Natural selection in the Health and Retirement Study

David Hugh-Jones Tobias Edwards

2023-11-01

We 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 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. We also estimate effect sizes corrected for noise in the polygenic scores. natural selection: among both white and black respondents,

Hugh-Jones and Abdellaoui (2022) explain patterns of natural selection on polygenic scores in the UK, using an economic theory of fertility derived from @becker1976child. 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, we 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 we 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. we 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 8827 genotyped white participants and 2319 genotyped black participants.

Genotyping took place in 2006, 2008 and subsequent years. PGS were taken from those pre-calculated by the HRS (Ware et al. 2020) and those produced by the Social Science Genetic Association Consortium, as part of their Polygenic Index Repository (Becker et al. 2021). Scores created by the HRS were provided for black and white participants, but Polygenic Index Repository scores were only created for white participants.

For the white participants, when scores from the two samples measured the same trait, we only used the PGS from the Polygenic Index Repository. For some traits, polygenic scores were created from European ancestry GWAS and GWAS with samples of different ancestry. We choose to use polygenic scores trained only on individuals of European ancestry. We discard obsolete PGS for which there is a newer, more accurate score targeting the same phenotype. we also discard PGS for number of children ever born (but keep scores for age at first birth). This leaves a total of 68 scores for the white participants and 47 for the black participants. 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, we 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 mean number of children of people born in the same year was calculated using sampling weights.

The HRS contains weights which match survey respondents to the US population. We 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, we 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 [@lumleysurvey2023].

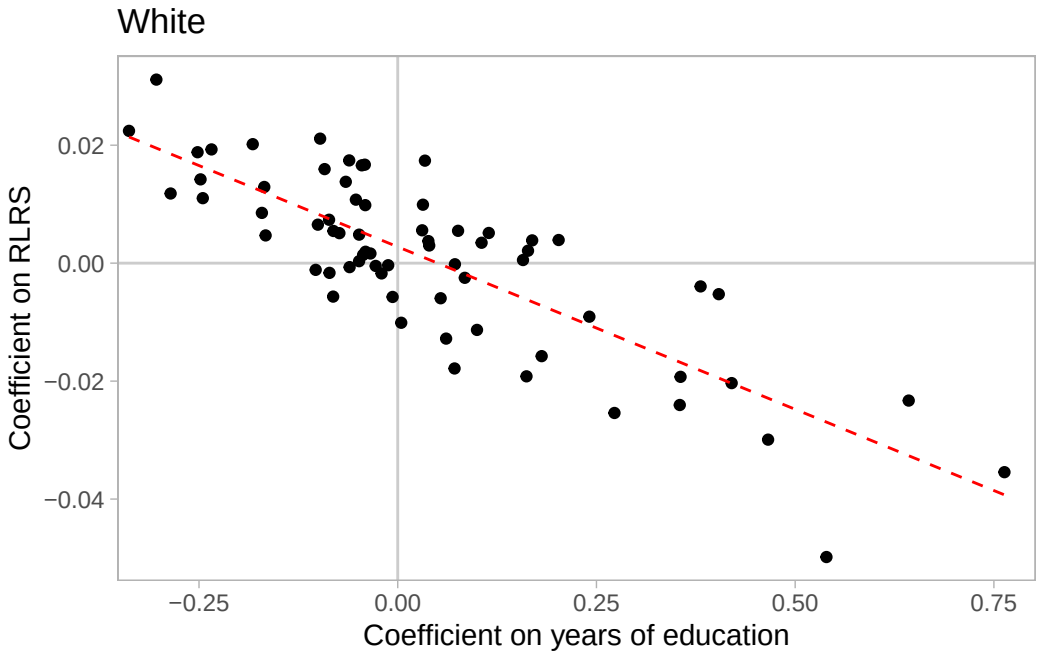
Results

We 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. 3 scores are significant at Bonferroni-corrected $p < 0.05/115$. The scores are age at first birth, educational attainment and attention deficit disorder. But we are most concerned with looking at patterns across scores rather than judging the significance of individual scores.

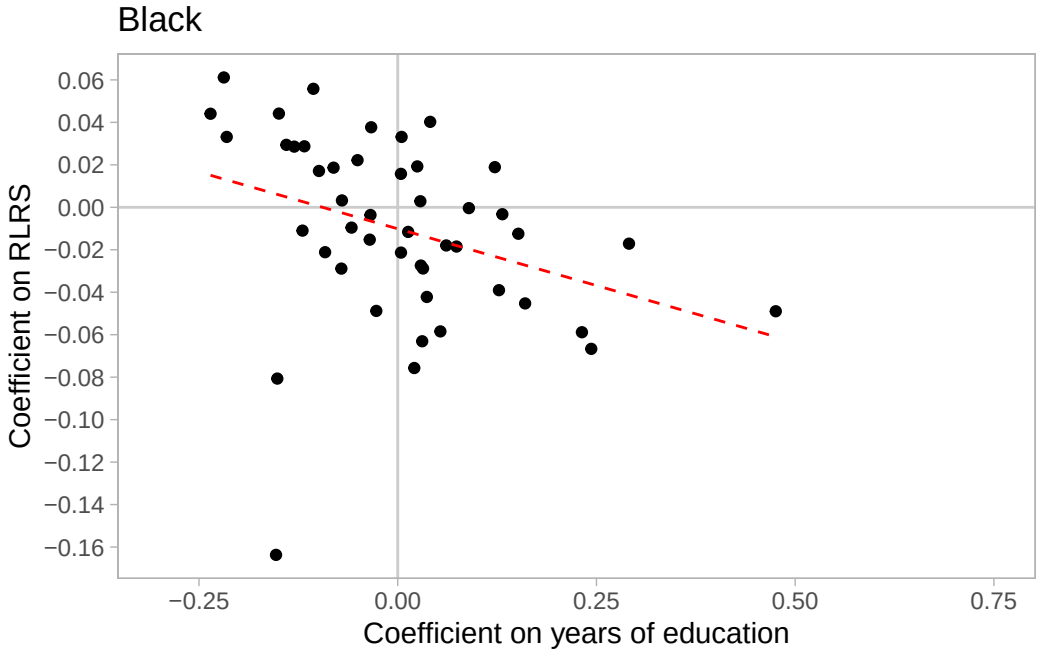
Figure 1 plots each PGS's partial correlation with RLRS against its partial correlation with educational attainment. The relationship is negative among whites (correlation -0.819, bootstrap 95% C.I. -0.99 to -0.648) and negative but insignificant among blacks (correlation -0.345, bootstrap 95% C.I. -0.721 to 0.03). Survey bootstraps (Canty and Davison 1999) are used so as to make inferences from the sample of respondents.

We can also examine natural selection in the preceding and succeeding generations, by using reported number of siblings and grandchildren respectively. We regress PGS on respondents' number of living siblings in 2010. Data for dead siblings has too many missing values to use. We 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 only significant and negative in the white respondents, although ethnicities have large standard errors (whites: correlation -0.387, bootstrapped 95% C.I. -0.691 to -0.084; blacks: correlation 0.19, bootstrapped 95% C.I. -0.274 to 0.654). There is a positive and significant correlation across generations, i.e. between PGS coefficients on number of siblings and number of children (among whites: 0.557,cbootstrapped 95% C.I. 0.306 to 0.808).

To examine selection in the succeeding generation, we divide the respondents' number of grandchildren by their number of children. In other words, we calculate the average number of children the respondent's children have. This measures the reproductive success in the second generation which we can call RS2. 245 white respondents and 38 black respondents report having grandchildren despite reporting having had no children. These individuals were removed for this test. In regression we reweight respondents by the number of children they have, since more fecund grandparents account for a larger proportion of the next generation. Older grandparents will have more time for their number of grandkids to accumulate. To deal with this time trend, we divide RS2 by the mean RS2 of respondents born in the same year, creating relative RS2 (RRS2).



(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.

We regress the respondent's (the grandparent's) PGS on RRS2. The resulting coefficient is only a proxy for selection in the next generation, not a perfect measure. The grandparent's PGS indicates, but does not determine the PGS of the parent. To know the expected parent's PGS, we would need both grandparent's PGS. Given that some respondents do not have a partner in the HRS and some have had children with multiple partners, such an approach is untenable in our sample. We expect the effect of the grandparent's PGS to also be dependent upon the level of assortative mating for the trait. For traits with high assortative mating, the grandparent's PGS will more strongly predict the parent's PGS, leading to a greater regression slope.

Appendix Figure 6 plots coefficients on RRS2 versus coefficients on years of education. Correlations are significant for whites but not blacks. Standard errors were moderate in whites, but large in blacks (whites: correlation -0.86, bootstrapped 95% C.I. -0.965 to -0.755; blacks: correlation 0.01, bootstrapped 95% C.I. -0.55 to 0.57). There is a strong positive correlation between the parent and grandchild regressions, i.e. between PGS coefficients on RRS2 and RLRS (among whites: 0.751, bootstrapped 95% C.I. 0.539 to 0.964).

We 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 we only use the white subsample: there are too few black respondents to be informative.

Appendix Figure 7 shows essentially no difference in the distribution of absolute coefficient sizes between respondents with more and less education. Appendix Figure 8, on the other hand, shows that coefficients appear larger among lower-income respondents. Appendix Figure 9 shows absolute coefficients for respondents who were married in 2010 against all other statuses. Coefficients appear larger for unmarried respondents.

Appendix Figure 10 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 (-0.199), and for 43 out of 68 PGS, coefficients have opposite signs. However, in bootstraps, these statistics are not significantly different from 0 and 34 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.

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. we 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: -0.076, 95% C.I. -0.093 to -0.058; whites: -0.029, 95% C.I. -0.035 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 (Becker et al. 2021). Moving from error correction in a univariate model to a multivariate model requires a slightly more complex correction, although it uses the same parameters. To do this we use the error-corrected estimator developed by @becker2021resource.

SNP- or chip-heritabilities and R^2_{PGS} are calculated by Becker et al. (2021) for scores in the Polygenic Index Repository. The authors use GCTA to estimate heritability. When the corresponding phenotype is not available in the HRS to estimate R^2_{PGS}, we use parameters estimated by the authors in the Wisconsin Longitudinal Study instead. We also perform error correction with twin heritabilites. To attain precise estimates we use heritabilities from a meta-analysis including over 14,000 twin pairs authored by Polderman et al. (2015)[2]. We remove polygenic scores with $R^2_{PGS} > 0.005$ to focus on PGS with adequate power. This removed the ADHD PGS which significantly predicted RLRS, even after bonferroni correction.

Figure 2 reports the error corrected estimates of selection. 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. Age of First Birth has a twin corrected heritability of 0.15, implying that it maybe under fast selection. The confidence bounds capture uncertainty from sampling variation, but not other sources, including uncertainty about the true h^2 , the true R^2_{PGS}, 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.

Results from a few traits show substantial uncertainty regarding the true effect size. For asthma, when chip heritability is used, the effect is positive and significant, but negative and significant when twin heritability is used! Such a large disparity is because Becker et al. (2021) estimate the chip heritability at 0.015 and we use a twin heritability 0f 0.55.

[2]For age of first birth, we could not find an appropriately close trait in Polderman et al. (2015). For this trait we used a twin heritability of 0.15 as estimated in the Midlife in the United States (MIDUS) sample (Briley, Tropf, and Mills 2017).

For personality traits, chip heritability is often low (e.g. < 0.05) but twin heritabilities are often much higher (≈ 0.4) implying substantial missing heritability. For these traits, the confidence intervals on twin heritabilities are enormous. Error-corrected effect sizes and the R^2 and h^2 parameters used can be found on the Github page for this paper.

To estimate how natural selection will contribute to changing the phenotype, we can multiply the change in the mean genetic value by the correlation between genetic values and phenotypes h, which is the square root of heritability. For cognitive performance we have estimated a genetic change of -0.066 standard deviations per generation, assuming a twin heritability of 0.51. This implies a phenotypic change of $-0.066 \times 0.51^{\frac{1}{2}} = -0.047$ standard deviations per generation, equivalent to -0.71 points in the units of IQ, where a standard deviation is equal to 15 points. This calculation assumes the heritability of the trait remains constant, that the genetic correlation across time is equal to one and it ignores the environmental contributions to changes in the phenotype. Given the assumptions required for this calculation, on top of the problems involved in estimating the genetic change, it can only be considered a guess.

```
[1] TRUE
```

Discussion

The economic theory of fertility is driven by the trade-off between children and income. On the one hand, a higher hourly wage makes time raising children more costly (a "substitution effect"). On the other hand, higher expected income makes children more affordable (an "income effect"). Prediction 1 of the theory holds when substitution effects dominate income effects. Prediction 2 is driven by the specific form of individuals' preferences for income: when utility for income is sharply curved, the substitution effect is stronger for those who expect to earn less.

The results here support prediction 1 but are more ambiguous for prediction 2. 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 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 theory can accommodate this non-result, if preferences for income are less curved in the US for whatever reason. But note that *any* theory with a negative relationship between education and fertility will give prediction 1.[3] In this sense, results here are less supportive of the economic theory specifically.

[3]See Balbo, Billari, and Mills (2013) for a broad review of fertility theories.

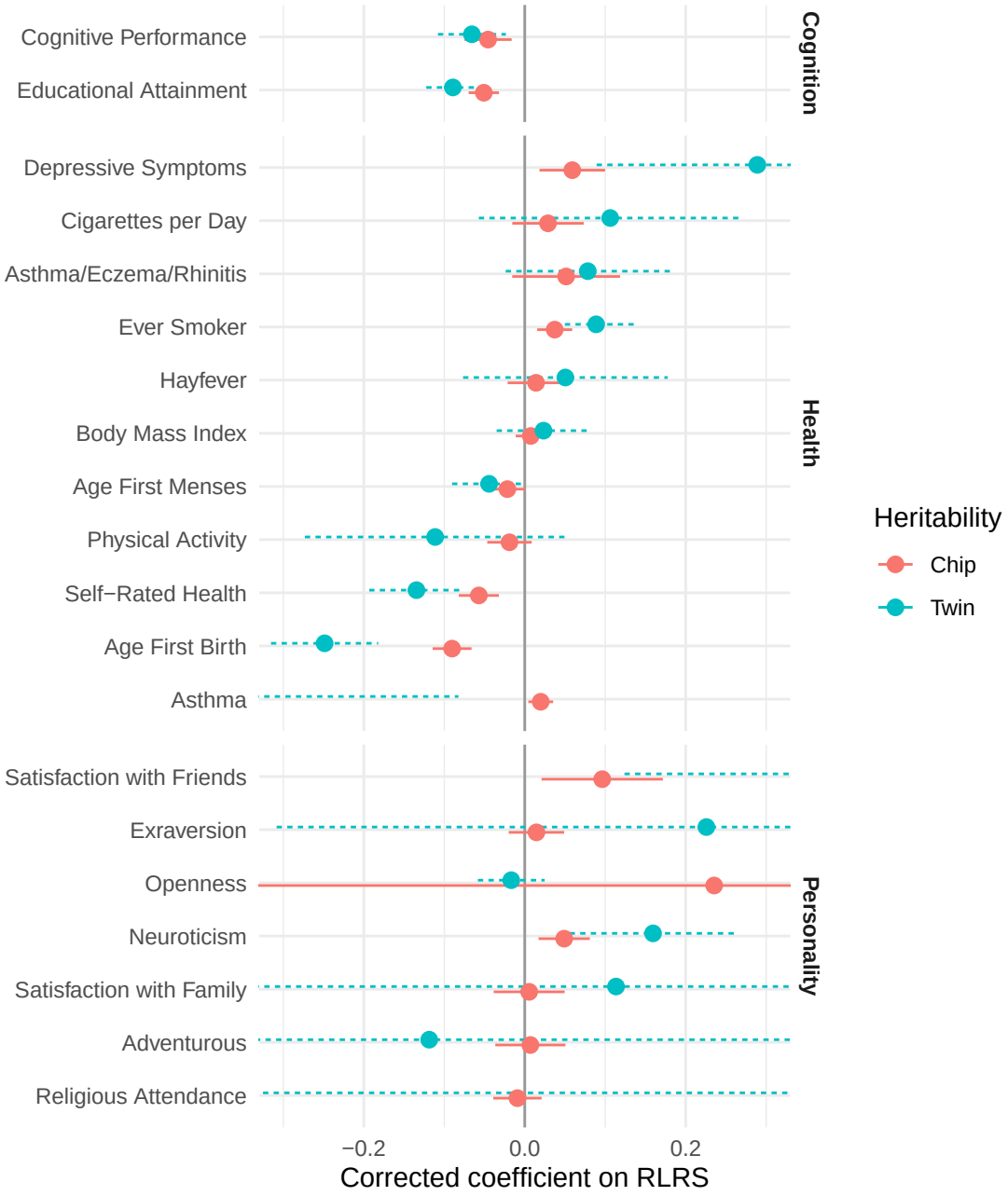


Figure 2: Estimated coefficients of true polygenic scores on RLRS calculated using estimates of chip- and twin-heritability and the most recent polygenic score.

The smaller black sample makes most tests inconclusive for this population: we can only say that the data do not reject a negative association between PGS correlations with RLRS and PGS correlations with education. Lastly, estimated effect sizes of "true" polygenic scores are typically between 0-0.2 standard deviations. To know more, we must await more accurate scores.

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

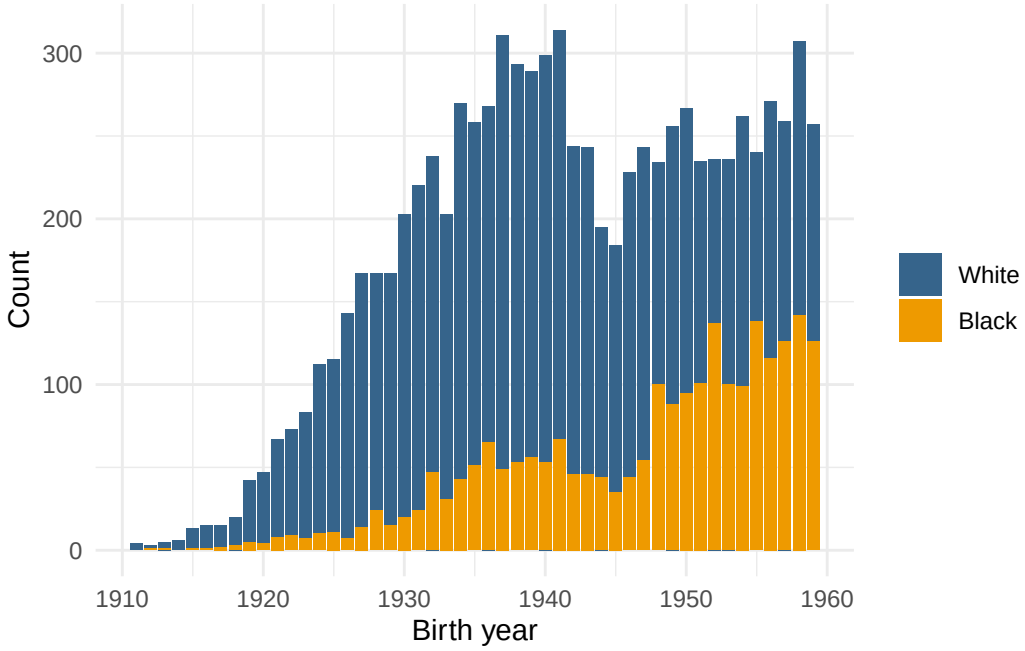


Figure 3: Distribution of birth years for the sample

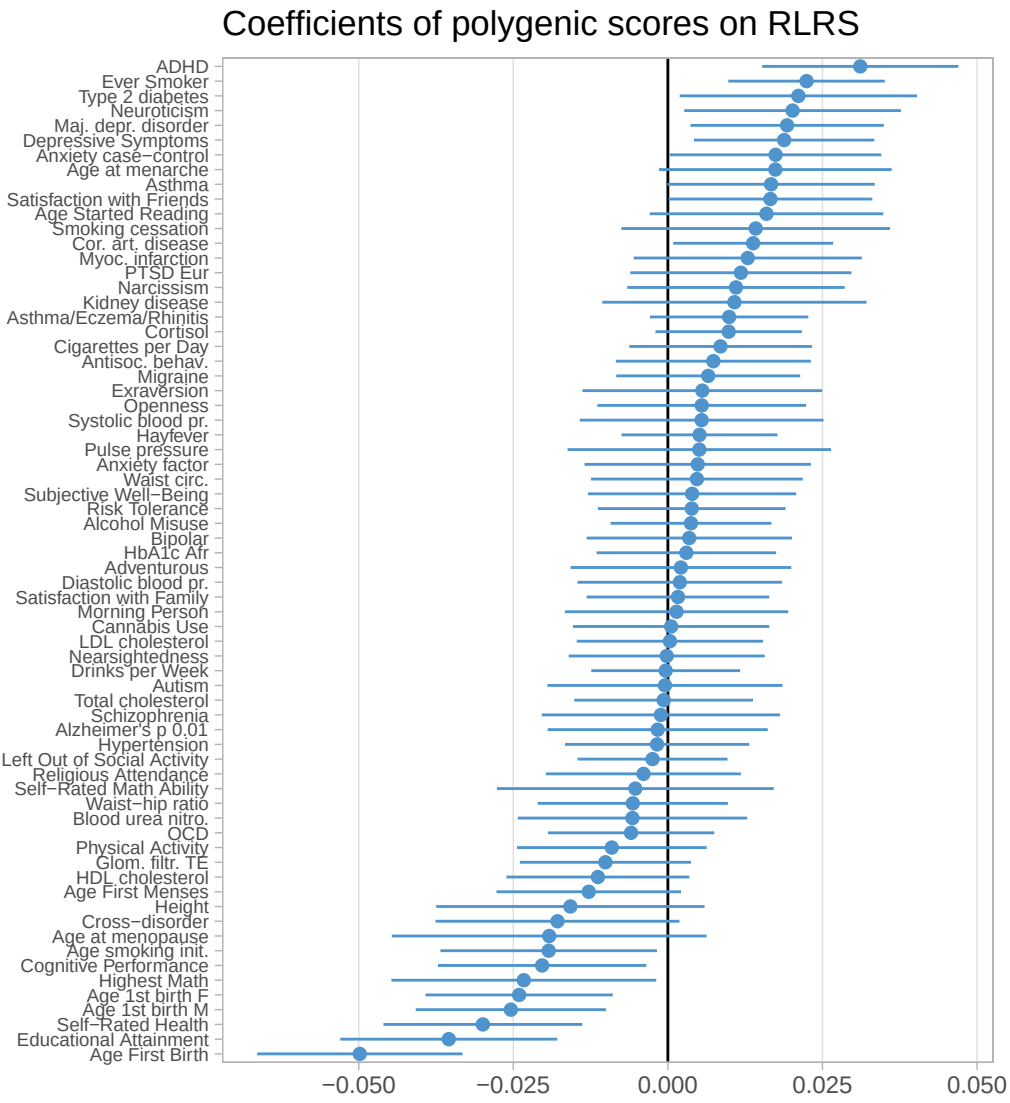
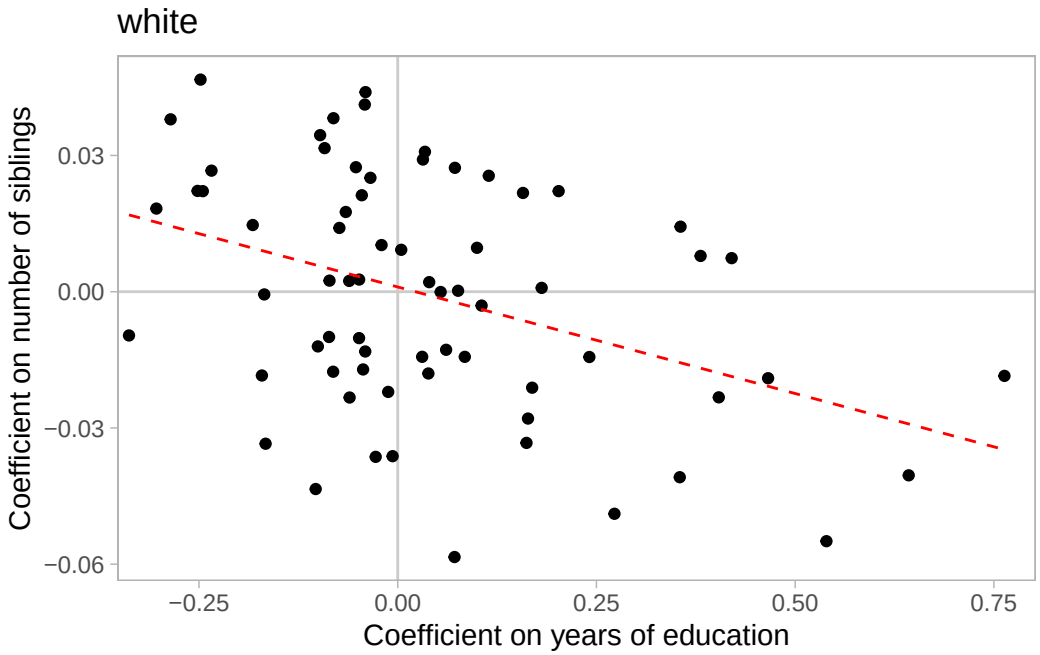
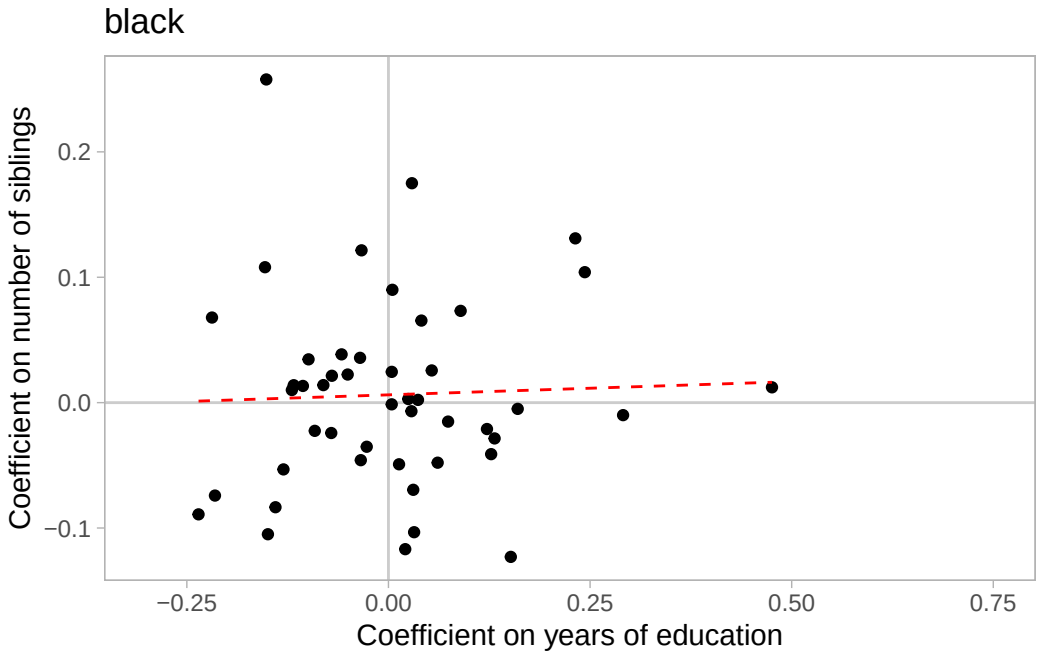


Figure 4: Coefficients of PGS on RLRS among white respondents, controlling for 10 principal components of genomic array data. Lines are 95% confidence intervals.

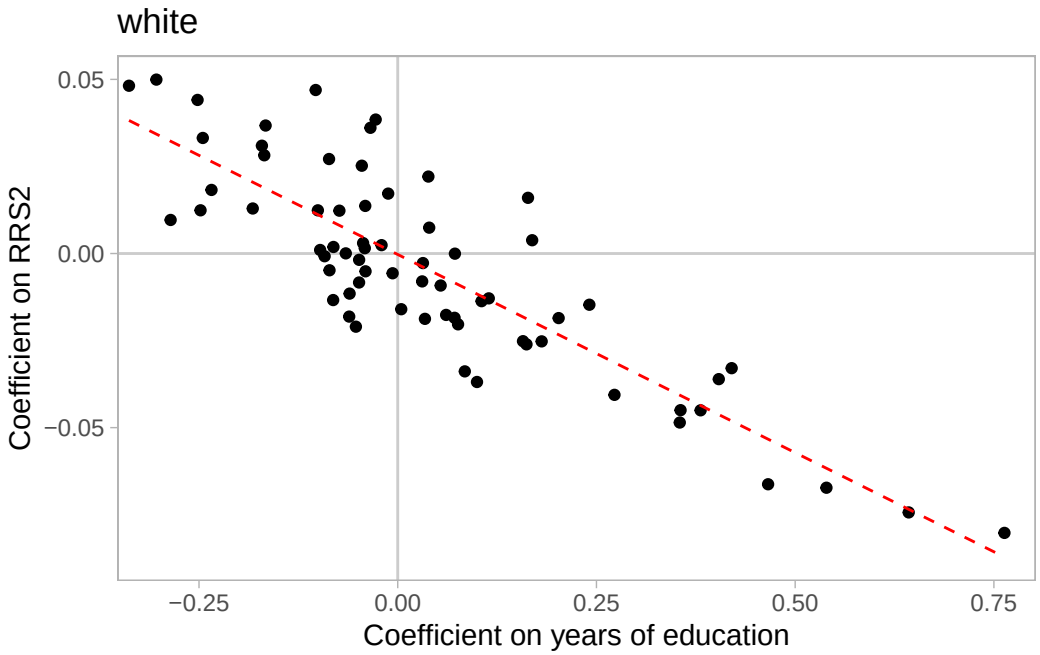


(a)

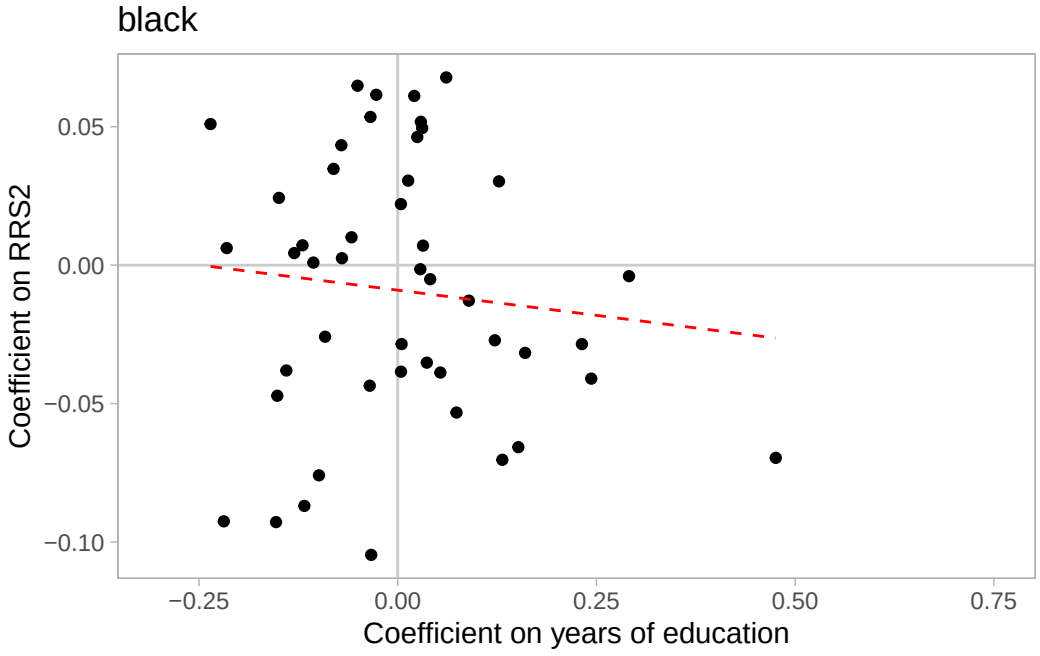


(b)

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.



(a)



(b)

Figure 6: Scatterplot of PGS coefficients on RLRS2 and years of education. RRS2 is the reproductive success of respondent's offspring relative to the success of other offspring. Controls include 10 principal components of genetic array data. Dashed lines show linear regressions.

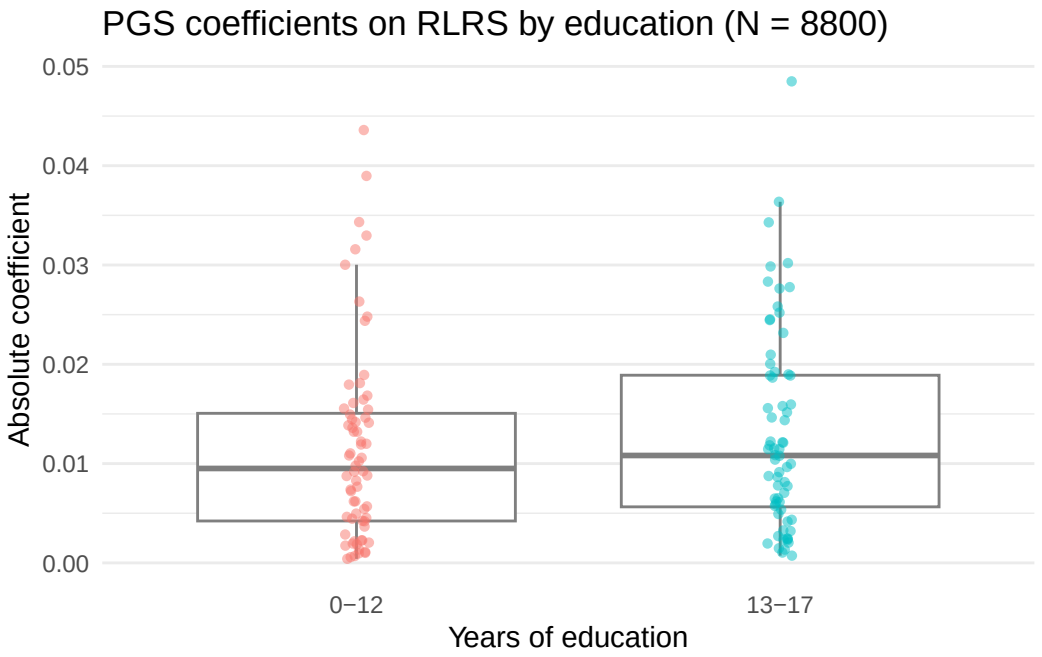


Figure 7: Boxplots of polygenic score coefficients on RLRS among white respondents, controlling for 10 principal components of genetic array data, estimated within the low/high education group. Boxes show quartiles.

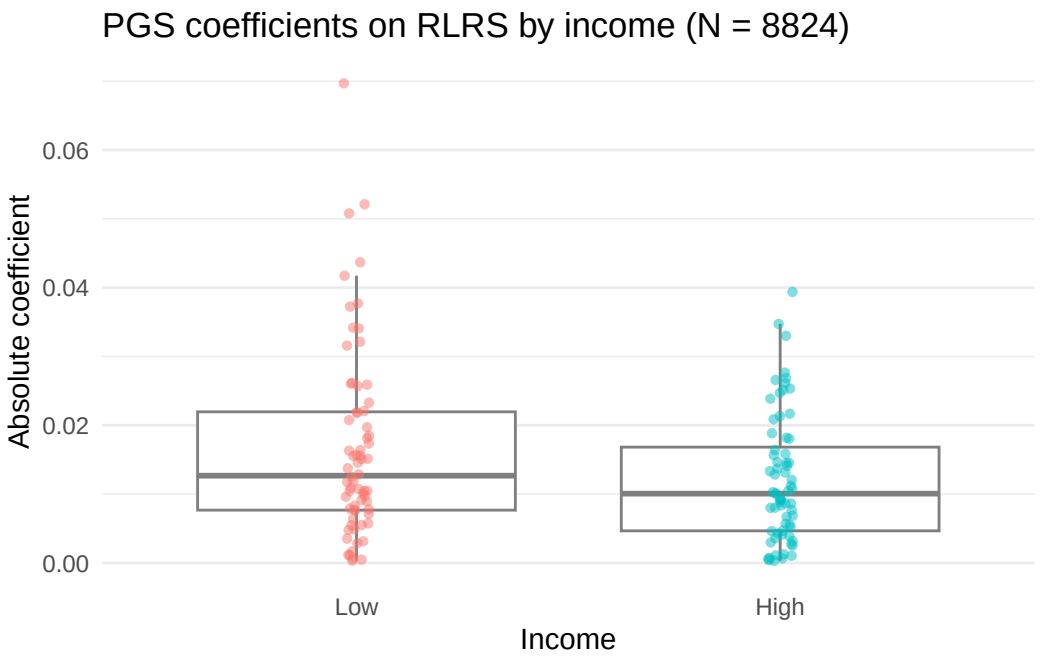


Figure 8: Boxplots of polygenic score coefficients on RLRS among white respondents, controlling for 10 principal components of genetic array data, estimated within the below/above median income group. Boxes show quartiles.

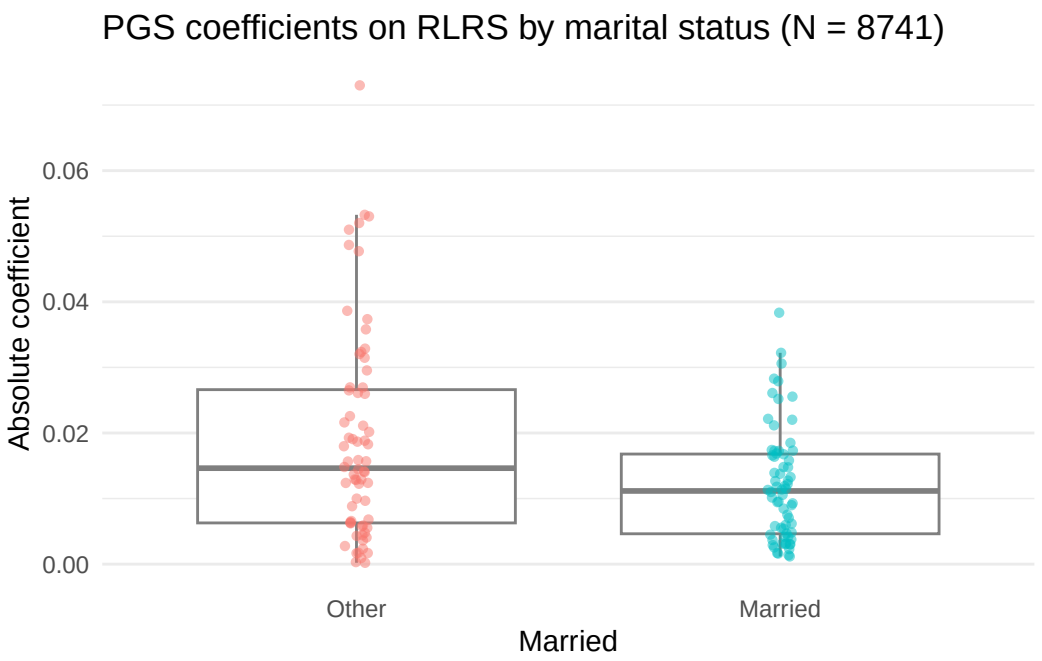


Figure 9: Boxplots of polygenic score coefficients on RLRS among white respondents, controlling for 10 principal components of genetic array data, estimated within married and other respondents. Boxes show quartiles.

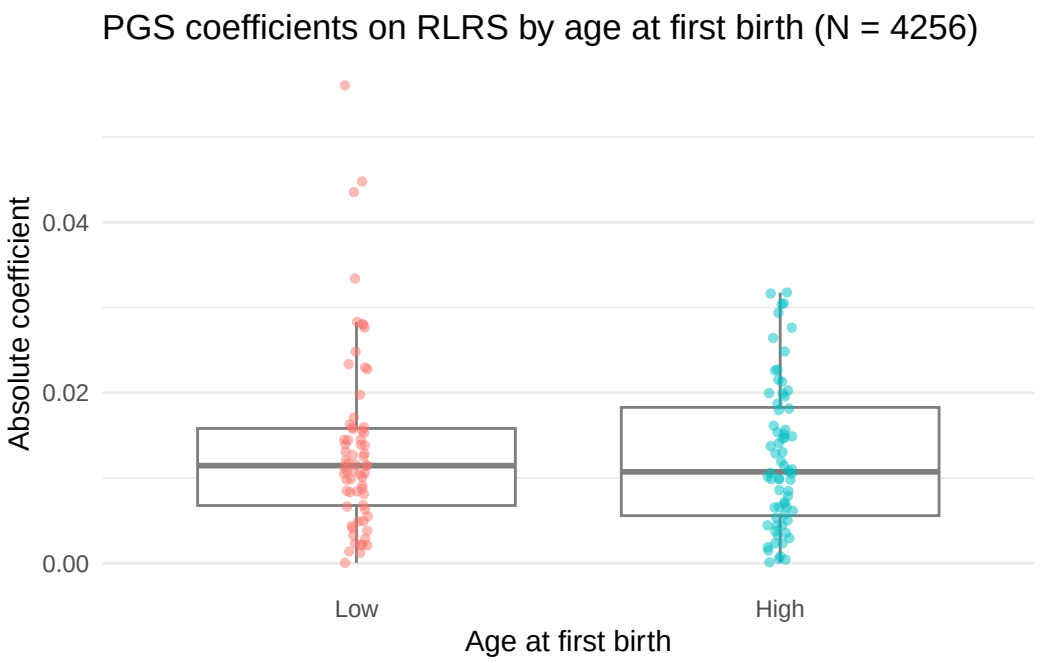


Figure 10: Boxplots of polygenic score coefficients on RLRS among white respondents, controlling for 10 principal components of genetic array data, estimated within below/above median age at first birth respondents. Boxes show quartiles.

Bibliography

Balbo, Nicoletta, Francesco C Billari, and Melinda Mills. 2013. "Fertility in Advanced Societies: A Review of Research: La fécondité Dans Les Sociétés Avancées: Un Examen Des Recherches." *European Journal of Population/Revue Europeenne de Demographie* 29: 1–38.

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.

Briley, Daniel A., Felix C. Tropf, and Melinda C. Mills. 2017. "What Explains the Heritability of Completed Fertility? Evidence from Two Large Twin Studies." *Behavior Genetics* 47: 36–51.

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.

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.

Polderman, Tinca J. C., Beben Benyamin, Christiaan A. De Leeuw, Patrick F. Sullivan, Arjen Van Bochoven, Peter M. Visscher, and Danielle Posthuma. 2015. "Meta-analysis of the heritability of human traits based on fifty years of twin studies." *Nature Genetics* 47 (7): 702–9. https://doi.org/10.1038/ng.3285.

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.

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.