medRxiv preprint doi: https://doi.org/10.1101/2020.02.01.20020008; this version posted February 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Health related effects of education: a Mendelian randomization study

Running head: Education and health outcomes

Shuai Yuan, M.D.[1,2], Ying Xiong, M.D.[3], Madeleine Michaëlsson, PhD[4], Karl Michaëlsson, PhD[1], Susanna C. Larsson, PhD[1,2]

Affiliations

[1] Department of Surgical Sciences, Uppsala University, Uppsala, Sweden

[2] Unit of Cardiovascular and Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

[3] Department of Public Health Sciences, Karolinska Institutet, Stockholm, Sweden

[4] Department of Education, Health and Social Studies, Dalarna University, Falun, Sweden

shuai.yuan@ki.se; ying.xiong@stud.ki.se; mmh@du.se

karl.michaelsson@surgsci.uu.se; susanna.larsson@surgsci.uu.se

Corresponding author:

Susanna C. Larsson, Department of Surgical Sciences, Uppsala University, Dag Hammarskjölds Väg 14B, Uppsala, 75185, Sweden; and Institute of Environmental Medicine, Karolinska Institutet, Nobelsväg 13, Stockholm, 17177, Sweden.
E-mail: susanna.larsson@surgsci.uu.se

Number of Tables: 2

Number of Figures: 2

Number of supplementary tables: 3

Number of supplementary figures: 6

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.01.20020008; this version posted February 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Abstract

A deeper understanding of the causal links from education level to health outcomes may shed a light for disease prevention. We conducted a wide-angled Mendelian randomization to disentangle the causal role of education level from intelligence for 31 health outcomes and explore to what extent body mass index and smoking mediate the associations. Genetically higher education level was associated with lower risk of major psychiatric disorders and most somatic diseases independent of intelligence, including suicide attempts, large artery stroke, heart failure, lung cancer, rheumatoid arthritis and type 2 diabetes, insomnia, major depressive disorder, total and ischemic stroke, coronary artery disease, breast cancer, ovarian cancer and gout. Adjustment for body mass index and smoking attenuated the associations between education and several outcomes, especially for type 2 diabetes and heart failure. These findings emphasize the importance of education to reduce the burden of common diseases.

Keywords: education; intelligence; Mendelian randomisation analysis; somatic disorder; psychiatric disorder.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.01.20020008; this version posted February 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Introduction

Education level is an important health social determinant and has been proposed as a modifiable risk factor for a number of disorders and diseases, such as depression [1], age-related cognitive decline [2], suicide [3], cardiovascular disease [4], cancer [5], and several other diseases [6-8]. However, it is unclear whether the associations are causal and independent of intelligence. Understanding the causal effects of education level on diseases can facilitate the aetiology pathway exploration of diseases as well as development of new strategies for disease prevention. Notwithstanding, randomized controlled trials are ethically and practically infeasible on this topic.

Exploiting genetic variants as instrumental variables for an exposure (e.g., education level), Mendelian randomization (MR) can strengthen the causal inference of an exposure-outcome association [9]. Comparing the risk of disease across individuals who have been classified by their genotype enables the causal effect of an exposure to be estimated with substantially less bias, such as confounding and reverse causality, than in a traditional observational analysis [9]. The rationale for diminished bias in MR studies is that genetic variants are randomly assorted and fixed at conception and therefore largely independent of confounders and cannot be modified by disease development [9].

We conducted an MR study to disentangle the causal role of education level from intelligence in major mental and neurological disorders and somatic diseases. A secondary aim was to explore whether intelligence is causally related to the same health outcomes independently of education. We additionally investigated the causal associations of education level and intelligence with modifiable health-related risk factors and whether main lifestyle factors mediate the pathway from education to health outcomes.

Methods

Study design

The design and hypothesis of the present study are displayed in **Supplementary Figure 1**. We used summary-level data from large genome-wide association studies (GWASs) and genetic consortia (**Table 1)**. Totally, our study included 11 mental and neurological disorders [10-20], 20 major somatic diseases [21-36], and 10 health-related risk factors [31,34,35,37-39]. A systematic review was conducted to find meta-analyses of observational studies of education level and diseases (**Supplementary table 1**).

Selection of instrumental variables

medRxiv preprint doi: https://doi.org/10.1101/2020.02.01.20020008; this version posted February 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Instrumental variables for education level and intelligence were identified from GWASs of, respectively, 1 131 881 and 269 867 individuals of European ancestries [40,41]. In total, 1271 and 205 single-nucleotide polymorphisms (SNPs) at the genome-wide significance threshold ($p<5\times10^{-8}$) were proposed as instrumental variables for education level and intelligence. All used SNPs explained around 12.2% and 5.2% variance for education level and intelligence, respectively. Education level was defined as number of years of education and was unified across included studies according to an International Standard Classification of Education category. Across all cohorts, the sample-size-weighted mean of education year was 16.8 years of schooling with a standard deviation (SD) of 4.2 years. For the definition of intelligence as an instrumental variable in our analysis, included cohorts extracted a single sum score, mean score, or factor score from a multidimensional set of cognitive performance tests in GWAS with linear model, with the exception of High-IQ/Health and Retirement Study where a logistic regression GWAS was run with "case" status (high intelligence) versus controls (normal intelligence level). All included GWASs adjusted for key covariates, such as age, sex and principal components for ancestry. The numbers of SNPs used as instrumental variables in each analysis are displayed in **Table 1**.

Outcome sources

Summary-level data for the associations of the education- and intelligence-associated SNPs with the outcomes were extracted from large-scale GWASs or genetic consortia. In the present MR study, we included 11 mental and neurological disorders [10-20], 9 cardiovascular diseases [21-25], 4 major cancers [26-29], 7 other diseases [30-36] and 10 established health-related risk factors for major diseases [31,34,35,37-39]. Detailed information, such as the number cases and controls, population structure and the source for each outcome, is presented in **Table 1**. Definitions of the diseases are presented in **Supplementary Table 2**.

Systematic review for meta-analysis of observational studies

A systematic literature search was conducted in the PubMed database before November 1st, 2019 to find meta-analyses of observational studies of education level in relation to diseases studied in the present MR study. We found latest published meta-analysis on 13 diseases and two risk factors, including major depressive disorders [42], suicide attempts [43], posttraumatic stress disorder [44], amyotrophic lateral sclerosis [45], Alzheimer's disease [46], coronary artery disease [47], heart failure [48], stroke [49], breast cancer [50], prostate cancer [51], lung cancer [52], type 2 diabetes [53], chronic kidney disease [54], body mass index [55] and hypertension (blood pressure) [56].

medRxiv preprint doi: https://doi.org/10.1101/2020.02.01.20020008; this version posted February 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

We extracted publication data (PMID number, the first author's name and year of publication), sample size, and risk estimates with their corresponding confidence intervals. Search strategy and characteristics of included meta-analyses are shown in **Supplementary Table 1**.

Statistical analysis

The random-effects inverse-variance weighted method was used to assess the associations of education and intelligence with the outcomes. The weighted median method and MR-Egger regression were used as sensitivity analyses to examine the consistency of results and to detect potential pleiotropy. The weighted median method gives accurate estimates if at least 50% of the instrumental variables are valid [57]. The MR-Egger regression can detect and adjust for pleiotropy albeit rendering low precision of the estimates [58]. Given the phenotypical and genetic correlation between education level and intelligence [59], we used a multivariable inverse-variance weighted method to disentangle the causal effect of education level on outcomes independent of intelligence and vice versa [60]. Considering that smoking and body mass index are modifiable risk factors linking education level to the most diseases, we also used multivariable-adjusted MR analysis to determine the effects of body mass index and smoking behaviour on health outcomes for associations reaching the conventional significance level ($p<0.05$) in both univariable and intelligence-adjusted inverse-variance weighted model to explore the mediation effects of body mass index and smoking, [61]. Proportions of attenuated effect size were calculated to present the magnitude of mediation effects. Odds ratios (ORs) and 95% confidence intervals (CIs) of diseases and changes of levels of risk factors were scaled to an SD increase in genetically predicted years of education (4.2 years) and intelligence. We calculated the power for the analyses of education level using a web-tool [62] and results are displayed in **Table 1**. All statistical analyses were two-sided and performed using the mrrobust package in Stata/SE 15.0 [63] and TwoSampleMR in R 3.6.0 software and MR-Base [64]. *P* values were not used strictly to define statistical significance; however, we interpreted the results based on the magnitude and strengths of the associations [65].

Results

Genetically predicted education level and diseases

Genetically predicted education level was causally associated with most diseases, including 8 out of 11 mental and neurological disorders, all 9 studied cardiovascular diseases, all 4 studied cancers, and 4 of 7 other common diseases in the univariable inverse-variance weighted MR analysis (**Figure 1**). In multivariable inverse-variance weighted analysis with adjustment for

medRxiv preprint doi: https://doi.org/10.1101/2020.02.01.20020008; this version posted February 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

intelligence, higher education level was additionally associated with higher odds of schizophrenia and anxiety (these associations were not observed in the univariable analysis) (**Figure 1**). However, the inverse associations of education level with amyotrophic lateral sclerosis, Alzheimer's disease, cardioembolic stroke, intracerebral haemorrhage and inflammatory bowel disease observed in the crude MR analysis did not remain after adjustment for intelligence (**Figure 1**). In the intelligence-adjusted model, higher education level was associated with lower odds of rheumatoid arthritis (OR 0.45; 0.33, 0.61), type 2 diabetes (OR 0.48; 0.43, 0.55), suicide attempts (OR 0.48; 0.37, 0.62), large artery stroke (OR 0.50; 0.36, 0.68), heart failure (OR 0.51; 0.42, 0.63), lung cancer (OR 0.52; 0.42, 0.65), ovarian cancer (OR 0.53; 0.43, 0.66), small vessel stroke (OR 0.62; 0.47, 0.81), any ischemic stroke (OR 0.69; 0.61, 0.78), insomnia (OR 0.69; 0.64, 0.75), gout (OR 0.71; 0.60, 0.84), major depressive disorder (OR 0.78; 0.72, 0.85) and breast cancer (OR 0.85; 0.77, 0.94). Conversely, higher education level adjusted for intelligence was associated with an increased risk of obsessive-compulsive disorder (OR 2.24; 1.47, 3.41), bipolar disorder (OR 2.04; 1.64, 2.54), schizophrenia (OR 1.88; 1.49, 2.36), anorexia nervosa (OR 1.88; 1.53, 2.30) and anxiety (OR 1.84; 1.33, 2.56) (**Figure 1**). Results of sensitivity analyses were directionally similar but with wider CIs (**Supplementary Figure 2**). We detected possible pleiotropy in the analysis of obsessive-compulsive disorder, inflammatory bowel disease and rheumatoid arthritis (*p* for intercept in MR-Egger <0.05).

Genetically predicted intelligence and diseases

The associations between intelligence and diseases are presented in **Supplementary Figure 3 and 4.** The effects of intelligence on psychiatric and somatic diseases showed the same patterns with that of education level with the exception for schizophrenia (OR 0.47; 95% CI, 0.33, 0.67), bipolar disorder (OR 0.72; 95% CI, 0.52, 1.00), and Alzheimer's disease (OR 0.77, 95% CI, 0.59, 1.01) after adjustment for education level (**Supplementary Figure 3**). Similar results were obtained from the weighted median method and pleiotropy was detected in the analysis of heart failure in the MR-Egger regression analysis (**Supplementary Figure 4**).

Education, intelligence and risk factors

Genetically predicted higher education level was associated with later age of smoking initiation (β=0.31; 95% CI, 0.29, 0.33), higher bone mineral density (β=0.05; 95% CI, 0.03, 0.08), lower serum urate levels (β=-0.12; 95% CI, -0.15, -0.09), lower systolic (β=-0.13; 95% CI, -0.15, -

medRxiv preprint doi: https://doi.org/10.1101/2020.02.01.20020008; this version posted February 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

0.11) and diastolic (β=-0.18; 95% CI, -0.20, -0.15) blood pressure, lower waist-to-hip ratio (β=-0.29; 95% CI, -0.31, -0.27), lower body mass index (β=-0.34; 95% CI, -0.37, -0.31) and fewer cigarettes smoked per day (β=-0.37; 95% CI, -0.42, -0.32) in the univariable model; the estimates were similar in the intelligence-adjusted model (**Figure 2**). Findings were consistent in sensitivity analyses and no pleiotropy was observed (**Supplementary Figure 5**). The same patterns of associations were observed for genetically predicted intelligence (**Supplementary Figure 6**). Nevertheless, after adjustment for education level, the magnitude of associations attenuated largely (**Supplementary Figure 6**).

Comparison with observational studies

The present MR findings were generally similar in the direction and magnitude to the estimates based on meta-analyses of observational studies (**Supplementary Table 3**). However, there were discrepancies concerning the effects of education level on suicide attempts, breast cancer and prostate cancer.

Mediation effects of body mass index and smoking

Table 2 shows the results of mediation analyses after adjusting for body mass index and smoking behaviour in the pathway from education to health outcomes. Although not apparent for all disease outcomes, body mass index and smoking partly mediated most associations between education and diseases (**Table 2**). A strong mediation effect of body mass index was observed in the associations of education with gout (68%), type 2 diabetes (57%) and heart failure (34%). With regard to smoking, a strong mediation effect was detected in the association of education with major depressive disorder (24%), type 2 diabetes (20%) and lung cancer (19%). After adjustment for both body mass index and smoking, the direct causal effect of education on outcomes was substantially attenuated for gout (64%), type 2 diabetes (63%), heart failure (36%), obsessive-compulsive disorder (32%), suicide attempts (31%), coronary artery disease (31%) and lung cancer (29%).

Discussion

In the present MR study with up to 1.3 million individuals, genetic predisposition to higher education level was causally associated with the majority of major health outcomes and related risk factors. Specifically, genetic predisposition to higher education level, independent of intelligence, was associated with lower risk of major depressive disorder, insomnia, suicide

medRxiv preprint doi: https://doi.org/10.1101/2020.02.01.20020008; this version posted February 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

attempts, stroke, heart failure, coronary artery disease, breast cancer, ovarian cancer, lung cancer, gout, type 2 diabetes, and rheumatoid arthritis. Conversely, higher education level was associated with higher risk of obsessive-compulsive disorder, anorexia nervosa, bipolar disorder, and prostate cancer. Genetically predicted higher intelligence, independent of education, was inversely related to bipolar disorder, schizophrenia and Alzheimer's disease. Body mass index and smoking displayed strongest mediation effects observed for gout, type 2 diabetes, obsessive-compulsive disorder, suicide attempts and heart failure.

Comparison with previous studies

Our findings are broadly in line with a vast body of observational studies showing a protective association of high educational level on major depressive disorder [42], amyotrophic lateral sclerosis [45], Alzheimer's disease [46], coronary heart disease [47], heart failure [48], stroke [49], lung cancer [52], type 2 diabetes [53], chronic kidney disease [54], hypertension [56] and obesity [55]. However, for suicide attempts, posttraumatic stress disorder, breast cancer and prostate cancer, our MR findings differ from observational findings. The discrepancies might be attributed by reverse causality in the observational studies, heterogeneity and small sample sizes in the meta-analyses. A substantial heterogeneity (I^2=85%; p<0.001) was observed among included observational studies in the meta-analysis of breast cancer [50], and the sample size was small for prostate cancer [51]. Some studies have proposed that the higher risk of prostate cancer among men with high education level was driven by higher prostate-specific antigen screening rate among educated men compared with men with low education level [66]. With regard to the inverse associations of higher education level with breast and ovarian cancer, these associations may in part be mediated by reproductive or hormone-related factors, or other health behaviours such as healthier diet and physical activity. We are not aware of any previous MR studies on education or intelligence in relation to prostate, breast, or ovarian cancer, but a protective causal effect of higher education on lung cancer risk has been reported recently [67].

Previous MR studies based on much fewer SNPs (up to 162) as predictors of education level showed a protective effect of higher educational level on coronary artery disease [68] and Alzheimer's disease [69]. The present study based on substantially more SNPs as instrumental variables more precisely verified the findings in previous research. Notably, the effects of high education level in the previous studies might be influenced by high intelligence given the tight phenotype and genetic correlation between intelligence and education level. In the present study, we used multivariable MR analysis to assess the direct effect of education level that is not mediated via intelligence with coronary artery disease and Alzheimer's disease. For

medRxiv preprint doi: https://doi.org/10.1101/2020.02.01.20020008; this version posted February 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Alzheimer's disease, we found that higher intelligence rather than education level may be the protective factor, whereas higher education level was the protective factor for coronary artery disease. In a previous MR study of the direct effect of education and intelligence on certain health outcomes, including diabetes, hypertension, heart attack, total stroke, total cancer, and depression, no significant association with education or intelligence was observed despite significant or suggestive associations of genetically predicted education with potential risk factors (blood pressure, smoking, alcohol consumption, body mass index, vigorous physical activity, and television watching) [59]. However, the genetic instrument used in that study comprised only 219 SNPs associated with education or intelligence and genetic associations with the outcomes were estimated in up to 138 670 UK Biobank participants, with relatively few outcomes. Thus, the power may have been insufficient to detect associations. Findings of other MR studies of education level in relation to health-related risk factors, such as obesity [70], cigarette smoking [71] and blood pressure [72], are consistent with our findings.

The inverse association between intelligence and schizophrenia was also observed in an observational study with 24 706 Swedish adults independent of overlapping genetic risks of two traits [73]. However, the effects of high intelligence on bipolar disorder were conflicting across observational studies. A large-scale cohort study with over 20-year follow-up duration proposed a "reversed-J" shape association between intelligence and risk of bipolar disorder, which indicated that individuals with the lowest and highest intelligence had the greatest risk of bipolar disorder) [74]. In a GWAS, it was revealed that bipolar disorder risk alleles were associated with better cognitive performance [75], which is opposite to our findings. Since education had a contradictory effect, which we observed in the present study, on bipolar disease to intelligence, the discrepancy among these studies might be caused by a mixed effect of education and intelligence at both aspects of phenotype and gene.

Possible mechanisms

Based on results of the present MR study and previous observational studies, there are three major possible pathways linking education level to health outcomes: 1) modifiable risk factors largely mediates the educational effects on diseases [72,76]; 2) there may be direct effects from education-related brain structures or function change via gene methylation, gene silencing etc. [77-79], especially for mental and neurological disorders; and 3) subjective well-being, happiness and meaning of life influenced by education level exerts effects on psychiatric and somatic diseases directly or indirectly [80-83]. Education, as measured in this study, can be defined as an institutionalized form of social resource, and more specifically a form of cultural capital

medRxiv preprint doi: https://doi.org/10.1101/2020.02.01.20020008; this version posted February 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

drawing on the terminology of the French sociologist Pierre Bourdieu. Related forms of cultural capital emerge as objectivized resources – such as books, art or scientific tools – or incorporated resources, such as knowledge, attitudes and practices [84,85]. Our study shows that education is a health relevant cultural capital whilst intelligence is not to the same degree related with health and risk of disease.

Observational studies have found that the associations between education level and diseases attenuated largely after adjustment for health-related risk factors. Compared with unadjusted model, the risk of cardiovascular diseases of low education attainment attenuated around 30-45% in statistical models adjusted for multiple risk factors, such as smoking, body mass index, hypertension and physical activity [86,87]. However, measurement error and misclassification of mediators in observational studies often underestimates the mediation effects. The mediation effects were also proved in previous MR analysis [72,76]. In the present study, genetically predicted education level was associated with a favourable risk factor profile: with improved smoking behaviours (postponed smoking initiation age and less cigarette per day) as well as lower adiposity (body mass index and waist-to-hip ratio), blood pressure and serum urate levels, which might mediate associations between education level and diseases. By conducting mediation analysis, we showed that body mass index and smoking behaviour partly or entirely mediated the pathway from education level to several health outcomes, in particular gout, type 2 diabetes, obsessive-compulsive disorder, suicide attempts, atrial fibrillation, heart failure, coronary artery disease and lung cancer.

Previous studies have found that low education level might influence the changes in biochemical response and risk-related brain function, such as inflammation [77], cardiometabolic traits [78] and amygdala reactivity [79], via gene methylation, thereby influencing disease risk. In addition, genetic studies have also revealed that improvement of subjective well-being [80,81], happiness [80,81], meaning of life [82], social interaction [83], possibly derived from high education level benefited human health directly and indirectly (e.g. influencing brain morphology, central nervous system and adrenal/pancreas tissues). There are other possible explanations, like followings: education level also could modify the risk of health outcomes through other diseases (comorbidity), the use of health care services, neighbourhood environment, occupations, income and marital status, which were amenable if education level was increased.

The results indicate that more than knowledge itself is affecting how people live their life, for instance through pathways regarding reduced smoking or alcohol habits among highly educated people. Therefore, we should consider further explanations, such as the relationship between high education on the one hand and the status and resources that follow it, on the other,

medRxiv preprint doi: https://doi.org/10.1101/2020.02.01.20020008; this version posted February 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

which could by itself have a positive health effect on the individual. A further explanation assumes that it is the process itself that can be associated with increased well-being. That is, the process of taking part of and acquiring external knowledge rather than remaining with one's own innate thinking or being kept oblivious. Should only a fraction of the disease burden be explained by this process of mental activity − given that education leads to a different kind of thinking, which is supported by the present study in that health is affected regardless of intelligence level − then increased knowledge through education might lead to longevity through mechanisms beyond health literacy pathways of late-onset diseases and beyond the influence of social and material factors.

Strengths and limitations

The present study is the first study that comprehensively investigated the causal effects of education and intelligence on a very broad range of major disease outcomes and associated risk factors using genetic data from large-scale GWASs and genetic consortia. We used much more SNPs deriving from a larger GWAS with around 1.1 million individuals as instrumental variables for education level and the latest GWASs with largest sample size for outcomes compared to previous MR studies, thereby assuring adequate statistical power to detect weak associations. In addition, we disentangled the independent effect of education level from intelligence using a multivariable MR approach. Thus, it is a straightforward approach to estimate the possible health benefits from education promotion among general population. We used mediation analysis to reveal the roles of body mass index and smoking behaviour as mediators in the pathway from education level to health outcomes. Even though there were genetic data for certain outcomes from GWASs with trans-ancestry populations, the majority of included participants were individuals with European ancestry thereby diminishing population stratification bias. However, population confinement limited the transferability of the present findings to populations of non-European ancestries.

The major limitation in the present study is the possible unbalanced horizontal pleiotropy aroused from used genetic variants marking more generic biological pathways. It has been found that lead SNPs related to education level and intelligence are significantly overexpressed in the central nervous system, such as hippocampus and cerebral cortex, but not other organs [40]. For cardiovascular disease, cancers and other physical diseases, we can minimize the possibility of pleiotropy from the global or systemic measures of fitness (such as mitochondrial function). It is more likely to conclude that the potential pleiotropy might exert a large to moderate effect via predominantly neurological pathways (for example, behaviours

medRxiv preprint doi: https://doi.org/10.1101/2020.02.01.20020008; this version posted February 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

associated with obesity or smoking) for somatic diseases. In this scenario, the vertical pleiotropy would not bias the total causal effect by a higher educational level on disease development. With regard to mental and neurological disorders, although gene overwhelmingly expressed in brain or central nervous system, studies found no, or at most a small, genetic correlation between lower education attainment and psychiatric and neurological disorders by using bivariate genomic-relationship-matrix restricted maximum likelihood analysis [1]. Thus, the associations between education level and mental or neurological diseases were not mainly because of measurable pleiotropic genetic effects, but because of education-related environmental factors. In addition, from a statistical perspective, we detected almost no pleiotropy in the results of MR-Egger regression and the estimates were consistent through sensitivity analyses, which indicated a negligible distortion by potential pleiotropy. Intergenerational effects from parents for certain disease, such as coronary artery diseases and type 2 diabetes, could not be assessed by using the data in the present MR study.

Future research direction and implications

The present study established the findings on the hypothesis that education level influenced the risk of mental and neurological disorders and somatic diseases are partly mediated via associated modifiable risk factors, including body mass index and smoking. It is of interest to investigate the possible mechanisms behind the associations between education level and diseases from the perspective of education-related risk factors as mediators in more detail. Whether there are interaction effects within the education-associated risk factors in the pathway from education level to diseases needs more investigation. Considering the limited feasibility of interaction analysis in MR design, more advanced statistical methods are warranted [61]. Education benefits on health in addition to direct effects on known risk factors need also to be evaluated. Molecular and genetic research are also needed to illuminate the potential mechanisms or pleiotropic limitations, and also able to shed light on disease management and drug discovery, especially for mental and neurological disorders. For example, given a genetic overlap between bipolar disorder and intelligence [75], it will be of interest to use genetic approaches to test whether the association between education level and bipolar disorder is because of causality or pleiotropy.

Given the lower risk of major diseases from increasing education level independent of intelligence, it is of benefit to encourage education across the whole population. Policies that promotes education may also generate social benefits, such as minimizing health inequality,

medRxiv preprint doi: https://doi.org/10.1101/2020.02.01.20020008; this version posted February 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

increasing economic productivity and lowering crime [88]. However, it needs to be thought deeply how to increase education level in an acceptable, affordable and sustainable way.

Conclusion

The present MR study strengthened the evidence of causal protective effects of high education level on the majority of psychiatric and somatic diseases independent of intelligence. Body mass index and smoking partly mediated several of the associations between education level and health outcomes. These findings strongly suggest increasing education level for overall health benefits.

Acknowledgments

Summary-level data for education level, intelligence and outcomes were obtained from the UK Biobank and genetic consortia, including Psychiatric Genomics Consortium (bipolar disorder, schizophrenia, obsessive-compulsive disorder, post-traumatic stress disorder, anxiety and anorexia nervosa), The International Genomics of Alzheimer's Project (IGAP), Project MinE, Center for Neurogenomics and Cognitive Research (CNCR), iPSYCH, CARDIoGRAMplusC4D Consortium, MEGASTROKE Consortium, Atrial Fibrillation Consortium (AFGen), International Stroke Genetics Conosrtium (ISGC), Breast Cancer Association Consortium (BCAC), The Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) consortium, The International Lung Cancer Consortium (ILCCO), The Ovarian Cancer Association Consortium (OCAC), The DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) consortium, The Global Urate Genetics Consortium (GUGC), Genetics and Allied research in Rheumatic diseases Networking (GARNET) consortium, UK Inflammatory Bowel Disease Genetics (IBD) Consortium, CKDGen Consortium, the EArly Genetics and Lifecourse Epidemiology (EAGLE) Consortium, GEnetic Factors for OSteoporosis (GEFOS) Consortium, GIANT consortium, Neale lab, Consortium of Alcohol and Nicotine use (GSCAN). The authors thank all investigators for sharing these data. The list of investigators of MEGASTROKE is available at http://megastroke.org/authors.html. Funding of the MEGASTROKE project are specified at megastroke.org/acknowledgements.html.

Additional information

Contributors: SY drafted the manuscript and conducted the statistical analyses. SY and YX conducted the systematic review. SCL is the guarantor of this work. All authors contributed to

medRxiv preprint doi: https://doi.org/10.1101/2020.02.01.20020008; this version posted February 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

the interpretation of the results and critical revision of the manuscript for important intellectual content and approved the final version of the manuscript.

Funding: Funding for this study came from the Swedish Research Council (Vetenskapsrådet; Grant Number 2019-00977) and the Swedish Research Council for Health, Working Life and Welfare (Forte; Grant Number 2018-00123).

Competing interests: All authors declared no support from any organisation for the submitted work; no financial relationships with any organisation that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: This MR study was approved by the Swedish Ethical Review Authority.

Data sharing: The datasets analysed in this study are publicly available summary statistics.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.01.20020008; this version posted February 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

References

1 Peyrot, W. J. *et al.* The association between lower educational attainment and depression owing to shared genetic effects? Results in ~25,000 subjects. *Mol Psychiatry* **20**, 735-743, doi:10.1038/mp.2015.50 (2015).

2 Lipnicki, D. M. *et al.* Age-related cognitive decline and associations with sex, education and apolipoprotein E genotype across ethnocultural groups and geographic regions: a collaborative cohort study. *PLoS Med* **14**, e1002261, doi:10.1371/journal.pmed.1002261 (2017).

3 Naghavi, M. Global, regional, and national burden of suicide mortality 1990 to 2016: systematic analysis for the Global Burden of Disease Study 2016. *Bmj* **364**, l94, doi:10.1136/bmj.l94 (2019).

4 Rosengren, A. *et al.* Socioeconomic status and risk of cardiovascular disease in 20 low-income, middle-income, and high-income countries: the Prospective Urban Rural Epidemiologic (PURE) study. *Lancet Glob Health* **7**, e748-e760, doi:10.1016/s2214-109x(19)30045-2 (2019).

5 Albano, J. D. *et al.* Cancer mortality in the United States by education level and race. *J Natl Cancer Inst* **99**, 1384-1394, doi:10.1093/jnci/djm127 (2007).

6 Putrik, P. *et al.* Lower education and living in countries with lower wealth are associated with higher disease activity in rheumatoid arthritis: results from the multinational COMORA study. *Ann Rheum Dis* **75**, 540-546, doi:10.1136/annrheumdis-2014-206737 (2016).

7 Patzer, R. E. & McClellan, W. M. Influence of race, ethnicity and socioeconomic status on kidney disease. *Nat Rev Nephrol* **8**, 533-541, doi:10.1038/nrneph.2012.117 (2012).

8 Mountjoy, E. *et al.* Education and myopia: assessing the direction of causality by mendelian randomisation. *Bmj* **361**, k2022, doi:10.1136/bmj.k2022 (2018).

9 Davey Smith, G. & Ebrahim, S. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? *Bmj* **330**, 1076-1079, doi:10.1136/bmj.330.7499.1076 (2005).

10 Stahl, E. A. *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet* **51**, 793-803, doi:10.1038/s41588-019-0397-8 (2019).

11 Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell* **173**, 1705-1715.e1716, doi:10.1016/j.cell.2018.05.046 (2018).

12 Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci* **22**, 343-352, doi:10.1038/s41593-018-0326-7 (2019).

13 Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet* **51**, 414-430, doi:10.1038/s41588-019-0358-2 (2019).

14 Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol Psychiatry* **23**, 1181-1188, doi:10.1038/mp.2017.154 (2018).

15 van Rheenen, W. *et al.* Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nat Genet* **48**, 1043-1048, doi:10.1038/ng.3622 (2016).

16 Nievergelt, C. M. *et al.* International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. *Nat Commun* **10**, 4558, doi:10.1038/s41467-019-12576-w (2019).

medRxiv preprint doi: https://doi.org/10.1101/2020.02.01.20020008; this version posted February 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

17 Otowa, T. *et al.* Meta-analysis of genome-wide association studies of anxiety disorders. *Mol Psychiatry* **21**, 1391-1399, doi:10.1038/mp.2015.197 (2016).

18 Jansen, P. R. *et al.* Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat Genet* **51**, 394-403, doi:10.1038/s41588-018-0333-3 (2019).

19 Watson, H. J. *et al.* Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat Genet* **51**, 1207-1214, doi:10.1038/s41588-019-0439-2 (2019).

20 Erlangsen, A. *et al.* Genetics of suicide attempts in individuals with and without mental disorders: a population-based genome-wide association study. *Mol Psychiatry*, doi:10.1038/s41380-018-0218-y (2018).

21 Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* **47**, 1121-1130, doi:10.1038/ng.3396 (2015).

22 Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* **50**, 524-537, doi:10.1038/s41588-018-0058-3 (2018).

23 Aragam, K. G. *et al.* Phenotypic Refinement of Heart Failure in a National Biobank Facilitates Genetic Discovery. *Circulation*, doi:10.1161/circulationaha.118.035774 (2018).

24 Roselli, C. *et al.* Multi-ethnic genome-wide association study for atrial fibrillation. *Nat Genet* **50**, 1225-1233, doi:10.1038/s41588-018-0133-9 (2018).

25 Woo, D. *et al.* Meta-analysis of genome-wide association studies identifies 1q22 as a susceptibility locus for intracerebral hemorrhage. *Am J Hum Genet* **94**, 511-521, doi:10.1016/j.ajhg.2014.02.012 (2014).

26 Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92-94, doi:10.1038/nature24284 (2017).

27 Schumacher, F. R. *et al.* Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet* **50**, 928-936, doi:10.1038/s41588-018-0142-8 (2018).

28 Wang, Y. *et al.* Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nat Genet* **46**, 736-741, doi:10.1038/ng.3002 (2014).

29 Phelan, C. M. *et al.* Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat Genet* **49**, 680-691, doi:10.1038/ng.3826 (2017).

30 Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* **50**, 1505-1513, doi:10.1038/s41588-018-0241-6 (2018).

31 Tin, A. *et al.* Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels. *Nat Genet*, doi:10.1038/s41588-019-0504-x (2019).

32 Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376-381, doi:10.1038/nature12873 (2014).

33 de Lange, K. M. *et al.* Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet* **49**, 256-261, doi:10.1038/ng.3760 (2017).

34 Morris, J. A. *et al.* An atlas of genetic influences on osteoporosis in humans and mice. *Nat Genet* **51**, 258-266, doi:10.1038/s41588-018-0302-x (2019).

medRxiv preprint doi: https://doi.org/10.1101/2020.02.01.20020008; this version posted February 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

35 Wuttke, M. *et al.* A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet* **51**, 957-972, doi:10.1038/s41588-019-0407-x (2019).

36 Paternoster, L. *et al.* Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nat Genet* **47**, 1449-1456, doi:10.1038/ng.3424 (2015).

37 Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet* **28**, 166-174, doi:10.1093/hmg/ddy327 (2019).

38 *Neale lab*, <http://www.nealelab.is> (2019).

39 Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* **51**, 237-244, doi:10.1038/s41588-018-0307-5 (2019).

40 Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* **50**, 1112-1121, doi:10.1038/s41588-018-0147-3 (2018).

41 Savage, J. E. *et al.* Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet* **50**, 912-919, doi:10.1038/s41588-018-0152-6 (2018).

42 Chang-Quan, H., Zheng-Rong, W., Yong-Hong, L., Yi-Zhou, X. & Qing-Xiu, L. Education and risk for late life depression: a meta-analysis of published literature. *Int J Psychiatry Med* **40**, 109-124, doi:10.2190/PM.40.1.i (2010).

43 Li, Y. & Cao, J. Factors associated with suicidal behaviors in mainland China: a meta-analysis. *BMC Public Health* **12**, 524, doi:10.1186/1471-2458-12-524 (2012).

44 Tang, B., Deng, Q., Glik, D., Dong, J. & Zhang, L. A Meta-Analysis of Risk Factors for Post-Traumatic Stress Disorder (PTSD) in Adults and Children after Earthquakes. *Int J Environ Res Public Health* **14**, doi:10.3390/ijerph14121537 (2017).

45 Wang, M. D., Little, J., Gomes, J., Cashman, N. R. & Krewski, D. Identification of risk factors associated with onset and progression of amyotrophic lateral sclerosis using systematic review and meta-analysis. *Neurotoxicology* **61**, 101-130, doi:10.1016/j.neuro.2016.06.015 (2017).

46 Xu, W. *et al.* Meta-analysis of modifiable risk factors for Alzheimer's disease. *J Neurol Neurosurg Psychiatry* **86**, 1299-1306, doi:10.1136/jnnp-2015-310548 (2015).

47 Khaing, W., Vallibhakara, S. A., Attia, J., McEvoy, M. & Thakkinstian, A. Effects of education and income on cardiovascular outcomes: A systematic review and meta-analysis. *Eur J Prev Cardiol* **24**, 1032-1042, doi:10.1177/2047487317705916 (2017).

48 Potter, E. L., Hopper, I., Sen, J., Salim, A. & Marwick, T. H. Impact of socioeconomic status on incident heart failure and left ventricular dysfunction: systematic review and meta-analysis. *Eur Heart J Qual Care Clin Outcomes* **5**, 169-179, doi:10.1093/ehjqcco/qcy047 (2019).

49 McHutchison, C. A., Backhouse, E. V., Cvoro, V., Shenkin, S. D. & Wardlaw, J. M. Education, Socioeconomic Status, and Intelligence in Childhood and Stroke Risk in Later Life: A Meta-analysis. *Epidemiology* **28**, 608-618, doi:10.1097/ede.0000000000000675 (2017).

50 Dong, J. Y. & Qin, L. Q. Education level and breast cancer incidence: a meta-analysis of cohort studies. *Menopause*, doi:10.1097/gme.0000000000001425 (2019).

medRxiv preprint doi: https://doi.org/10.1101/2020.02.01.20020008; this version posted February 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

51 Brown, C. R. *et al.* Social determinants of prostate cancer in the Caribbean: a systematic review and meta-analysis. *BMC Public Health* **18**, 900, doi:10.1186/s12889-018-5696-y (2018).

52 Sidorchuk, A. *et al.* Socioeconomic differences in lung cancer incidence: a systematic review and meta-analysis. *Cancer Causes Control* **20**, 459-471, doi:10.1007/s10552-009-9300-8 (2009).

53 Bellou, V., Belbasis, L., Tzoulaki, I. & Evangelou, E. Risk factors for type 2 diabetes mellitus: An exposure-wide umbrella review of meta-analyses. *PLoS One* **13**, e0194127, doi:10.1371/journal.pone.0194127 (2018).

54 Zeng, X. *et al.* Associations between socioeconomic status and chronic kidney disease: a meta-analysis. *J Epidemiol Community Health* **72**, 270-279, doi:10.1136/jech-2017-209815 (2018).

55 He, J., Chen, X., Fan, X., Cai, Z. & Huang, F. Is there a relationship between body mass index and academic achievement? A meta-analysis. *Public Health* **167**, 111-124, doi:10.1016/j.puhe.2018.11.002 (2019).

56 Leng, B., Jin, Y., Li, G., Chen, L. & Jin, N. Socioeconomic status and hypertension: a meta-analysis. *J Hypertens* **33**, 221-229, doi:10.1097/hjh.0000000000000428 (2015).

57 Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol* **40**, 304-314, doi:10.1002/gepi.21965 (2016).

58 Burgess, S. & Thompson, S. G. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol* **32**, 377-389, doi:10.1007/s10654-017-0255-x (2017).

59 Davies, N. M. *et al.* Multivariable two-sample Mendelian randomization estimates of the effects of intelligence and education on health. *Elife* **8**, doi:10.7554/eLife.43990 (2019).

60 Rees, J. M. B., Wood, A. M. & Burgess, S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. *Stat Med* **36**, 4705-4718, doi:10.1002/sim.7492 (2017).

61 Carter, A. R. *et al.* Mendelian randomisation for mediation analysis: current methods and challenges for implementation. doi:10.1101/835819 (2019).

62 Brion, M. J., Shakhbazov, K. & Visscher, P. M. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol* **42**, 1497-1501, doi:10.1093/ije/dyt179 (2013).

63 Spiller, W., Davies, N. M. & Palmer, T. M. Software Application Profile: mrrobust - a tool for performing two-sample summary Mendelian randomization analyses. doi:10.1101/142125 (2017).

64 Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, doi:10.7554/eLife.34408 (2018).

65 Sterne, J. A. & Davey Smith, G. Sifting the evidence-what's wrong with significance tests? *Bmj* **322**, 226-231, doi:10.1136/bmj.322.7280.226 (2001).

66 Lund Nilsen, T. I., Johnsen, R. & Vatten, L. J. Socio-economic and lifestyle factors associated with the risk of prostate cancer. *Br J Cancer* **82**, 1358-1363, doi:10.1054/bjoc.1999.1105 (2000).

67 Zhou, H. *et al.* Education and lung cancer: a Mendelian randomization study. *Int J Epidemiol* **48**, 743-750, doi:10.1093/ije/dyz121 (2019).

medRxiv preprint doi: https://doi.org/10.1101/2020.02.01.20020008; this version posted February 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

68 Tillmann, T. *et al.* Education and coronary heart disease: mendelian randomisation study. *Bmj* **358**, j3542, doi:10.1136/bmj.j3542 (2017).

69 Larsson, S. C. *et al.* Modifiable pathways in Alzheimer's disease: Mendelian randomisation analysis. *Bmj* **359**, j5375, doi:10.1136/bmj.j5375 (2017).

70 Bockerman, P. *et al.* Does higher education protect against obesity? Evidence using Mendelian randomization. *Prev Med* **101**, 195-198, doi:10.1016/j.ypmed.2017.06.015 (2017).

71 Gage, S. H., Bowden, J., Davey Smith, G. & Munafo, M. R. Investigating causality in associations between education and smoking: a two-sample Mendelian randomization study. *Int J Epidemiol* **47**, 1131-1140, doi:10.1093/ije/dyy131 (2018).

72 Carter, A. R. *et al.* Understanding the consequences of education inequality on cardiovascular disease: mendelian randomisation study. *Bmj* **365**, l1855, doi:10.1136/bmj.l1855 (2019).

73 Kowalec, K. *et al.* Increased schizophrenia family history burden and reduced premorbid IQ in treatment-resistant schizophrenia: a Swedish National Register and Genomic Study. *Mol Psychiatry*, doi:10.1038/s41380-019-0575-1 (2019).

74 Gale, C. R. *et al.* Is bipolar disorder more common in highly intelligent people? A cohort study of a million men. *Mol Psychiatry* **18**, 190-194, doi:10.1038/mp.2012.26 (2013).

75 Smeland, O. B. *et al.* Genome-wide analysis reveals extensive genetic overlap between schizophrenia, bipolar disorder, and intelligence. *Mol Psychiatry*, doi:10.1038/s41380-018-0332-x (2019).

76 Zeng, L. *et al.* Genetically modulated educational attainment and coronary disease risk. *Eur Heart J* **40**, 2413-2420, doi:10.1093/eurheartj/ehz328 (2019).

77 Stringhini, S. *et al.* Life-course socioeconomic status and DNA methylation of genes regulating inflammation. *Int J Epidemiol* **44**, 1320-1330, doi:10.1093/ije/dyv060 (2015).

78 Huang, J. Y. *et al.* Accounting for Life-Course Exposures in Epigenetic Biomarker Association Studies: Early Life Socioeconomic Position, Candidate Gene DNA Methylation, and Adult Cardiometabolic Risk. *Am J Epidemiol* **184**, 520-531, doi:10.1093/aje/kww014 (2016).

79 Swartz, J. R., Hariri, A. R. & Williamson, D. E. An epigenetic mechanism links socioeconomic status to changes in depression-related brain function in high-risk adolescents. *Mol Psychiatry* **22**, 209-214, doi:10.1038/mp.2016.82 (2017).

80 Song, L. *et al.* Polygenic Score of Subjective Well-Being Is Associated with the Brain Morphology in Superior Temporal Gyrus and Insula. *Neuroscience* **414**, 210-218, doi:10.1016/j.neuroscience.2019.05.055 (2019).

81 Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet* **48**, 624-633, doi:10.1038/ng.3552 (2016).

82 Baselmans, B. M. L. & Bartels, M. A genetic perspective on the relationship between eudaimonic -and hedonic well-being. *Sci Rep* **8**, 14610, doi:10.1038/s41598-018-32638-1 (2018).

83 Day, F. R., Ong, K. K. & Perry, J. R. B. Elucidating the genetic basis of social interaction and isolation. *Nat Commun* **9**, 2457, doi:10.1038/s41467-018-04930-1 (2018).

84 Richardson, J. G. *Handbook of theory and research for the sociology of education*. (Greenwood Press, 1986).

85 Bourdieu, P. *Distinction : a social critique of the judgement of taste*. (Routledge, 2010).

medRxiv preprint doi: https://doi.org/10.1101/2020.02.01.20020008; this version posted February 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

86 Manrique-Garcia, E., Sidorchuk, A., Hallqvist, J. & Moradi, T. Socioeconomic position and incidence of acute myocardial infarction: a meta-analysis. *J Epidemiol Community Health* **65**, 301-309, doi:10.1136/jech.2009.104075 (2011).

87 Stringhini, S. *et al.* Association of socioeconomic position with health behaviors and mortality. *Jama* **303**, 1159-1166, doi:10.1001/jama.2010.297 (2010).

88 Stephen Machin, O. M., Sunčica Vujić. Youth Crime and Education Expansion. *German Economic Review*, doi:10.1111/j.1468-0475.2012.00576.x (2012).

Table 1. Characteristics of included studies of psychiatric and somatic disorders

Disease	Cases, No.	Controls, No.	Population	Used SNPs, No.	Source	Consortium
Anorexia nervosa	16 992	55 525	European	1227	Watson HJ et al	Psychiatric Genomics Consortium
Anxiety	7016	14 745	European	1207	Otowa T et al	Psychiatric Genomics Consortium
Bipolar disorder	20 352	31 358	European	1271	Stahl EA et al	Psychiatric Genomics Consortium
Insomnia	397 959	933 051	European	1194	Jansen PR et al	CNCR
Major depressive disorder	170 756	329 443	European	1244	Stahl EA et al	UK Biobank
Obsessive-compulsive disorder	2688	7037	European	1265	IOCDF-GC and OCGAS	Psychiatric Genomics Consortium
Posttraumatic stress disorder	30 000	170 000	Mix	1271	Nievergelt CM et al	Psychiatric Genomics Consortium
Suicide attempts	6024	44 240	European	1072	Erlangsen A et al	iPSYCH
Schizophrenia	33 426	54 065	European	1253	Psychiatric Genomics Consortium	Psychiatric Genomics Consortium
Neurological disorder						
Amyotrophic lateral sclerosis	21 982	41 944	European	1268	Kunkle BW et al	Project MinE
Alzheimer's disease	12 577	23 475	European	1260	van Rheenen W et al	IGAP
Atrial fibrillation	65 446	522 000	Mix	1261	Roselli C et al	AFGen
Coronary artery disease	60 801	123 504	Mix	1266	Nikpay M et al	CARDIoGRAMplusC4D Consortium
Heart failure	7382	387 652	European	1255	Aragam KG et al	UK Biobank
Total stroke	67 162	454 450	Mix	1258	Malik R et al	MEGASTROKE Consortium
Any ischemic stroke	60 341	NA	Mix	1270	Malik R et al	MEGASTROKE Consortium
Large artery stroke	6688	146 392	Mix	1270	Malik R et al	MEGASTROKE Consortium
Small vessel stroke	11 710	192 662	Mix	1268	Malik R et al	MEGASTROKE Consortium
Cardioembolic stroke	9006	204 570	Mix	1269	Malik R et al	MEGASTROKE Consortium
Intracerebral haemorrhage	1545	1481	Mix	852	Woo D et al	ISGC
Cancer						
Breast cancer	122 977	105 974	Mix	1269	Michailidou K et al	BCAC
Breast cancer ER+	69 501	NA	Mix	1269	Michailidou K et al	BCAC
Breast cancer ER-	21 468	NA	Mix	1269	Michailidou K et al	BCAC
Lung cancer	11 348	15 861	European	1230	Wang Y et al	ILCCO
Ovarian cancer	25 509	40 941	European	1213	Phelan CM et al	OCAC

Prostate cancer	79 194	61 112	European	1175	Schumacher FR et al	PRACTICAL Consortium
Atopic dermatitis	21 399	95 464	Mix	1269	Paternoster L et al	EAGLE Consortium
Chronic kidney disease	41 395	439 303	European	1263	Wuttke M et al	CKDGen Consortium
Fracture	53 184	373 611	European	1259	Morris JA et al	GEFOS Consortium
Gout	13 179	750 634	Mix	1271	Tin A et al	GUGC
Inflammatory bowel disease	25 042	34 915	European	1213	de Lange KM et al	UK IBD consortium
Rheumatoid arthritis	29 880	73 758	Mix	1254	Okada Y et al	GARNET consortium
Type 2 diabetes	74 124	824 006	European	1263	Mahajan A et al	DIAGRAM consortium
Body mass index	NA	694 649	Mix	1260	Pulit SL et al	GIANT consortium
Waist-to-hip ratio	NA	697 734	Mix	1260	Pulit SL et al	GIANT consortium
Blood pressure, systolic	NA	317 754	European	1263	Neale lab	UK Biobank
Blood pressure, diastolic	NA	317 756	European	1263	Neale lab	UK Biobank
Age of smoking initiation	NA	341 427	European	1255	Liu M et al	GSCAN
Cigarettes per day	NA	337 334	European	1255	Liu M et al	GSCAN
Alcohol intake per week	NA	941 280	European	1255	Liu M et al	GSCAN
Bone mineral density	NA	426 824	European	1208	Morris JA et al	GEFOS Consortium
Estimated glomerular filtration rate	NA	765 348	Mix	1264	Wuttke M et al	CKDGen Consortium
Serum urate	NA	288 649	European	1264	Tin A et al	GUGC

AFGen stands for Atrial Fibrillation Consortium; BCAC, Breast Cancer Association Consortium; CNCR, Center for Neurogenomics and Cognitive Research; DIAGRAM, The DIAbetes Genetics Replication And Meta-analysis; EAGLE, The EArly Genetics and Lifecourse Epidemiology; ER, estrogen receptor; GARNET, Genetics and Allied research in Rheumatic diseases Networking; GEFOS, GEnetic Factors for Osteoporosis, GUGC, The Global Urate Genetics Consortium; GSCAN, Consortium of Alcohol and Nicotine use; IGAP, The International Genomics of Alzheimer's Project; ILCCO, The International Lung Cancer Consortium; ISGC, International Stroke Genetics Consortium; NA, Not available; OCAC, The Ovarian Cancer Association Consortium; PRACTICAL, The Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome; SNP, single-nucleotide polymorphism; UK IBD consortium, UK Inflammatory Bowel Disease Genetics Consortium.

Table 2. Mediation analysis to disentangle the effects of body mass index and smoking in the pathway from education level to health outcomes

Health outcome	Total effect of education			Effect after adjusting for BMI				Effect after adjusting for smoking				Effect after adjusting for both			
	OR[a]	95% CI	P	OR[b]	95% CI	P	%*	OR[c]	95% CI	P	%*	OR[d]	95% CI	P	%*
Mental disorder															
Anorexia nervosa	1.78	(1.60, 1.98)	1.56E-26	1.54	(1.36, 1.75)	1.70E-11	25	1.83	(1.63, 2.05)	3.00E-24	0	1.58	(1.39, 1.80)	4.90E-12	21
Insomnia	0.72	(0.69, 0.75)	4.15E-58	0.74	(0.71, 0.78)	2.80E-34	8	0.75	(0.72, 0.79)	1.10E-36	12	0.76	(0.73, 0.80)	1.80E-26	16
Major depressive disorder	0.77	(0.74, 0.80)	1.28E-36	0.80	(0.76, 0.84)	9.60E-20	15	0.82	(0.78, 0.85)	9.40E-19	24	0.83	(0.79, 0.87)	2.80E-13	29
Obsessive-compulsive disorder	2.12	(1.71, 2.62)	5.66E-12	1.70	(1.31, 2.19)	5.50E-05	29	1.99	(1.57, 2.53)	1.10E-08	8	1.67	(1.28, 2.17)	2.00E-04	32
Suicide attempts	0.56	(0.49, 0.64)	5.18E-18	0.64	(0.54, 0.75)	3.60E-08	23	0.61	(0.52, 0.70)	2.10E-11	15	0.67	(0.57, 0.79)	2.50E-06	31
Neurological disease															
Amyotrophic lateral sclerosis	0.88	(0.79, 0.98)	1.50E-02	0.88	(0.78, 1.00)	4.60E-02	0	0.83	(0.74, 0.93)	1.00E-03	0	0.84	(0.74, 0.95)	8.00E-03	0
Alzheimer’s disease	0.72	(0.66, 0.79)	6.60E-13	0.73	(0.65, 0.81)	4.17E-09	4	0.72	(0.65, 0.79)	3.00E-11	0	0.72	(0.64, 0.80)	5.01E-09	0
Cardiovascular disease															
Coronary artery disease	0.62	(0.58, 0.66)	8.54E-53	0.70	(0.65, 0.75)	2.30E-21	25	0.66	(0.61, 0.70)	9.30E-34	13	0.72	(0.67, 0.78)	2.50E-17	31
Heart failure	0.51	(0.46, 0.57)	4.16E-37	0.64	(0.56, 0.72)	3.50E-13	34	0.54	(0.48, 0.60)	1.80E-26	8	0.65	(0.57, 0.73)	1.20E-11	36
Total stroke	0.70	(0.67, 0.74)	1.10E-41	0.75	(0.71, 0.80)	1.10E-19	19	0.71	(0.67, 0.75)	5.40E-33	4	0.76	(0.71, 0.80)	3.70E-18	23
Any ischemic stroke	0.68	(0.64, 0.72)	4.86E-33	0.70	(0.65, 0.76)	8.70E-20	8	0.68	(0.64, 0.73)	2.40E-27	0	0.70	(0.65, 0.76)	1.70E-18	8
Large artery stroke	0.51	(0.43, 0.60)	7.33E-17	0.55	(0.46, 0.67)	1.60E-09	11	0.53	(0.44, 0.63)	9.00E-13	6	0.56	(0.46, 0.69)	1.40E-08	14
Small vessel stroke	0.59	(0.51, 0.68)	3.07E-13	0.61	(0.52, 0.73)	1.40E-08	6	0.59	(0.50, 0.69)	2.00E-11	0	0.61	(0.51, 0.72)	2.60E-08	6
Cancer															
Breast cancer	0.87	(0.83, 0.91)	2.12E-08	0.88	(0.83, 0.94)	4.00E-05	8	0.88	(0.84, 0.93)	8.80E-06	8	0.89	(0.84, 0.95)	3.00E-04	16
Breast cancer ER+	0.89	(0.84, 0.94)	2.34E-05	0.90	(0.84, 0.97)	4.00E-03	10	0.90	(0.85, 0.96)	1.00E-03	10	0.91	(0.85, 0.98)	1.20E-02	19
Breast cancer ER-	0.71	(0.66, 0.77)	6.97E-17	0.70	(0.64, 0.77)	1.80E-13	0	0.72	(0.66, 0.78)	4.20E-17	4	0.71	(0.64, 0.78)	2.50E-12	0
Lung cancer	0.52	(0.47, 0.58)	3.20E-30	0.58	(0.50, 0.66)	1.30E-15	17	0.59	(0.53, 0.67)	6.30E-17	19	0.63	(0.55, 0.72)	2.90E-11	29
Ovarian cancer	0.84	(0.78, 0.92)	7.30E-05	0.90	(0.81, 0.99)	3.60E-02	40	0.82	(0.75, 0.90)	4.00E-05	0	0.88	(0.79, 0.98)	1.40E-02	27
Prostate cancer	1.12	(1.05, 1.20)	1.00E-03	1.06	(0.98, 1.15)	1.45E-01	49	1.07	(1.00, 1.15)	6.00E-02	40	1.04	(0.96, 1.13)	3.23E-01	65
Other disease															

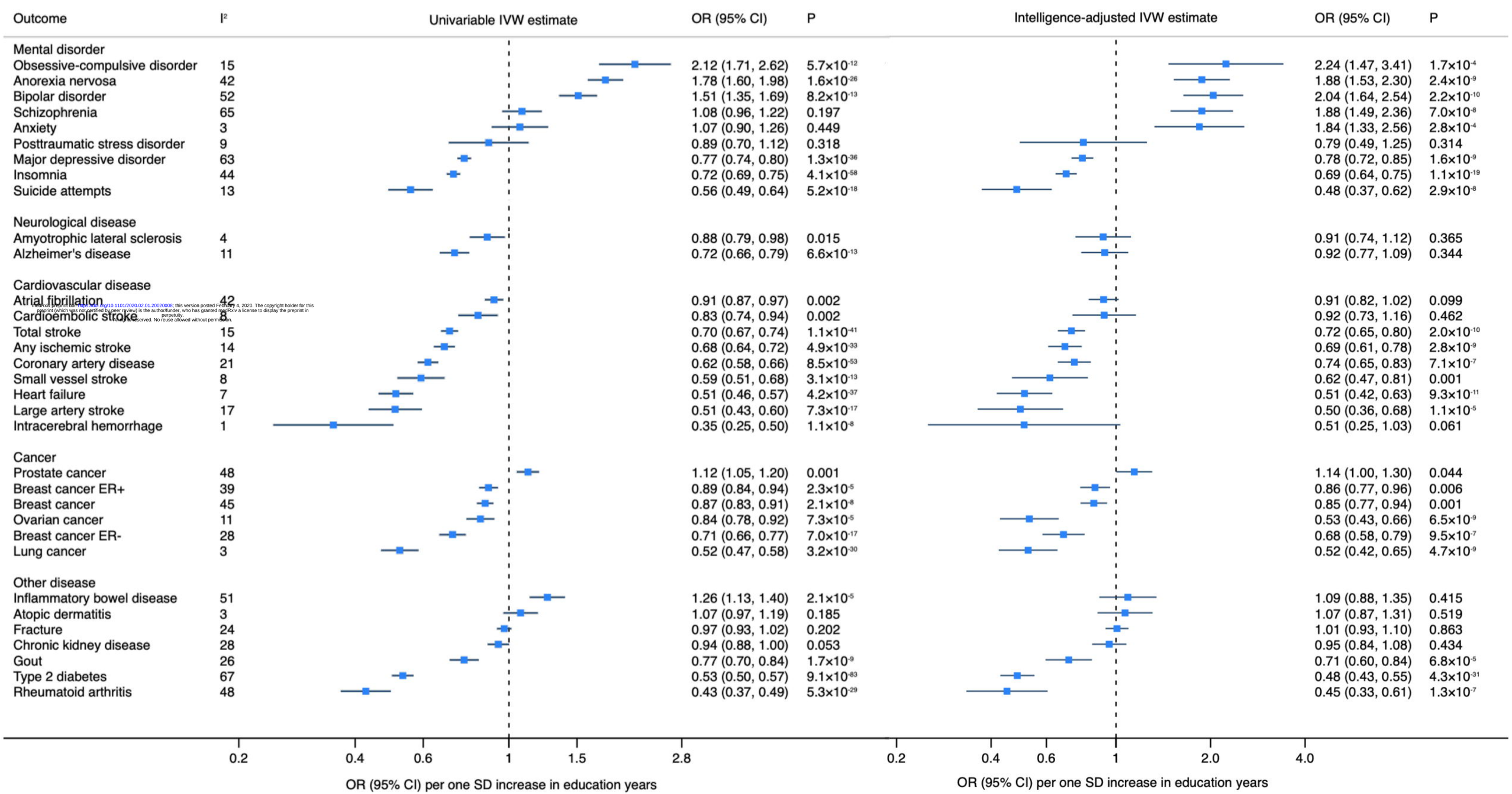
Gout	0.77	(0.70, 0.84)	1.75E-09	0.92	(0.84, 1.02)	1.35E-01	68	0.76	(0.69, 0.84)	4.40E-08	0	0.91	(0.81, 1.01)	6.70E-02	64
Rheumatoid arthritis	0.43	(0.37, 0.49)	5.29E-29	0.41	(0.34, 0.49)	1.50E-21	0	0.44	(0.37, 0.51)	7.90E-23	3	0.42	(0.35, 0.51)	1.40E-19	0
Type 2 diabetes	0.53	(0.50, 0.57)	9.07E-83	0.76	(0.71, 0.81)	1.20E-12	57	0.60	(0.56, 0.64)	3.30E-48	20	0.79	(0.74, 0.85)	6.30E-11	63

BMI indicates body mass index; ER, estrogen receptor.
[a] total effect without any adjustment;
[b] adjusted for the effect of body mass index;
[c] adjusted for the effect of smoking (cigarettes per day);
[d] adjusted for the effects of both body mass index and smoking behaviors;
*Percentage of the effect of education on the health outcome that is mediated by body mass index, smoking, or both (Formula: log(OR_total)-log(OR_adjusted)/log(OR_total)*100)). We replaced the values with zero for those percentage below zero.

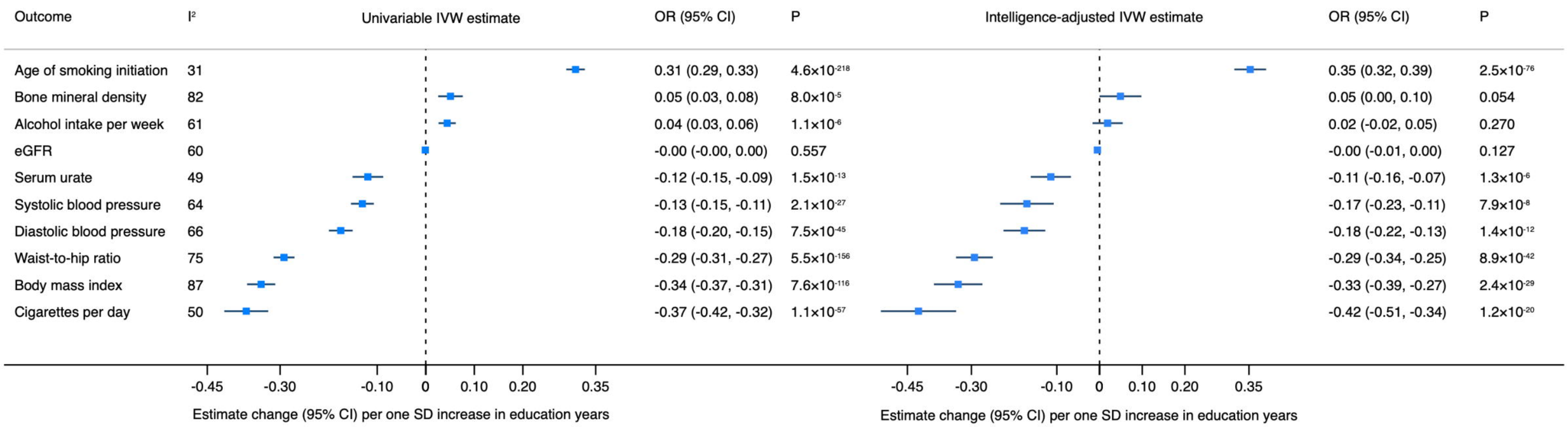
Figure legends

Figure 1. Associations of genetic predisposition to higher education level with diseases in MR analyses without and with adjustment for genetically predicted intelligence. CI indicates confidence interval; ER, estrogen receptor; IVW, inverse-variance weighted; OR, odds ratio; SD, standard deviation. I^2 represents the degree of heterogeneity among included SNPs for education.

Figure 2. Associations genetic predisposition to higher education level and health-related risk factors in MR analyses without and with adjustment for genetically predicted intelligence. CI indicates confidence interval; eGFR, estimated glomerular filtration rate; IVW, inverse-variance weighted; SD, standard deviation. I^2 represents the degree of heterogeneity among included SNPs for education.



medRxiv preprint doi: https://doi.org/10.1101/2020.02.01.20020008; this version posted February 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.


Outcome
I²
Univariable IVW estimate
OR (95% CI)
P
Intelligence-adjusted IVW estimate
OR (95% CI)
P
Age of smoking initiation 31 0.31 (0.29, 0.33) 4.6×10-218 0.35 (0.32, 0.39) 2.5×10-76
Bone mineral density 82 0.05 (0.03, 0.08) 8.0×10-5 0.05 (0.00, 0.10) 0.054
Alcohol intake per week 61 0.04 (0.03, 0.06) 1.1×10-6 0.02 (-0.02, 0.05) 0.270
eGFR 60 -0.00 (-0.00, 0.00) 0.557 -0.00 (-0.01, 0.00) 0.127
Serum urate 49 -0.12 (-0.15, -0.09) 1.5×10-13 -0.11 (-0.16, -0.07) 1.3×10-6
Systolic blood pressure 64 -0.13 (-0.15, -0.11) 2.1×10-27 -0.17 (-0.23, -0.11) 7.9×10-8
Diastolic blood pressure 66 -0.18 (-0.20, -0.15) 7.5×10-45 -0.18 (-0.22, -0.13) 1.4×10-12
Waist-to-hip ratio 75 -0.29 (-0.31, -0.27) 5.5×10-156 -0.29 (-0.34, -0.25) 8.9×10-42
Body mass index 87 -0.34 (-0.37, -0.31) 7.6×10-116 -0.33 (-0.39, -0.27) 2.4×10-29
Cigarettes per day 50 -0.37 (-0.42, -0.32) 1.1×10-57 -0.42 (-0.51, -0.34) 1.2×10-20
-0.45 -0.30 -0.10 0 0.10 0.20 0.35
Estimate change (95% CI) per one SD increase in education years
-0.45 -0.30 -0.10 0 0.10 0.20 0.35
Estimate change (95% CI) per one SD increase in education years