

# Natural selection in the Health and Retirement Study

David Hugh-Jones

2023-09-01

Hugh-Jones and Abdellaoui (2022) explain patterns of natural selection on polygenic scores in the UK, using an economic theory of fertility derived from Becker (XXX). 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.
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.

The evidence for point 1 is that polygenic scores which correlate positively with human capital correlate negatively with number of children, i.e. they are being selected against. And vice versa, scores which correlate positively with human capital are being selected for. The evidence for point 2 is that correlations with number of children are larger and more significant among people with lower income, less education and single parents.

Here, I test the same theory in the US population, using the Health and Retirement Survey (HRS; XXX). This gives an independent test of the theory. The HRS is more representative of the population than UK Biobank, which addresses one potential weakness of the previous paper. Also, the HRS provides precalculated polygenic scores (XXX Aysu) for both black and (non-hispanic) white participants, so I can check whether the human capital theory of natural selection works in both these ethnicities.

Economic theory treats all humans as the same. Once you have controlled appropriately for people's preferences, resources and beliefs, there should be no difference

between ethnicities. So, the economic theory of fertility predicts that patterns of natural selection should be the same in both white and black ethnic groups. Of course, ethnic groups may have different economic characteristics. Also, most of the polygenic scores were created using people of European ancestry. This means they are less predictive for people with African ancestry. If natural selection works via scores' correlation with phenotypes, then we would expect these correlations to be smaller among black people.

## 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.

TODO: I think for all these guys, the raeborn question will have been asked by the time they were indeed over 45/40. Double check.

The resulting sample contains 12046 genotyped white participants and 3067 genotyped black participants. Genotyping took place in 2006, 2008 and subsequent years. I use polygenic scores (PGS) precalculated by XXX. 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).

PGS are rescaled to zero mean and unit variance within each ethnic group, so effect 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 principal components of the DNA array data.

I use realised fertility (number of children ever born) as the key dependent variable. I use this rather than relative lifetime reproductive success for ease of interpretation and because I don't have data for dates when children were born. My focus is on examining patterns of natural selection across PGS, rather than estimating their effects on PGS over time; raw fertility is good enough for this.

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.

## Results

Figure 1 shows effect sizes of PGS correlations with realised fertility among each ethnic group. Standard errors are large because of the relatively low sample sizes. But I am most concerned with looking at patterns across scores rather than judging the significance of individual scores. (Also, because all scores are differently weighted sums of the same set of alleles, if we can reject the null for any one score we should probably reject it for all of them; in other words, effect sizes are more interesting than p values here.)

Figure 2 plots each PGS's correlation with fertility against its correlation with educational attainment. The relationship is negative among whites (correlation -0.785, bootstrapped 95% C.I. -0.805 to -0.412) and negative but insignificant among blacks (correlation -0.393, bootstrapped 95% C.I. -0.694 to 0.072). I use bootstraps to make inferences from the sample of individuals, rather than the “sample” of different PGS.

The relationship is imprecisely estimated in the smaller black sample, but it is also steeper. Put another way, if we measured PGS in units of “years of associated education”, estimated selection effects would be larger among blacks than whites. One reason for this may be that the correlation between phenotypic years of education and fertility is larger for blacks than for whites. In bivariate regressions, an extra year of education reduces fertility by -0.1 for whites and -0.229 for blacks (cf. Goldscheider and Uhlenberg 1969; Johnson 1979; Yang and Morgan 2003). In turn, the simplest explanation for this fact would be point 2 of the theory. Because black people on average have lower incomes and fewer years of education, and are more likely to be single parents, the fertility-wage tradeoff is especially acute for them. TODO what does the existing literature say?

TODO if so then presumably controlling for education and income would reduce these differences.

## Bibliography

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

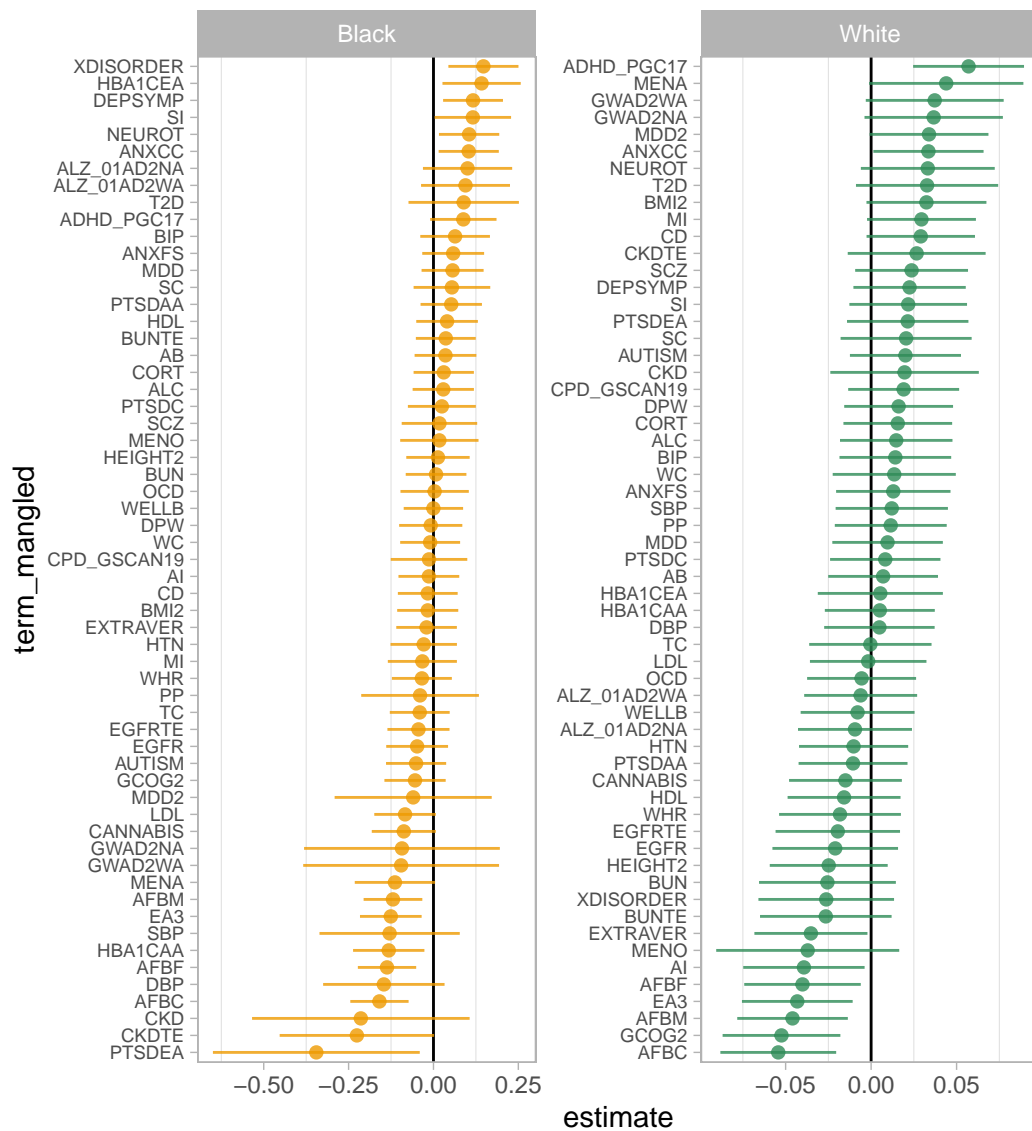
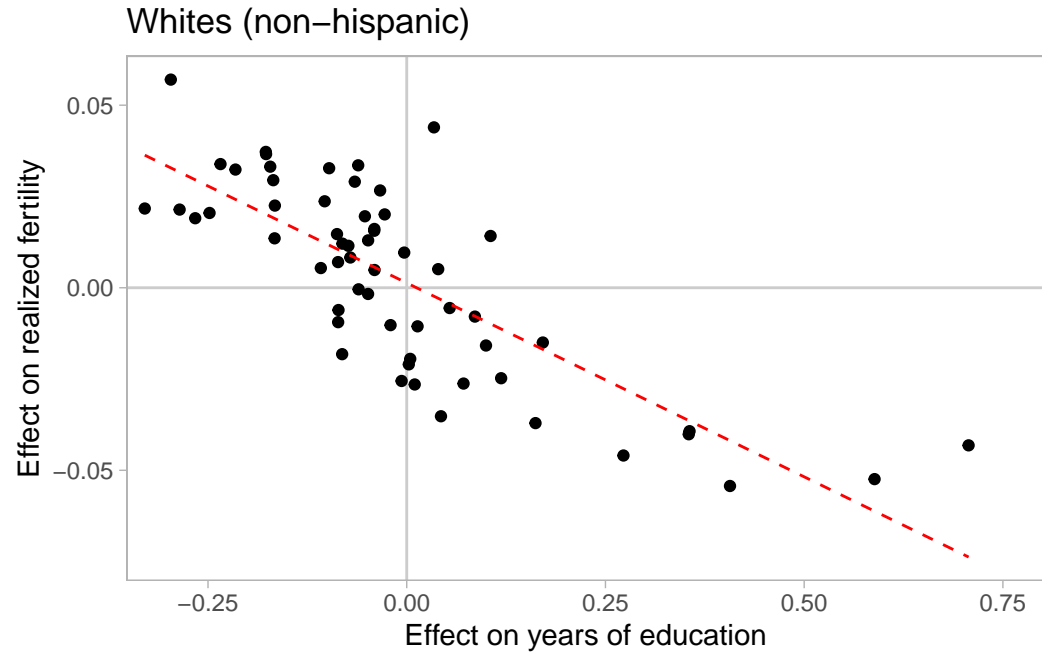
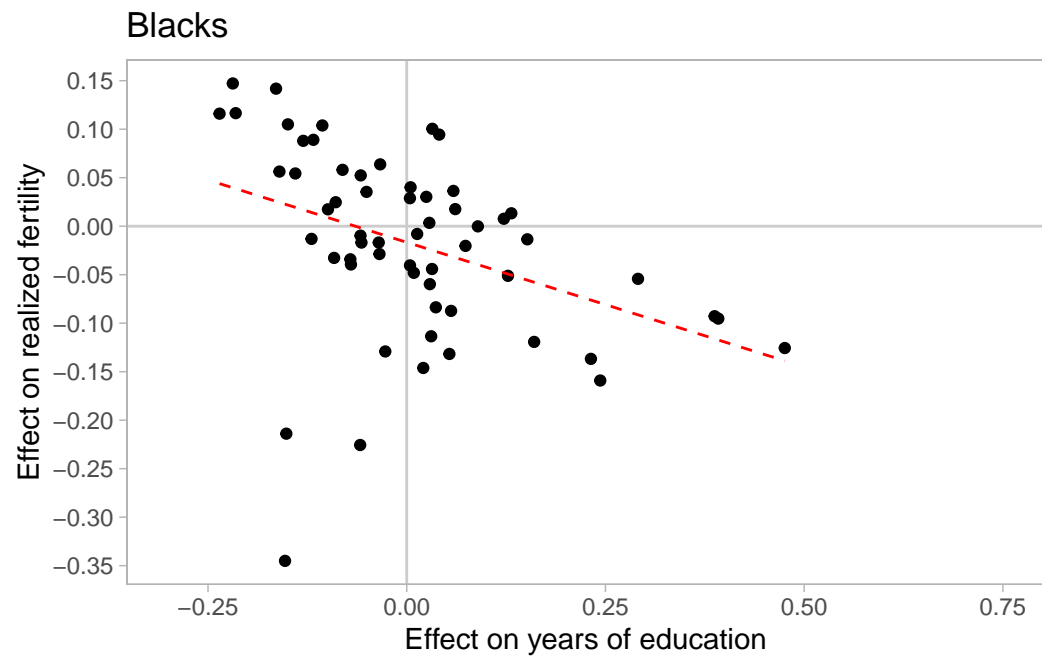


Figure 1: Effects of PGS on realized fertility. Values are effects of a one standard deviation (within ethnicity) change of the PGS on fertility, controlling for 10 principal components of genomic array data. Scales differ between the two charts.



(a)



(b)

Figure 2: Scatterplot of PGS partial correlations with fertility and years of education. Controls include 10 principal components of genetic array data. Dashed lines show linear regressions.

- 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.
- 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.