more information about MR see [3-6].

Any findings from analyses that use PGSs as explanatory variables require careful interpretation. Users are strongly advised to read the corresponding GWAS paper, a reference to which will be included in the README file. It is important to note aspects of the GWAS cohort(s) which may be relevant to the measurement of the trait, such as age and health, and it is especially crucial to note the ancestral population(s), as single-ancestry PGSs have limited portability across different populations due to factors such as differences in allele frequencies and patterns of linkage disequilibrium. It is also important to note any measures of the PGS's explained variance and predictive power included in the GWAS

paper. Finally, genetic associations can occur due to environmental confounding, especially in traits that are more behavioural, social or psychological. More discussion on the application of PGS in social research can be found in [7-12].

4. Data Access

These Data are released through the UK Data Service. The Special Licence (SL) or safeguarded version, SN7587, can be found here:
https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=7587.

These datasets are categorized as Special Licence and therefore have more restrictive access conditions, details of which can be found in the "Access data" tab at the above links. In addition, access to these Special Licence datasets requires users to complete forms justifying why the data is needed. Further details on the application process can be found here: https://ukdataservice.ac.uk/find-data/access-conditions/

5. Citation

If you use *Understanding Society* data you must cite every study that you use.

The citation for the data can be found at
https://www.understandingsociety.ac.uk/documentation/citation

All works which use or refer to these materials should acknowledge these sources by means of bibliographic citation. To ensure that such source attributions are captured for bibliographic indexes, citations must appear in footnotes or in the reference section of publications.

Citing this User Guide

When citing this User Guide, you can use the citation of this particular version quoted below. Note that where an online version is available on the Understanding Society website it is always the most up to date.

Institute for Social and Economic Research. (2022), *Understanding Society:* Polygenic Scores, 2019*, User Guide, Version 1, September, 2022*, Colchester: University of Essex.

References

[1] Hughes A, Bao Y, Smart M, Kumari M. Body mass index, earnings and partnership: genetic instrumental variable analysis in two nationally representative UK samples. bioRxiv 2019. DOI: 10.1101/608588

[2] Howe LD, Kanayalal R, Beaumont R, Davies AR, Frayling TM, Harrison S, et al. Effects of body mass index on relationship status, social contact, and socioeconomic position: Mendelian Randomization study in UK Biobank. 2019:524488

[3] Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. International Journal of Epidemiology. 2013;42(4):1134-44.

[4] Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;362.

[5] Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. Epidemiology. 2017;28(1):30-42.

[6] Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. Statistics in Medicine. 2017;36(11):1783-802.

[7] Morris TT, Davies NM, Hemani G, Smith GD. Population phenomena inflate genetic associations of complex social traits. Science advances 6 (16), eaay0328

[8] Barth D, Papageorge N W, Thom K, Genetic Ability, Wealth, and Financial Decision-Making. IZA discussion paper series, 2017. IZA DP No. 10567.

[9] Beauchamp, J P. Genetic evidence for natural selection in humans in the contemporary United States. PNAS, 2016. 113 (**28**): 7774-7779.

[10] Belsky D W, Moffitt T E, Corcoran D L, et al. The genetics of success: how single nucleotide polymorphisms associated with educational attainment relate to life-course development. Psychological Science, 2016. 27 (**7**): 957-972.

[11] Munafò, M. , Davies, N. M., & Davey Smith, G. (2020). Can genetics reveal the causes and consequences of educational attainment? Journal of the Royal Statistical Society: Series A, 183(2), 681-688. https://doi.org/10.1111/rssa.12543

[12] Abdellaoui, A., Verweij, K.J.H. Dissecting polygenic signals from genome-wide association studies on human behaviour. Nat Hum Behav 5, 686–694 (2021). https://doi.org/10.1038/s41562-021-01110-y