Human capital mediates natural selection in contemporary humans

David Hugh-Jones (University of East Anglia)
Abdel Abdellaoui (Amsterdam UMC)

Contemporary humans are undergoing natural selection on polygenic scores (Barban et al. 2016; Beauchamp 2016; Kong et al. 2017; Sanjak et al. 2018).

But we have no theory of why.

- We examine natural selection on 33 polygenic scores in UK Biobank.
- Patterns in the data can be explained by the economic theory of fertility (Becker 1960).
- Natural selection has substantial effects on genetic inequality (Harden 2021).

DNA

Human DNA has about 3 billion base pairs.

Each pair consists of two molecules of adenine (A), cytosine (C), guanine (G), and thymine (T). The pairs are A-T or C-G.

Most of these are identical across all humans, but we differ from each other at about 4-5 million sites.

One kind of variation is the Single Nucleotide Polymorphism (SNP): a difference at a single base pair. Humans may carry 0, 1 or 2 of the minor allele (the less common of the possible pairs).

DNA arrays capture SNP data at a large number of locations.

Polygenic scores

A polygenic score is a summary statistic derived from DNA.

It is a weighted sum of SNP alleles. Weights are derived from bivariate correlations with the target *phenotype* (e.g. years of education, coronary artery disease, age started smoking).

Only correlations below a given p-value threshold may be used, and weights may be adjusted for correlation among different SNPs.

We create 33 polygenic scores for individuals in UK Biobank.

Summary statistics (weights) for the polygenic scores were calculated excluding any data from UK Biobank.

Scores are normalized to mean 0, variance 1.

To estimate natural selection, we calculate bivariate correlations with fertility.

Scores which correlate positively (negatively) with fertility are being selected for (against).

The correlation equals the expected polygenic score in the children's generation.

Note: no claim about causality here!

Sample

UK Biobank: a health survey comprising about 500,000 individuals, born 1934-1970.

Respondents are selected geographically (22 assessment centres around Great Britain); and by health, income, etc. due to different response rates.

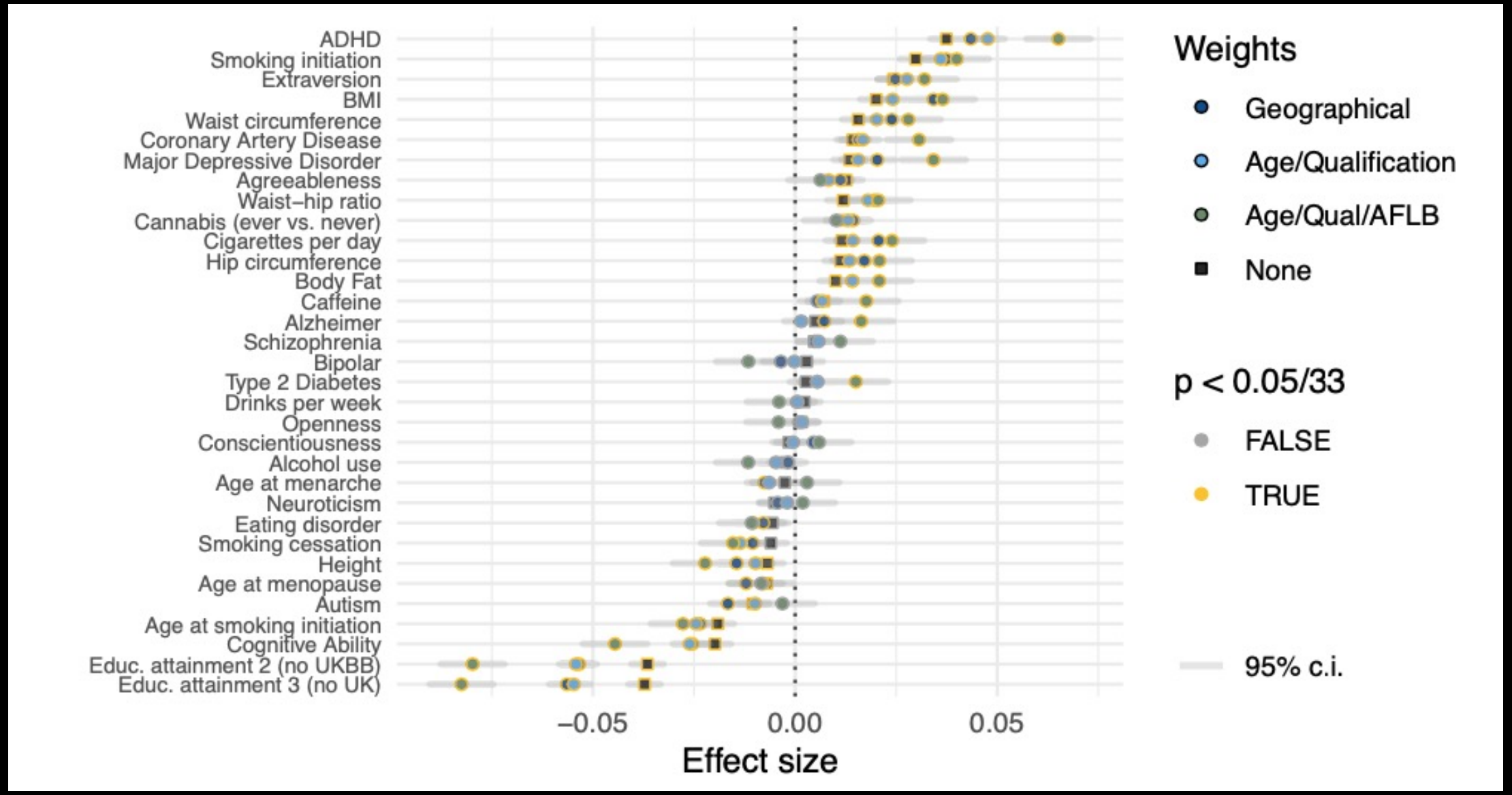
Includes DNA data, health questionnaire and some “social science” variables.

We use:

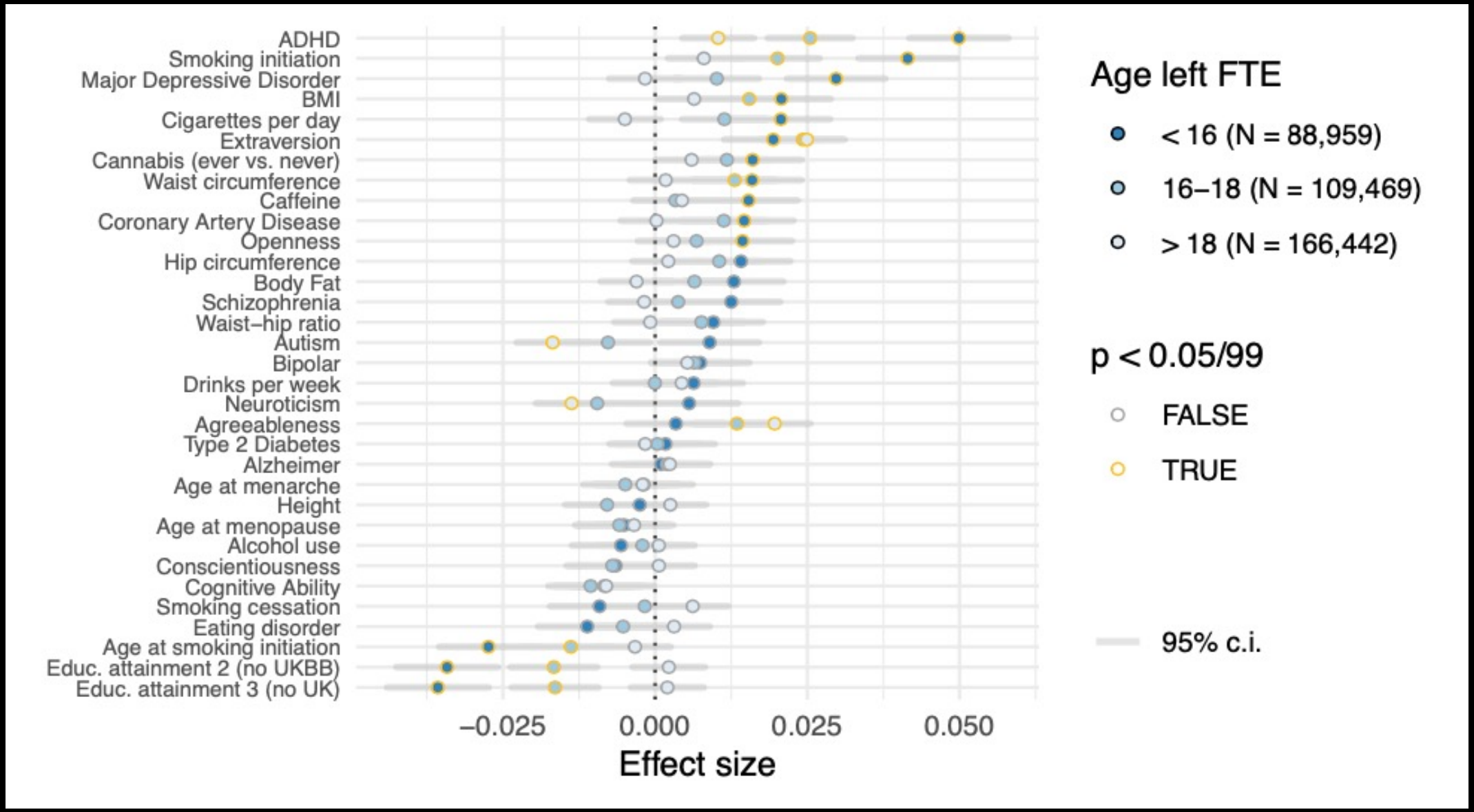
- Number of children
- Number of siblings (to estimate natural selection in the previous generation)

For number of children we use individuals aged 45+.

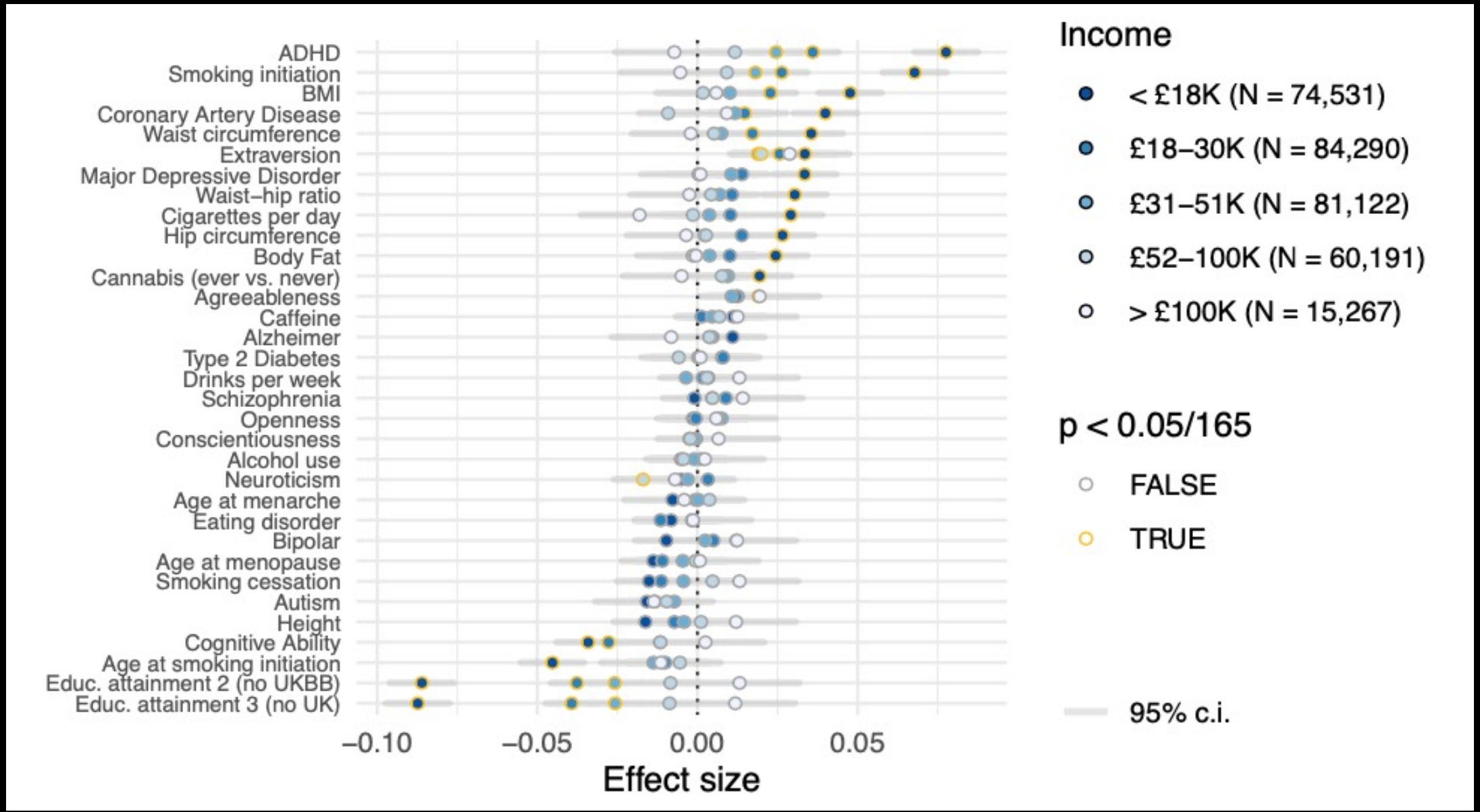
Correlations with completed fertility



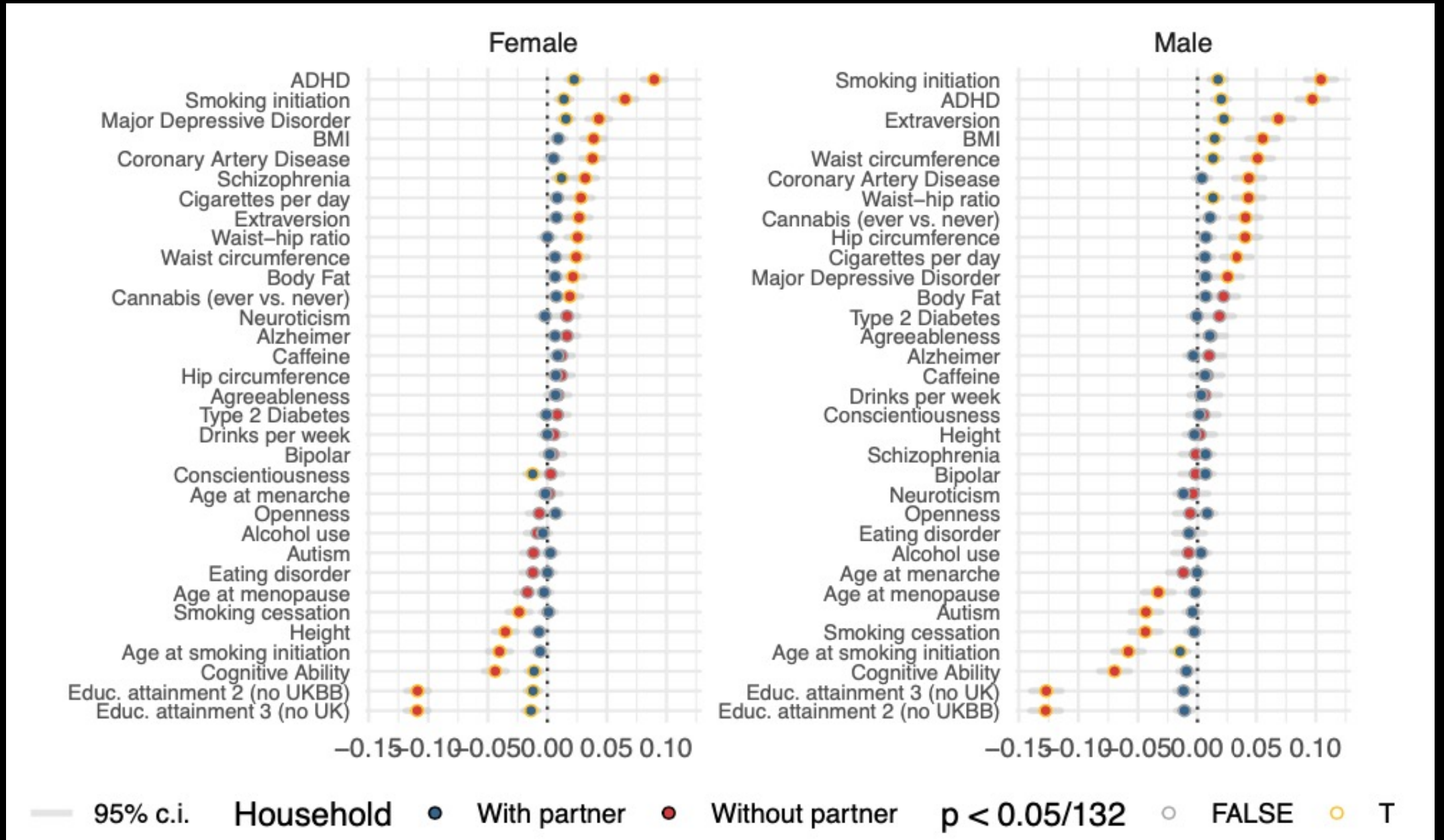
Education


ADHD
Smoking initiation
Major Depressive Disorder
BMI
Cigarettes per day
Extraversion
Cannabis (ever vs. never)
Waist circumference
Caffeine
Coronary Artery Disease
Openness
Hip circumference
Body Fat
Schizophrenia
Waist–hip ratio
Autism
Bipolar
Drinks per week
Neuroticism
Agreeableness
Type 2 Diabetes
Alzheimer
Age at menarche
Height
Age at menopause
Alcohol use
Conscientiousness
Cognitive Ability
Smoking cessation
Eating disorder
Age at smoking initiation
Educ. attainment 2 (no UKBB)
Educ. attainment 3 (no UK)
−0.025
0.000
0.025
0.050
Effect size
Age left FTE
< 16 (N = 88,959)
16–18 (N = 109,469)
> 18 (N = 166,442)
p < 0.05/99
FALSE
TRUE
95% c.i.

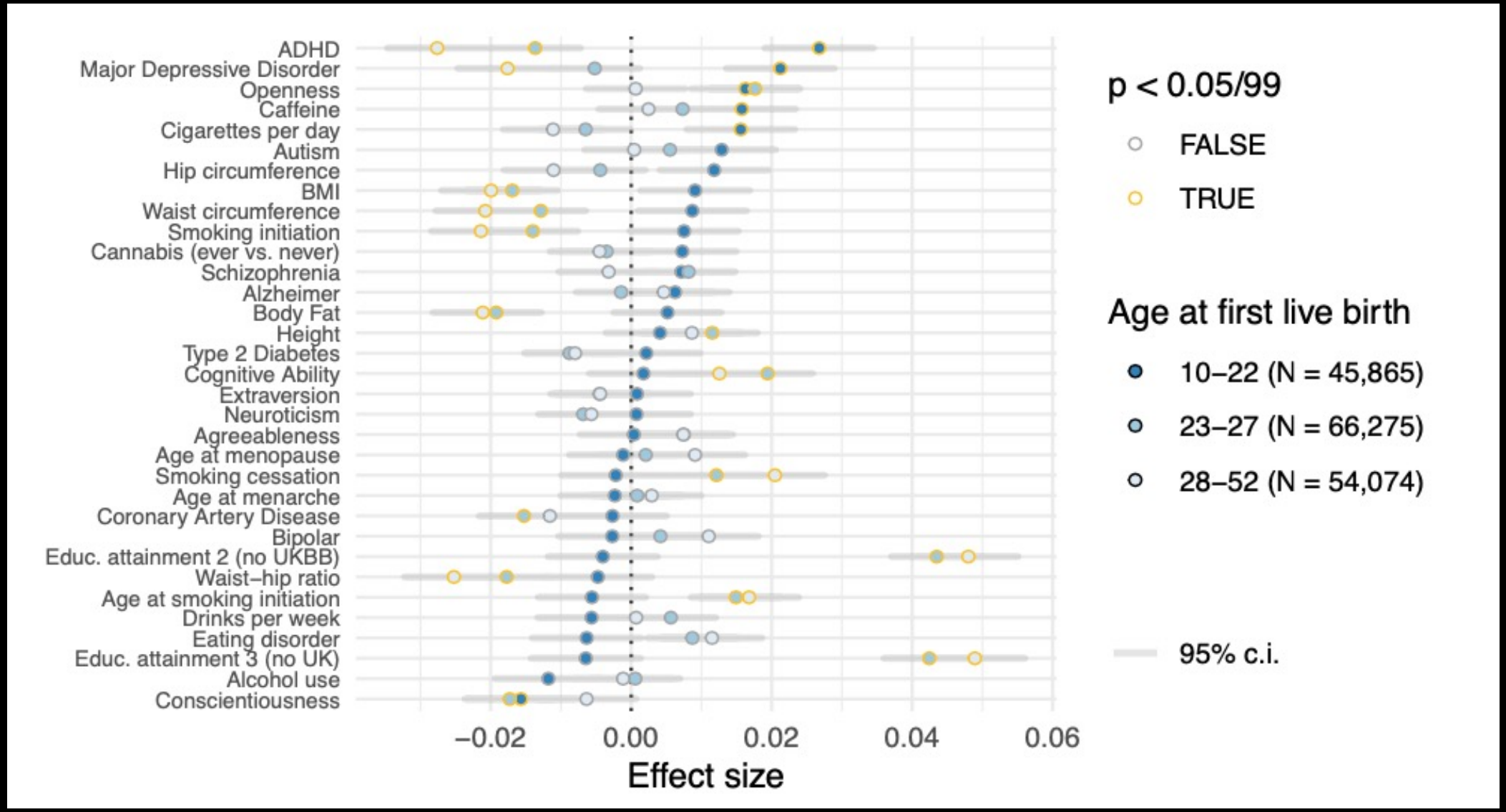
Income


ADHD
Smoking initiation
BMI
Coronary Artery Disease
Waist circumference
Extraversion
Major Depressive Disorder
Waist–hip ratio
Cigarettes per day
Hip circumference
Body Fat
Cannabis (ever vs. never)
Agreeableness
Caffeine
Alzheimer
Type 2 Diabetes
Drinks per week
Schizophrenia
Openness
Conscientiousness
Alcohol use
Neuroticism
Age at menarche
Eating disorder
Bipolar
Age at menopause
Smoking cessation
Autism
Height
Cognitive Ability
Age at smoking initiation
Educ. attainment 2 (no UKBB)
Educ. attainment 3 (no UK)
−0.10
−0.05
0.00
0.05
Effect size
Income
< £18K (N = 74,531)
£18–30K (N = 84,290)
£31–51K (N = 81,122)
£52–100K (N = 60,191)
> £100K (N = 15,267)
p < 0.05/165
FALSE
TRUE
95% c.i.

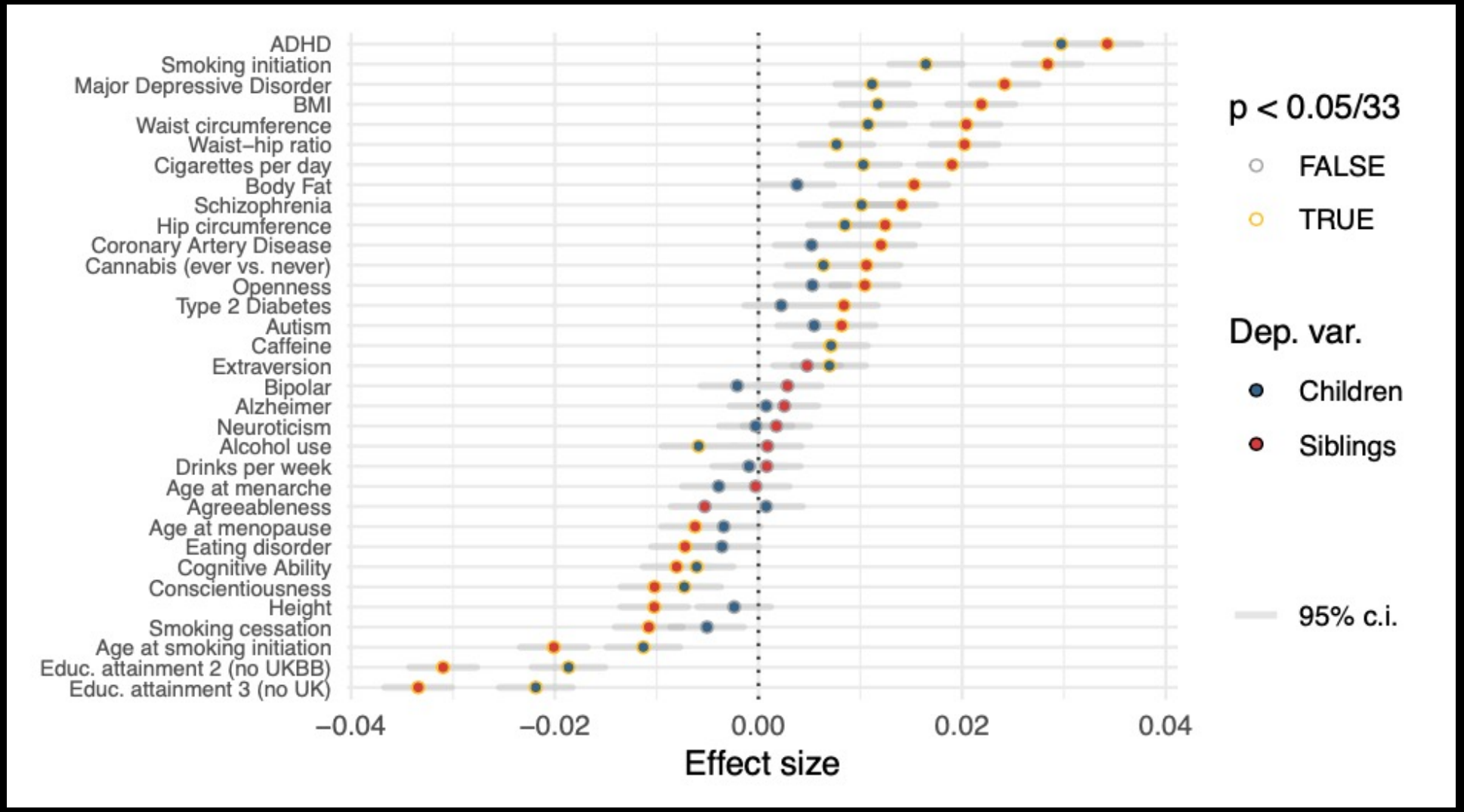
Household structure



Age at first birth (women)



Parents' generation (dep. var.: N siblings)



Is there change over time?

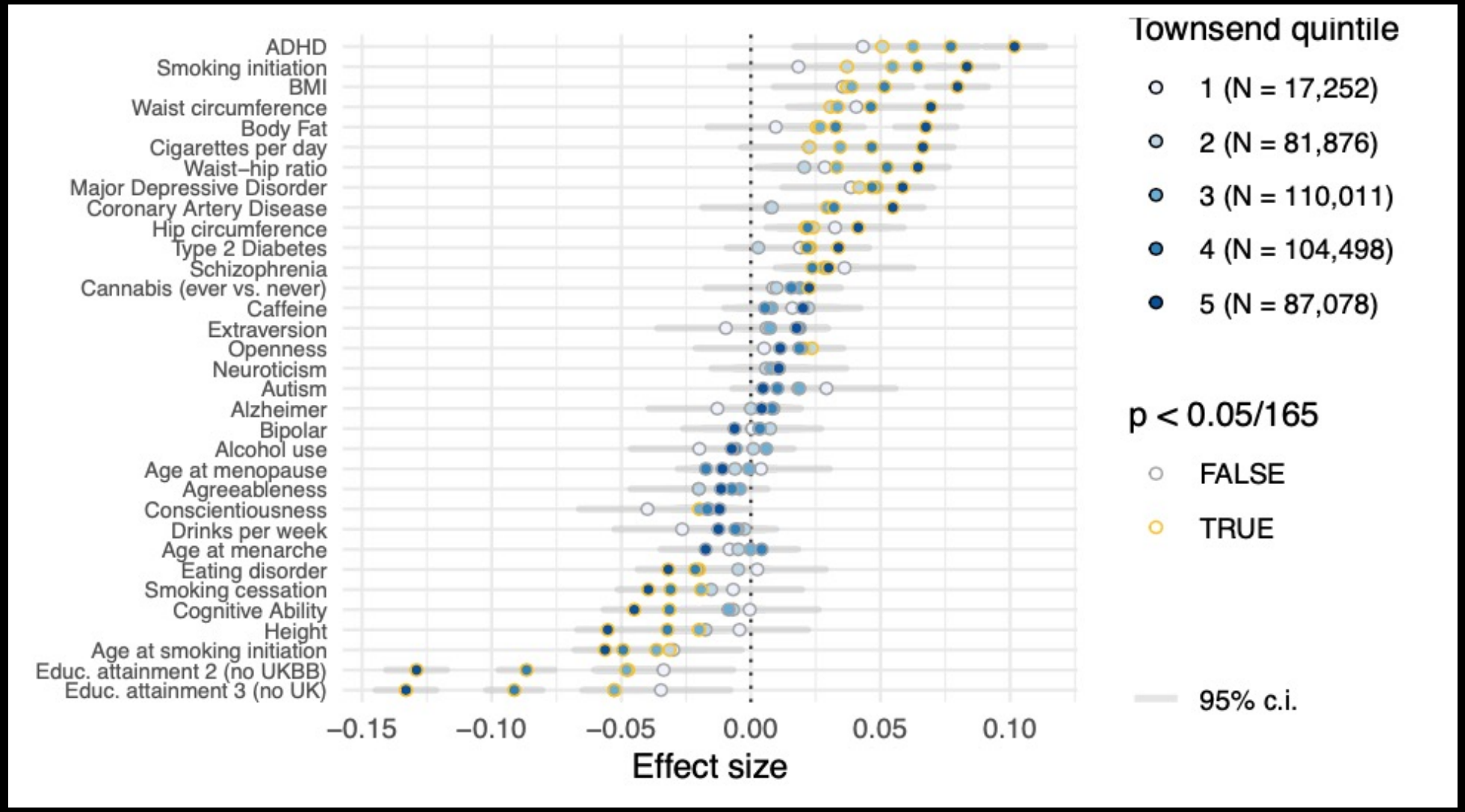
Comparing siblings to children regressions is hard:

- Parents with 0 children can’t be parents of anyone in the sample.
- Effects of childlessness on estimates may vary across generations.

Instead, we median-split the sample at YOB 1950, and compare effect sizes among these groups.

We find little evidence for change in effect sizes.

Townsend deprivation quintile, parents' generation



Model

By increasing a person's expected wage, human capital has opposing effects on fertility (Becker 1960):

- Children become more affordable (*income effect*)…
- … but cost more in foregone wages (*substitution effect*)

We write down a simple model of human capital and fertility choice.

Model

Utility is given by

$$U(N1, N2) = u(Y1) + u(Y2) + aN1 + aN2$$

Where

- $N1$, $N2$ are the number of children in each period;
- $Y1$, $Y2$ is income in each period.

Period 1 wages are 1.

In period 1, individuals choose a level of education at time cost s. Period 2 wages are $w(s,h) = sh$ where h is exogenous human capital.

Raising a child takes time b.

Utility over income is CRRA:

$$u(y) = \frac{y^{1-\sigma} - 1}{\sigma - 1}$$

Results

For $\sigma < 1$ and close enough to 1, the fertility-human capital relationship dN^*/dh is:

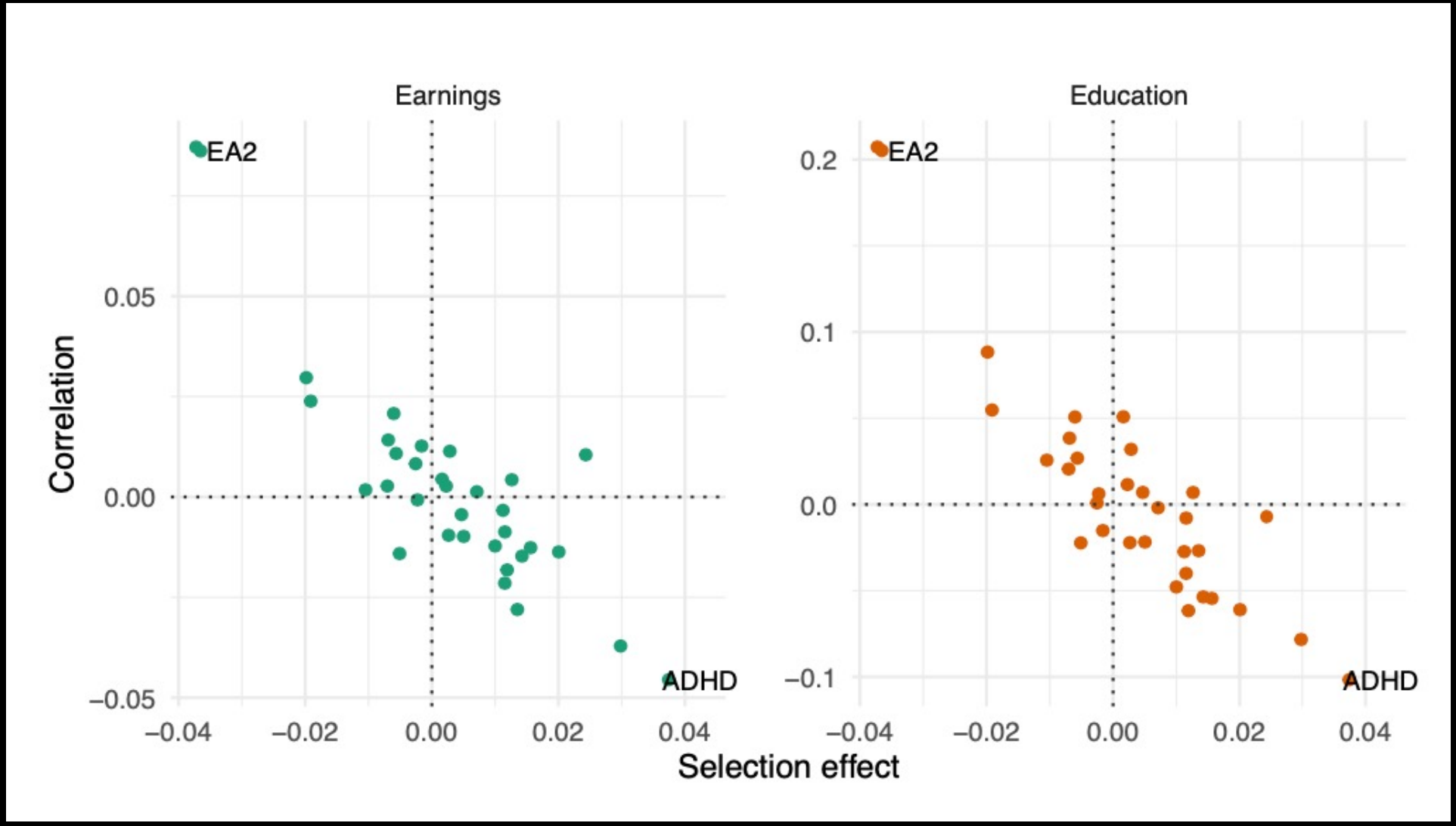
1. Negative.
2. Closer to zero at higher wages and/or levels of human capital.
3. Closer to zero at higher levels of education *s*.
4. More negative when the time burden of children *b* is larger.
5. Closer to zero among those who start fertility in period 2 ($N^*1 = 0$) than among those who start fertility in period 1 ($N^*1 > 0$).

Results

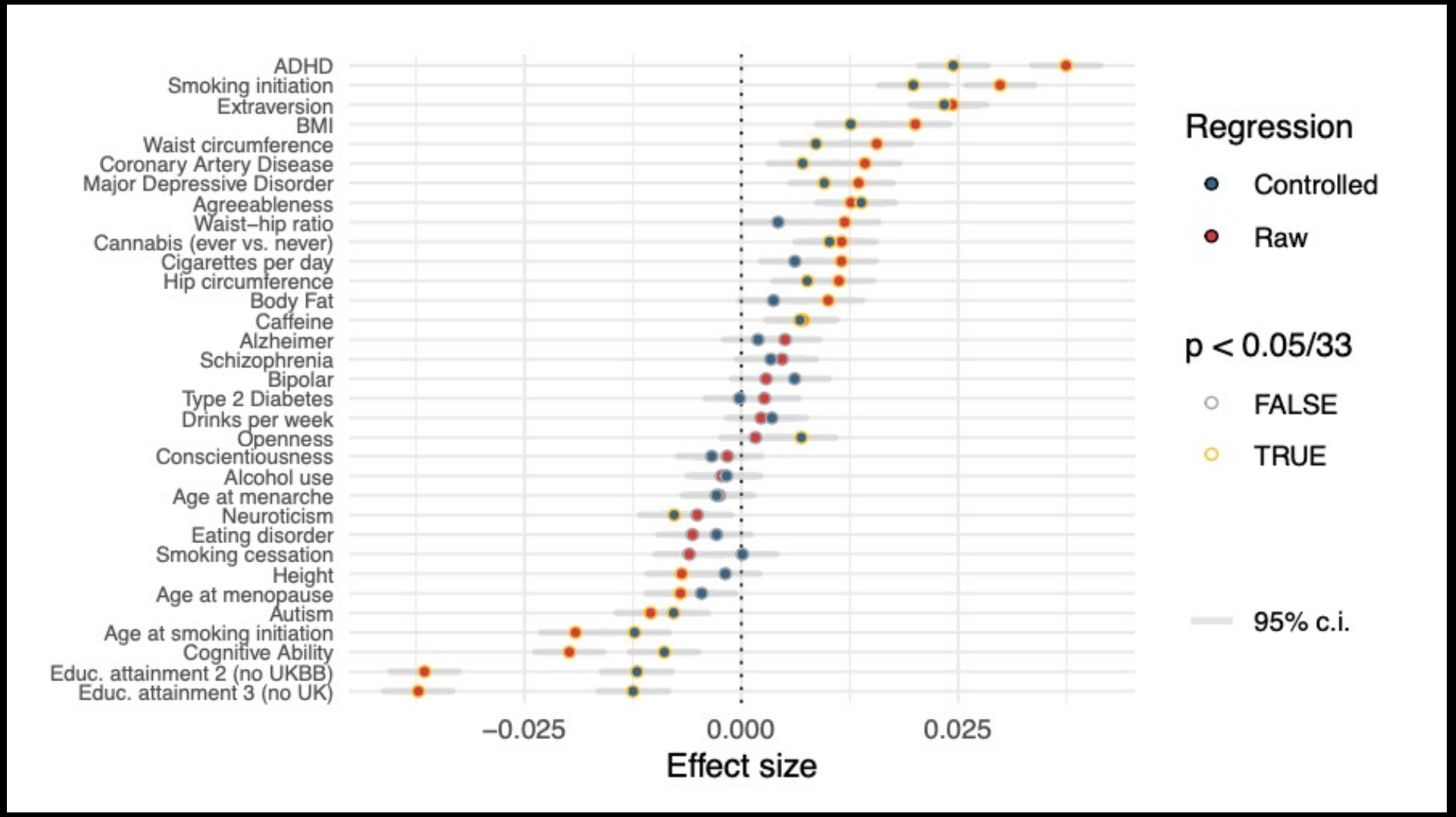
For $\sigma < 1$ and close enough to 1, the fertility-human capital relationship dN^*/dh is:

1. Negative.
2. Closer to zero at higher wages and/or levels of human capital. *Weaker effects on fertility at higher income.*
3. Closer to zero at higher levels of education *s*. *Weaker effects at higher education levels.*
4. More negative when the time burden of children *b* is larger. *Stronger effects among those living without a partner.*
5. Closer to zero among those who start fertility in period 2 ($N^*1 = 0$) than among those who start fertility in period 1 ($N^*1 > 0$). *Weaker effects for later age at first birth.*

PGS which correlate negatively with earnings and education are selected for



Controlling for earnings and education



Within-siblings regressions

Within siblings, genetic variation is assigned randomly.

We can use this to test for causal effects of genes on fertility among ~ 36,000 individuals in a sibling group.

No results are significant after Bonferroni correction.

Results correlate with pooled effect sizes, and are about 1/3 smaller (regression of within-sibling on pooled, b = 0.71).

Controlling for earnings and education has little effect, though educational attainment effect size is reduced by ~ 20%.

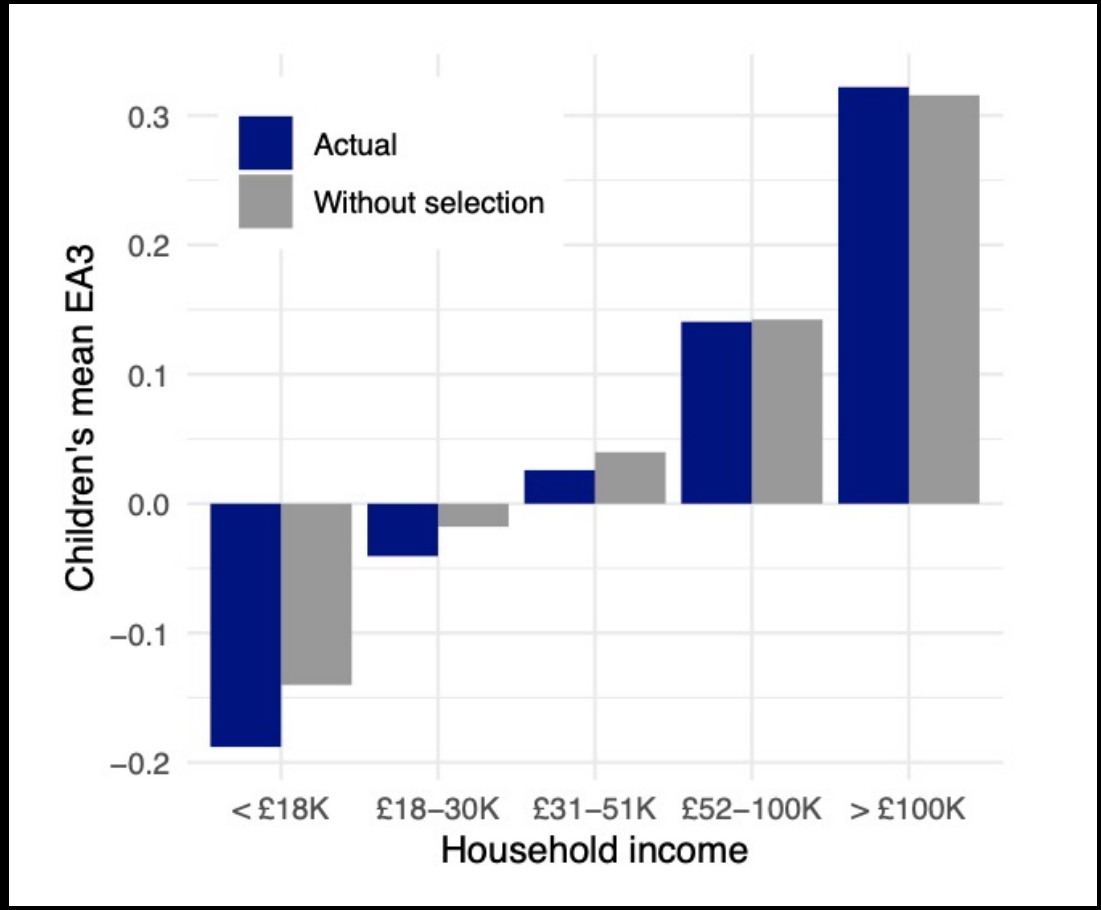
Effects

Effects on population means are small.

- But noisy PGS imply errors-in-variables.
- And sample selection in UKBB seems to bias results towards 0.

Effects on inequality are substantive.

- Results increase inequality for 29 out of 33 scores.
- Median increase in difference between highest- and lowest-income groups: 14% in 1 generation.



SAPERE AUDE

Thank you!