Author of target article
Callie H Burt

Word counts
ABSTRACT 57
MAIN TEXT 984
REFERENCES 246
ENTIRE TEXT 1392

Title
Social scientists would do well to steer clear of polygenic scores

Full name
David Curtis

Institutions
UCL Genetics Institute, University College London
Centre for Psychiatry, Queen Mary University of London

Institutional address
David Curtis, UCL Genetics Institute, UCL, Darwin Building, Gower Street, London WC1E 6BT.

Institutional telephone number
00 44 20 7679 2212

Email
d.curtis@ucl.ac.uk

Web page
http://www.davecurtis.net/dcurtis.html

Abstract

The problems with polygenic scores (PGSs) have been understated. The fact that they are ancestry-specific means that biases related to sociodemographic factors would be impossible to avoid. Additionally, the requirement to obtain DNA would have profound impacts for study design and required resources, as well as likely introducing recruitment bias. PGSs are unhelpful for social science research.

Main text

Burt does an excellent job of debunking some of the hype attaching to sociogenomics and the field of polygenic scores (PGSs) in general (Burt, 2022). While she concludes that PGSs may be not very useful for social science, in fact there are good reasons to regard them as perhaps being worse than useless.

Why should social scientists feel quite comfortable not incorporating PGSs into their research? There is no doubt that genetic factors can have substantial effects on relevant outcomes. For example, countless variants in DNA sequence have been identified which lead to profound intellectual disability, effectively reducing educational attainment to zero (Ilyas et al., 2020). Likewise, it is not up

for debate that the effect of some genetic variation will be moderated by environment. Genetic factors increasing athletic ability will be expected to be associated with increased educational attainment if colleges recruit students on sports scholarships and less so if admission is based only on intellectual capability. So the issue is not that genetic factors do not impact outcomes of interest but rather, as Burt explains, that PGSs are so poor at capturing the genetic variation which is biologically relevant while at the same time being profoundly influenced by exactly the kind of confounders social scientists do not want contaminating their research, such as race, socioeconomic status and parental characteristics.

Although Burt does touch on many of the relevant issues, I would argue that the situation is even more problematic than she presents it to be. In my view, what she refers to as population stratification produces effects of such magnitude and malignancy as to render the proposal to routinely incorporate PGSs as covariates in social science research a complete non-starter. There are two related phenomena. One is that the absolute magnitude of PGSs varies with ancestry and the other is that the strength of the association between a PGS and the trait it is supposed to predict also varies with ancestry (Martin et al., 2019). These are not small effects. The PGS for schizophrenia is much more strongly associated with ancestry than it is with schizophrenia (Curtis, 2018). Researchers working with PGSs now routinely utilise ancestry-specific PGSs produced by carrying out genome wide association studies (GWAS) in relevant cohorts. A PGS for white Europeans will need to be derived from a GWAS of an exclusively white European cohort, a PGS for Asians will be derived from a GWAS of an exclusively Asian cohort (Ho et al., 2022). And so on, except that because Africans have more genetic diversity than other populations a PGS derived from a GWAS of an African cohort will always perform less well than its counterparts for other ancestries.

Given these now well-recognised properties of PGSs it is truly challenging to see how one could consider incorporating a PGS as a covariate in a social science research project. The value of a subject's PGS would be profoundly influenced by their ancestry. If one went down the route of attempting to use an ancestry-specific PGS then a prerequisite would be that a GWAS of the trait in question should have been performed on every relevant ancestry group. Knowing which one to use would require determining the ancestry of each subject. For subjects of mixed ancestry, an attempt would need to be made to combine PGSs (Marnetto et al., 2020). For subjects with African ancestry the PGS would capture less of the genetic risk than for other subjects. Thus, the PGS represents a variable which not only performs badly in terms of measuring genetic risk but also performs more badly for some subjects than others. Such an obvious source of systematic bias would make it difficult or impossible to draw useful conclusions from studies which incorporated it.

There is another way in which Burt's treatment is too kind to PGSs. She has not presented a full account of the difficulties of obtaining them for participants in a social science research project. Once one has obtained SNP genotypes then producing a PGS is a trivial exercise. But obtaining SNP genotypes cannot be done by having the subject fill out a questionnaire or go through a structured interview – they have to actually donate a DNA sample and it has to be processed by a laboratory. Incorporating a PGS into social science research involves adding a whole new biological dimension with a very substantial impact on the overall shape of the project. It also requires that subjects voluntarily provide a DNA sample. While some may be happy to do this, it is unarguably the case that a DNA sample represents a large quantity of personal information which is potentially sensitive in a number of ways (Alsaffar et al., 2022). An individual's genetic profile provides at least some information about their risk of a large number of health conditions. It could potentially be of use to police and security forces who might seek to identify the perpetrator of a crime, or at least one of their relatives. While safeguards may be in place which attempt to prevent the misuse of genetic data some individuals may feel reluctant to provide DNA for reasons which are not wholly irrational. Of especial concern is that one might well expect that factors influencing an individual's enthusiasm

for donating DNA would include a number of factors which might be of interest to social scientists, such as education, race, health, substance misuse and criminality. Thus, introducing DNA sampling as a routine aspect of social science research seems certain to introduce systematic bias into recruitment. And as far as research involving children is concerned, I would argue that the privacy concerns about possible misuse of genetic data would mean that it could not be ethical to obtain DNA even if their parents consented.

The inclusion of PGSs into social science research is impractical and highly likely to introduce bias. For these reasons and others, I believe that PGSs have a negative utility.

Conflict of interest statement

Conflicts of interest: none.

Funding Statement

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

References

Alsaffar, M. M., Hasan, M., McStay, G. P., & Sedky, M. (2022). Digital DNA lifecycle security and privacy: an overview. *Briefings in Bioinformatics*, *23*(2), 1–16. https://doi.org/10.1093/BIB/BBAB607

Burt, C. (2022). Challenging the Utility of Polygenic Scores for Social Science: Environmental Confounding, Downward Causation, and Unknown Biology. *Behavioral and Brain Sciences*.

Curtis, D. (2018). Polygenic risk score for schizophrenia is more strongly associated with ancestry than with schizophrenia. *Psychiatric Genetics*, *28*(5), 85–89. https://doi.org/10.1097/YPG.0000000000000206

Ho, W. K., Tai, M. C., Dennis, J., Shu, X., Li, J., Ho, P. J., Millwood, I. Y., Lin, K., Jee, Y. H., Lee, S. H., Mavaddat, N., Bolla, M. K., Wang, Q., Michailidou, K., Long, J., Wijaya, E. A., Hassan, T., Rahmat, K., Tan, V. K. M., … Teo, S. H. (2022). Polygenic risk scores for prediction of breast cancer risk in Asian populations. *Genetics in Medicine*, *24*(3), 586–600. https://doi.org/10.1016/J.GIM.2021.11.008

Ilyas, M., Mir, A., Efthymiou, S., & Houlden, H. (2020). The genetics of intellectual disability: advancing technology and gene editing. *F1000Research*, *9*. https://doi.org/10.12688/F1000RESEARCH.16315.1

Marnetto, D., Pärna, K., Läll, K., Molinaro, L., Montinaro, F., Haller, T., Metspalu, M., Mägi, R., Fischer, K., & Pagani, L. (2020). Ancestry deconvolution and partial polygenic score can improve susceptibility predictions in recently admixed individuals. *Nature Communications 2020 11:1*, *11*(1), 1–9. https://doi.org/10.1038/s41467-020-15464-w

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*, *51*(4), 584–591. https://doi.org/10.1038/s41588-019-0379-x