Negative Selection in the 20th Century

David Hugh-Jones, Abdel Abdellaoui

April 2020

```
# TODO
* Improve imputation of parents' YOB.
* Selection bias.
  - You could weight subjects by education level; sex; birth region?

## Data to gather
* f.2139 - age first had sex (includes "never had sex" which may explain
  some of the many NAs for f.2141, num sex partners)
```

1 Data

Data is taken from UK Biobank. Polygenic scores were normalized to mean 0, variance 1.

2 Results

We run regressions on two dependent variables:

- *siblings*, the number of full siblings in the respondent's sib (including himself or herself).
- The number of *children* ever born to/fathered by the respondent.

We run regressions both with and without controls for the 40 top principal components of the genetic data.

Figure 1 shows effect sizes of a one-standard deviation shift in each polygenic score.

Estimates are broadly consistent across generations. For 26 out of 33 polygenic scores, all 4 estimates have the same sign.

However, effect sizes are much smaller for *children* than *siblings* regressions. Among consistently-signed estimates, the median effect size for children as a proportion of the effect size for siblings is 0.27, or 0.41 with controls.

In siblings regressions, effect sizes are smaller when controlling for principal components – sometimes much smaller, as in the case of height. 27 out of 33 "controlled" effect sizes have a smaller absolute value than the corresponding "raw" effect size. The median proportion between raw and controlled effect sizes is 0.83. Among the children regressions, this no longer holds. Effect sizes are barely affected by controlling for principal components.

To get a further insight into this we regress *siblings* and *children* on individual principal components themselves. As Figure 2 shows, there are many more large and significant effects in siblings regressions.

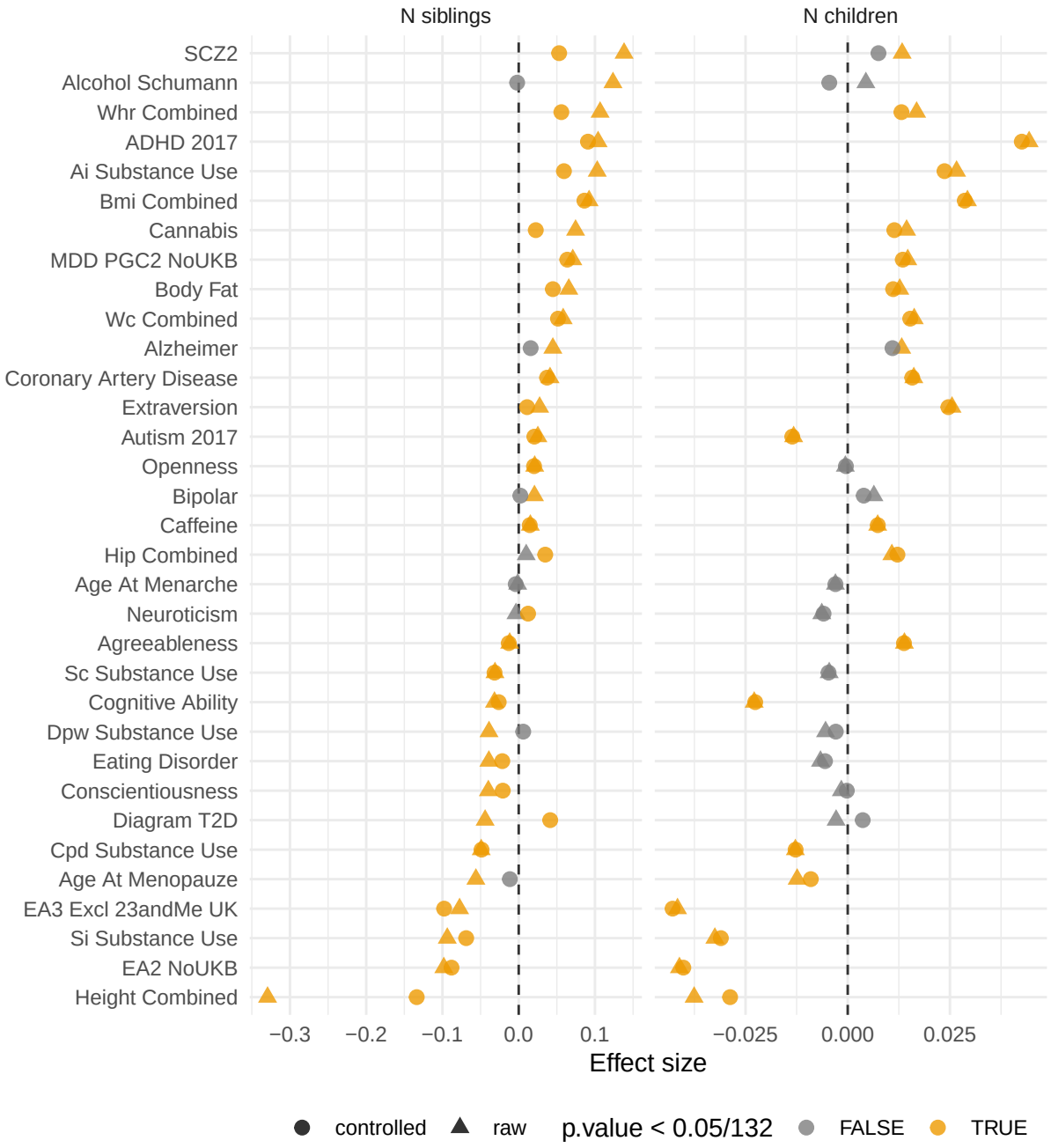

Figure 1: Effects of polygenic scores on number of siblings/children. Scores are plotted on different scales.

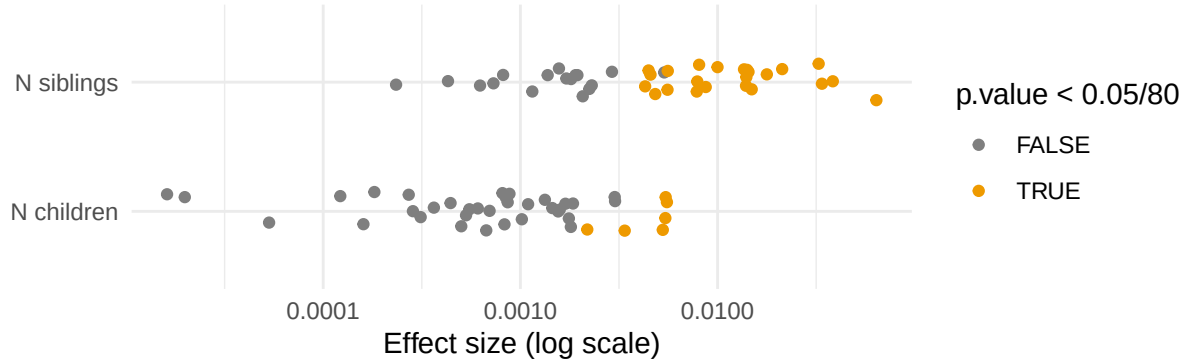

Figure 2: Effect of principal components on number of siblings/children. Each dot represents one principal component. Points are jittered on the Y axis.

3 Selection over time

Negative selection seems to decrease over time. Figure 3 shows effect sizes for *number of siblings* and *number of children*, median-split by parents' year of birth and own year of birth respectively. Parents' year of birth is imputed, which is likely to produce some bias.

By definition, the sibling regressions exclude members of the parents' generation who had no children. This is likely to bias results towards zero, since much of the effect in children regressions is due to respondents with high scores being more likely to have no children. So, we cannot directly compare effect sizes for the two sets of regressions. Within the sibling regressions, the most common pattern is that negative effects shrink in absolute size (Table 1).

Table 1: Change in effect sizes between early and late born parents, 'sibling' regressions

Change	**Number of scores**
Change sign to -	2
Change sign to +	1
Decreasing -	1
Decreasing towards 0	7
Increasing +	2
Increasing towards 0	1
Insignificant	19

Significance is measured at $p < 0.05/66$

In children regressions, no clear pattern is visible (Table 2).

Despite this, effect sizes tend to be smaller for children regressions. This could be caused by ascertainment bias in the UK Biobank sample – e.g., if respondents themselves are a more selected sample than the respondents' parents.

4 Causality

Different polygenic scores are correlated. Table 3 shows the top correlations in the sample. Because of this, bivariate correlations between PGS and number of children might be driven by other genetic scores. To explore which polygenic scores are driving negative selection, we run a single omnibus regression of *number*

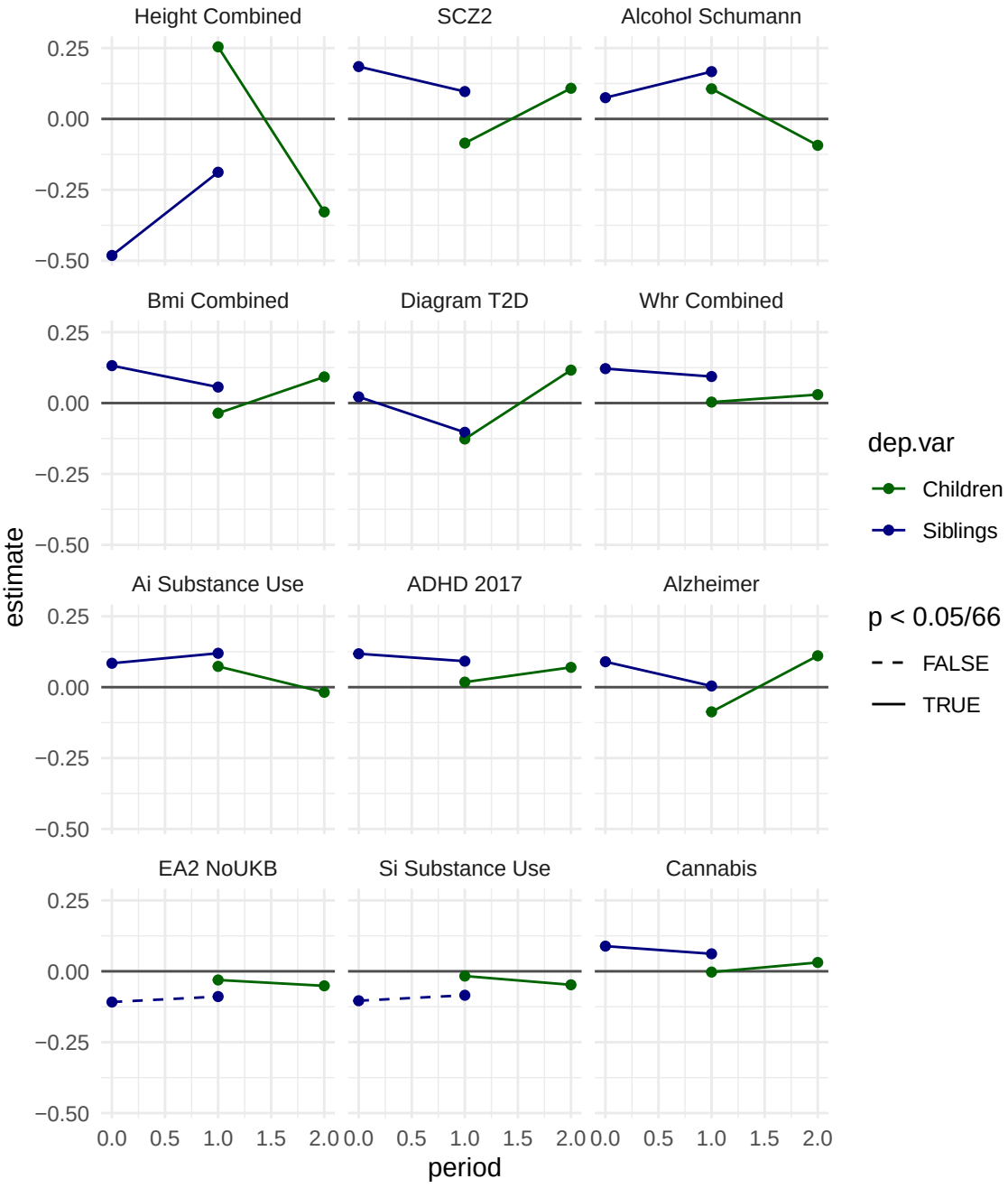

Figure 3: Effect sizes of PGS on number of children/siblings by own/parents' year of birth. PGS with the largest mean effect sizes are shown.

Table 2: Change in effect sizes between early and late born respondents, 'children' regressions

Change	**Number of scores**
Change sign to -	8
Change sign to +	12
Decreasing -	4
Decreasing towards 0	2
Increasing +	5
Insignificant	2

Significance is measured at $p < 0.05/66$

Table 3: Top 10 correlations between polygenic scores

PGS	**PGS**	**Correlation**
EA2 NoUKB	EA3 Excl 23andMe UK	0.89
Hip Combined	Wc Combined	0.807
Bmi Combined	Wc Combined	0.753
Wc Combined	Whr Combined	0.711
Bmi Combined	Hip Combined	0.697
Body Fat	Wc Combined	0.435
Bmi Combined	Body Fat	0.425
Bmi Combined	Whr Combined	0.425
Body Fat	Hip Combined	0.385
ADHD 2017	Autism 2017	0.328

of children on all the PGS. We exclude EA2, waist-hip ratio, WC (XXX what is it?) and "Hip combined" since they are highly correlated with other scores, which could make our estimates unstable. Figure 4 shows the results. Interestingly, several PGS remain independently significant, although effect sizes are reduced.

5 Subgroups

We next examine how different subgroups contribute to natural selection.

5.1 Males and females

Figure 5 shows effect sizes of PGS on number of children separately for males and females. For 16 out of 33 PGS, selection is more negative for women than for men. Differences are particularly large for educational attainment and height PGS.

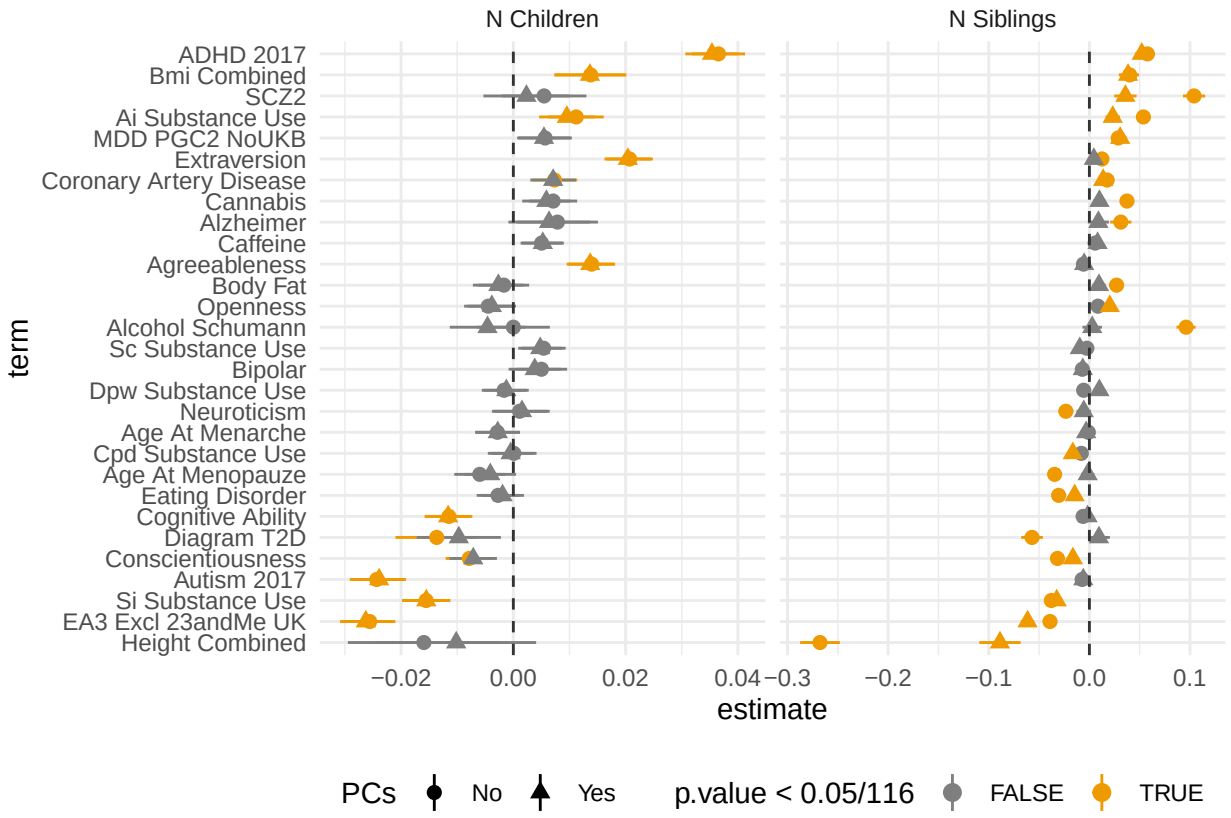

Figure 4: Partial correlations with number of children

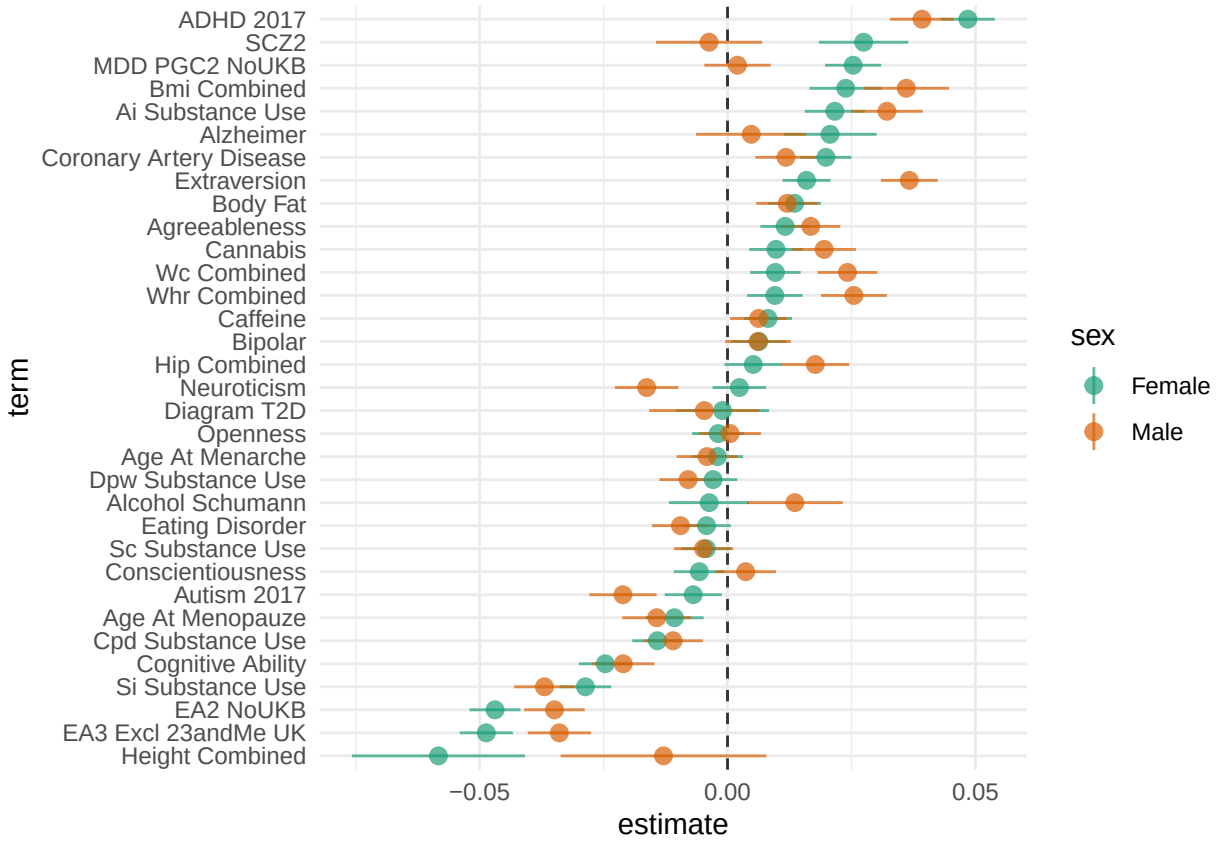

Figure 5: Effect sizes on number of children by sex

5.2 Number of sexual partners

Figure 6 splits males and females up by lifetime number of sexual partners. Remarkably, across both sexes, negative selection is strongly reversed for respondents who had 3 or fewer sexual partners in their lifetime.

5.3 Education levels

Figure 7 splits respondents up by education levels. Both negative and positive selection are typically larger and more significant for those who left school before 16. Table 4 summarizes the results.

Table 4: Negative selection by education level

Age left FTE	**% PGS significant**
< 16	63.6
16-18	27.3
> 18	6.06

6 Number of children

Figure 8 shows the full distribution of number of children born for different ventiles of the EA3 polygenic score. The strongest relationship seems to be for having 0 children versus 1 or more.

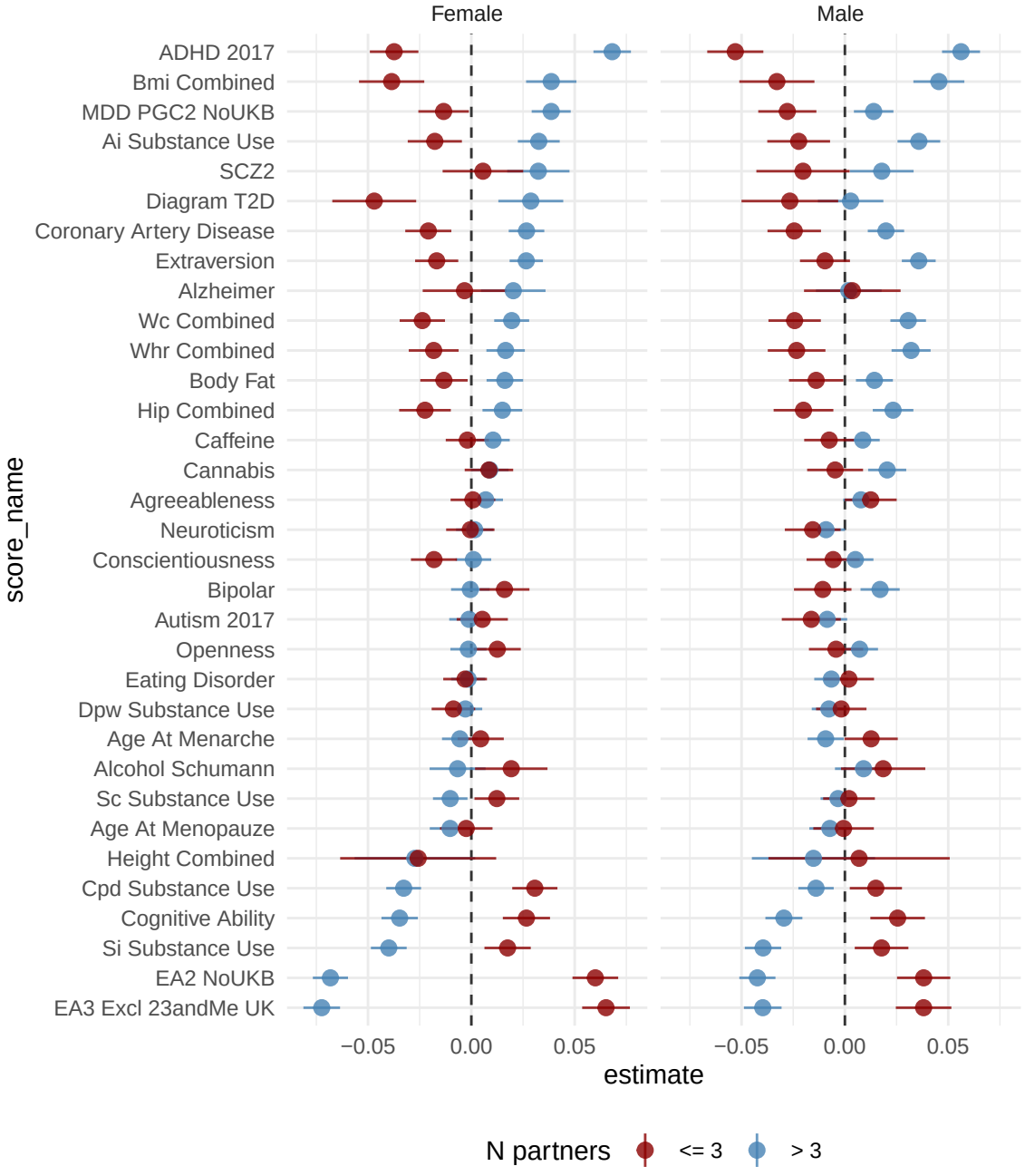

Figure 6: Effect sizes on number of children by number of sexual partners

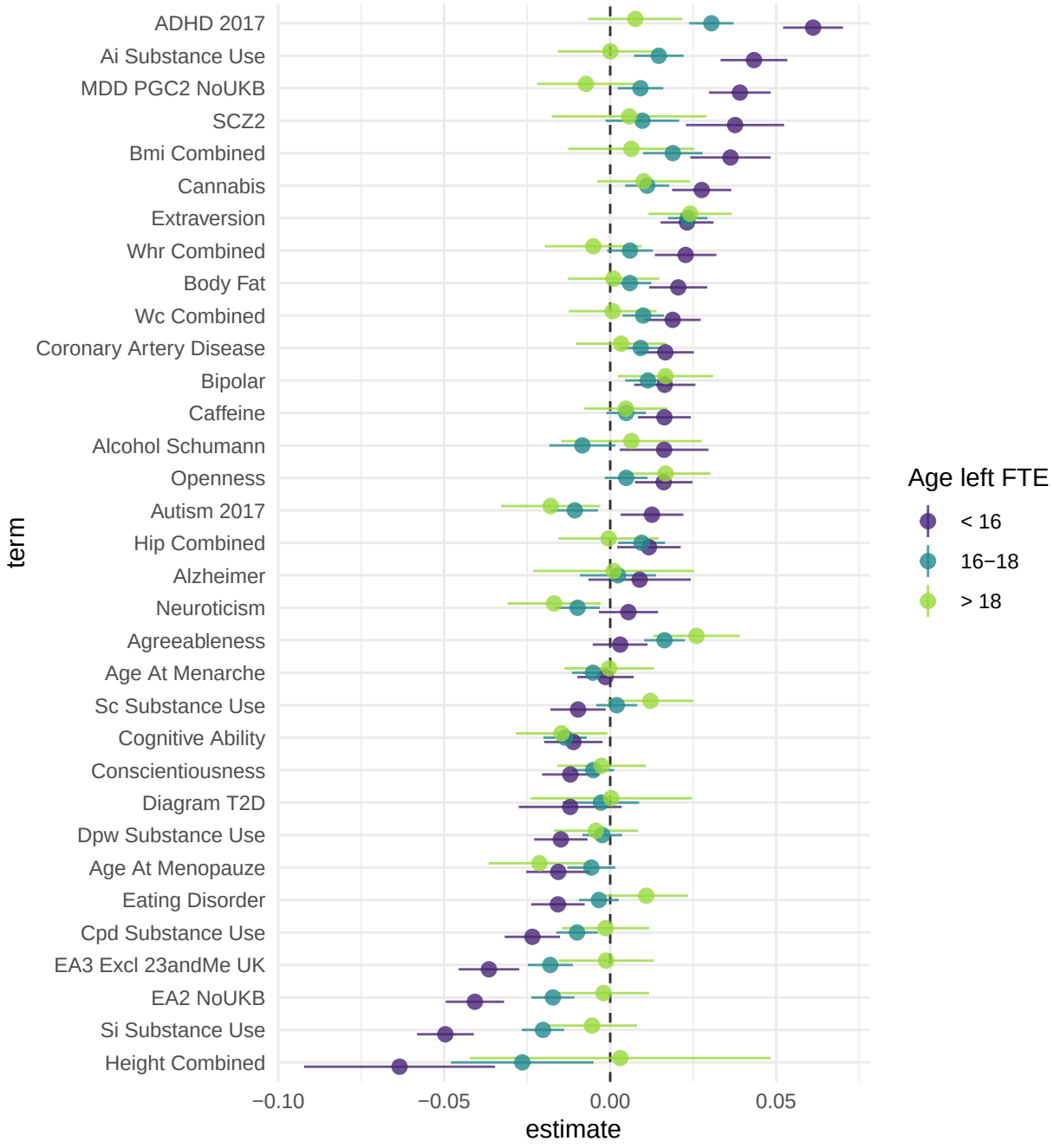

Figure 7: Effect sizes on number of children by age left full-time education

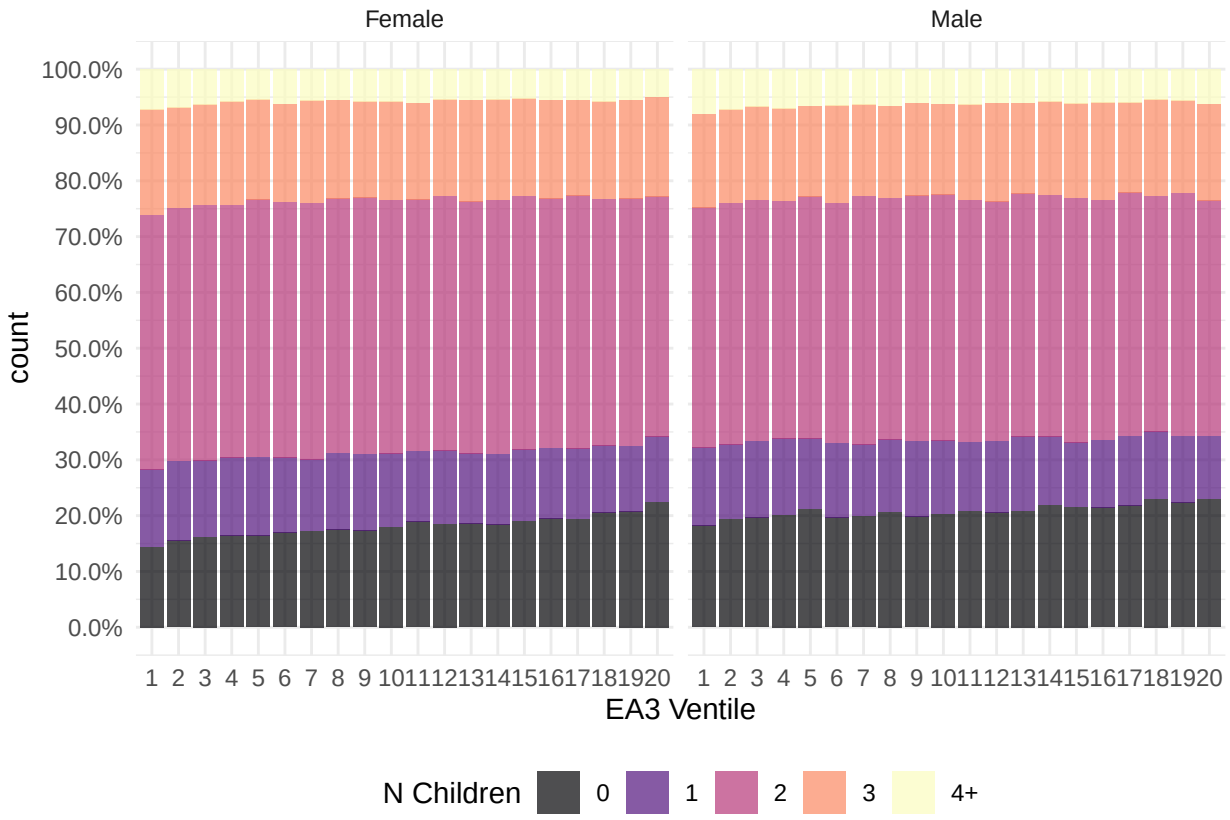

Figure 8: Number of children by ventiles of EA3 PGS