bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Cognitive ability and physical health: a Mendelian randomization study

Saskia P Hagenaars[1,2,3], Catharine R Gale[1,2,4], Ian J Deary[1,2] and Sarah E Harris[1,5]

1 Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK

2 Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ, UK

3 Division of Psychiatry, University of Edinburgh, Edinburgh, EH10 5HF, UK

4 MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK.

5 Medical Genetics Section, University of Edinburgh Centre for Genomic and Experimental Medicine and MRC Institute of Genetics and Molecular Medicine, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

Corresponding author: Sarah E Harris, Medical Genetics Section, University of Edinburgh Centre for Genomic and Experimental Medicine and MRC Institute of Genetics and Molecular Medicine, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK, Email: sarah.harris@igmm.ed.ac.uk

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Background

Causes of the association between lower cognitive ability and poorer health remain unknown, but may reflect a shared genetic aetiology as indicated by previous research. This study examines the causal genetic associations between cognitive ability and physical health outcomes.

Method

We carried out Mendelian randomization analyses using the inverse variance weighted method to test for causality between later life cognitive ability, educational attainment (as a proxy for cognitive ability in youth), BMI, height, systolic blood pressure, coronary artery disease, and type 2 diabetes in the UK Biobank sample (N = 112 151). Sensitivity analyses were performed using MR-Egger regression.

Results

BMI, systolic blood pressure, coronary artery disease and type 2 diabetes showed negative associations with cognitive ability, while height was positively associated with cognitive ability. The Mendelian randomization analyses provided no evidence for a casual association from health to cognitive ability. In the other direction, higher educational attainment predicted lower BMI, systolic blood pressure, coronary artery disease, type 2 diabetes, and taller stature. The Mendelian randomization analyses indicated partly causal associations from educational attainment to health, however when adjusting for bias using the MR-Egger regression, these effects disappeared.

Conclusions

The lack of consistent evidence for causal associations between cognitive ability, educational attainment, and physical health could be explained by violations of the Mendelian randomization assumptions, including biological pleiotropy.

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Key messages

- Cognitive ability and physical health outcomes are positively associated.
- Mendelian randomization analyses indicated that educational attainment influenced physical health outcomes.
- Sensitivity analyses, using MR-Egger regression, indicated that these associations were biased due to violations of the Mendelian randomization assumptions.

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Lower cognitive ability, lower educational attainment and greater cognitive decline are all associated with poorer health outcomes[1-3]. Some of these associations possibly arise because of the effect of lower cognitive ability in childhood on later life health, others because illnesses may lower cognitive ability in later life. The causes of these associations are unclear, but may reflect a shared genetic aetiology. Recent papers have reported genetic associations between cognitive ability and educational attainment, and a number of physical and mental health traits and diseases[4-6]. These[4, 6], and other[7-9], papers have shown successful use of educational attainment as a proxy for cognitive ability, showing phenotypic correlations between educational attainment and general cognitive ability around 0.50[9] and a genetic correlation of 0.72[4].

Some of the reciprocal phenotypic associations between cognitive and physical health variables, and their genetic correlations, are as follows. Short stature has been consistently linked with lower cognitive ability[10, 11]. Molecular genetic studies have indicated positive genetic correlations between height and cognitive ability[4, 12], as well as between height and education attainment[4, 5]. Higher polygenic scores for height have been associated with better cognitive ability in adulthood[4].

Multiple studies have shown associations between cognitive ability and cardiovascular risk factors. For example, lower childhood cognitive ability is associated with subsequent high blood pressure[13] and obesity[14]. However, higher BMI in mid-life[15] and both hypertension and hypotension[16] are associated with lower cognitive ability and greater cognitive decline in later life. A negative genetic correlation has been identified between BMI, but not blood pressure, and educational attainment and cognitive ability in mid to late life[4, 5] and a polygenic score

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

for higher BMI is associated with lower cognitive ability in mid to late life and lower educational attainment[4], whereas a polygenic score for higher systolic blood pressure is associated with lower educational attainment, but higher cognitive ability in mid to late life[4].

Similarly, associations have been identified between cognitive ability and cardio metabolic diseases. Childhood cognitive ability has been associated with developing diabetes[17] and coronary artery disease[18] later in life. Diabetes[19] and coronary artery disease[20, 21] in midlife have been associated with greater cognitive decline later in life. A polygenic risk score for type 2 diabetes is associated with lower educational attainment, but not with cognitive ability in mid to late life[4], although one has been associated with reduced cognitive decline[22]. To date no significant genetic correlation between diabetes and cognitive ability has been identified[4, 5]. A polygenic risk score for coronary artery disease is associated with lower educational attainment and lower mid to late life cognitive ability[4], and a negative genetic correlation was identified between coronary artery disease and educational attainment[4, 5], but not cognitive ability in mid to late life[4].

The question arises, are the genetic associations caused by: 1) genes influencing health traits/diseases, and then those health traits/diseases subsequently influencing cognitive ability; 2) genes influencing cognitive ability, and then cognitive ability subsequently influencing health traits/diseases; 3) genes influencing general bodily system integrity[23] that influences both cognitive ability and health traits/diseases?

To try to make some progress in understanding causality of the correlation between cognitive ability and a number of physical and mental health traits we used a bi-directional Mendelian randomization (MR) with Egger regression approach[24]. MR uses genetic variants as proxies

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

for environmental exposures and is subject to the following assumptions: 1) the genetic variants are associated with the exposure; 2) the genetic variants are only associated with the outcome of interest via their effect on the exposure (i.e., there is no biological pleiotropy, also called the exclusion restriction); and 3) the genetic variants are independent of confounders. Figure 1 shows the Mendelian randomization study model; the instrumental variable, based on genome-wide significant SNPs from independent studies for the exposure, is used to estimate if the exposure (e.g. BMI) causally influences the outcome (e.g. cognitive ability). Individual single nucleotide polymorphisms (SNPs) are often found to be weak instruments for investigating causality because they often have small effect sizes. Using multiple SNPs can increase the strength of the instrument. However, this increases the chance of violating the MR assumptions, specifically violation of the assumption that the genetic variant affects the outcome via a different pathway than via the exposure. In the present study, Egger regression was used as a sensitivity analysis to test for such violations. We used multiple genetic variants for a number of health-related traits and diseases, previously identified in a genome-wide association study, as instrumental variables to see if they predicted cognitive ability (verbal-numerical reasoning) in mid to later life in the UK Biobank. We then used genome-wide significant educational attainment SNPs as an instrumental variable to test whether genetic differences associated with educational attainment (a proxy measure of cognitive ability in early life[6, 8]) predicts later life health outcomes in the UK Biobank.

Methods

Sample

This study uses baseline data from the UK Biobank Study, a large resource for identifying determinants of human diseases in middle aged and older individuals[25]. UK Biobank received ethical approval from the Research Ethics Committee (reference 11/NW/0382). This study

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

has been completed under UK Biobank application 10279. Around 500 000 community-dwelling participants aged between 37 and 73 years were recruited and underwent assessments between 2006 and 2010 in the United Kingdom. This included cognitive and physical assessments, providing blood, urine and saliva samples for future analysis, and giving detailed information about their backgrounds and lifestyles, and agreeing to have their health followed longitudinally. For the present study, genome-wide genotyping data were available on 112 151 individuals (58 914 females) aged 40–70 years (mean age=56.9 years, s.d.=7.9) after the quality control process which is described in more detail elsewhere[4].

Measures

Body mass index

Body mass index (BMI) was calculated as weight(kg)/height(m)^2, and measured using an impedance measure, i.e. a Tanita BC418MA body composition analyser, to estimate body composition. We used the average of the two methods when both measures were available (r = 0.99); if only one measure was available, that measure was used (N = 1629). 291 individuals did not have information on BMI. One outlier was excluded based on visual inspection of the BMI distribution (BMI > 50). 111 712 individuals had valid BMI and genetic data.

Height

Standing and sitting height (cm) were measured using a Seca 202 device. We used standing height and excluded one individual based on the visual inspection of the height distribution with a standing height < 125 cm and a sitting/standing height ratio < 0.75. 111 959 had valid height and genetic data.

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Systolic blood pressure

Systolic blood pressure was measured twice, a few moments apart, using the Omron Digital blood pressure monitor. A manual sphygmomanometer was used if the digital blood pressure monitor could not be employed (N = 6652). Systolic blood pressure was calculated as the average of measures at the two time points (for either automated or manual readings). Individuals with a history of coronary artery disease were excluded from the analysis (N = 2513). Following the recommendation by Tobin, Sheehan et al[26], 15 mmHg was added to the average systolic blood pressure of individuals taking antihypertensive medication (N = 10 988). Individuals with a systolic blood pressure (after correcting for medication) more than 4 SD from the mean were excluded from future analyses (N = 75). After all exclusions, 106 759 individuals remained with valid blood pressure and genetic data.

Coronary artery disease

UK Biobank participants completed a touch screen questionnaire on past and current health, which included the question "Has a doctor ever told you that you have had any of the following conditions? heart attack/angina/stroke/high blood pressure/none of the above/prefer not to answer". This was followed by a verbal interview with a trained nurse who was made aware if the participant had a history of certain illnesses and confirmed these diagnoses with the participant. For the present study, coronary artery disease was defined as a diagnosis of myocardial infarct or angina, reported during both the touchscreen and the verbal interview (N = 5288). The control group (N = 104 784) consisted of participants who reported none of the following diseases (based on the non-cancer illness code provided by UK Biobank): myocardial infarction, angina, heart failure, cerebrovascular disease, stroke, transient ischaemic attack, subdural haemorrhage, cerebral aneurysm, peripheral vascular disease, leg claudication/intermittent claudication, arterial embolism.

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Type 2 diabetes

Type 2 diabetes case-control status was created using the same method as described by Wood et al[27]. Cases included participants who reported type 2 diabetes or generic diabetes during the nurse interview, started insulin treatment at least one year after diagnosis, were older than 35 years at the time of diagnosis, and did not receive a diagnosis one year prior to baseline testing (N = 3764). The control group consisted of participants who did not fulfil these criteria, and did not report a diagnosis of type 1 diabetes, diabetes insipidus and gestational diabetes (N = 108 015).

Years of education

As part of the sociodemographic questionnaire in the study, participants were asked, "Which of the following qualifications do you have? (You can select more than one)". Possible answers were: "College or University Degree/A levels or AS levels or equivalent/O levels or GCSE or equivalent/CSEs or equivalent/NVQ or HND or HNC or equivalent/Other professional qualifications e.g. nursing, teaching/None of the above/Prefer not to answer". For the present study, a new continuous variable was created measuring 'years of education completed'. This was based on the ISCED coding, using the 1997 International Standard Classification of Education (ISCED) of the United Nations Educational, Scientific and Cultural Organization[28]. See the Table 1 for further details. For the current study, years of education was used a proxy phenotype for cognitive ability[4, 6, 8]. A total of 111 114 individuals had valid data for the years of education variable.

Cognitive ability

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Cognitive ability was measured using a 13-item touchscreen computerized verbal-numerical reasoning test. The test included six verbal and seven numerical questions, all with multiple-choice answers, with a two-minute time limit. An example verbal item is: 'Stop means the same as?' (possible answers: 'Pause/Close/Cease/Break/Rest/do not know/prefer not to answer'). An example numerical item is: Which number is the largest' (possible answers: '642/308/987/714/253/do not know/prefer not to answer'). The cognitive ability score was the total score out of 13 (further detail can be found in Hagenaars et al.[4]). A total of 36 035 had valid cognitive ability and genetic data.

Covariates

All analyses were adjusted for the following covariates: age when attending assessment centre, sex, genetic batch and array, and the first ten genetic principal components for population stratification.

Instrumental variables

SNPs used in the instrumental variables were extracted from the imputed UK Biobank genotypes interim release including 112 151 individuals after quality control. Details on the quality control process have been published previously[4]. All instrumental variables were created based on SNPs that reached genome-wide significance in the largest available GWAS in European samples for the variables of interest (BMI[29], height[27], systolic blood pressure[30], coronary artery disease[31], type 2 diabetes[32] and educational attainment[33]). SNPs out of Hardy-Weinberg equilibrium (HWE, $p < 1\times10^{-6}$), with an imputation quality below 0.9, or individual genotypes with a genotype probability below 0.9 were excluded from the instrumental variables. The individual variants were recoded as 0, 1 or 2 according to the number of trait increasing alleles. Table 1 includes information on the number of SNPs

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

included, the reference paper, and the amount of variance explained by the instrumental variables for the corresponding variable of interest. Supplementary Table 1a-f provides details of the included SNPs.

Statistical analysis

Phenotypic associations

We performed linear regression analysis using BMI, height, systolic blood pressure, coronary artery disease, and type 2 diabetes to predict cognitive ability. We regressed BMI, height, and systolic blood pressure against educational attainment in a linear regression model; coronary artery disease and type 2 diabetes were regressed against educational attainment in logistic regression models.

Mendelian randomization analysis

The Mendelian randomization analysis was performed using inverse variance weighted regression analysis based on SNP level data, with each instrumental variable (IV) consisting of multiple SNPs[24]. The inverse variance weighted method is based on a regression of two vectors with the intercept constrained to zero; the genetic variant with the exposure association, and the genetic variant with the outcome association (Figure 1). By constraining the intercept to zero, this method assumes that all variants are valid instrumental variables. We performed an association analysis between each SNP in the instrumental variable for the exposure and the exposure itself (IV - exposure), as well as between the instrumental variable for the exposure and the outcome (IV - outcome). We then used the vector of the instrumental variable-outcome association analyses against the vector of the instrumental variable-exposure analyses. This association (vector IV - outcome ~ vector IV - exposure) was weighted by the standard error of the original IV-outcome association, to correct for minor

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

allele frequency, as described by Bowden et al[24]. If the inverse variance weighted method indicated a significant effect, we used MR-Egger regression as a sensitivity analysis to test for violations of the instrumental variable assumptions[24]. This method is similar to the inverse variance weighted method, but uses an unconstrained intercept. The MR-Egger regression will give a bias-adjusted estimate, as it uses a weaker version of the exclusion restriction, by assuming that the associations of the genetic variants with the exposure, and the direct effect of the genetic variants on the outcome are independent of each other. This tests for directional pleiotropy, as the instrumental variable is not constrained by the instrumental variable assumptions.

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

Health outcomes predicting cognitive ability

BMI, height, systolic blood pressure, and coronary artery disease predicted performance on the verbal-numerical reasoning test of cognitive ability (Table 3). A 1 SD higher BMI was associated with a 0.05 SD lower score for cognitive ability. A 1 SD greater height was associated with a 0.18 SD higher score for cognitive ability. A 1 SD higher systolic blood pressure was associated with a 0.05 SD lower score for cognitive ability. Individuals with coronary artery disease had, on average, a 0.27 SD lower score for cognitive ability. Individuals with type 2 diabetes had, on average, a 0.06 SD lower score for cognitive ability. The inverse variance weighted analyses, with the five health outcomes as the exposures, and cognitive ability as the outcome, did not provide any genetic causal evidence for any of these associations.

Education predicting health outcomes

Educational attainment, as measured by years of education, predicted BMI, height, systolic blood pressure, type 2 diabetes and coronary artery disease (Table 4, Figure 2). In every case, the inverse variance weighted method showed a causal effect of educational attainment on the health outcomes. However, the bias adjusted effect from the MR Egger regression indicated that the association between educational attainment and the five health outcomes was influenced by a violation of the MR assumptions. This suggested that there is no genetic causal association from educational attainment to health outcomes. The full results can be found in Table 4 and Figure 2.

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

This study was designed to investigate causes of the well replicated finding that lower cognitive ability is associated with poorer health outcomes[1-3]. It used a bidirectional MR approach to investigate this. We found no consistent evidence for causal association between cognitive ability, in middle and older age, and several health outcomes. Inverse weighted MR analysis indicated that number of years of education (a proxy measure of cognitive ability in early life[4, 8]) may lead to lower BMI, greater height, lower systolic blood pressure and a reduced chance of coronary artery disease. However, the MR Egger sensitivity analyses indicated that these associations were driven by a violation of the MR assumptions.

The lack of causal associations could possibly indicate pleiotropy between health and cognitive ability, meaning that sets of genetic variants have independent effects on different phenotypes, rather than a set of genetic variants causally influencing cognitive ability via the related health exposure or vice versa. This would violate the assumption that the genetic variants only influence the outcome via the exposure, and not via other pathways. The idea of pleiotropy between health and cognitive ability is consistent with the theoretical construct of bodily system integrity[23], whereby a latent trait is manifest as individual differences in how effectively people meet cognitive and health challenges from the environment, and which has some genetic aetiology.

The MR Egger regression results highlight the importance of performing sensitivity analyses in an MR framework. For example, the paper by Bowden et al.[24] applied MR-Egger regression to data previously published on the causal effects of blood pressure on coronary artery disease[30]. The original study showed causal effect of genetic variants for blood

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

pressure on coronary artery disease, which indicated directional pleiotropy. The MR Egger regression by Bowden et al.[24] subsequently showed that the bias-adjusted estimates were closer to null and non-significant. Many of the original causal estimates in our study between educational attainment and health outcomes indicated directional pleiotropy, however the MR Egger regression indicated that the instrumental variable assumptions were violated.

Strengths of this study include the large sample size of UK Biobank, the participants of which all took the same cognitive tests, completed the same questionnaires and answered the same interview questions, in contrast to most genetic studies, where assessments across different cohorts often vary. A further strength is the fact that all of the UK Biobank genetic data were processed in a consistent matter, on the same platform and at the same location. The genetic variants on which the instrumental variables originated used the largest available GWAS at moment of testing.

Limitations of this study include the fact that cognitive ability was only measured on a subset of the UK Biobank participants and that it was a bespoke test. A second major limitation was that there is no published large genome-wide association study of cognitive ability in early life from which we could obtain genetic variants to use as an instrumental variable. Therefore, we used genome-wide significant SNPs associated with educational attainment as our early life cognitive ability instrument.

Instrumental variables for cardiovascular disease, type 2 diabetes, blood pressure, and educational attainment explain a small amount of the variance in the exposure. A good/strong instrumental variable would be expected to explain a substantial amount of the variance of the exposure. The current study used the same dataset (UK Biobank) for estimating the

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

association of the variants with the exposure and outcome. This could lead to bias in the results, as the association between the variants and the exposure is likely to be overestimated, because variants are chosen due to the association with the exposure in the analysed dataset. This could lead to overestimation of the association with the outcome, also called the Beavis effect or winner's curse[34].

Overall, this study found phenotypic cognitive-physical health associations, but did not find consistent evidence for causal associations between cognitive ability and physical health. This could be due to biological pleiotropy or violations of the instrumental variable assumptions. Future work should focus on stronger instrumental variables, as well as better case-control ascertainment.

Funding

This work was supported by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and Medical Research Council (MRC) is gratefully acknowledged. This research was conducted using the UK Biobank Resource.

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

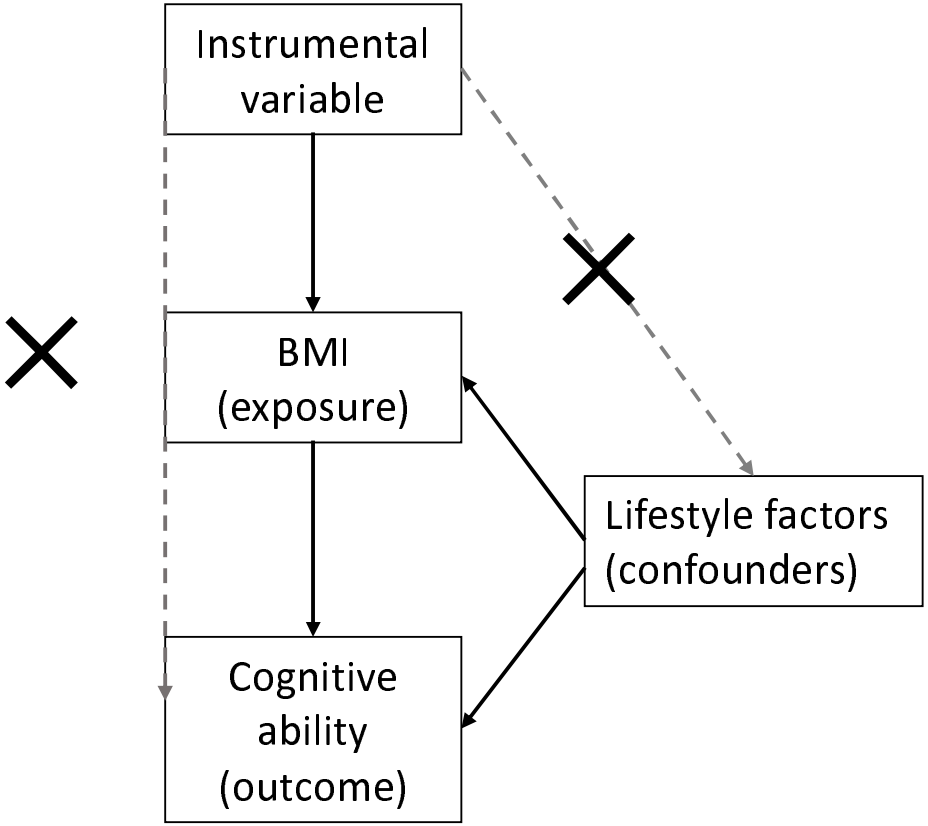
References

1. Deary IJ, Weiss A, Batty GD. Intelligence and Personality as Predictors of Illness and Death: How Researchers in Differential Psychology and Chronic Disease Epidemiology Are Collaborating to Understand and Address Health Inequalities. *Psychological Science in the Public Interest* 2010; **11**: 53-79.
2. Calvin CM, Deary IJ, Fenton C, et al. Intelligence in youth and all-cause-mortality: systematic review with meta-analysis. *International Journal of Epidemiology* 2011; **40**: 626-44.
3. Wraw C, Deary IJ, Gale CR, Der G. Intelligence in youth and health at age 50. *Intelligence* 2015; **53**: 23-32.
4. Hagenaars SP, Harris SE, Davies G, et al. Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N=112 151) and 24 GWAS consortia. *Mol Psychiatry* 2016.
5. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015; **47**: 1236-41.
6. Hill WD, Davies G, Liewald DC, McIntosh AM, Deary IJ. Age-Dependent Pleiotropy Between General Cognitive Function and Major Psychiatric Disorders. *Biological Psychiatry* 2016; **80**: 266-73.
7. Trampush JW, Lencz T, Knowles E, et al. Independent evidence for an association between general cognitive ability and a genetic locus for educational attainment. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* 2015; **168**: 363-73.
8. Rietveld CA, Esko T, Davies G, et al. Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proceedings of the National Academy of Sciences* 2014; **111**: 13790-4.
9. Banks J, Mazzonna F. The Effect of Education on Old Age Cognitive Abilities: Evidence from a Regression Discontinuity Design*. *The Economic Journal* 2012; **122**: 418-48.
10. Silventoinen K, Posthuma D, Van Beijsterveldt T, Bartels M, Boomsma DI. Genetic contributions to the association between height and intelligence: evidence from Dutch twin data from childhood to middle age. *Genes, Brain and Behavior* 2006; **5**: 585-95.
11. Keller MC, Garver-Apgar CE, Wright MJ, et al. The Genetic Correlation between Height and IQ: Shared Genes or Assortative Mating? *PLoS Genet* 2013; **9**: e1003451.
12. Marioni RE, Batty GD, Hayward C, et al. Common Genetic Variants Explain the Majority of the Correlation Between Height and Intelligence: The Generation Scotland Study. *Behavior Genetics* 2014; **44**: 91-6.
13. Starr JM, Taylor MD, Hart CL, et al. Childhood mental ability and blood pressure at midlife: linking the Scottish Mental Survey 1932 and the Midspan studies. *Journal of Hypertension* 2004; **22**: 893-7.
14. Belsky DW, Caspi A, Goldman-Mellor S, et al. Is Obesity Associated With a Decline in Intelligence Quotient During the First Half of the Life Course? *American Journal of Epidemiology* 2013; **178**: 1461-8.
15. Dahl A, Hassing LB, Fransson E, et al. Being Overweight in Midlife Is Associated With Lower Cognitive Ability and Steeper Cognitive Decline in Late Life. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 2010; **65A**: 57-62.
16. Novak V, Hajjar I. The relationship between blood pressure and cognitive function. *Nat Rev Cardiol* 2010; **7**: 686-98.
17. Mõttus R, Luciano M, Starr JM, Deary IJ. Diabetes and life-long cognitive ability. *Journal of Psychosomatic Research* 2013; **75**: 275-8.
18. Lawlor DA, David Batty G, Clark H, McIntyre S, Leon DA. Association of childhood intelligence with risk of coronary heart disease and stroke: findings from the Aberdeen Children of the 1950s cohort study. *European Journal of Epidemiology* 2008; **23**: 695-706.

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

19. Rawlings AM, Sharrett AR, Schneider ALC, et al. Diabetes in Midlife and Cognitive Change Over 20 YearsA Cohort StudyDiabetes in Midlife and Cognitive Change Over 20 Years. *Annals of Internal Medicine* 2014; **161**: 785-93.
20. Kovacic JC, Castellano JM, Fuster V. The links between complex coronary disease, cerebrovascular disease, and degenerative brain disease. *Annals of the New York Academy of Sciences* 2012; **1254**: 99-105.
21. Eggermont LHP, de Boer K, Muller M, Jaschke AC, Kamp O, Scherder EJA. Cardiac disease and cognitive impairment: a systematic review. *Heart* 2012.
22. Luciano M, Mõttus R, Harris SE, et al. Predicting cognitive ability in ageing cohorts using Type 2 diabetes genetic risk. *Diabetic Medicine* 2014; **31**: 714-20.
23. Deary IJ. Looking for 'System Integrity' in Cognitive Epidemiology. *Gerontology* 2012; **58**: 545-53.
24. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology* 2015; **44**: 512-25.
25. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med* 2015; **12**: e1001779.
26. Tobin MD, Sheehan NA, Scurrah KJ, Burton PR. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Statistics in Medicine* 2005; **24**: 2911-35.
27. Wood AR, Esko T, Yang J, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* 2014; **46**: 1173-86.
28. UNESCO. *International Standard Classification of Education ISCED 1997*. 2006 [cited 2016 30 September]; Available from: http://www.unesco.org/education/information/nfsunesco/doc/isced_1997.htm
29. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015; **518**: 197-206.
30. The International Consortium for Blood Pressure Genome-Wide Association Studies. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 2011; **478**: 103-9.
31. Schunkert H, Konig IR, Kathiresan S, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* 2011; **43**: 333-8.
32. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 2012; **44**: 981-90.
33. Okbay A, Beauchamp JP, Fontana MA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 2016; **533**: 539-42.
34. Taylor AE, Davies NM, Ware JJ, VanderWeele T, Smith GD, Munafò MR. Mendelian randomization in health research: Using appropriate genetic variants and avoiding biased estimates. *Economics & Human Biology* 2014; **13**: 99-106.

Figure 1. Model for Mendelian randomization study. The instrumental variable, based on genome-wide significant SNPs from independent studies for the exposure, is used to estimate if the exposure (e.g. BMI) causally influences the outcome (e.g. cognitive ability). The instrumental variable should be unrelated to potential confounders of the exposure-outcome association and should only affect the outcome via the exposure.



bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1. Coding for years of education in UK Biobank based on the ISCED coding.

UK Biobank degree level	UK Biobank code	ISCED code	Years of education	N
College or university degree	1	5	20 (19 +1)	33852
A levels/AS levels or equivalent	2	3	13	12560
O levels/GCSEs or equivalent	3	2	10	24802
CSEs or equivalent	4	2	10	6064
NVQ or HND or HNC or equivalent	5	NA	NA	7788
Other professional qualification eg: nursing, teaching	6	NA	NA	5776
None of the above	-7	1	7	20272
Prefer not to answer	-3	NA	NA	953

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 2. Information about instrumental variables.

Phenotype	SNPs included	total SNPs	N	Reference	Unavailable in UK Biobank	HWE $p < 1 \times 10^{-6}$	imputation < 0.9	R^2 %
BMI	72	75	236,231	Locke et al. Nature 2015; 518: 197-206. PMID: 25673413	1 (rs2033529)	3 (rs17001654, rs2075650, rs9925964)	0	1.50%
Height	395	405	253,288	Wood et al. Nat Genet 2014; 11: 1173-86. PMID: 25282103	3 (rs1420023, rs2413143, rs567401)	5 (rs4953951, rs7692995, rs915506, rs3790086, rs1401795)	2 (rs7534365, rs11683207)	13.68%
Systolic blood pressure	25	25	69,395	Ehret et al. Nature 2011; 478: 103-109. PMID: 21909115	0	0	0	0.65%
Coronary artery disease	25	25	22,233 cases; 64,762 controls	Schunkert et al. Nat Genet 2011; 43: 333-338. PMID: 21378990	0	0	0	0.28%
Type 2 diabetes	18	18	12,171 cases; 56,862 controls	Morris et al. Nat Genet 2012; 44: 981-990. PMID: 22885922	0	0	0	0.15%
Educational attainment	68	74	293,723	Okbay et al. Nature 2016; 533: 539-542. PMID: 27225129	2 (rs6799130, rs61160187)	4 (rs6799130, rs61160187, rs7131944, rs7306755)	0	0.76%

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 3. Phenotypic and genetic associations, using Mendelian randomization analysis, between five health instrumental variables and cognitive ability, using the verbal-numerical reasoning test. Significant associations are in bold. OR, odds ratio; MR-IVW, Mendelian randomization - inverse variance weighted method.

Cognitive ability	**Phenotypic: health outcomes – cognitive ability**			**MR-IVW: health SNPs – cognitive ability**		
SNPs (nr)	Beta	SE	p	Beta	SE	p
BMI (72)	-0.0490	0.005	**1.51×10^{-20}**	-0.0664	0.064	0.3035
Height (395)	0.1816	0.008	**5.53×10^{-124}**	0.020	0.017	0.2288
Systolic blood pressure (25)	-0.0492	0.006	**2.24×10^{-17}**	-0.0089	0.089	0.9207
Coronary artery disease (25)	-0.2651	0.026	**4.62×10^{-25}**	-0.0188	0.099	0.8507
Type 2 diabetes (18)	-0.0634	0.029	**0.0292**	0.0287	0.070	0.6882

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 4. Phenotypic and genetic associations, using Mendelian randomization analysis, between the educational attainment instrumental variable and five health outcomes. Significant associations are in bold. OR, odds ratio; MR-IVW, Mendelian randomization - inverse variance weighted method; MR-Egger, Mendelian randomization – Egger regression.

	Educational attainment – health outcomes			**Educational attainment SNPs (68) – health outcomes**					
	Phenotypic			**MR-IVW**			**MR-Egger**		
	Beta	SE	p	Beta	SE	p	Beta	SE	p
BMI	-0.1376	0.003	**< 1.00×10^{-300}**	-0.3102	0.051	**6.38×10^{-8}**	-0.1940	0.100	0.0559
Height	0.1040	0.002	**< 1.00×10^{-300}**	0.2192	0.057	**0.0003**	0.1359	0.112	0.2299
Systolic blood pressure	-0.0751	0.003	**7.12×10^{-130}**	-0.1616	0.042	**0.0003**	-0.1040	0.083	0.2166
Type 2 diabetes	OR: 0.8094	0.019	**1.88×10^{-27}**	-0.1172	0.033	**0.0006**	0.0218	0.062	0.7258
Coronary artery disease	OR: 0.7018	0.018	**6.79×10^{-91}**	-0.0653	0.030	**0.0318**	-0.0167	0.059	0.7761

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

bioRxiv preprint doi: https://doi.org/10.1101/084798; this version posted November 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

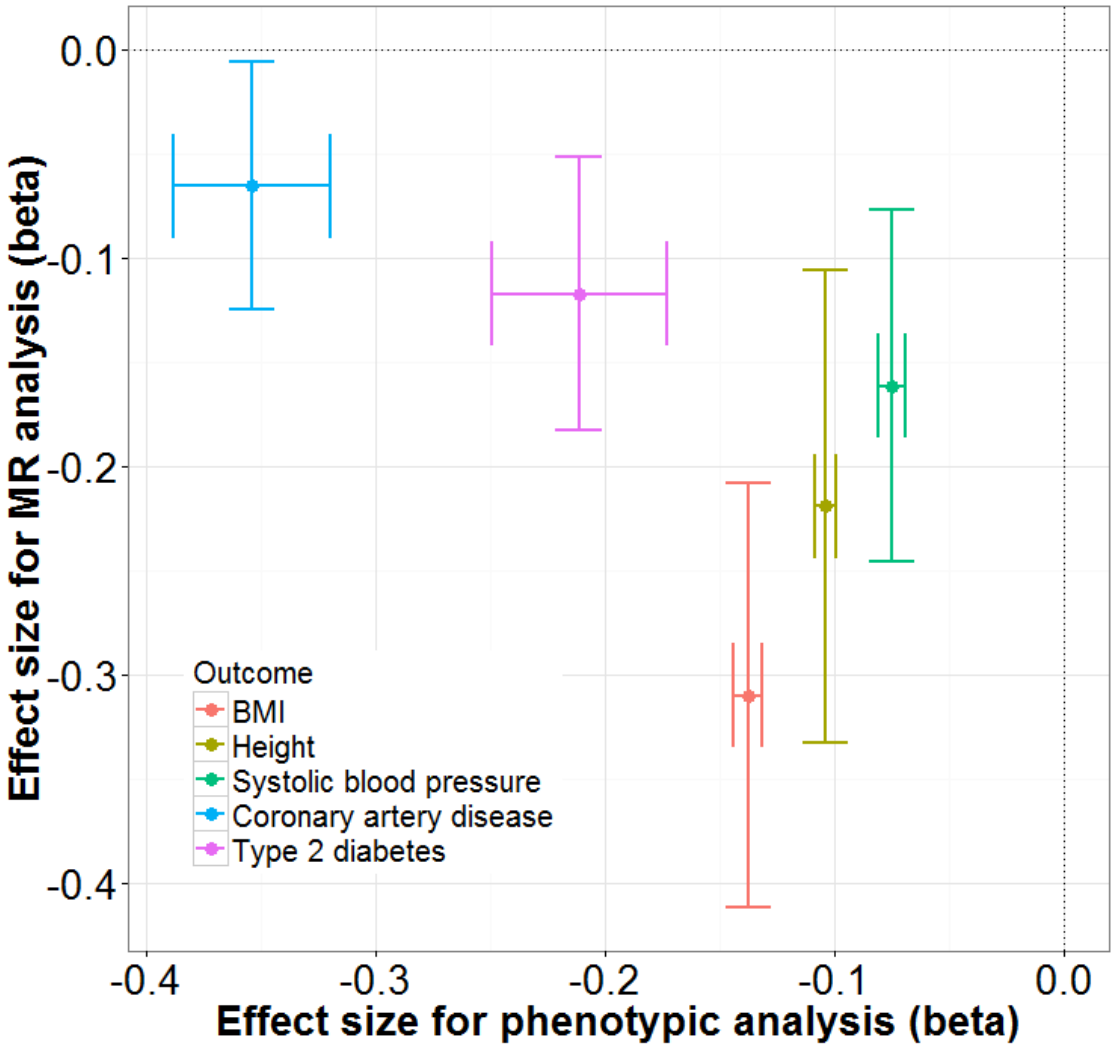



Figure 2. Comparison of phenotypic (x-axis) and genetic (y-axis) associations with 95% confidence intervals between educational attainment (exposure) and five health variables (outcomes). For the purpose of the figure, the assocation between height and educational attainment have been flipped, with a negative association indicating lower height and lower educational attainment.