Natural selection in the Health and Retirement Study

David Hugh-Jones

2023-09-01

I investigate natural selection on polygenic scores in the contemporary US, using the Health and Retirement Study. Results partially support the economic theory of fertility as an explanation for natural selection: among both white and black respondents, scores which correlate negatively (positively) with education are selected for (against). Selection coefficients are larger among low-income and unmarried parents, but not among younger parents or those with less education. I also estimate effect sizes corrected for noise in the polygenic scores.

Hugh-Jones and Abdellaoui (2022) explain patterns of natural selection on polygenic scores in the UK, using an economic theory of fertility derived from G. S. Becker and Tomes (1976). The theory has two components.

1. There is a trade-off between time spent working and raising children. This leads people with more human capital and higher expected wages to have fewer children. Evidence for this is that polygenic scores which correlate positively with human capital correlate negatively with number of children, i.e. they are being selected against; conversely, scores which correlate positively with human capital are being selected for.
2. The trade-off is sharper for low-income people, people with low human capital, and single parents. As a result, natural selection is stronger among these groups. Evidence for this is that scores' regression coefficients on number of children are larger among people with lower income or less education, and single parents.

Here, I make an independent test of the theory in the US population, using the Health and Retirement Survey (HRS 2023a, 2023b). The HRS is more representative of the population than UK Biobank, which addresses one potential weakness of the previous paper. The HRS also provides precalculated polygenic scores (Ware et al. 2020) for both black and (non-hispanic) white participants, so I can check whether the theory predicts patterns of natural selection in both these ethnicities.

Data

The HRS sample focuses on cohorts born between 1920 and 1960, but contains some younger and older participants. I include only male participants born before 1965 and female participants born before 1970, which guarantees that most will have completed their fertility by 2010. The resulting sample contains 10619 genotyped white participants and 2803 genotyped black participants.

Genotyping took place in 2006, 2008 and subsequent years. I discard obsolete PGS for which there is a newer, more accurate score targeting the same phenotype. I also discard PGS for number of children ever born (but keep scores for age at first birth). This leaves a total of 58 scores. PGS are rescaled to zero mean and unit variance within each ethnic group, so coefficient sizes are not directly comparable between ethnic groups, but are measured in standard deviations of the within-ethnic-group score. In all regressions using PGS, I control for ten within-ethnicity principal components of the DNA array data.

The key dependent variable is relative lifetime reproductive success (RLRS): number of children ever born, divided by mean number of children of people born in the same year. RLRS is calculated pooling ethnicities, i.e. treating them as members of the same biological population.

The HRS contains weights which match survey respondents to the US population. I use weights for the biomarker subsample (*BIOWGTR in the HRS tracker file). Since half the sample enters the extended interview including biomarker data in each biannual survey, I weight individuals by either their 2010 weight or their 2012 weight. This maximizes the available sample of both black and white respondents, and should approximately match the US population of the sample cohorts between 2010 and 2012. Statistical tests are adjusted for clustering and stratification using the R "survey" package (Lumley 2023).

Results

I estimated coefficients of PGS on RLRS among black and white respondents separately. These are not meant to identify causal effects; recall that natural selection involves correlation, not necessarily causation, between selected characteristics and fertility. Appendix Figure 4 shows coefficients for white respondents only; power is too low for individual PGS estimates to be informative in the black sample. Standard errors are large because of the relatively low sample sizes, and only 1 score is significant at Bonferroni-corrected $p < 0.05/116$. But I am most concerned with looking at patterns across scores rather than judging the significance of individual scores.

Figure 1 plots each PGS's partial correlation with RLRS against its partial correlation with educational attainment. The relationship is negative among whites (correlation -0.824, bootstrap 95% C.I. -0.984 to -0.664) and negative but insignificant among blacks (correlation -0.294, bootstrap 95% C.I. -0.758 to 0.171). Survey bootstraps (Canty and Davison 1999) are used so as to make inferences from the sample of respondents.

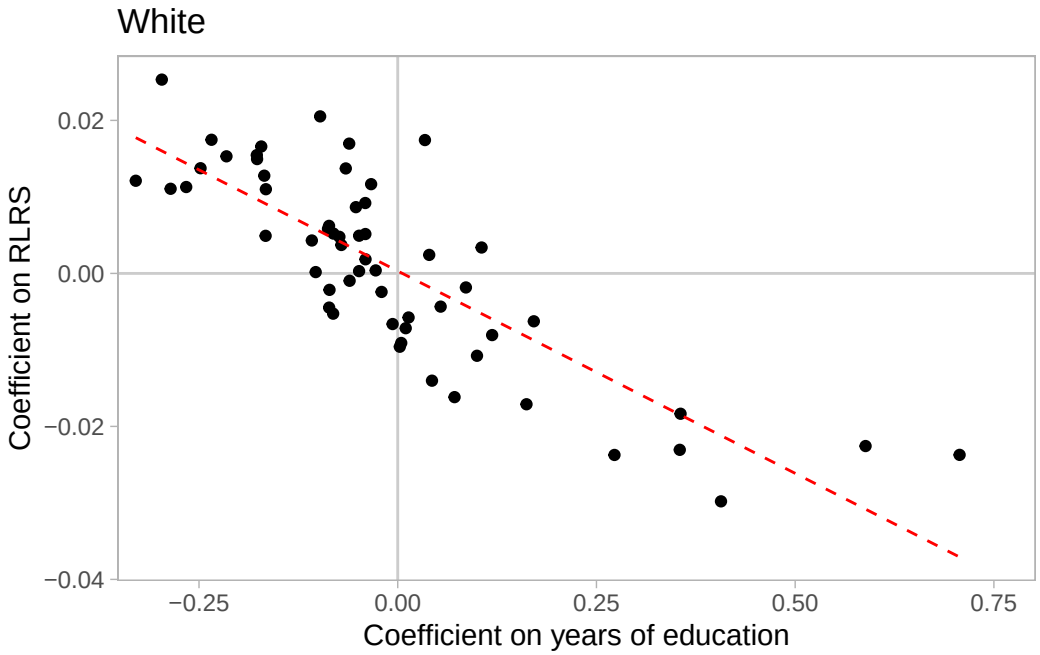
I can also examine natural selection in the previous generation, by regressing PGS on respondents' number of living siblings in 2010. (Data for dead siblings has too many missing values to use.) I reweight respondents by the reciprocal of their number of siblings, to account for parents of many siblings being more likely to be a parent of a respondent. Parents of no siblings cannot be included, so coefficient sizes are not comparable across the generations. Appendix Figure 5 plots coefficients on number of siblings versus coefficients on years of education. Correlations are insignificant for both ethnic groups, with large standard errors (whites: correlation -0.225, bootstrapped 95% C.I. -0.546 to 0.096; blacks: correlation 0.373, bootstrapped 95% C.I. -0.129 to 0.874). However, there is a positive and significant correlation across generations, i.e. between PGS coefficients on number of siblings and number of children (among whites: 0.417, bootstrapped 95% C.I. 0.109 to 0.724.

I next test part 2 of the theory by interacting PGS with measures of education, income, marital status, and age at first birth. Education is years of education, split at 12 years. Income is respondent's mean wage income over all surveys, residualized on a full set of birth year dummies, and median-split. From here on I only use the white subsample: there are too few black respondents to be informative.

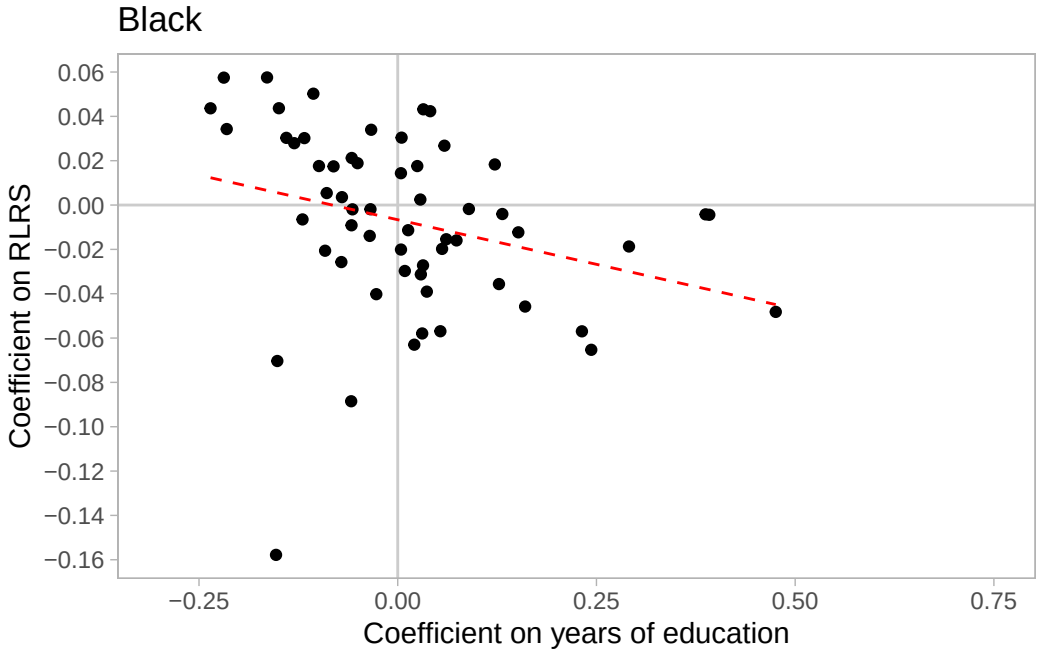
Appendix Figure 6 shows essentially no difference in the distribution of absolute coefficient sizes between respondents with more and less education. Appendix Figure 7, on the other hand, shows that coefficients appear larger among lower-income respondents. Appendix Figure 8 shows absolute coefficients for respondents who were married in 2010 against all other statuses. Coefficients appear larger for unmarried respondents.

Appendix Figure 9 splits respondents by age at first birth (median within each gender). Note that the N is lower here due to missing values. There are no large differences in absolute effect size between younger and older parents. As in the UK Biobank sample, correlations of PGS coefficients between younger and older parents are negative (-0.319), and for 42 out of 58 PGS, coefficients have opposite signs. However, in bootstraps, these statistics are not significantly different from 0 and 29 respectively. Also, in the UK Biobank sample, PGS coefficients on fertility controlling for age at first birth were negatively correlated with the uncontrolled coefficients; here, that isn't true, with coefficients remaining positively correlated.

Table 1 shows means and 95% confidence intervals for mean absolute coefficient sizes in each pair of groups, and the difference between them, using bootstraps. Differences for income and marriage are significant.



(a)



(b)

Figure 1: Scatterplot of PGS regression coefficients on RLRS against coefficients on years of education. Each dot is one polygenic score. Controls include 10 principal components of genetic array data. Dashed lines are fitted from linear regressions.

Table 1: Bootstrap estimates and 95% confidence intervals for mean absolute coefficients of PGS on RLRS in "low" and "high" groups among white respondents. Groups are: 0-12 years education vs. 13-17 years; below vs. above median income; all others vs. married; below vs. above median age at first birth; born before vs. after 1942. 199 bootstraps.

	Low group	High group	Difference
Education	0.011 (0.008 to 0.014)	0.012 (0.007 to 0.016)	-0.001 (-0.005 to 0.004)
Income	0.017 (0.014 to 0.020)	0.011 (0.007 to 0.014)	0.006 (0.002 to 0.010)
Marriage	0.017 (0.013 to 0.021)	0.011 (0.008 to 0.015)	0.006 (0.001 to 0.010)
Age 1st birth	0.013 (0.007 to 0.019)	0.013 (0.009 to 0.018)	0.000 (-0.005 to 0.005)
Birth year	0.011 (0.007 to 0.015)	0.014 (0.010 to 0.017)	-0.003 (-0.007 to 0.002)

The economic theory of fertility also implies that correlations between education and RLRS should be more negative for lower-income/education people, single parents and people who have children earlier.[1] This in turn implies that correlations between PGS coefficients on education and on RLRS should be more negative among these groups. I tested this with bootstraps, but confidence intervals were always too wide to be informative.

Why does the US data show fewer differences by socio-economic status (SES) than the UK? One possibility is that SES maps on to race in the US, so that ethnic differences here capture some of the variation seen in the UK. The regression coefficient of phenotypic educational attainment on RLRS is more negative among black than white respondents (blacks: -0.074, 95% C.I. -0.091 to -0.056; whites: -0.029, 95% C.I. -0.034 to -0.023; cf. Goldscheider and Uhlenberg (1969), Johnson (1979), Yang and Morgan (2003)). And the slope of PGS education coefficients on fertility coefficients is larger among black respondents, though imprecisely estimated (see Figure 1). But comparisons of PGS selection coefficients between the ethnic groups are hard because of the smaller sample size and differences in the scores' predictive power, so this hypothesis can only be speculative. Looking within the sample, changes in PGS by birth year are small for both groups, and are probably mostly driven by selective mortality. Another possibility is that the US cohort were exposed to a smaller welfare state than the UK cohort, since many of them had children before the "Great Society" programs of the 1960s. The last line of Table 1 shows that effect sizes are larger for respondents born after 1942, but the difference is imprecisely estimated and not significant.

Lastly, we would like to know natural selection's effect sizes. The bivariate correlations of PGS with RLRS gives the change in one generation in the PGS due to natural selection, measured in standard deviations. Polygenic scores contain error, so estimated correlations

[1] See Hugh-Jones and Abdellaoui (2022) equation (6) and following.

are biased towards zero compared to the correlation of the true PGS. They can be scaled up by

$$\hat{\beta}_{TRUE} = \hat{\beta}_{PGS}\sqrt{\frac{h^2}{R^2_{PGS}}}$$

where h^2 is the heritability of the PGS target phenotype and R^2_{PGS} is the R^2 of the measured PGS on the target phenotype (J. Becker et al. 2021). I make this calculation using heritabilities from twin studies, and SNP- or chip-heritabilities (the total variance explained by SNPs recorded on array chips).[2]

Figure 2 reports estimates for phenotypes where selection effects were nominally significant and a measure of the phenotype was available in-sample and was significantly predicted by the PGS. As a rule of thumb, a 0.1 standard deviation change in a polygenic score over a generation might count as "serious": about 54% of the new generation will be below the parents' mean. Many upper confidence bounds meet that threshold, but lower bounds are often small or include zero. The confidence bounds capture uncertainty from sampling variation, but not other sources, including uncertainty about the true h^2 , limitations of the within-sample phenotypes, noise from correlated environments, and for twin-heritability, different relationships with fertility among variants not measured on the chip. Given all this, the estimates mostly show the limits of our knowledge, and should be treated as best guesses only.

The HRS contains multiple PGS for several phenotypes, typically with the R^2 increasing over time. Appendix Figure 10 runs a sanity check in which I predict the coefficient of later scores on fertility, using earlier scores and the above formula. Mostly, earlier PGS do a good job at predicting the effect of later PGS on fertility, with the later PGS's point estimate within the earlier PGS's 95% confidence interval.

Discussion

The results here provide qualified support for the economic theory of fertility as an explanation for contemporary natural selection in humans. PGS which predict less education are being selected for, and PGS which predict more education are being selected against. PGS coefficients on RLRS also appear larger for low-income groups and unmarried respondents. But there is little evidence for larger coefficients among people with lower education, or

[2]Heritability estimates are from Marenberg et al. (1994) and Nikpay, Stewart, and McPherson (2017) for coronary heart disease); Willemsen et al. (2015) and Mahajan et al. (2018) for Type II diabetes; Evans et al. (2021) for age of smoking initiation; and Abdellaoui and Verweij (2021) for everything else. I used low-end heritability estimates for Type II diabetes (22%) and heart disease (40%).

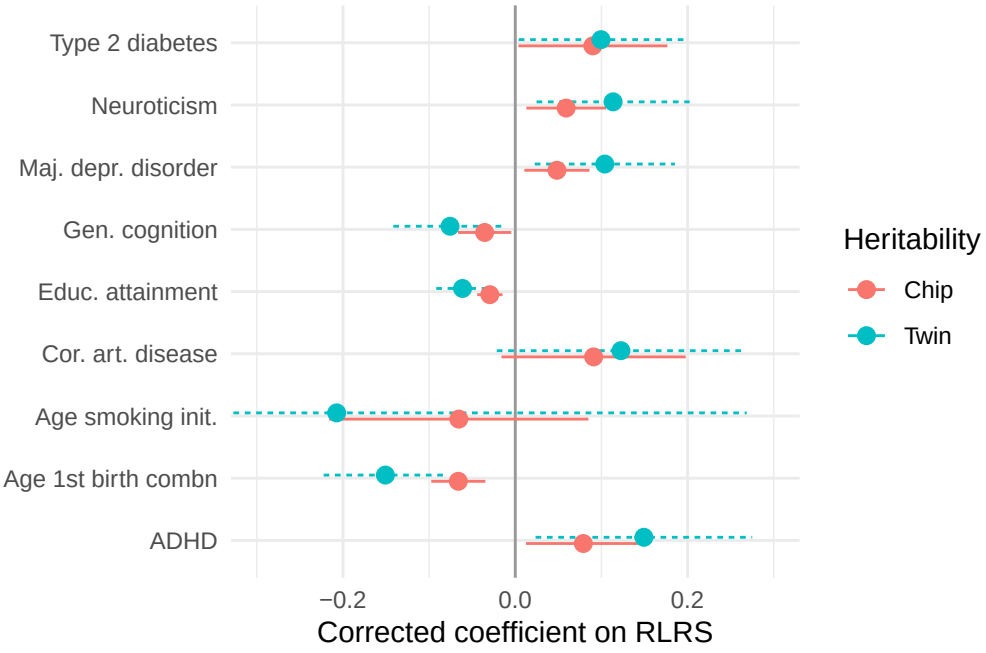


Figure 2: Estimated coefficients of true polygenic scores on RLRS calculated using estimates of chip- and twin-heritability and the most recent polygenic score.

younger parents. This may be due to the low sample size. But in the UK, the between-group differences were large (Hugh-Jones and Abdellaoui 2022); differences that big would surely have been visible here. The smaller black sample makes most tests inconclusive for this population: I can only say that the data do not reject a negative association between PGS correlations with RLRS and with education. Lastly, true effect sizes are estimated and are typically between 0-0.2 standard deviations.

Appendix

Acknowledgements

The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.

Code to reproduce this paper is available at https://github.com/hughjonesd/hrs-selection.

Figures

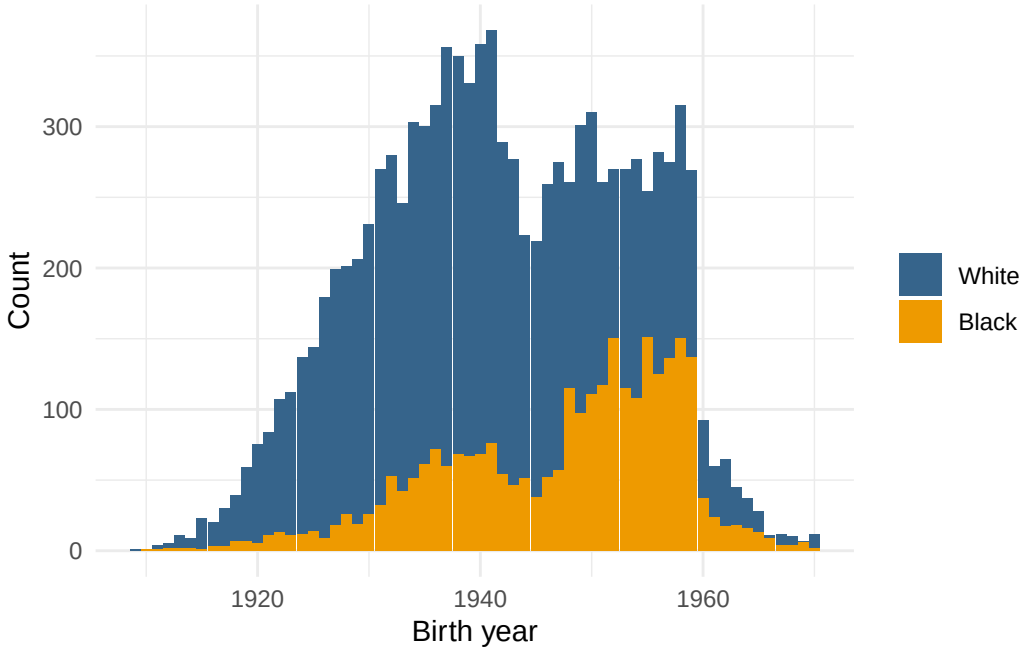


Figure 3: Distribution of birth years for the sample

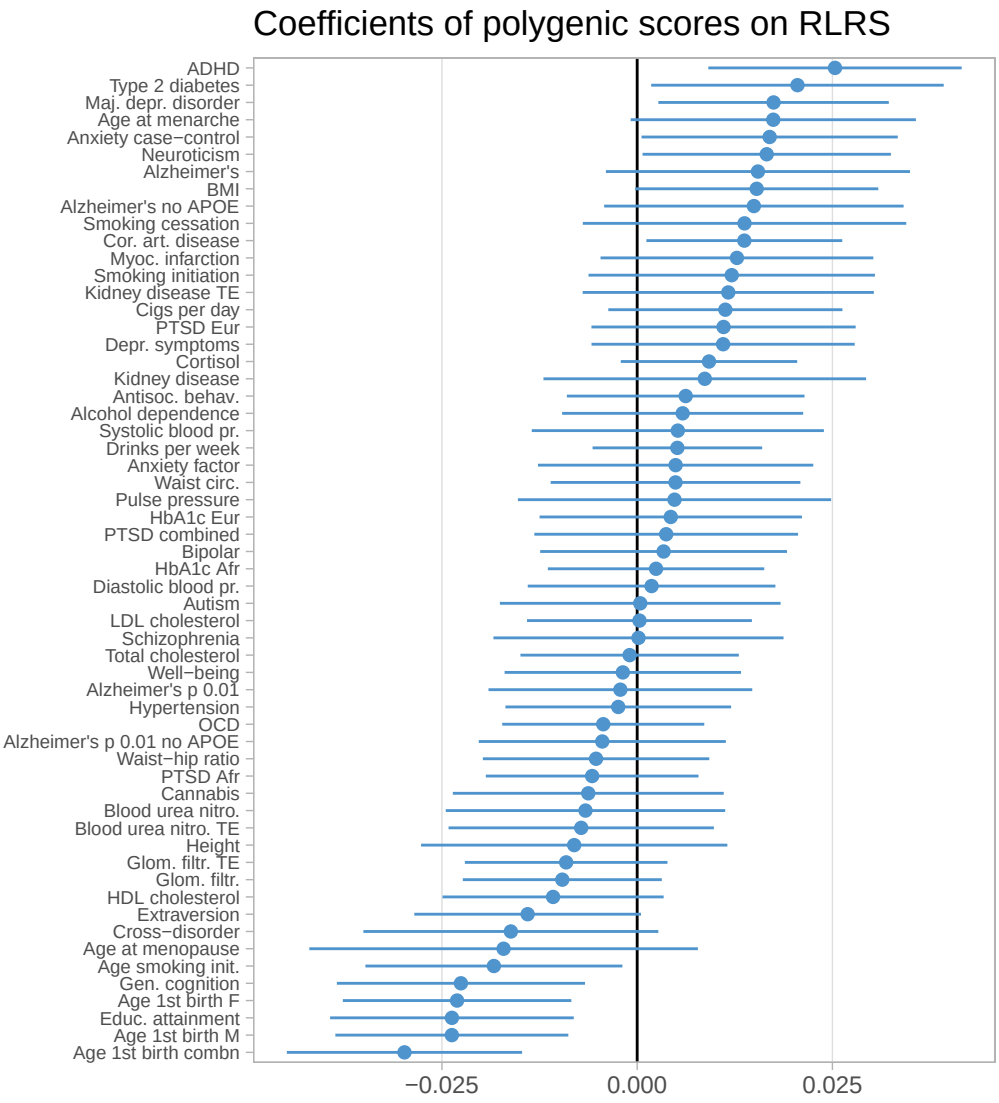
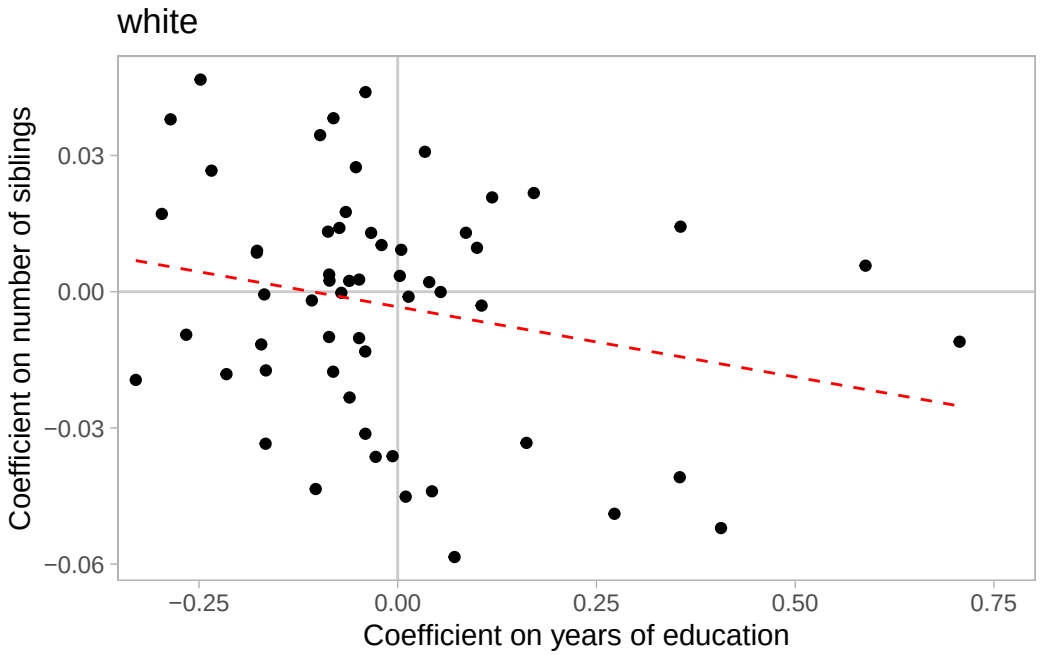
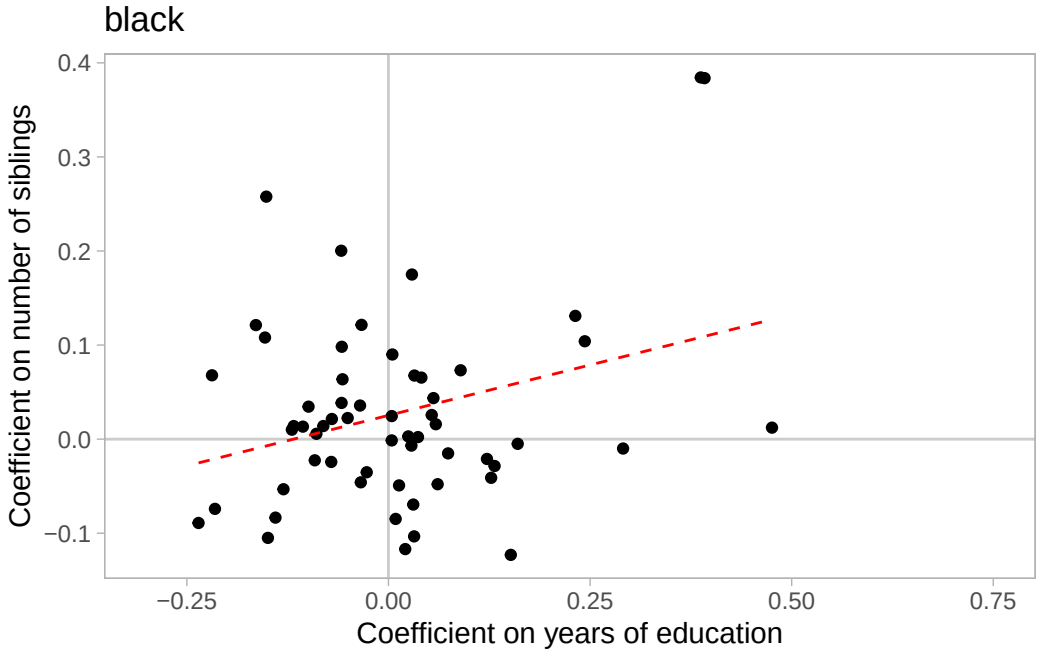


Figure 4: Coefficients of PGS on RLRS among white respondents, controlling for 10 principal components of genomic array data. Lines are 95% confidence intervals.



(a)



(b)

Figure 5: Scatterplot of PGS coefficients on number of live siblings and years of education. Controls include 10 principal components of genetic array data. Dashed lines show linear regressions.

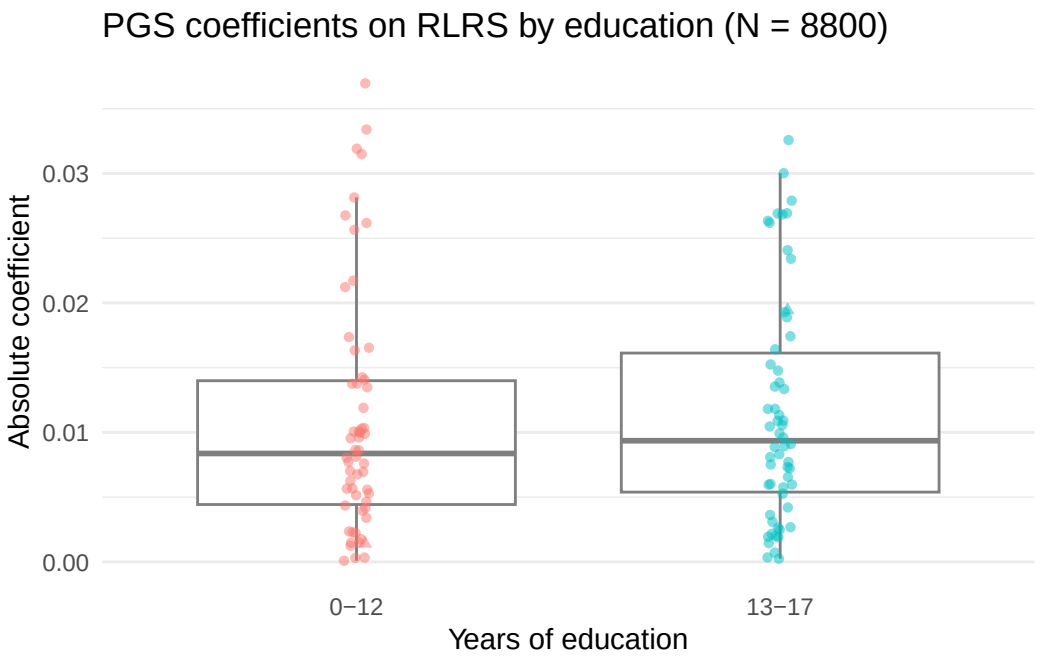


Figure 6: Boxplots of polygenic score coefficients on RLRS among white respondents, controlling for 10 principal components of genetic array data, estimated within the low/high education group. Boxes show quartiles.

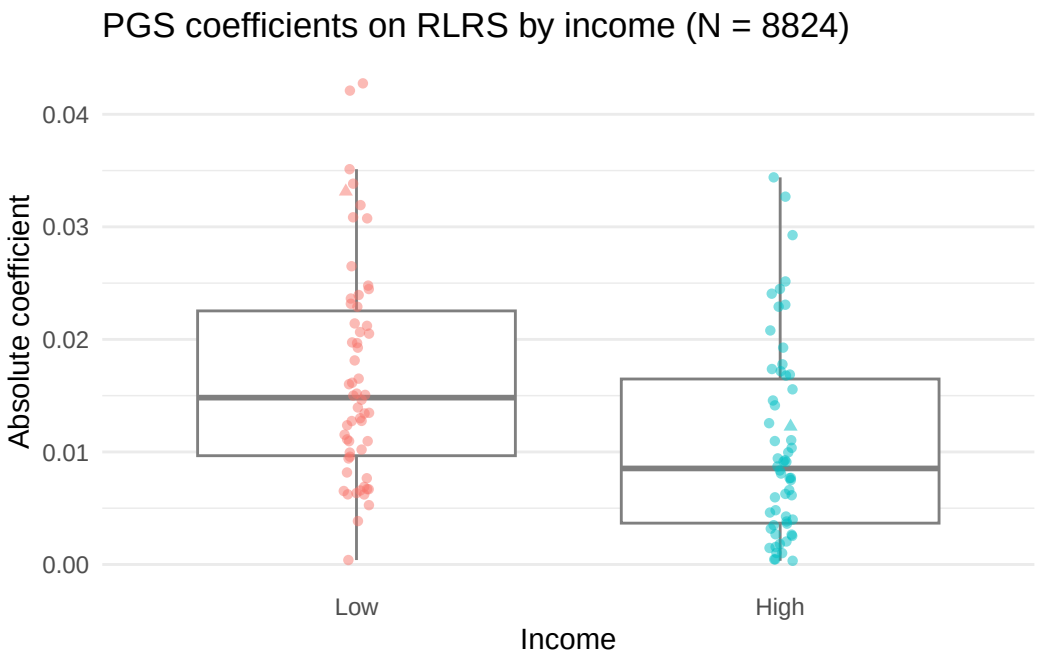


Figure 7: Boxplots of polygenic score coefficients on RLRS among white respondents, controlling for 10 principal components of genetic array data, estimated within the below/above median income group. Boxes show quartiles.

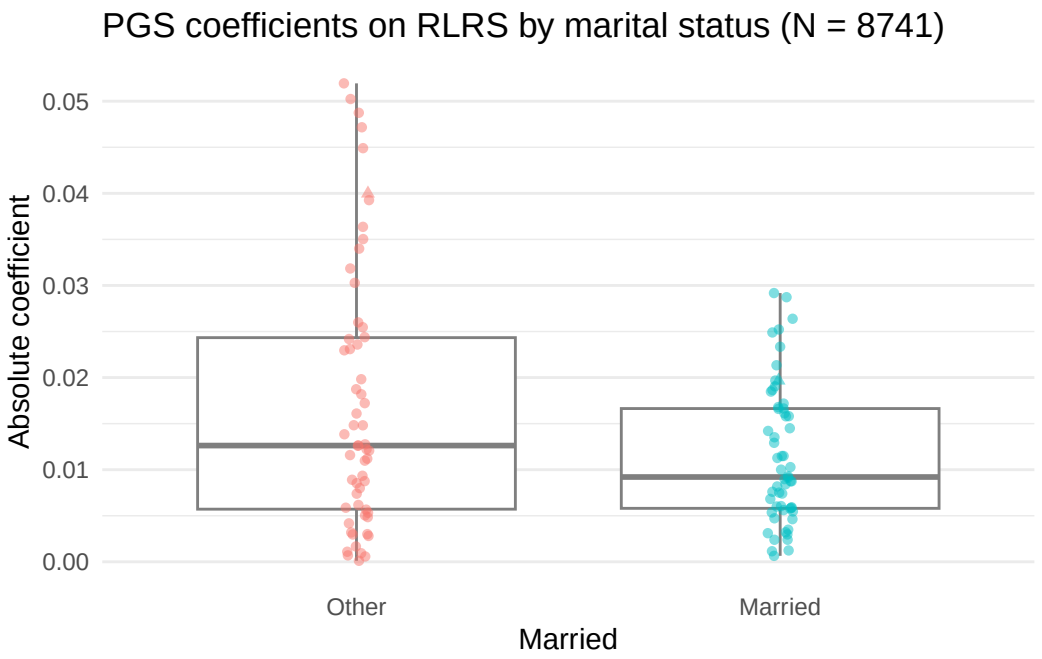


Figure 8: Boxplots of polygenic score coefficients on RLRS among white respondents, controlling for 10 principal components of genetic array data, estimated within married and other respondents. Boxes show quartiles.

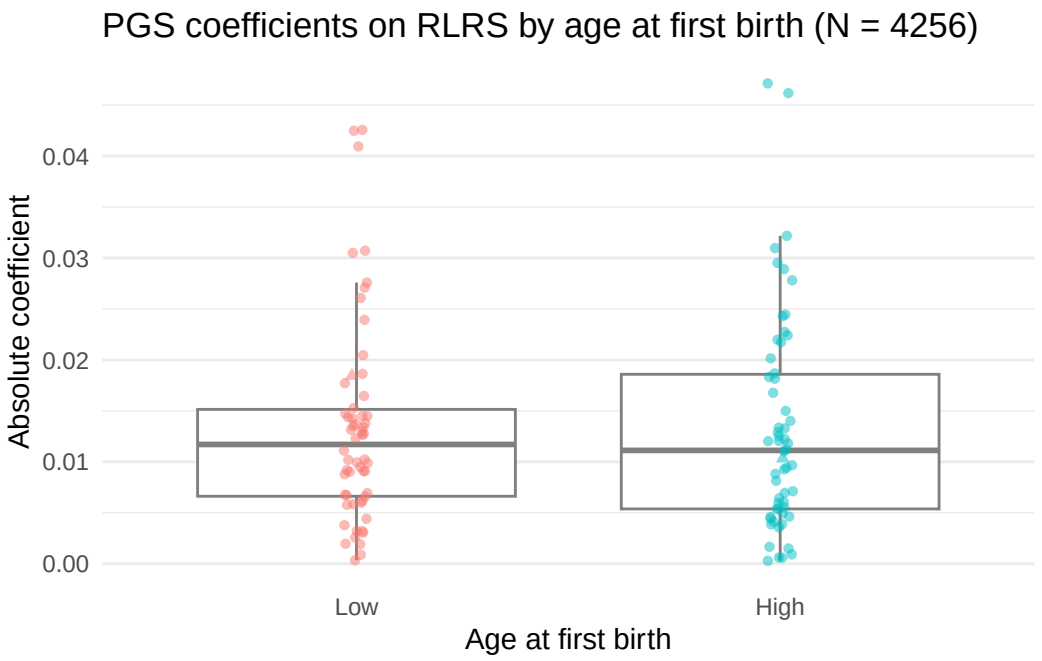


Figure 9: Boxplots of polygenic score coefficients on RLRS among white respondents, controlling for 10 principal components of genetic array data, estimated within below/above median age at first birth respondents. Boxes show quartiles.

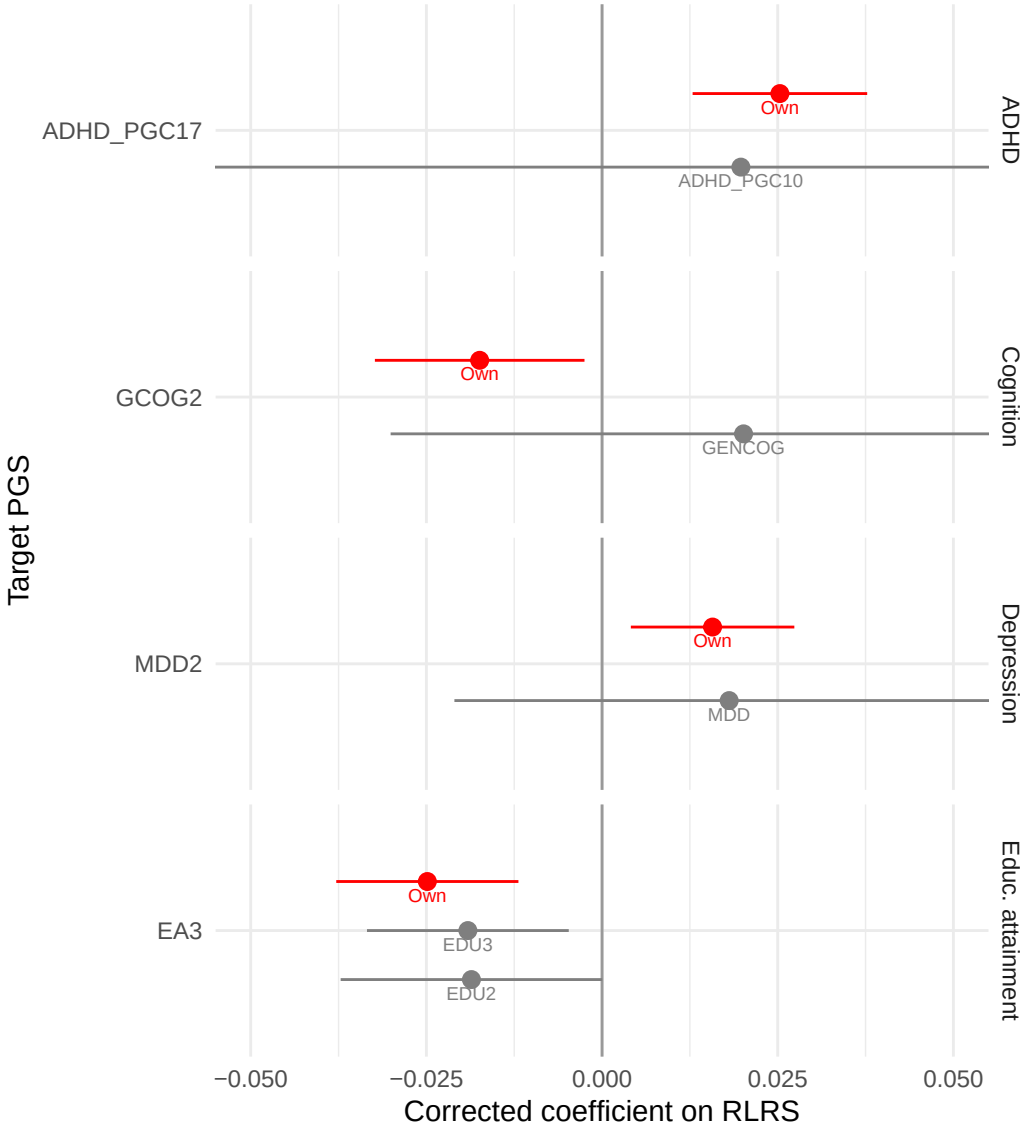


Figure 10: Estimated coefficients of series of PGS on fertility using errors-in-variables correction. "Own" is the uncorrected estimate using the PGS itself. Other estimates are from earlier PGS targeting the same phenotype, corrected via the method in J. Becker et al. (2021).

Bibliography

Abdellaoui, Abdel, and Karin JH Verweij. 2021. "Dissecting Polygenic Signals from Genome-Wide Association Studies on Human Behaviour." *Nature Human Behaviour* 5 (6): 686–94.

Becker, Gary S, and Nigel Tomes. 1976. "Child Endowments and the Quantity and Quality of Children." *Journal of Political Economy* 84 (4, Part 2): S143–62.

Becker, Joel, Casper AP Burik, Grant Goldman, Nancy Wang, Hariharan Jayashankar, Michael Bennett, Daniel W Belsky, et al. 2021. "Resource Profile and User Guide of the Polygenic Index Repository." *Nature Human Behaviour* 5 (12): 1744–58.

Canty, Angelo J, and Anthony C Davison. 1999. "Resampling-Based Variance Estimation for Labour Force Surveys." *Journal of the Royal Statistical Society Series D: The Statistician* 48 (3): 379–91.

Evans, Luke M, Seonkyeong Jang, Dana B Hancock, Marissa A Ehringer, Jacqueline M Otto, Scott I Vrieze, and Matthew C Keller. 2021. "Genetic Architecture of Four Smoking Behaviors Using Partitioned SNP Heritability." *Addiction* 116 (9): 2498–508.

Goldscheider, Calvin, and Peter R Uhlenberg. 1969. "Minority Group Status and Fertility." *American Journal of Sociology* 74 (4): 361–72.

HRS. 2023a. "Health and Retirement Study - Rand HRS Family Data 2018." Ann Arbor, Michigan: University of Michigan; Survey Research Center, Institute for Social Research.

———. 2023b. "Health and Retirement Study - Rand HRS Longitudinal File." Ann Arbor, Michigan: University of Michigan.

Hugh-Jones, David, and Abdel Abdellaoui. 2022. "Human Capital Mediates Natural Selection in Contemporary Humans." *Behavior Genetics* 52 (4-5): 205–34.

Johnson, Nan E. 1979. "Minority-Group Status and the Fertility of Black Americans, 1970: A New Look." *American Journal of Sociology* 84 (6): 1386–1400.

Lumley, Thomas. 2023. "Survey: Analysis of Complex Survey Samples."

Mahajan, Anubha, Daniel Taliun, Matthias Thurner, Neil R Robertson, Jason M Torres, N William Rayner, Anthony J Payne, et al. 2018. "Fine-Mapping Type 2 Diabetes Loci to Single-Variant Resolution Using High-Density Imputation and Islet-Specific Epigenome Maps." *Nature Genetics* 50 (11): 1505–13.

Marenberg, Marjorie E, Neil Risch, Lisa F Berkman, Birgitta Floderus, and Ulf de Faire. 1994. "Genetic Susceptibility to Death from Coronary Heart Disease in a Study of Twins." *New England Journal of Medicine* 330 (15): 1041–46.

Nikpay, Majid, Alexandre FR Stewart, and Ruth McPherson. 2017. "Partitioning the Heritability of Coronary Artery Disease Highlights the Importance of Immune-Mediated Processes and Epigenetic Sites Associated with Transcriptional Activity." *Cardiovascular Research* 113 (8): 973–83.

Ware, E. B., A. M. Gard, L. L. Schmitz, and J. D. Faul. 2020. "HRS Polygenic Scores – Release 4." Ann Arbor, Michigan: University of Michigan; Survey Research Center,

Institute for Social Research.

Willemsen, Gonneke, Kirsten J Ward, Christopher G Bell, Kaare Christensen, Jocelyn Bowden, Christine Dalgård, Jennifer R Harris, et al. 2015. "The Concordance and Heritability of Type 2 Diabetes in 34,166 Twin Pairs from International Twin Registers: The Discordant Twin (DISCOTWIN) Consortium." *Twin Research and Human Genetics* 18 (6): 762–71.

Yang, Yang, and S Philip Morgan. 2003. "How Big Are Educational and Racial Fertility Differentials in the US?" *Social Biology* 50 (3-4): 167–87.