The causal effects of education on health: evidence from Mendelian randomization and the raising of the school leaving age.

Neil M Davies, Matt Dickson, George Davey Smith,
Gerard van den Berg, and Frank Windmeijer

Danish Epidemiological Society Meeting

Aarhus 16th April 2018

This is work in progress all comments, questions very welcome!

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

MRC Integrative Epidemiology Unit

Motivation

- Educated people are on average healthier, wealthier, and have higher life expectancy.
- Are these differences caused by education? Or merely correlated?
- Two potential instruments
    1. Raising of the school leaving age ("the ROSLA") in the UK in 1972
    2. Genetic variants associated with education (Mendelian randomization - MR)
- The ROSLA widely accepted as valid instrument for education

- What about Mendelian randomization?
    - Nguyen et al. (2016). Effect of education on cognitive impairment.
    - Hagenaars et al (2017). Effect of education on cognition and other traits later in life.

Background: The causal effects of education

The Relationship Between Education and Adult Mortality in the United States

ADRIANA LLERAS-MUNEY
Princeton University

First version received May 2002; *final version accepted February* 2004 (*Eds.*)

Prior research has uncovered a large and positive correlation between education and health. This paper examines whether education has a causal impact on health. I follow synthetic cohorts using successive U.S. censuses to estimate the impact of educational attainment on mortality rates. I use compulsory education laws from 1915 to 1939 as instruments for education. The results suggest that education has a causal impact on mortality, and that this effect is perhaps larger than has been previously estimated in the literature.

1. INTRODUCTION

Access to healthcare insurance[1] and expenditures on healthcare,[2] have been shown to have little effect on health. On the other hand, there is a large and positive correlation between education and health (Grossman and Kaestner, 1997). This correlation is strong and significant

- Instruments: Changes to compulsory education laws in the US
- Outcome: 1% random sample of census data
  - 10 year mortality aged 50 to 60
- Education: NHANES + NHEFS
- Partial F-stats: 4.69 to 14.93
- WLS estimate:
  - 1.7 (95%CI: 2.4 to 4.8) percentage point (pp) drop in mortality per year of education
- IV estimate:
  - 3.6 (95%CI: 0.9 to 2.5) pp reduction in mortality per year of education.

Background: The causal effects of education

American Economic Review 2013, 103(6): 2087–2120
http://dx.doi.org/10.1257/aer.103.6.2087

The Effect of Education on Adult Mortality and Health: Evidence from Britain†

By DAMON CLARK AND HEATHER ROYER*

There is a strong, positive, and well-documented correlation between education and health outcomes. In this paper, we attempt to understand to what extent this relationship is causal. Our approach exploits two changes to British compulsory schooling laws that generated sharp across-cohort differences in educational attainment. Using regression discontinuity methods, we find the reforms did not affect health although the reforms impacted educational attainment and wages. Our results suggest caution as to the likely health returns to educational interventions focused on increasing educational attainment among those at risk of dropping out of high school, a target of recent health policy efforts. (*JEL* H52, I12, I21, I28)

The causal effect of education on health is a key parameter. It is central to models of the demand for health capital (Grossman 1972) and models of the influence of childhood development on adult outcomes (Heckman 2007; Conti, Heckman, and Urzua 2010). It is also relevant to macroeconomic growth models that incorporate mortality and human capital accumulation (Acemoglu and Johnson 2007; Cervellati

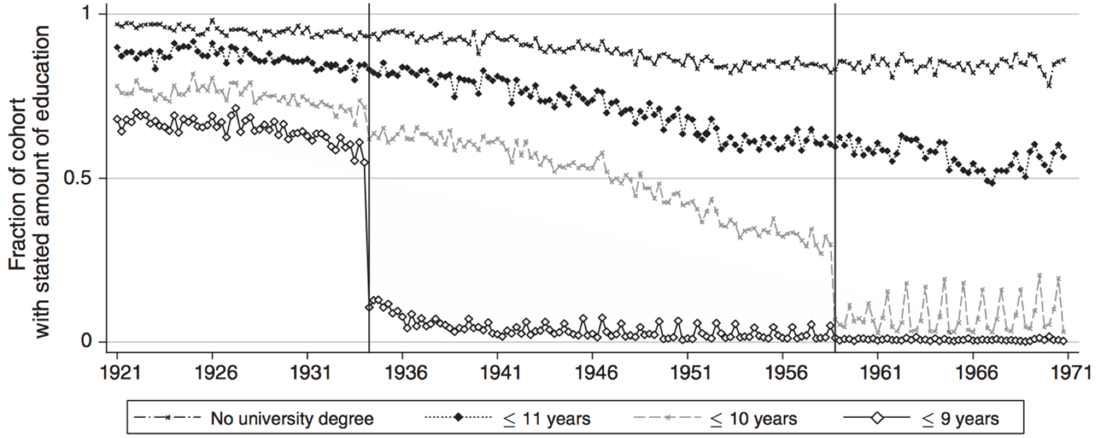


FIGURE 1. YEARS OF FULL-TIME EDUCATION BY QUARTER OF BIRTH

- Instruments: Changes to compulsory education laws in the UK in 1947 & 1972
- Outcome: full sample of census data
  - For 1972 reform 5 year mortality individuals aged 40 to 44
- Education: Health Survey for England and General Household Survey.
- Partial F-stats: Not provided
- IV estimate:
  - Assuming mortality risk of 0.79%
  - 0.12 (95%CI: 0.01 to 0.24) pp reduction in mortality per year of education.

nature
human behaviour

Letter

The causal effects of education on health outcomes in the UK Biobank

Neil M. Davies, Matt Dickson, George Davey Smith, Gerard J. van den Berg & Frank Windmeijer

Nature Human Behaviour **2**, 117–125 (2018)
doi:10.1038/s41562-017-0279-y
Download Citation
Economics Education

Received: 22 November 2016
Accepted: 05 December 2017
Published online: 29 January 2018

Abstract

Educated people are generally healthier, have fewer comorbidities and live longer than people with less education[1,2,3]. Much of the evidence

Potential instrumental variable 1

- Raising of the school leaving age
  - Minimum school leaving age increased from 15 to 16 in September 1972
  - Forced some students to remain in school for an additional year
  - The students affected were lower ability on average
  - Had few detectable affects on other educational choices (e.g. getting a degree, or remaining in school till 18.
  - Widely used in the literature (e.g. Harmon and Walker 1995, Clark and Royer 2013, Dickson 2013).

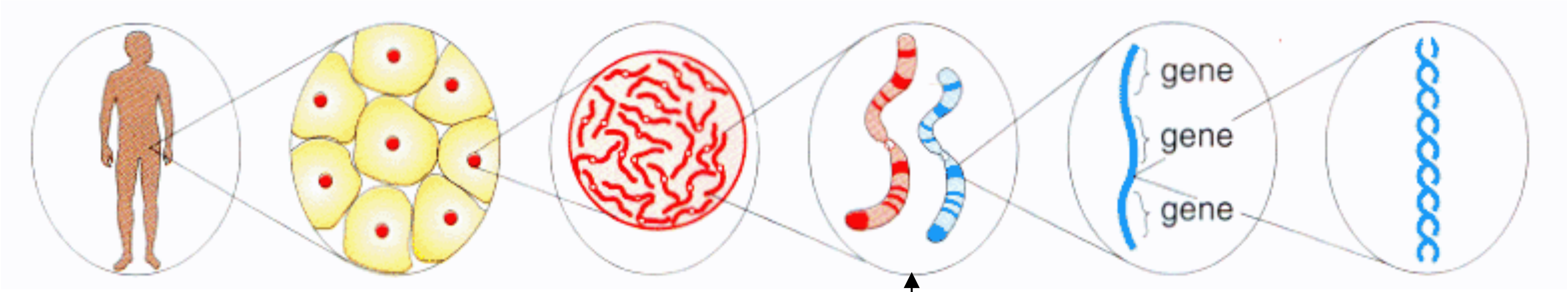
Potential instrumental variable 2

- Mendelian randomization
  - Okbay et al. (2016) 74 SNPs associated with educational attainment at genome-wide significance levels in the discovery sample
  - No overlap with the UK Biobank
  - Constructed weighted allele scores for educational attainment
  - Harmonized the alleles

Structure of DNA and chromosomes

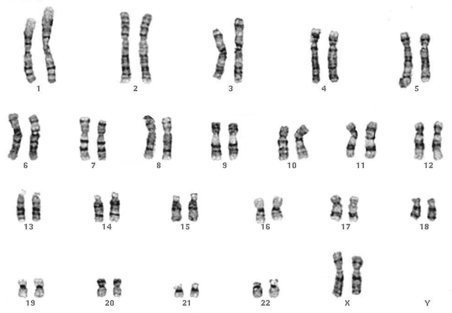
Cell nucleus

Genes

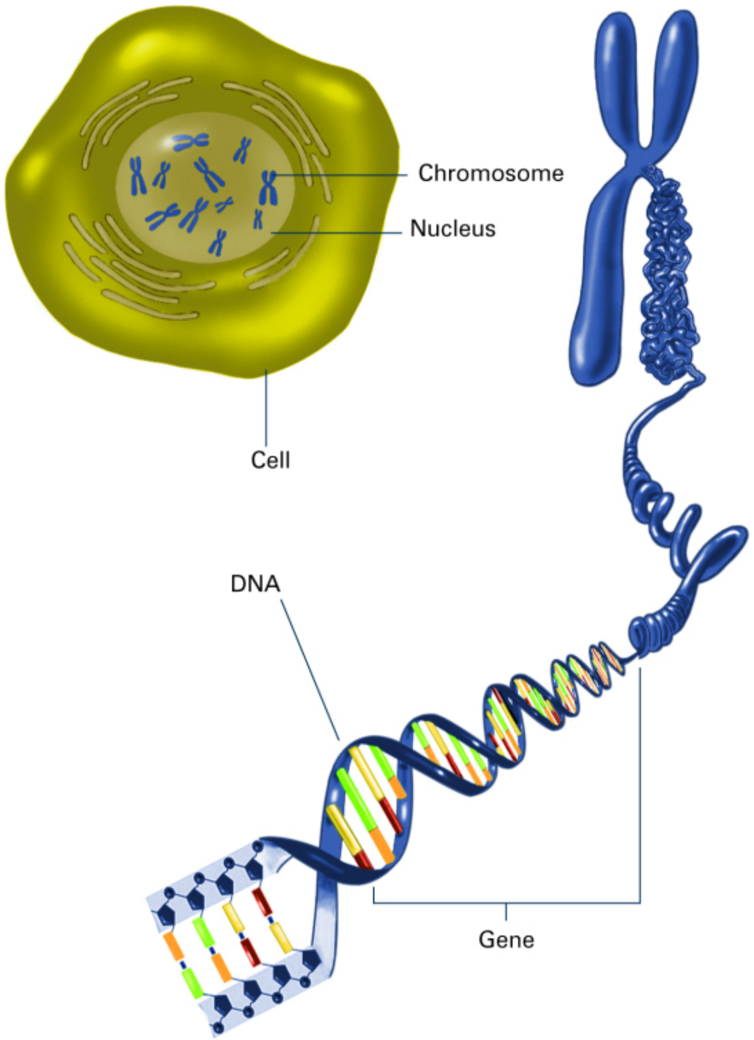


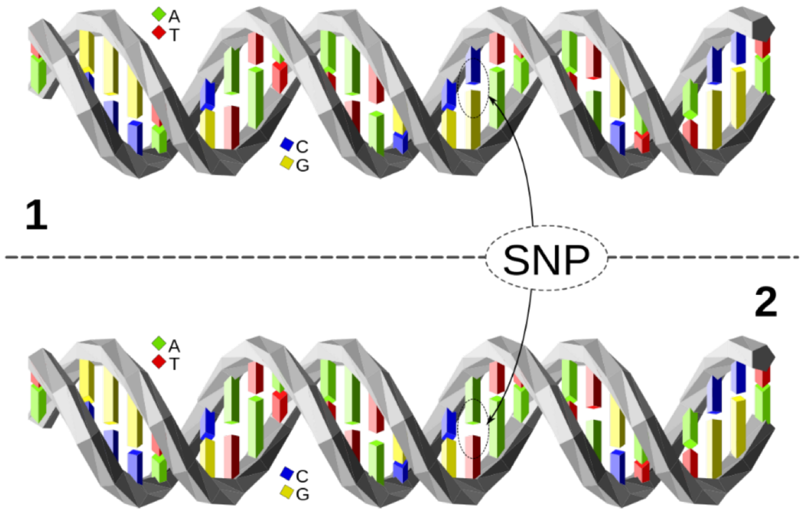
Chromosomes

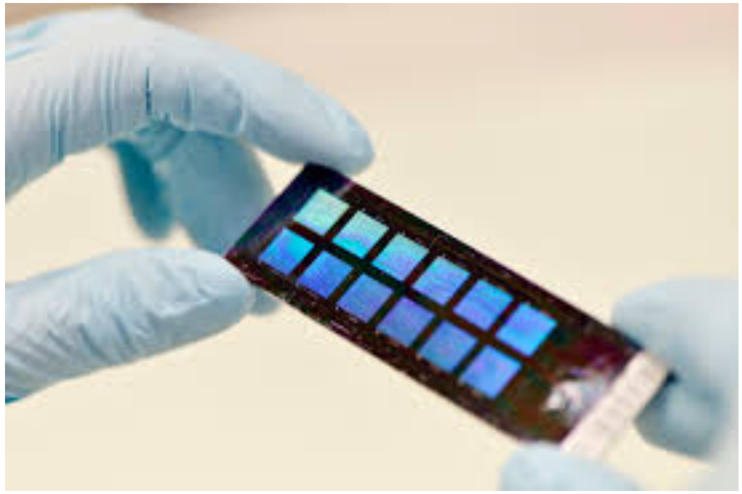
DNA



How do we measure genetics?







Sources of bias in Mendelian randomization

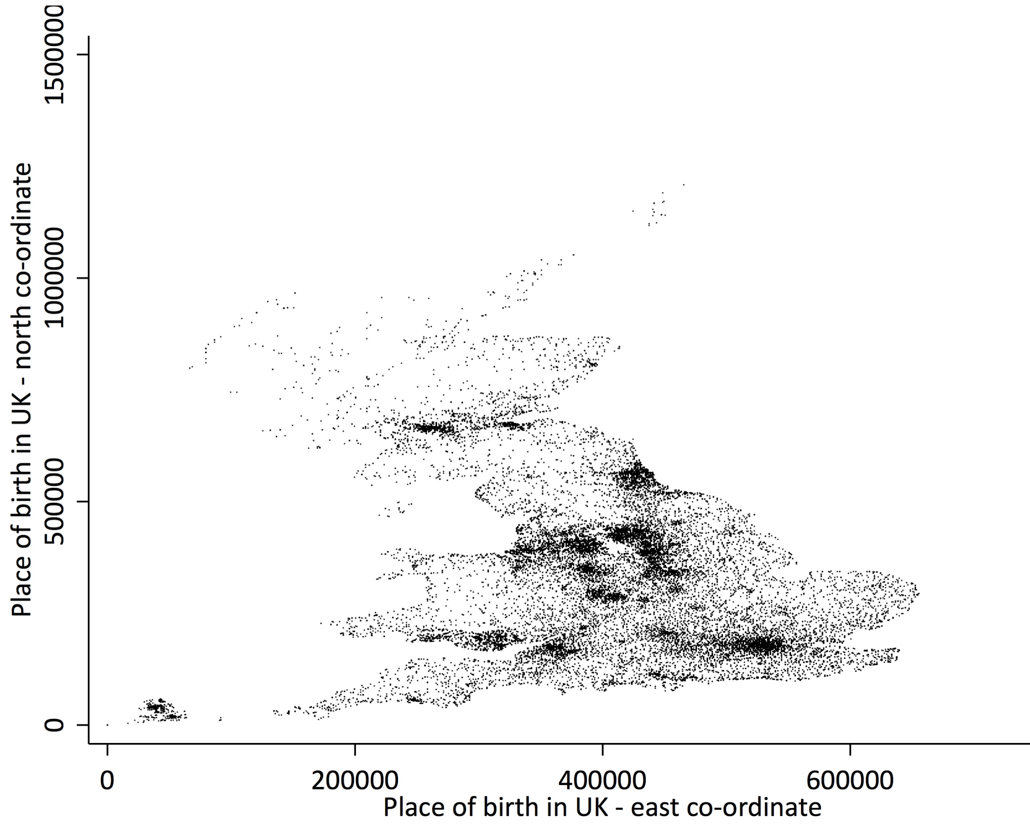
- Weak instrument
- Pleiotropy
- Residual population stratification
- Assortative mating
- Dynastic effects
- Sample selection

Methods: the UK Biobank

The UK Biobank

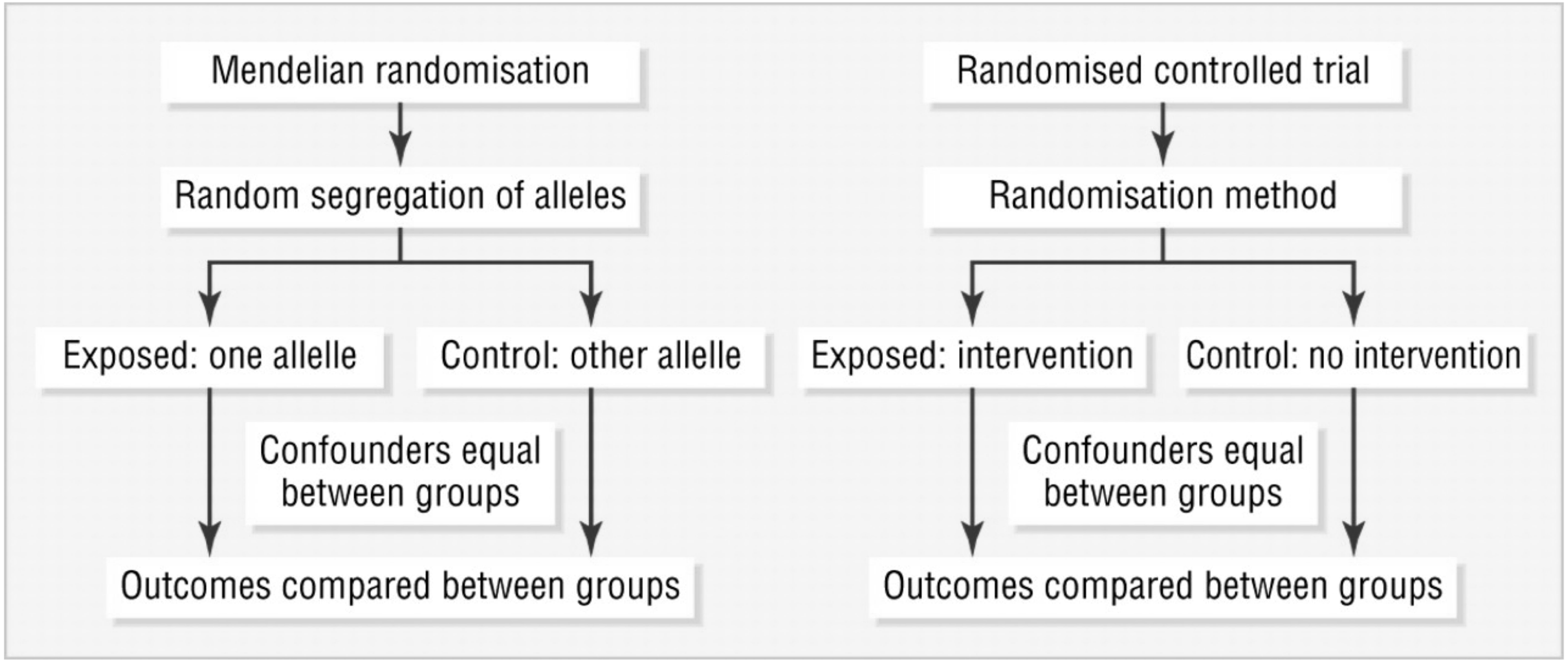
- Sent 9.2 million invitations
- Sampled 503,325 individuals
- Detailed phenotypic measurements
- Long term follow-up via record linkage
- Genome wide data

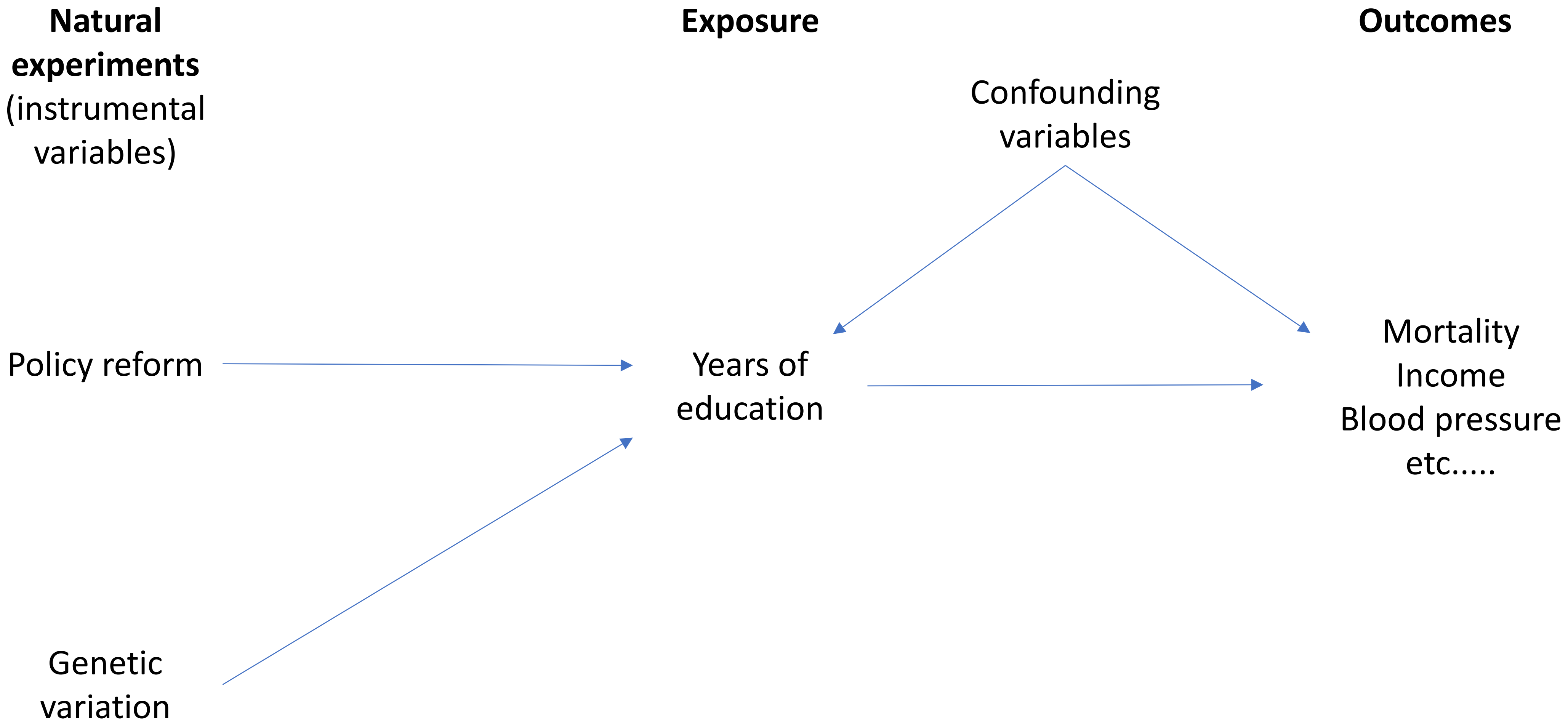
Figure 1: Raw birth location



Mendelian randomization

Randomized controlled trial





Natural experiments (instrumental variables)
Exposure
Outcomes
Confounding variables
Policy reform
Years of education
Mortality
Income
Blood pressure
etc.....
Genetic variation

Exposure: Educational attainment

- For the ROSLA: remained in school after at age 15.
  - Derived from question “what age did you leave full time education?”
  - Participants with college degrees assumed to leave after age 15.
- For Mendelian randomization
  - Derived from question “what qualifications do you have?”
  - Mapped to International Standard Classification for Education (ISCED)
  - Four levels mapped to:
    - Left age 15
    - Left age 16 (GCSE)
    - Left age 18 (A-levels)
    - Left age 20 (Post secondary vocational training)
    - Left age 21 (college degree)

25 Outcomes

- Not a “PheWAS”, but chosen a priori
- Morbidity
  - Self reported hypertension, diabetes, stroke, depressive episodes
  - Registry linked cancer diagnosis
- Mortality
  - NHS linked mortality records (7.75 years of follow-up)
- Health behaviors
  - Smoking (current and ever)
  - Alcohol consumption
  - Exercise (weekly vigorous or moderate)
  - Hours watching TV per day

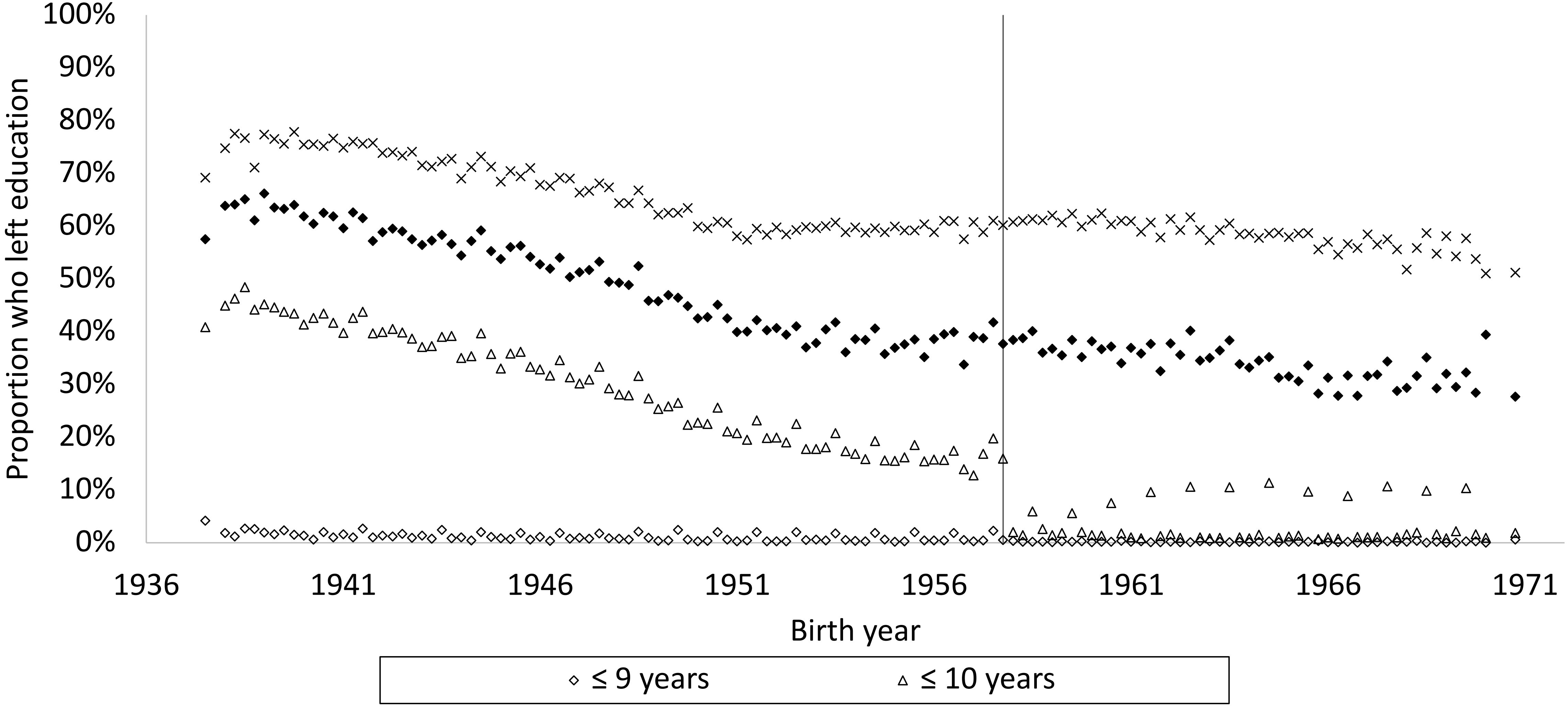
25 Outcomes (continued)

- Income
  - Household, 4 item scale
- Indicators of aging
  - Grip strength
  - Arterial stiffness
- Blood pressure
  - Systolic and diastolic
- Neurocognitive
  - Intelligence
  - Happiness

Methods: Statistical analysis 1

- <u>Evaluate the IV assumptions</u>
  1. Relevance (strength of association of IV and educational attainment)
  2. Independence assumption (no confounders of IV and educational attainment association).
  3. Exclusion restriction (IVs only affect outcome via their effect on educational attainment)
- Assumption 1 evaluated using partial F-statistics
- Assumption 2 evaluated using covariate balance tests
  - Jackson and Swanson 2016, and Pischke and Schwandt 2016.
- These assume a linear model and constant effects of the exposure
- Account for the relative strength of the instruments

Figure 2: Effect of the raising of the school leaving age on educational attainment.



Davies et al. (2018) Nature Human Behaviour. 2 (2), 117-125.

Methods: Statistical analysis 2

- Estimate effect of education
  - 2SLS for continuous outcomes
  - Additive structural mean models for binary outcomes Clarke and Windmeijer (2012)
- Baseline results include gender and month of birth, and the 10 genetic principal components as covariates
- Standard errors clustered by month of birth
- Less educated much less likely to take part
  - Census indicates that ~30% of the population left school at age 15.
  - In UK Biobank only 17.5% say they left school at age 15.
- This is accounted for by using inverse probability weights

New Results

Selection bias in instrumental variable analyses

Rachael A Hughes, Neil M Davies, George Davey Smith, Kate Tilling

doi: https://doi.org/10.1101/192237

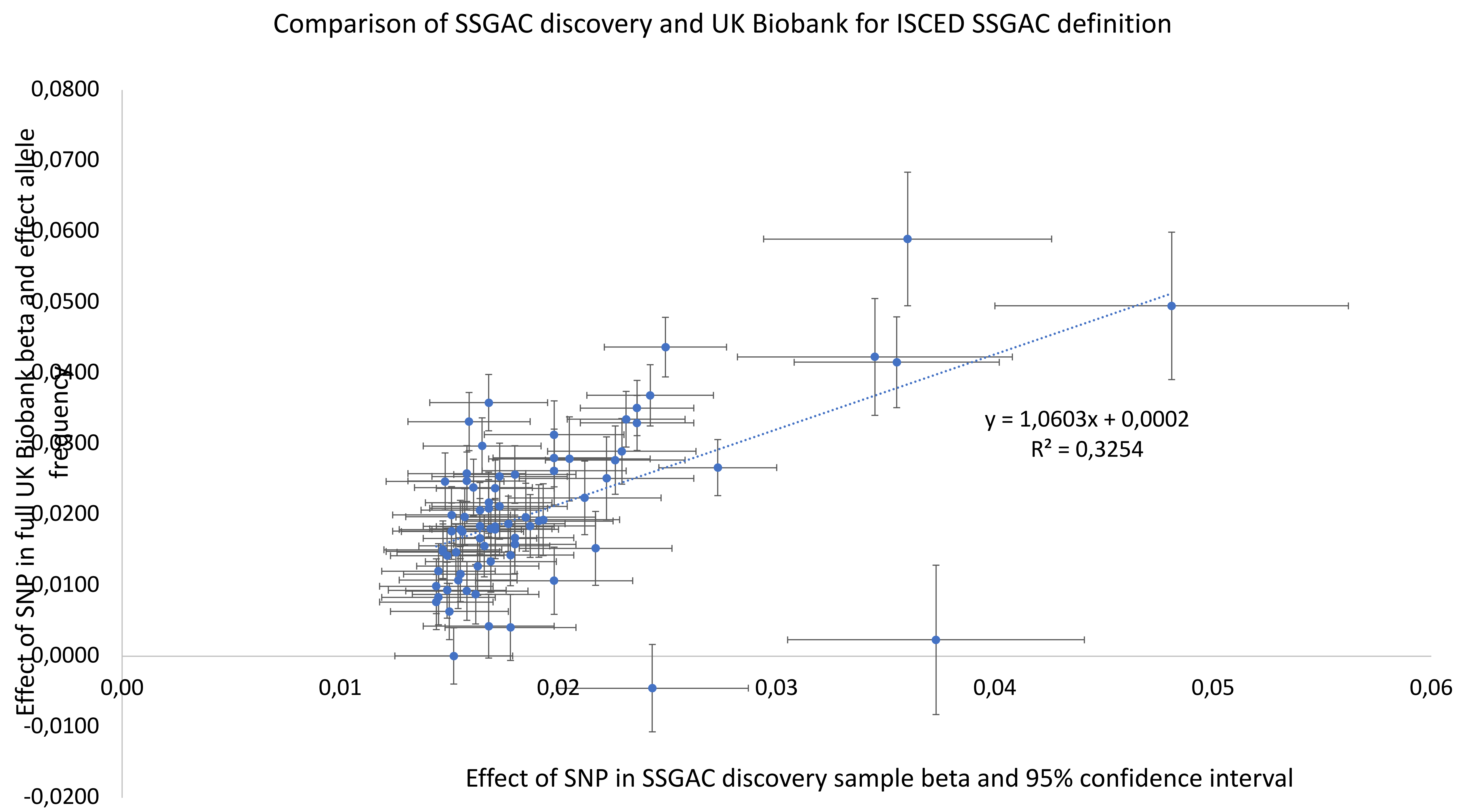
Abstract

Participants in epidemiological and genetic studies are rarely truly random samples of the populations they are intended to represent, and both known and unknown factors can influence participation in a study

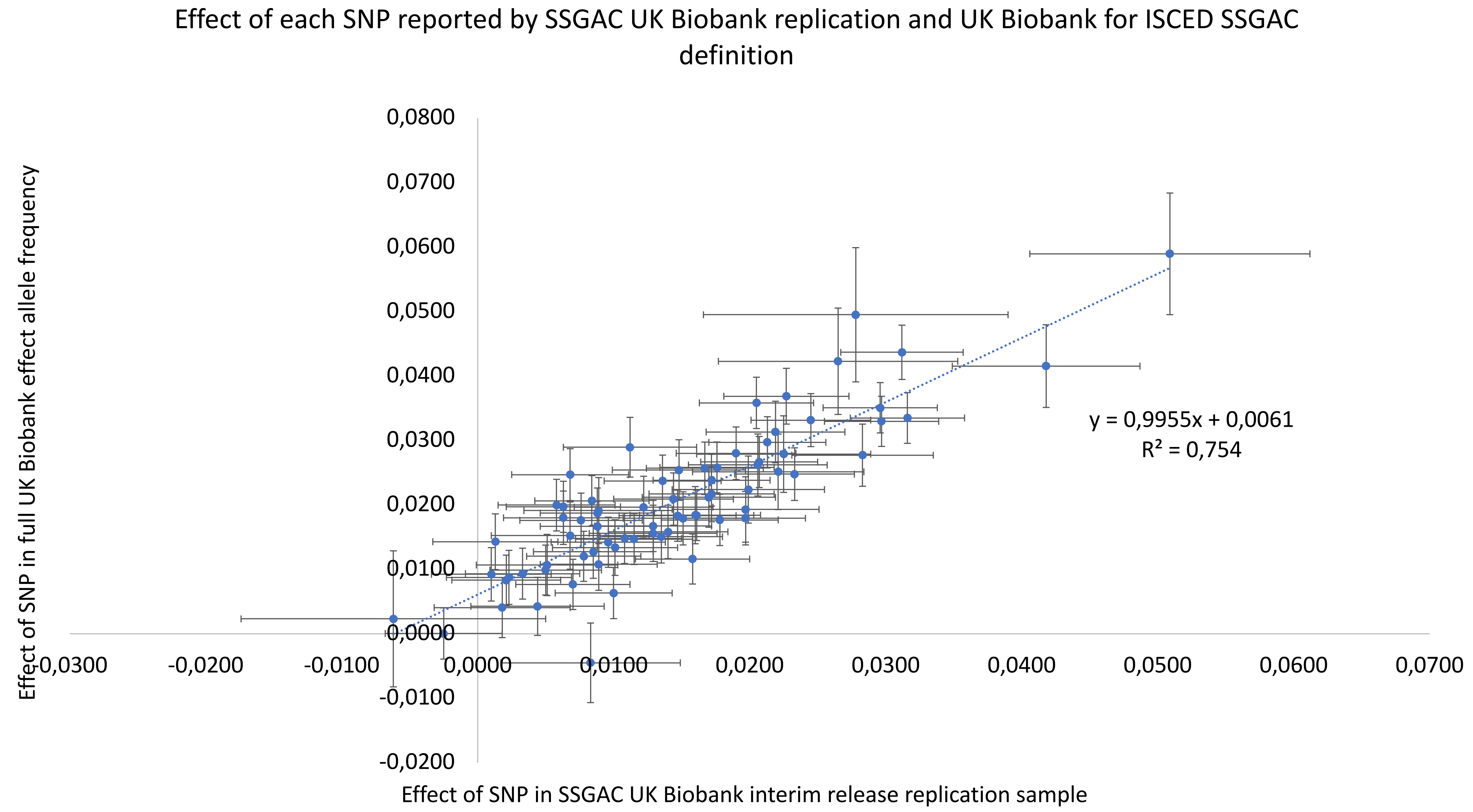
Results – association of the instruments and educational attainment

- Participants affected by the ROSLA were 23.0 (95% confidence interval (95%CI): 21.7 to 24.4) more likely to remain in school.

- Each unit increase in the Okbay score was associated with a 1.36 (95%CI) 1.29 to 1.44 additional years of education.
  - Minimum partial F-statistic= 288, maximum=1118



Comparison of SSGAC discovery and UK Biobank for ISCED SSGAC definition
Effect of SNP in full UK Biobank beta and effect allele frequency
Effect of SNP in SSGAC discovery sample beta and 95% confidence interval
y = 1,0603x + 0,0002
R² = 0,3254
0,0800
0,0700
0,0600
0,0500
0,0400
0,0300
0,0200
0,0100
0,0000
-0,0100
-0,0200
0,00
0,01
0,02
0,03
0,04
0,05
0,06

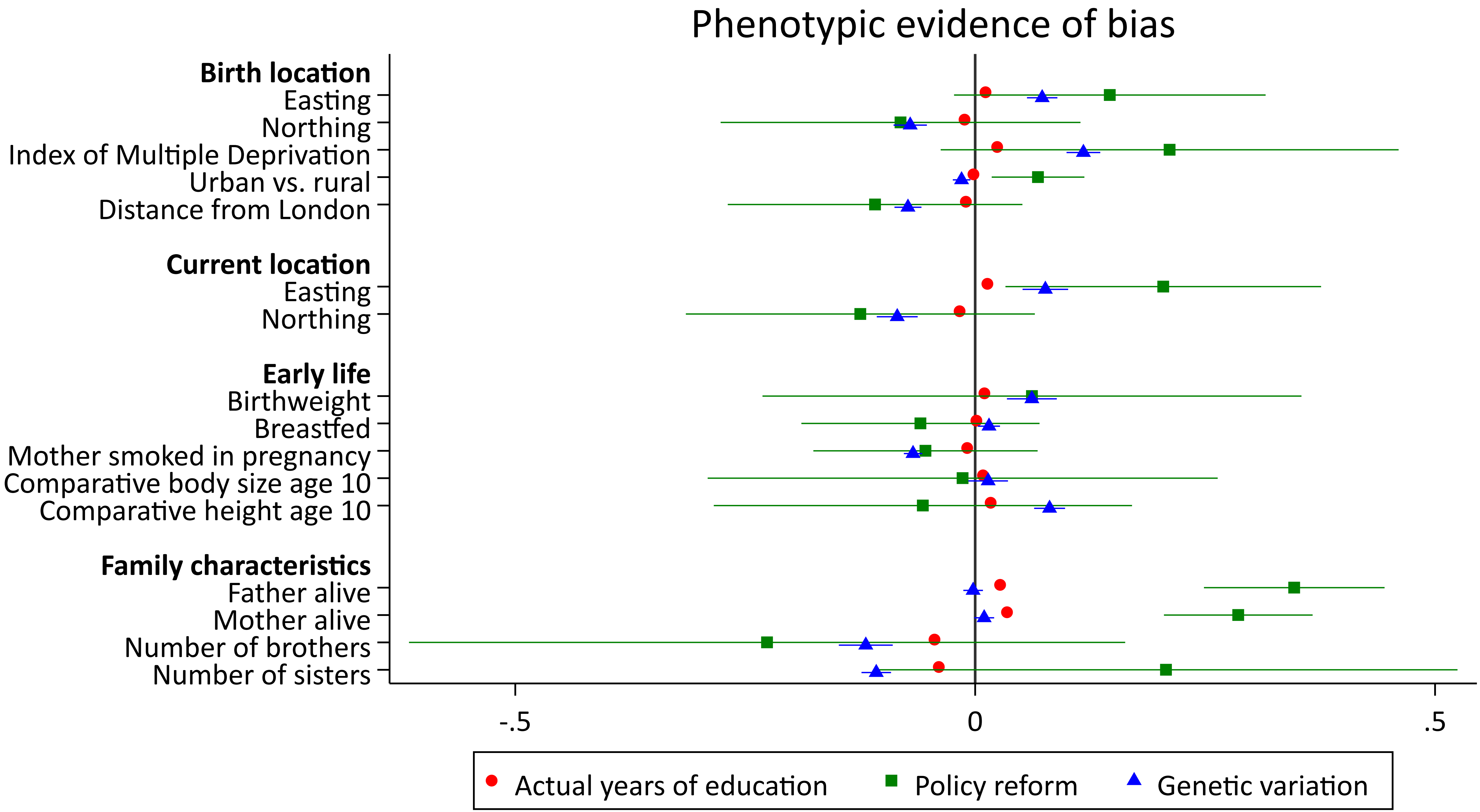


Effect of each SNP reported by SSGAC UK Biobank replication and UK Biobank for ISCED SSGAC definition
Effect of SNP in full UK Biobank effect allele frequency
Effect of SNP in SSGAC UK Biobank interim release replication sample
y = 0,9955x + 0,0061
R² = 0,754
0,0800
0,0700
0,0600
0,0500
0,0400
0,0300
0,0200
0,0100
0,0000
-0,0100
-0,0200
-0,0300
-0,0200
-0,0100
0,0000
0,0100
0,0200
0,0300
0,0400
0,0500
0,0600
0,0700

Results: validating the instruments

Bias assessment

- Phenotypic
  - Little information about pre-conception or family background
  - Place of birth, breast fed, mother smoked in pregnancy
- Genotypic
  - Constructed weighted allele scores 45 traits ($p<5e-5$) – liberal threshold
  - Excluded variants within 500kb of the 74 Okbay SNPs
  - LD pruned $r^2>0.001$
  - Harmonized alleles to UKBB



Phenotypic evidence of bias
Birth location
Easting
Northing
Index of Multiple Deprivation
Urban vs. rural
Distance from London
Current location
Easting
Northing
Early life
Birthweight
Breastfed
Mother smoked in pregnancy
Comparative body size age 10
Comparative height age 10
Family characteristics
Father alive
Mother alive
Number of brothers
Number of sisters
-.5
0
.5
Actual years of education
Policy reform
Genetic variation

New Results

Common genetic variants and health outcomes appear geographically structured in the UK Biobank sample: Old concerns returning and their implications.

Posted April 11, 2018.

Simon Haworth, Ruth Mitchell, Laura Corbin, Kaitlin H Wade, Tom Dudding, Ashley Budu-Aggrey, David Carslake, Gibran Hemani, Lavinia Paternoster, George Davey Smith, Neil Davies, Dan Lawson, Nicholas Timpson

doi: https://doi.org/10.1101/294876

Subject Area

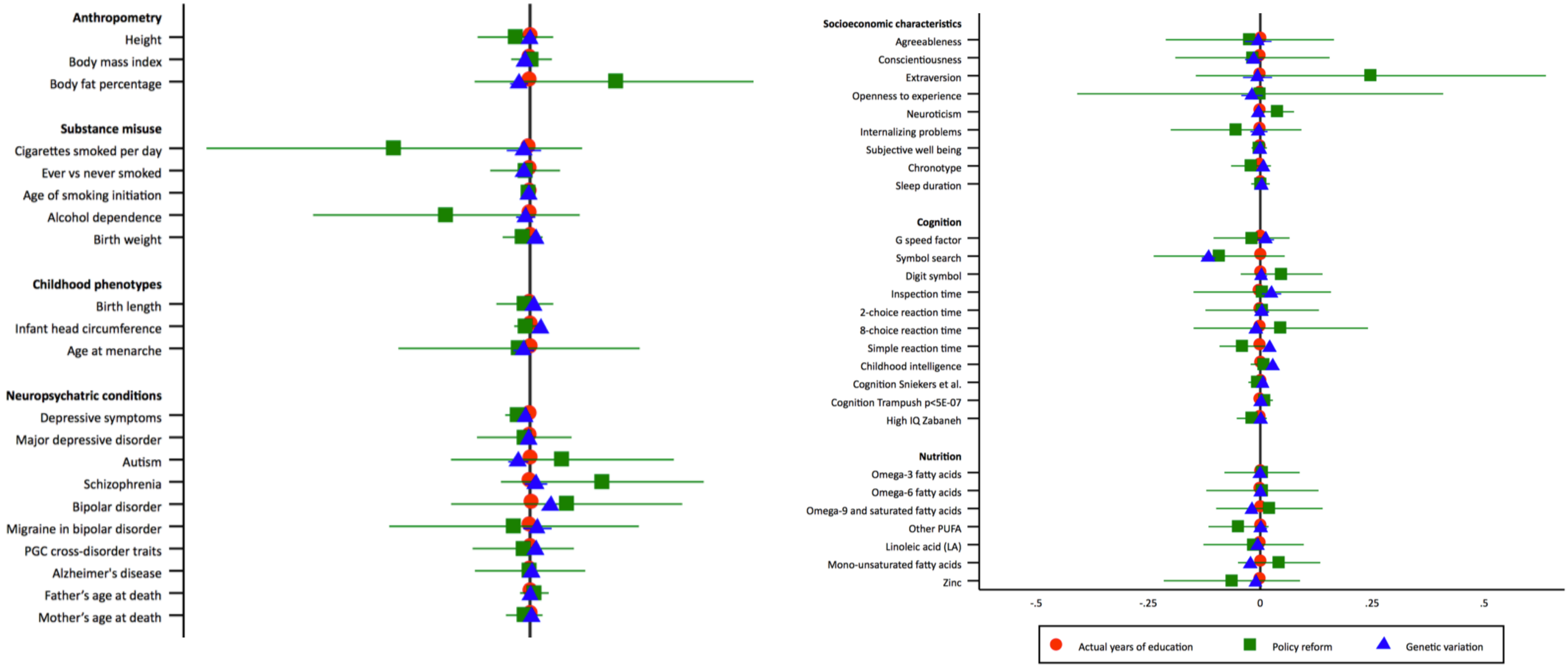
Genetics

Subject Areas

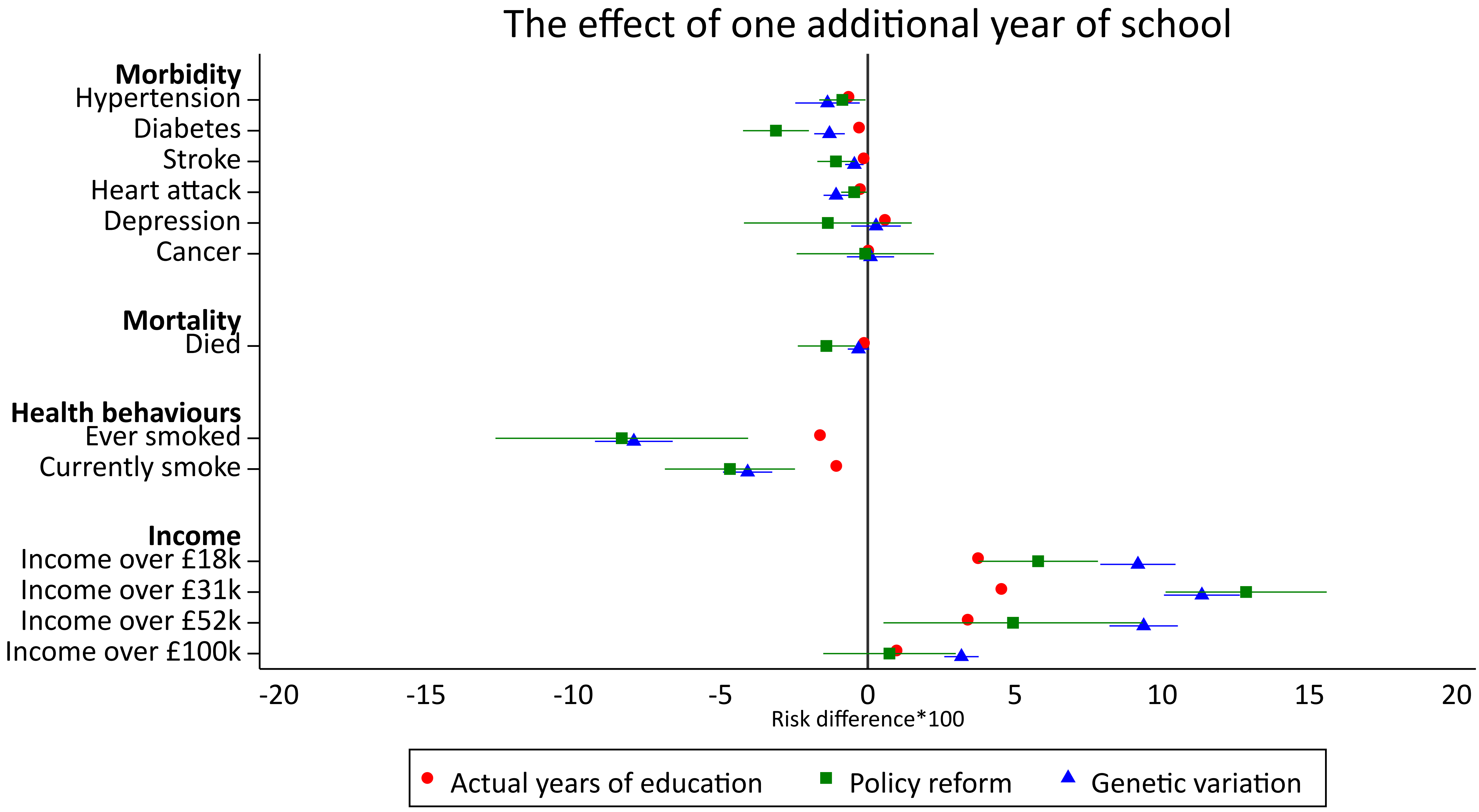
Abstract

The inclusion of genetic data in large studies has enabled the discovery of genetic contributions to complex traits and their application in applied

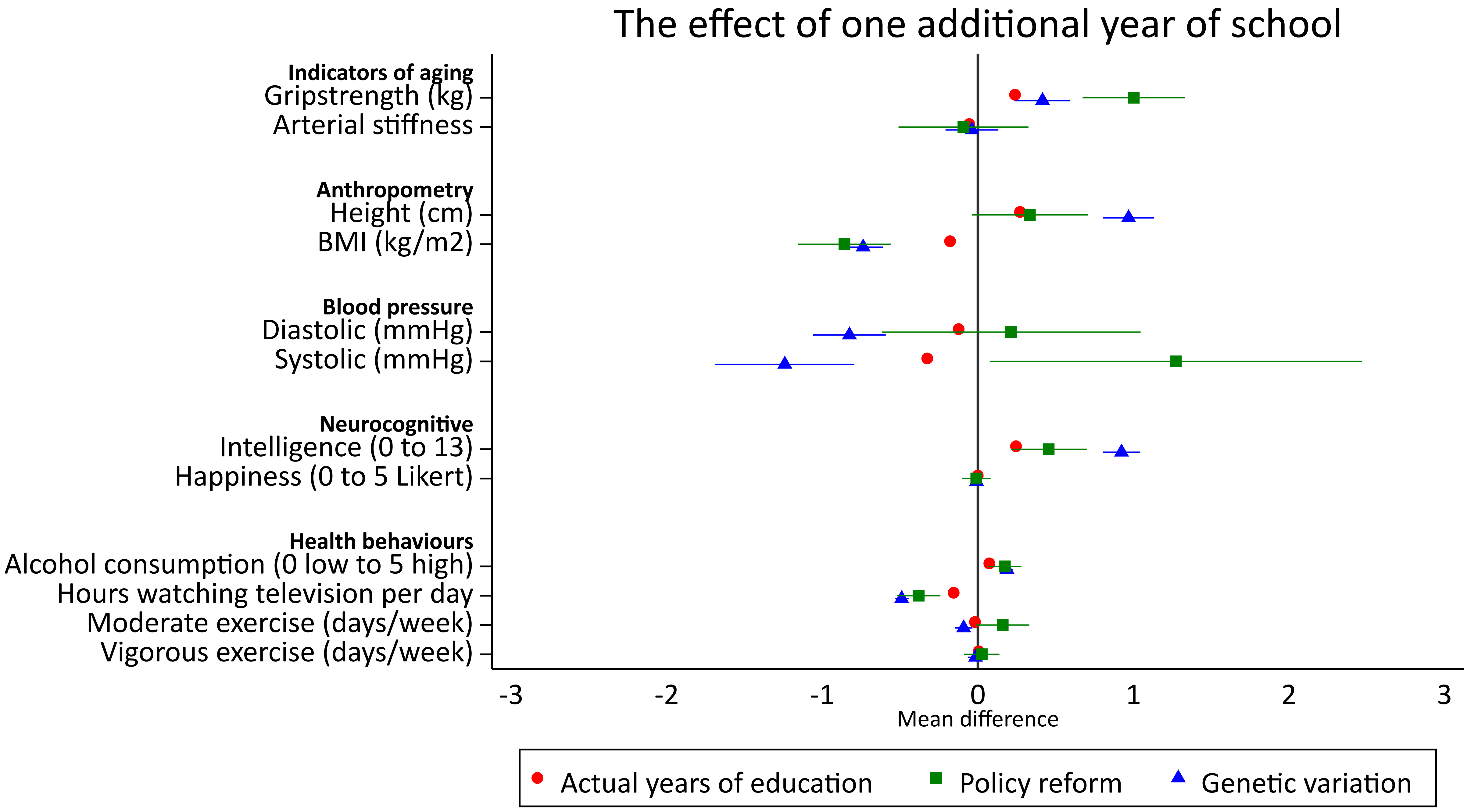
Genotypic evidence of bias



Results: The effects of education



The effect of one additional year of school
Morbidity
Hypertension
Diabetes
Stroke
Heart attack
Depression
Cancer
Mortality
Died
Health behaviours
Ever smoked
Currently smoke
Income
Income over £18k
Income over £31k
Income over £52k
Income over £100k
-20
-15
-10
-5
0
5
10
15
20
Risk difference*100
Actual years of education
Policy reform
Genetic variation



The effect of one additional year of school
Indicators of aging
Gripstrength (kg)
Arterial stiffness
Anthropometry
Height (cm)
BMI (kg/m2)
Blood pressure
Diastolic (mmHg)
Systolic (mmHg)
Neurocognitive
Intelligence (0 to 13)
Happiness (0 to 5 Likert)
Health behaviours
Alcohol consumption (0 low to 5 high)
Hours watching television per day
Moderate exercise (days/week)
Vigorous exercise (days/week)
-3
-2
-1
0
1
2
3
Mean difference
Actual years of education
Policy reform
Genetic variation

Figure 6a: Univariate MR effects of education and cognition on outcomes

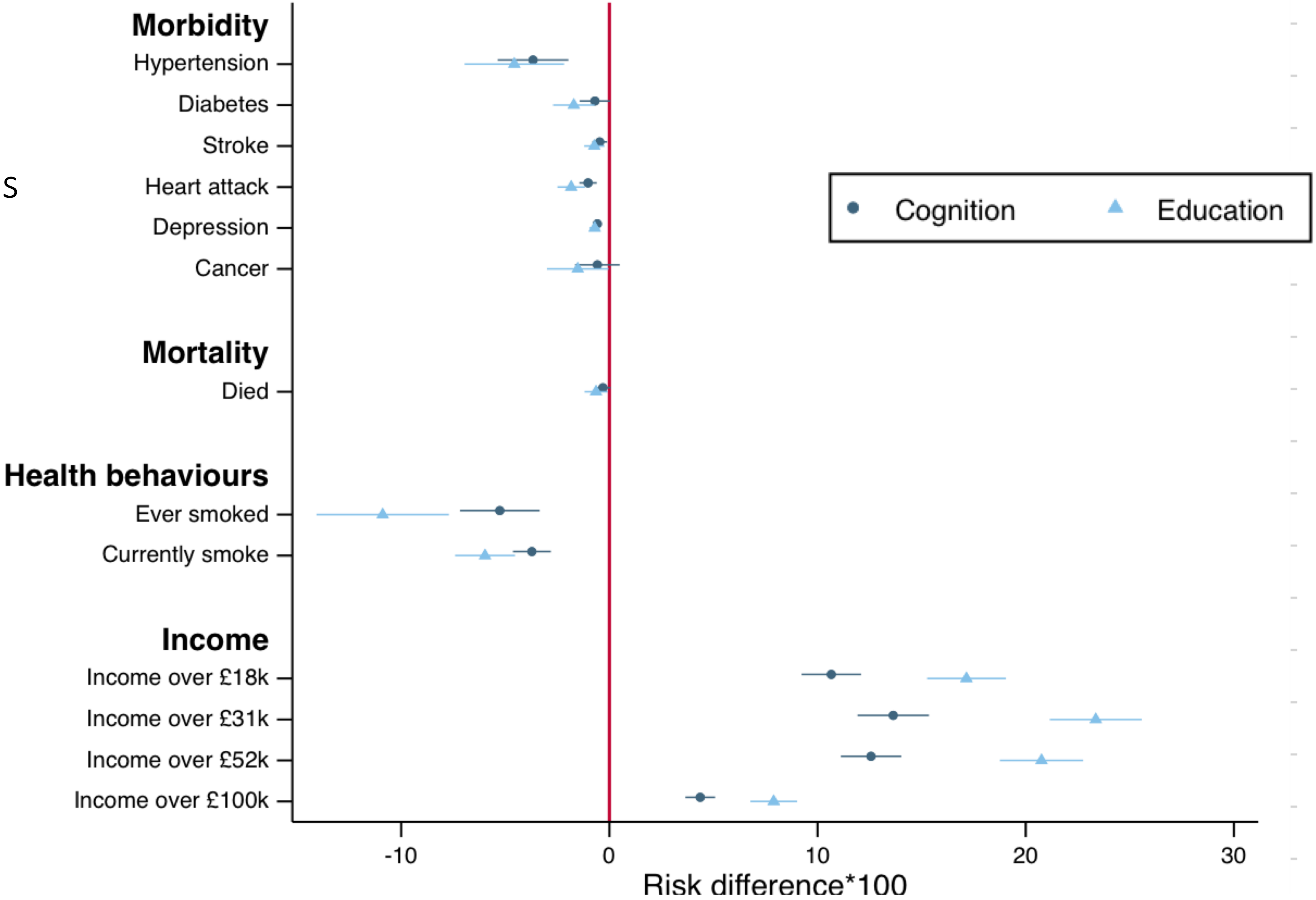


Figure 6b: **Bivariate MR** effects of education and cognition on outcomes

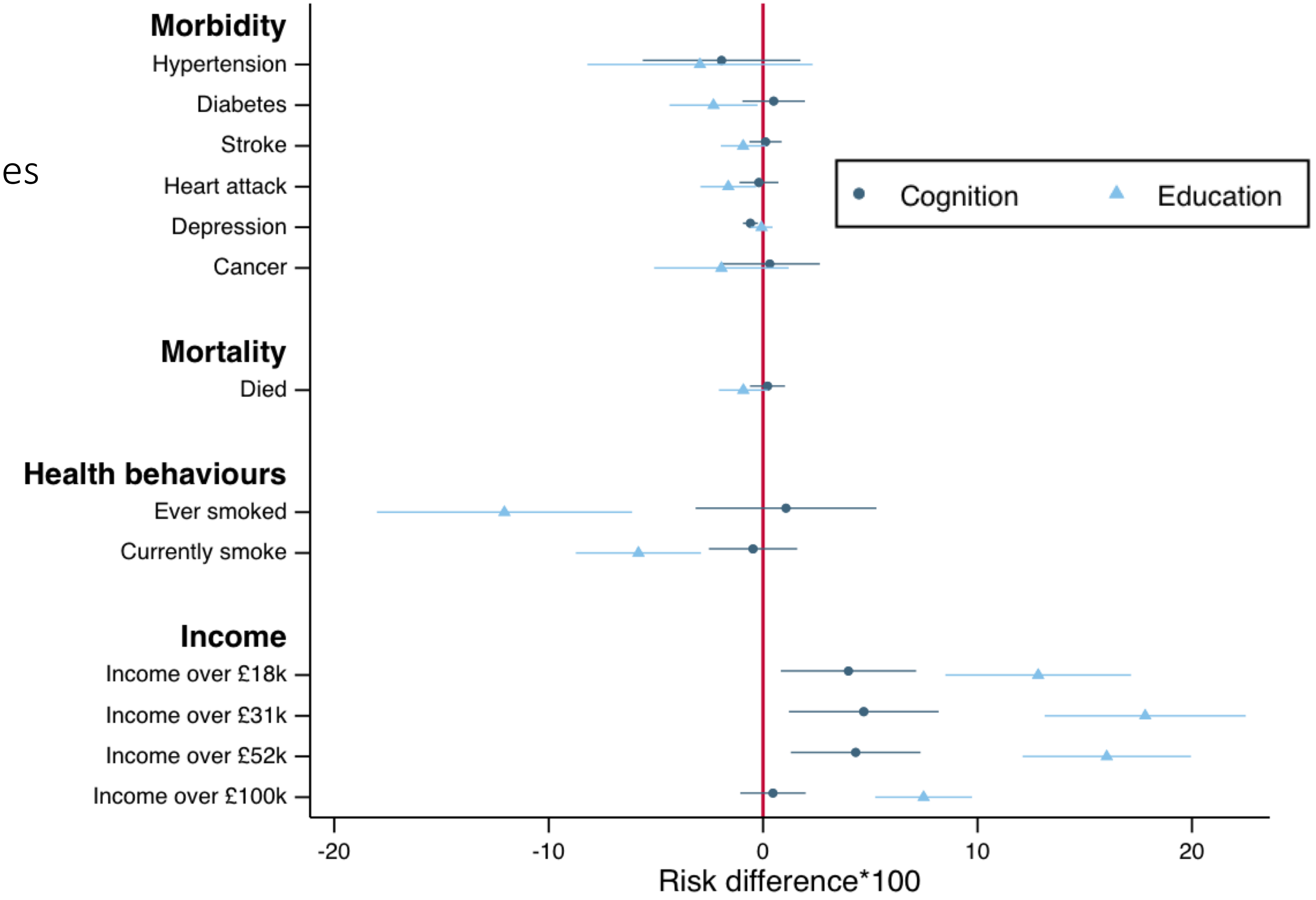


Figure 7a: <u>Univariate MR</u> effect of education and cognition on outcomes

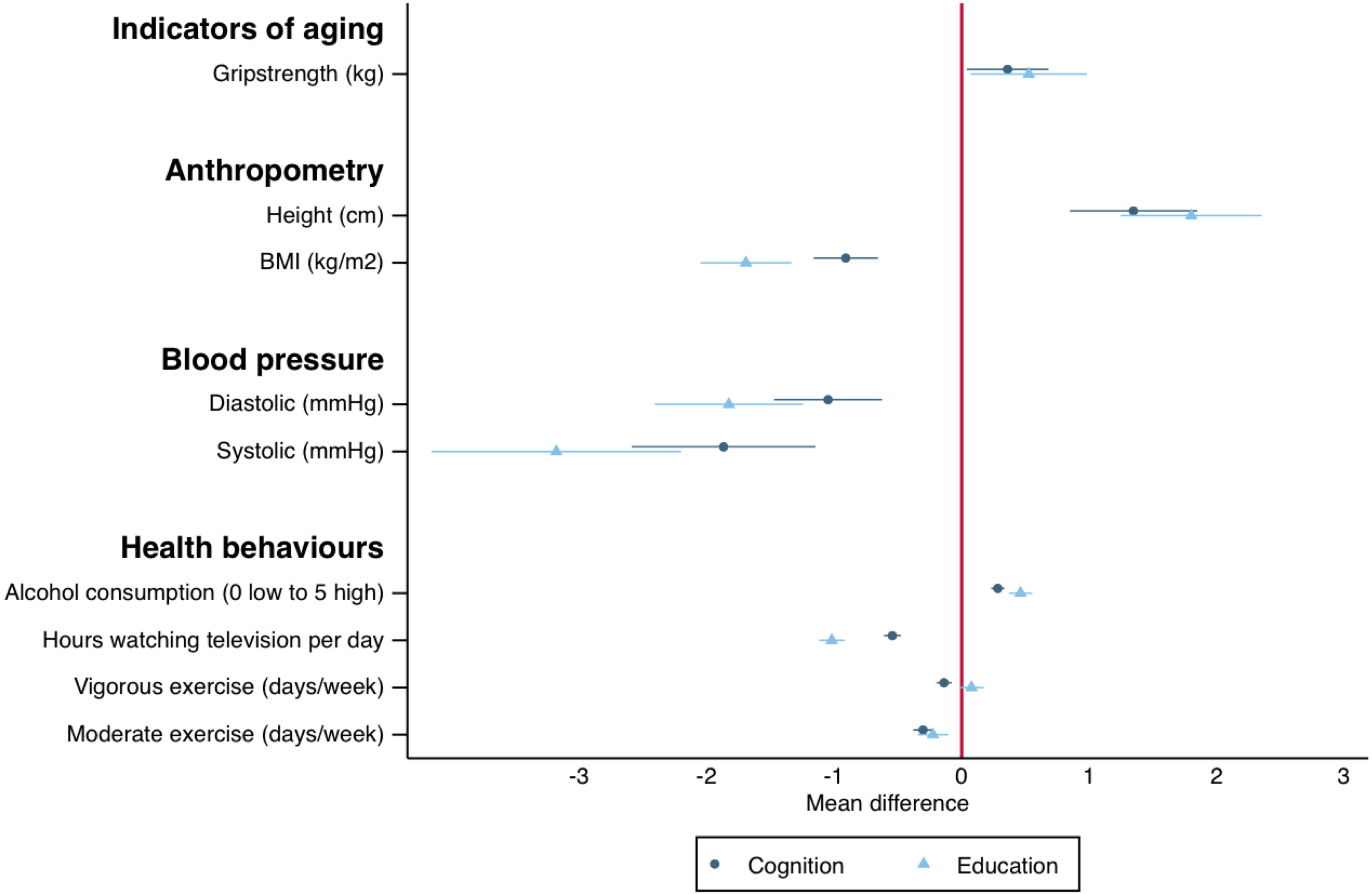
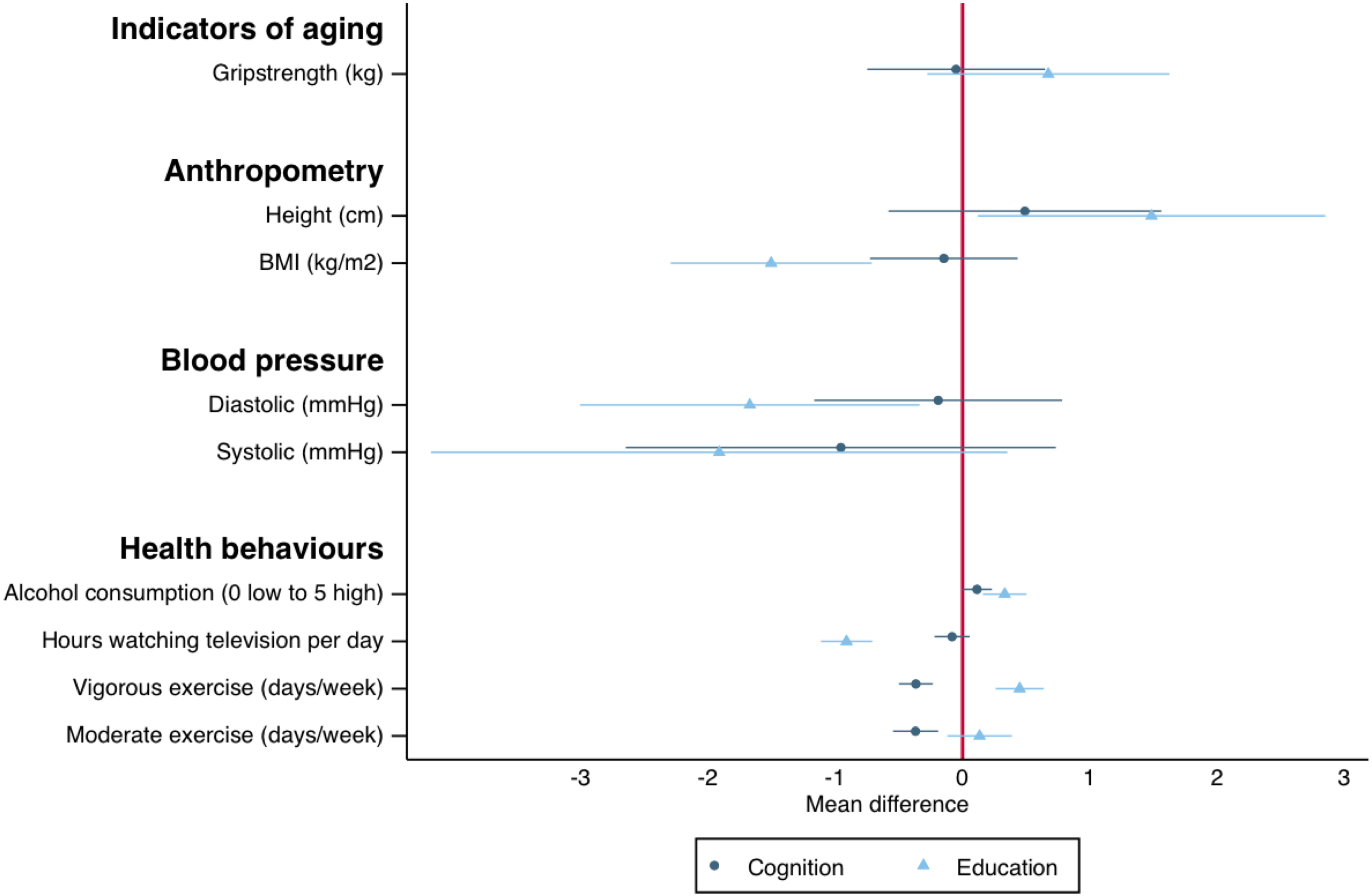


Figure 7b: <u>Bivariate MR</u> effect of education and cognition on outcomes



Potential sources of bias

- Assortative mating
- Dynastic effects (genetic nurture)
- Pleiotropy
  - Horizontal (a problem)
  - Vertical (not a problem)
- Population stratification

Science Home News Journals Topics Careers

RESEARCH ARTICLE

The nature of nurture: Effects of parental genotypes

Augustine Kong[1,2,3,*], Gudmar Thorleifsson[1], Michael L. Frigge[1], Bjarni J. Vilhjalmsson[4,5], Alexander I. Yo...
+ See all authors and affiliations

Science 26 Jan 2018:
Vol. 359, Issue 6374, pp. 424-428
DOI: 10.1126/science.aan6877

Article | Figures & Data | Info & Metrics | eLetters | PDF

Sensitivity analyses

- No time, but presented in the paper
- Pleiotropy robust
  - MR-Egger, weighted median, and mode
- Weighted vs. unweighted
- Not adjusted, partially adjusted, vs. fully adjusted
- Reduced form

- All consistent with these results, no major changes

Future research:

- Cross country natural experiments
- Long term follow-up
- Bivariate Mendelian randomization
- Effect heterogeneity
- Randomized controlled trials

Questions:

- What are the confounders of education-health associations?
- How to model non-linear or clinical thresholds?

Conclusions

- Strength and balance of confounders similar for MR as ROSLA
- Limitations of MR
  - Horizontal pleiotropy
  - Assortative mating
  - Dynastic effects
- Limitations of both
  - Sample selection
- Can either explain the results?
- Results remarkably consistent for MR and ROSLA

References

- Nguyen TT, Tchetgen EJT, Kawachi I, Gilman SE, Walter S, Liu SY, et al. Instrumental variable approaches to identifying the causal effect of educational attainment on dementia risk. Annals of Epidemiology. 2016 Jan;26(1):71–76.e3.
- Hagenaars SP, Harris SE, Davies G, Hill WD, Liewald DC, Ritchie SJ, et al. Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N = 112 151) and 24 GWAS consortia. [Internet]. 2015 Nov [cited 2015 Nov 11]. Report No.: 31120. Available from: http://biorxiv.org/lookup/doi/10.1101/031120
- Harmon C, Walker I. Estimates of the Economic Return to Schooling for the United Kingdom. The American Economic Review. 1995 Dec 1;85(5):1278–86.
- Clark D, Royer H. The Effect of Education on Adult Mortality and Health: Evidence from Britain. American Economic Review. 2013 Oct;103(6):2087–120.
- Dickson M. The Causal Effect of Education on Wages Revisited*: *The causal effect of education*. Oxford Bulletin of Economics and Statistics. 2013 Aug;75(4):477–98.
- Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature. 2016 May 11;533(7604):539–42.
- Jackson JW, Swanson SA. Toward a Clearer Portrayal of Confounding Bias in Instrumental Variable Applications. Epidemiology. 2015 Jul;26(4):498–504.
- Pischke J-S, Schwandt H. Poorly Measured Confounders are More Useful on the Left Than on the Right [Internet]. Working Paper, London School of Economics; 2016 [cited 2015 Apr 9]. Available from: http://econ.lse.ac.uk/staff/spischke/ec533/C_var_note.pdf

Thanks for listening! Thoughts comments?

- Acknowledgements:
- George Davey Smith
- Frank Windmeijer
- Matt Dickson
- Gerard van den Berg
- Gib Hemani
- Aysu Okbay
- SSGAC

- The Medical Research Council and the University of Bristol fund the MRC Integrative Epidemiology Unit [MC_UU_12013/1, MC_UU_12013/9].
- NMD is supported by the Economics and Social Research Council (ESRC) via a Future Research Leaders Fellowship [ES/N000757/1].

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

MRC Integrative Epidemiology Unit