



OPEN

Genetically predicted education attainment in relation to somatic and mental health

Shuai Yuan¹, Ying Xiong², Madeleine Michaëlsson³, Karl Michaëlsson⁴ & Susanna C. Larsson^{1,4}✉

A deeper understanding of the causal links from education level to health outcomes may shed a light for disease prevention. In the present Mendelian randomization study, we found that genetically higher education level was associated with lower risk of major mental disorders and most somatic diseases, independent of intelligence. Higher education level adjusted for intelligence was associated with lower risk of suicide attempts, insomnia, major depressive disorder, heart failure, stroke, coronary artery disease, lung cancer, breast cancer, type 2 diabetes and rheumatoid arthritis but with higher risk of obsessive–compulsive disorder, anorexia nervosa, anxiety, bipolar disorder and prostate cancer. Higher education level was associated with reduced obesity and smoking, which mediated quite an extent of the associations between education level and health outcomes. These findings emphasize the importance of education to reduce the burden of common diseases.

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 depression¹, age-related cognitive decline², suicide³, cardiovascular disease⁴, cancer⁵, 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 (i.e., education level), Mendelian randomization (MR) can strengthen the causal inference of an exposure–outcome association⁹. 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 analysis⁹. 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 development⁹.

Several previous MR studies revealed possible causal associations of genetically higher education level with health outcomes, such as Alzheimer's disease^{10,11}, diabetes¹², cardiovascular disease^{13–15}, cancer¹⁶, myopia⁸, chronic kidney disease¹⁷, amyotrophic lateral sclerosis¹⁸, and longevity¹⁹. In a Mendelian randomization study based on UK Biobank, however, the pattern of the protective effect of higher genetically predicted education level on a broad range of health endpoints was unclear²⁰. Education is an upstream health determinant that influences social and community networks and individual lifestyle factors, thereby affecting the risk of various health outcomes. Higher educational attainment has been established to have direct effects on income, alcohol consumption, and physical activity, and inverse associations with smoking, BMI and sedentary behavior. Although genetically proxied higher education level has been associated with several diseases²⁰, whether education exerts causal effects on a wide spectrum of health outcomes remains unknown.

Here, 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 associated with the same health outcomes independently of education. We additionally investigated the associations of education level and intelligence with modifiable health-related risk factors. Given that obesity and

¹Unit of Cardiovascular and Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Nobelsväg 13, 17177 Stockholm, Sweden. ²Department of Public Health Sciences, Karolinska Institutet, Stockholm, Sweden. ³Department of Education, Health and Social Studies, Dalarna University, Falun, Sweden. ⁴Unit of Medical Epidemiology, Department of Surgical Sciences, Uppsala University, Dag Hammarskjölds Väg 14B, 75185 Uppsala, Sweden. ✉email: susanna.larsson@surgsci.uu.se

smoking influence the risk of many diseases^{21–25}, we examined whether these two factors mediate the pathway from education to health outcomes.

Materials and methods

Study design. The design and hypothesis of the present study are displayed in Supplementary Fig. 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, 19 major somatic diseases, body mass index and cigarette smoking. A systematic review was conducted to find meta-analyses of observational studies of education level and diseases (Supplementary Table 1). All GWASs had been approved by a relevant ethical review board and participants had given informed consent. No individual-level data were used in the present MR study. This MR study was approved by the Swedish Ethical Review Authority.

Selection of instrumental variables. Instrumental variables for education level and intelligence were identified from GWASs of, respectively, 1,131,881 and 269,867 individuals of European ancestries^{26,27}. In total, 1271 and 205 single-nucleotide polymorphisms (SNPs) at the genome-wide significance threshold ($p < 5 \times 10^{-8}$) were identified to be associated with education level and intelligence, respectively. Independent SNPs ($r^2 < 0.01$ and clump window > 10 kb) without linkage disequilibrium were proposed as instrumental variables. Linkage disequilibrium among SNPs was calculated based on 1000 genomes LD reference panel (European population) using the PLINK clumping method. Possible palindromic SNPs were excluded. We used 663 and 178 SNPs as instrumental variables for education level and intelligence, respectively. The same instrumental variables were used in the multivariable MR analyses. 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. 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, 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.

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^{28–38}, 9 cardiovascular diseases^{39–43}, 3 major cancers^{44–46}, 7 other diseases^{47–53}, body mass index⁵⁴ and cigarette smoking⁵⁵. We did not find proxies for exposure-associated SNPs that were unavailable in the outcome datasets given that the percentage of missing SNPs for most outcomes were minimal and a few missing SNPs was not likely to bias the results based on hundreds of SNPs. 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 outcomes 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 disorders⁵⁶, suicide attempts⁵⁷, posttraumatic stress disorder⁵⁸, amyotrophic lateral sclerosis⁵⁹, Alzheimer’s disease⁶⁰, coronary artery disease⁶¹, heart failure⁶², stroke⁶³, breast cancer⁶⁴, prostate cancer⁶⁵, lung cancer⁶⁶, type 2 diabetes⁶⁷, chronic kidney disease⁶⁸ and body mass index⁶⁹. We extracted publication data (PubMed identifier 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 valid⁷⁰. The MR-Egger regression can detect and adjust for pleiotropy albeit rendering low precision of the estimates⁷¹. The false discovery rate method was used to adjust for multiple testing (Supplementary Tables 3 and 4). For associations that survived multiple testing, we used the multivariable MR method⁷² to disentangle the causal effect of education level on outcomes independent of intelligence and vice versa. For intelligence-adjusted inverse associations that survived multiple testing, we also used the multivariable MR analysis with adjustment for body mass index and smoking to explore the mediation effects of these factors on the associations between education and health outcomes. We performed several multivariable MR analyses to test the mediation effect from education, intelligence, body mass index or smoking, rather than allowing for independent effects as well as mediations by these factors in one MVMR model simultaneously.

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. All statistical analyses were two-sided and performed using the *mrrobust* package⁷³ in Stata/SE 15.0 (StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC.) and *TwoSampleMR*⁷⁴ in R Software 3.6.0 (R Core Team. R Foundation for Statistical Computing, Vienna, Austria. 2019. <https://www.R-project.org>).

Outcome	Cases, No	Controls, No	Population	Study source	Data source
Mental disorder					
Anorexia nervosa	16,992	55,525	European	Watson HJ et al.	Psychiatric Genomics Consortium
Anxiety	7016	14,745	European	Otowa T et al.	Psychiatric Genomics Consortium
Bipolar disorder	20,352	31,358	European	Stahl EA et al.	Psychiatric Genomics Consortium
Insomnia	397,959	933,051	European	Jansen PR et al.	CNCR
Major depressive disorder	170,756	329,443	European	Stahl EA et al.	UK Biobank
Obsessive–compulsive disorder	2688	7037	European	IOCDF-GC and OCGAS	Psychiatric Genomics Consortium
Posttraumatic stress disorder	30,000	170,000	Mix	Nievergelt CM et al.	Psychiatric Genomics Consortium
Suicide attempts	6024	44,240	European	Erlangen A et al.	iPSYCH
Schizophrenia	33,426	54,065	European	Psychiatric Genomics Consortium	Psychiatric Genomics Consortium
Neurological disease					
Amyotrophic lateral sclerosis	21,982	41,944	European	Kunkle BW et al.	Project MinE
Alzheimer's disease	12,577	23,475	European	van Rheenen W et al.	IGAP
Cardiovascular disease					
Atrial fibrillation	65,446	522,000	Mix	Roselli C et al.	AFGen
Coronary artery disease	60,801	123,504	Mix	Nikpay M et al.	CARDIoGRAMplusC4D Consortium
Heart failure	7382	387,652	European	Aragam KG et al.	UK Biobank
Total stroke	67,162	454,450	Mix	Malik R et al.	MEGASTROKE Consortium
Any ischemic stroke	60,341	NA	Mix	Malik R et al.	MEGASTROKE Consortium
Large artery stroke	6688	146,392	Mix	Malik R et al.	MEGASTROKE Consortium
Small vessel stroke	11,710	192,662	Mix	Malik R et al.	MEGASTROKE Consortium
Cardioembolic stroke	9006	204,570	Mix	Malik R et al.	MEGASTROKE Consortium
Intracerebral haemorrhage	1545	1481	Mix	Woo D et al.	ISGC
Cancer					
Breast cancer	122,977	105,974	Mix	Michailidou K et al.	BCAC
Breast cancer ER+	69,501	NA	Mix	Michailidou K et al.	BCAC
Breast cancer ER–	21,468	NA	Mix	Michailidou K et al.	BCAC
Lung cancer	11,348	15,861	European	Wang Y et al.	ILCCO
Prostate cancer	79,194	61,112	European	Schumacher FR et al.	PRACTICAL Consortium
Other disease					
Atopic dermatitis	21,399	95,464	Mix	Paternoster L et al.	EAGLE Consortium
Chronic kidney disease	41,395	439,303	European	Wuttke M et al.	CKDGen Consortium
Fracture	53,184	373,611	European	Morris JA et al.	GEFOS Consortium
Gout	13,179	750,634	Mix	Tin A et al.	GUGC
Inflammatory bowel disease	25,042	34,915	European	de Lange KM et al.	UK IBD consortium
Rheumatoid arthritis	29,880	73,758	Mix	Okada Y et al.	GARNET consortium
Type 2 diabetes	74,124	824,006	European	Mahajan A et al.	DIAGRAM consortium
Risk factor					
Body mass index	NA	694,649	Mix	Pulit SL et al.	GIANT consortium
Cigarettes per day	NA	337,334	European	Liu M et al.	GSCAN

Table 1. Characteristics of included studies of mental disorders, somatic diseases, and health-related risk factors. *AFGen* 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, *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.

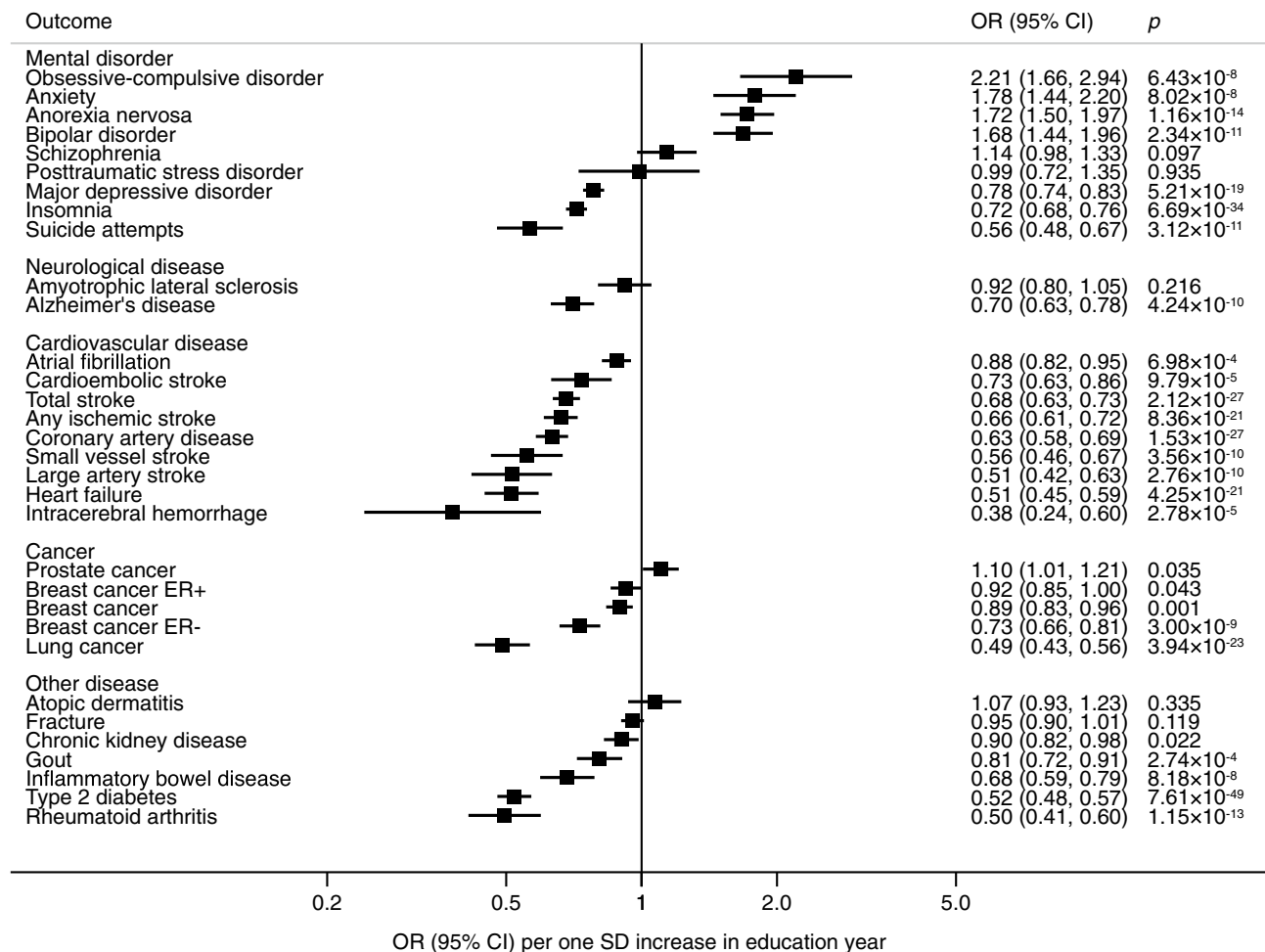


Figure 1. Associations of genetic predisposition to higher education level with health outcomes in univariable MR analyses. *CI* confidence interval, *ER* oestrogen receptor, *IVW* inverse-variance weighted, *OR* odds ratio, *SD* standard deviation.

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

Informed consent. All participants included in the genome-wide association studies gave informed consent.

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 3 studied cancers, and 5 out of 7 other common diseases in the univariable inverse-variance weighted MR analysis (Fig. 1 and Supplementary Table 3). In the multivariable inverse-variance weighted analysis, the associations of education level with Alzheimer's disease, atrial fibrillation, cardioembolic stroke, intracerebral haemorrhage, chronic kidney disease, gout, and inflammatory bowel disease did not remain after adjustment for intelligence (Fig. 2 and Supplementary Table 3). Results of sensitivity analyses were directionally similar but with wider CIs (Supplementary Table 5).

Genetically predicted intelligence and diseases. The associations between intelligence and outcomes are presented in Supplementary Tables 4, 6 and 7. Genetically predicted intelligence showed associations with obsessive-compulsive disorder, anorexia nervosa, schizophrenia, insomnia, suicide attempts, Alzheimer's disease, coronary artery disease, breast cancer (ER-), lung cancer, type 2 diabetes and rheumatoid arthritis in the univariable MR analyses (Supplementary Tables 4 and 6). After adjustment for genetically predicted education level, only the association with schizophrenia persisted (Supplementary Tables 4 and 7).

Education, intelligence, body mass index and smoking. Genetically predicted higher education level was associated with lower body mass index and fewer cigarettes smoked per day in the univariable model; the estimates were similar in the intelligence-adjusted model (Fig. 3 and Supplementary Tables 3 and 8). Find-

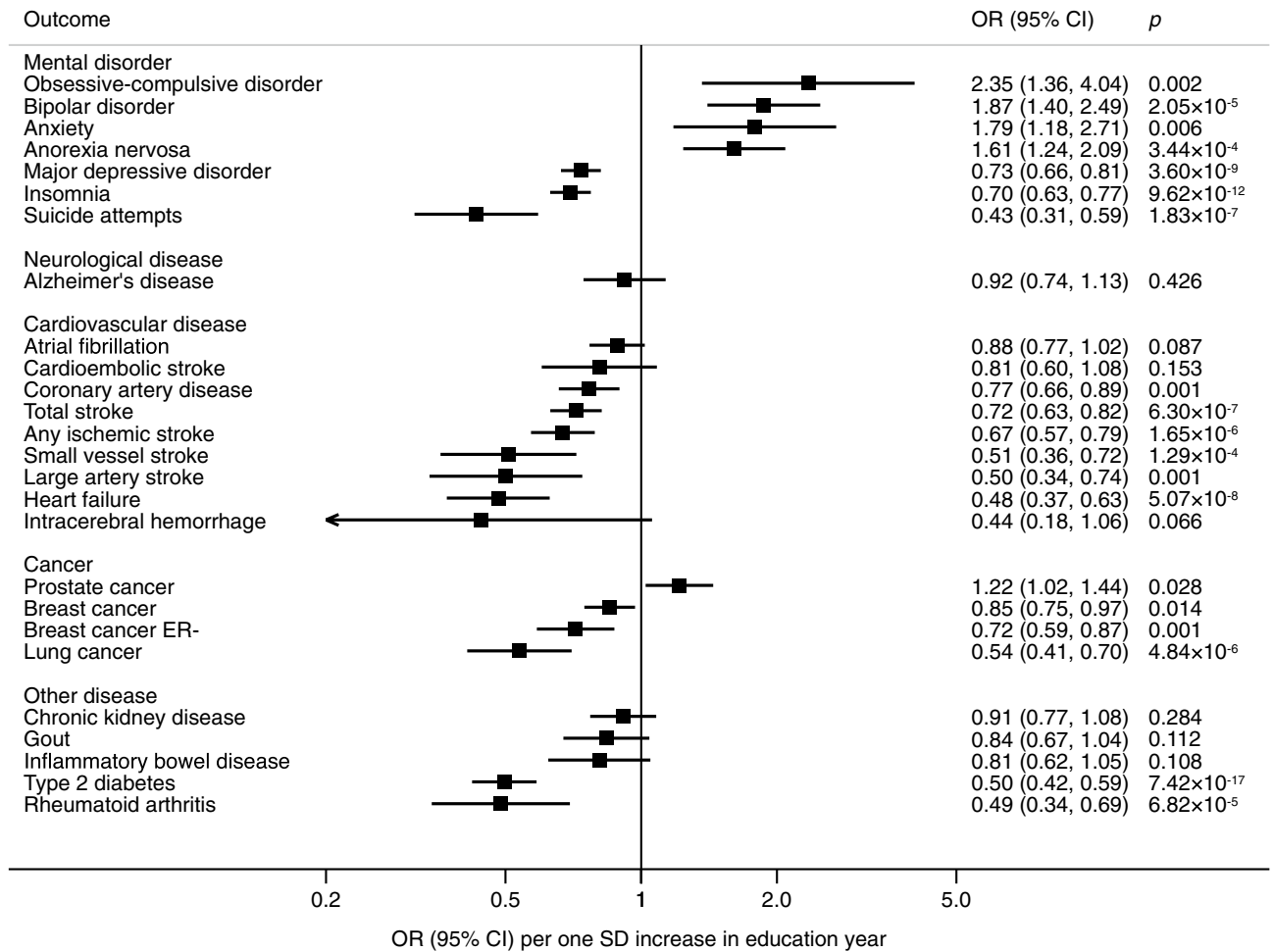


Figure 2. Associations of genetic predisposition to higher education level with health outcomes in multivariable MR analyses with adjustment for genetically predicted intelligence. *CI* confidence interval, *ER* oestrogen receptor, *IVW* inverse-variance weighted, *OR* odds ratio, *SD* standard deviation.

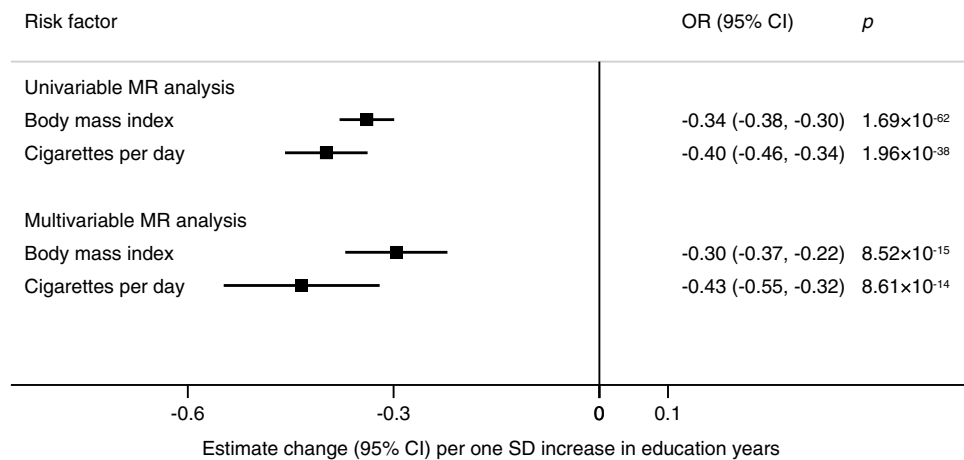


Figure 3. Associations of genetic predisposition to higher education level with body mass index and smoking in MR analyses without and with adjustment for genetically predicted intelligence. *CI* confidence interval, *IVW* inverse-variance weighted, *SD* standard deviation.

Outcomes	Total effect of education		Effect after adjusting for BMI			Effect after adjusting for smoking			Effect after adjusting for both		
	OR ^a	95% CI	OR ^b	95% CI	%*	OR ^c	95% CI	%*	OR ^d	95% CI	%*
Mental disorder											
Insomnia	0.72	0.68, 0.76	0.75	0.71, 0.80	12	0.74	0.69, 0.79	8	0.77	0.72, 0.82	20
Major depressive disorder	0.78	0.74, 0.83	0.85	0.80, 0.90	35	0.83	0.77, 0.88	25	0.87	0.81, 0.93	44
Suicide attempts	0.56	0.48, 0.67	0.60	0.50, 0.72	12	0.67	0.54, 0.82	31	0.68	0.55, 0.84	33
Cardiovascular disease											
Coronary artery disease	0.63	0.58, 0.69	0.67	0.61, 0.73	13	0.72	0.65, 0.79	29	0.74	0.66, 0.82	35
Heart failure	0.51	0.45, 0.59	0.55	0.47, 0.64	11	0.64	0.54, 0.75	34	0.65	0.55, 0.77	36
Total stroke	0.68	0.63, 0.73	0.70	0.64, 0.75	8	0.75	0.69, 0.81	25	0.76	0.69, 0.82	29
Any ischemic stroke	0.66	0.61, 0.72	0.67	0.61, 0.74	4	0.69	0.63, 0.77	11	0.70	0.63, 0.78	14
Large artery stroke	0.51	0.42, 0.63	0.54	0.43, 0.68	8	0.59	0.46, 0.76	22	0.59	0.46, 0.77	22
Small vessel stroke	0.56	0.46, 0.67	0.54	0.44, 0.67	0	0.54	0.43, 0.67	0	0.53	0.42, 0.66	0
Cancer											
Breast cancer	0.89	0.83, 0.96	0.92	0.85, 1.00	28	0.93	0.85, 1.01	38	0.94	0.86, 1.03	47
Breast cancer ER–	0.73	0.66, 0.81	0.74	0.66, 0.83	4	0.71	0.63, 0.80	0	0.71	0.63, 0.81	0
Lung cancer	0.49	0.43, 0.56	0.56	0.48, 0.65	19	0.53	0.45, 0.63	11	0.59	0.49, 0.70	26
Other diseases											
Rheumatoid arthritis	0.50	0.41, 0.60	0.44	0.36, 0.53	0	0.41	0.34, 0.50	0	0.42	0.34, 0.52	0
Type 2 diabetes	0.52	0.48, 0.57	0.60	0.55, 0.64	22	0.75	0.70, 0.81	56	0.79	0.73, 0.85	64

Table 2. Mediation analysis to disentangle the effects of body mass index and smoking in the pathway from education level to health outcomes. *BMI* body mass index, *ER* estrogen receptor. *Percentage of the effect of education on the health outcome that is mediated by body mass index, smoking, or both (Formula: $(\log(\text{OR}_{\text{total}}) - \log(\text{OR}_{\text{adjusted}})) / \log(\text{OR}_{\text{total}}) \times 100$). We replaced the values with zero for those percentage below zero. ^aTotal effect without any adjustment. ^bAdjusted for the effect of body mass index. ^cAdjusted for the effect of smoking (cigarettes per day). ^dAdjusted for the effects of both body mass index and smoking behaviours.

ings were consistent in sensitivity analyses and no pleiotropy was observed (Supplementary Table 5). Genetically predicted intelligence was not associated with body mass index or smoking (Supplementary Tables 4, 6 and 7).

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 9). 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. After adjustment for both body mass index and smoking, the direct causal effect of education on the outcomes was substantially attenuated for type 2 diabetes (64%), major depressive disease (44%), heart failure (36%) and coronary artery disease (35%).

Discussion

In the present MR study, genetic predisposition to higher education level was causally associated with the majority of major health outcomes, body mass index and smoking. Specifically, genetic predisposition to higher education level, independent of intelligence, was associated with lower risk of major depressive disorder, insomnia, suicide attempts, coronary artery disease, stroke, heart failure, breast cancer, lung cancer, type 2 diabetes, and rheumatoid arthritis. Conversely, higher education level was associated with higher risk of obsessive-compulsive disorder, bipolar disorder, anxiety, anorexia nervosa and prostate cancer. Genetically predicted higher intelligence, independent of education, was inversely related to schizophrenia. Body mass index and smoking displayed strongest mediation effects observed for type 2 diabetes, major depressive disease, heart failure and coronary artery disease.

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 disorder⁵⁶, Alzheimer's disease⁶⁰, coronary heart disease⁶¹, heart failure⁶², stroke⁶³, lung cancer⁶⁶, type 2 diabetes⁶⁷, chronic kidney disease⁶⁸, and obesity⁶⁹. 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 cancer⁶⁴, and the sample size was small for prostate cancer⁶⁵. 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 level⁷⁵. With regard to the inverse association of higher education level with breast cancer, the association 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 or breast cancer, but a protective causal effect of higher education on lung cancer risk has been reported recently¹⁶.

Previous MR studies showed a protective effect of higher educational level on Alzheimer's disease^{10,11}, type 2 diabetes¹², cardiovascular disease^{13–15}, lung cancer¹⁶, myopia⁸, chronic kidney disease¹⁷ and amyotrophic lateral sclerosis¹⁸. The present study using a larger body of SNPs as instrumental variables more precisely verified these findings and expanded the map of other health benefits of improved education level. Notably, the effects of high education level in some 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. For Alzheimer's disease, we found that higher intelligence rather than education level may be the protective factor. 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 factors²⁰. Findings of other MR studies of education level in relation to obesity⁷⁶ and cigarette smoking⁷⁷ are consistent with our findings.

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^{15,78}; (2) there may be direct effects from education-related brain structures or function change via gene methylation, gene silencing etc.^{79–81}, especially for mental and neurological disorders; and (3) subjective well-being, happiness and meaning of life influenced by education level exerts effects on mental and somatic diseases directly or indirectly^{82–85}. Education, as measured in this study, can be defined as an institutionalized form of social resource, and more specifically a form of cultural capital 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^{86,87}. 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^{88,89}. In the present study, genetically predicted education level was associated with a favourable risk factor profile: with improved smoking behaviours as well as lower adiposity, 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.

Previous studies have found that low education level might influence the changes in biochemical response and risk-related brain function, such as inflammation⁷⁹, cardiometabolic traits⁸⁰, and amygdala reactivity⁸¹, via gene methylation, thereby influencing disease risk. In addition, genetic studies have also revealed that improvement of subjective well-being^{82,83}, happiness^{82,83}, meaning of life⁸⁴, social interaction⁸⁵, 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 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, 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 may lead to longevity through mechanisms beyond health literacy pathways of late-onset diseases and beyond the influence of social and material factors.

Strengths. 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 using genetic data from large-scale GWASs and genetic consortia. We used SNPs deriving from a larger GWAS with around 1.1 million individuals as instrumental variables for education level, 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.

Limitations. The major limitation of the present study is the possible unbalanced horizontal pleiotropy aroused from used genetic variants marking more generic biological pathways. It has been found that the lead SNPs related to education level and intelligence are significantly overexpressed in the central nervous system, such as hippocampus and cerebral cortex, but not in other organs²⁶. 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 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 the brain or central nervous system, studies found no, or at most a small, genetic correlation between lower education attainment and mental and neurological disorders by using bivariate genomic-relationship-matrix restricted maximum likelihood analysis¹. 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 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.

There was sample overlap in some analyses, which might have resulted in model overfitting if the SNP-education associations were estimated in studies that were included in the outcome dataset. In addition, the present MR study based on summary-level data could not assess potential non-linear associations of genetically predicted education level and intelligence with outcomes. The effect size of our MR findings could not be quantitatively comparable to other studies given that we scaled the associations to one standard deviation increase in genetically predicted education and intelligence, which might differ from study to study.

In summary, the present MR study strengthened the evidence of protective role of high education level on the majority of mental disorders and somatic diseases independent of intelligence. Body mass index and smoking partly mediated several of the associations between education level and health outcomes.

Data availability

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

Received: 9 October 2020; Accepted: 8 February 2021

Published online: 22 February 2021

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. <https://doi.org/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. <https://doi.org/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**, 194. <https://doi.org/10.1136/bmj.194> (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. [https://doi.org/10.1016/s2214-109x\(19\)30045-2](https://doi.org/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/10.1136/bmj.330.7499.1076> (2005).
10. Larsson, S. C. *et al.* Modifiable pathways in Alzheimer's disease: Mendelian randomisation analysis. *BMJ* **359**, j5375. <https://doi.org/10.1136/bmj.j5375> (2017).
11. Anderson, E. L. *et al.* Education, intelligence and Alzheimer's disease: Evidence from a multivariable two-sample Mendelian randomization study. *Int. J. Epidemiol.* **49**, 1163–1172. <https://doi.org/10.1093/ije/dyz280> (2020).
12. Liang, J. *et al.* Educational attainment protects against type 2 diabetes independently of cognitive performance: A Mendelian randomization study. *Acta Diabetol.* <https://doi.org/10.1007/s00592-020-01647-w> (2021).
13. Liao, L. Z., Zhuang, X. D., Zhang, S. Z., Liao, X. X. & Li, W. D. Education and heart failure: New insights from the atherosclerosis risk in communities study and mendelian randomization study. *Int. J. Cardiol.* <https://doi.org/10.1016/j.ijcard.2020.09.068> (2020).
14. Gill, D., Efstathiadou, A., Cawood, K., Tzoulaki, I. & Dehghan, A. Education protects against coronary heart disease and stroke independently of cognitive function: Evidence from Mendelian randomization. *Int. J. Epidemiol.* **48**, 1468–1477. <https://doi.org/10.1093/ije/dyz200> (2019).

15. Carter, A. R. *et al.* Understanding the consequences of education inequality on cardiovascular disease: Mendelian randomisation study. *BMJ* **365**, 1855. <https://doi.org/10.1136/bmj.11855> (2019).
16. Zhou, H. *et al.* Education and lung cancer: A Mendelian randomization study. *Int. J. Epidemiol.* **48**, 743–750. <https://doi.org/10.1093/ije/dyz121> (2019).
17. Park, S. *et al.* Causal effects of education on chronic kidney disease: A Mendelian randomization study. *Clin. Kidney J.* <https://doi.org/10.1093/ckj/sfaa240> (2020).
18. Zhang, L., Tang, L., Xia, K., Huang, T. & Fan, D. Education, intelligence, and amyotrophic lateral sclerosis: A Mendelian randomization study. *Ann. Clin. Transl. Neurol.* **7**, 1642–1647. <https://doi.org/10.1002/acn3.51156> (2020).
19. van Oort, S., Beulens, J. W. J., van Ballegooijen, A. J., Burgess, S. & Larsson, S. C. Cardiovascular risk factors and lifestyle behaviours in relation to longevity: A Mendelian randomization study. *J. Intern. Med.* <https://doi.org/10.1111/joim.13196> (2020).
20. Davies, N. M. *et al.* Multivariable two-sample Mendelian randomization estimates of the effects of intelligence and education on health. *Elife.* <https://doi.org/10.7554/eLife.43990> (2019).
21. Larsson, S. C. *et al.* Smoking, alcohol consumption, and cancer: A mendelian randomisation study in UK Biobank and international genetic consortia participants. *PLoS Med.* **17**, e1003178. <https://doi.org/10.1371/journal.pmed.1003178> (2020).
22. Larsson, S. C. *et al.* Genetic predisposition to smoking in relation to 14 cardiovascular diseases. *Eur. Heart J.* **41**, 3304–3310. <https://doi.org/10.1093/eurheartj/ehaa193> (2020).
23. Yuan, S., Yao, H. & Larsson, S. C. Associations of cigarette smoking with psychiatric disorders: Evidence from a two-sample Mendelian randomization study. *Sci. Rep.* **10**, 13807. <https://doi.org/10.1038/s41598-020-70458-4> (2020).
24. Larsson, S. C., Bäck, M., Rees, J. M. B., Mason, A. M. & Burgess, S. Body mass index and body composition in relation to 14 cardiovascular conditions in UK Biobank: A Mendelian randomization study. *Eur. Heart J.* **41**, 221–226. <https://doi.org/10.1093/eurheartj/ehz388> (2020).
25. Millard, L. A. C., Davies, N. M., Tilling, K., Gaunt, T. R. & Davey Smith, G. Searching for the causal effects of body mass index in over 300,000 participants in UK Biobank, using Mendelian randomization. *PLoS Genet.* **15**, e1007951. <https://doi.org/10.1371/journal.pgen.1007951> (2019).
26. 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. <https://doi.org/10.1038/s41588-018-0147-3> (2018).
27. 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. <https://doi.org/10.1038/s41588-018-0152-6> (2018).
28. Stahl, E. A. *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803. <https://doi.org/10.1038/s41588-019-0397-8> (2019).
29. Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell* **173**, 1705–1715.e1716. <https://doi.org/10.1016/j.cell.2018.05.046> (2018).
30. 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. <https://doi.org/10.1038/s41593-018-0326-7> (2019).
31. 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. <https://doi.org/10.1038/s41588-019-0358-2> (2019).
32. Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol. Psychiatry* **23**, 1181–1188. <https://doi.org/10.1038/mp.2017.154> (2018).
33. 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. <https://doi.org/10.1038/ng.3622> (2016).
34. 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. <https://doi.org/10.1038/s41467-019-12576-w> (2019).
35. Otowa, T. *et al.* Meta-analysis of genome-wide association studies of anxiety disorders. *Mol. Psychiatry* **21**, 1391–1399. <https://doi.org/10.1038/mp.2015.197> (2016).
36. 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. <https://doi.org/10.1038/s41588-018-0333-3> (2019).
37. 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. <https://doi.org/10.1038/s41588-019-0439-2> (2019).
38. Erlangsen, A. *et al.* Genetics of suicide attempts in individuals with and without mental disorders: A population-based genome-wide association study. *Mol. Psychiatry* <https://doi.org/10.1038/s41380-018-0218-y> (2018).
39. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130. <https://doi.org/10.1038/ng.3396> (2015).
40. Malik, R. *et al.* Multi-ancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537. <https://doi.org/10.1038/s41588-018-0058-3> (2018).
41. Aragam, K. G. *et al.* Phenotypic refinement of heart failure in a national biobank facilitates genetic discovery. *Circulation* <https://doi.org/10.1161/circulationaha.118.035774> (2018).
42. Roselli, C. *et al.* Multi-ethnic genome-wide association study for atrial fibrillation. *Nat. Genet.* **50**, 1225–1233. <https://doi.org/10.1038/s41588-018-0133-9> (2018).
43. Woo, D. *et al.* Meta-analysis of genome-wide association studies identifies Iq22 as a susceptibility locus for intracerebral hemorrhage. *Am. J. Hum. Genet.* **94**, 511–521. <https://doi.org/10.1016/j.ajhg.2014.02.012> (2014).
44. Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92–94. <https://doi.org/10.1038/nature24284> (2017).
45. 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. <https://doi.org/10.1038/s41588-018-0142-8> (2018).
46. Wang, Y. *et al.* Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nat. Genet.* **46**, 736–741. <https://doi.org/10.1038/ng.3002> (2014).
47. 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. <https://doi.org/10.1038/s41588-018-0241-6> (2018).
48. Tin, A. *et al.* Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels. *Nat. Genet.* <https://doi.org/10.1038/s41588-019-0504-x> (2019).
49. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381. <https://doi.org/10.1038/nature12873> (2014).
50. 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. <https://doi.org/10.1038/ng.3760> (2017).
51. Morris, J. A. *et al.* An atlas of genetic influences on osteoporosis in humans and mice. *Nat. Genet.* **51**, 258–266. <https://doi.org/10.1038/s41588-018-0302-x> (2019).
52. Wuttke, M. *et al.* A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat. Genet.* **51**, 957–972. <https://doi.org/10.1038/s41588-019-0407-x> (2019).
53. 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. <https://doi.org/10.1038/ng.3424> (2015).

54. 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. <https://doi.org/10.1093/hmg/ddy327> (2019).
55. 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. <https://doi.org/10.1038/s41588-018-0307-5> (2019).
56. 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. <https://doi.org/10.2190/PM.40.1.i> (2010).
57. Li, Y. & Cao, J. Factors associated with suicidal behaviors in mainland China: A meta-analysis. *BMC Public Health* **12**, 524. <https://doi.org/10.1186/1471-2458-12-524> (2012).
58. 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.* <https://doi.org/10.3390/ijerph14121537> (2017).
59. 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. <https://doi.org/10.1016/j.neuro.2016.06.015> (2017).
60. Xu, W. *et al.* Meta-analysis of modifiable risk factors for Alzheimer's disease. *J. Neurol. Neurosurg. Psychiatry* **86**, 1299–1306. <https://doi.org/10.1136/jnnp-2015-310548> (2015).
61. 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. <https://doi.org/10.1177/2047487317705916> (2017).
62. 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. <https://doi.org/10.1093/ehjqcco/qcy047> (2019).
63. McHutchison, C. A., Backhouse, E. V., Cvorov, 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. <https://doi.org/10.1097/ede.0000000000000675> (2017).
64. Dong, J. Y. & Qin, L. Q. Education level and breast cancer incidence: A meta-analysis of cohort studies. *Menopause* <https://doi.org/10.1097/gme.0000000000001425> (2019).
65. Brown, C. R. *et al.* Social determinants of prostate cancer in the Caribbean: A systematic review and meta-analysis. *BMC Public Health* **18**, 900. <https://doi.org/10.1186/s12889-018-5696-y> (2018).
66. Sidorchuk, A. *et al.* Socioeconomic differences in lung cancer incidence: A systematic review and meta-analysis. *Cancer Causes Control* **20**, 459–471. <https://doi.org/10.1007/s10552-009-9300-8> (2009).
67. 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. <https://doi.org/10.1371/journal.pone.0194127> (2018).
68. Zeng, X. *et al.* Associations between socioeconomic status and chronic kidney disease: A meta-analysis. *J. Epidemiol. Community Health* **72**, 270–279. <https://doi.org/10.1136/jech-2017-209815> (2018).
69. 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. <https://doi.org/10.1016/j.puhe.2018.11.002> (2019).
70. 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. <https://doi.org/10.1002/gepi.21965> (2016).
71. Burgess, S. & Thompson, S. G. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur. J. Epidemiol.* **32**, 377–389. <https://doi.org/10.1007/s10654-017-0255-x> (2017).
72. 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. <https://doi.org/10.1002/sim.7492> (2017).
73. Spiller, W., Davies, N. M. & Palmer, T. M. Software Application Profile: mrrobust—a tool for performing two-sample summary Mendelian randomization analyses. *Int. J. Epidemiol.* <https://doi.org/10.1101/142125> (2017).
74. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. <https://doi.org/10.7554/eLife.34408> (2018).
75. 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. <https://doi.org/10.1054/bjoc.1999.1105> (2000).
76. Bockerman, P. *et al.* Does higher education protect against obesity? Evidence using Mendelian randomization. *Prev. Med.* **101**, 195–198. <https://doi.org/10.1016/j.ypmed.2017.06.015> (2017).
77. Gage, S. H., Bowden, J., Davey Smith, G. & Munafò, M. R. Investigating causality in associations between education and smoking: A two-sample Mendelian randomization study. *Int. J. Epidemiol.* **47**, 1131–1140. <https://doi.org/10.1093/ije/dyy131> (2018).
78. Zeng, L. *et al.* Genetically modulated educational attainment and coronary disease risk. *Eur. Heart J.* **40**, 2413–2420. <https://doi.org/10.1093/eurheartj/ehz328> (2019).
79. Stringhini, S. *et al.* Life-course socioeconomic status and DNA methylation of genes regulating inflammation. *Int. J. Epidemiol.* **44**, 1320–1330. <https://doi.org/10.1093/ije/dyv060> (2015).
80. 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. <https://doi.org/10.1093/aje/kww014> (2016).
81. 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. <https://doi.org/10.1038/mp.2016.82> (2017).
82. 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. <https://doi.org/10.1016/j.neuroscience.2019.05.055> (2019).
83. 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. <https://doi.org/10.1038/ng.3552> (2016).
84. Baselmans, B. M. L. & Bartels, M. A genetic perspective on the relationship between eudaemonic and hedonic well-being. *Sci. Rep.* **8**, 14610. <https://doi.org/10.1038/s41598-018-32638-1> (2018).
85. Day, F. R., Ong, K. K. & Perry, J. R. B. Elucidating the genetic basis of social interaction and isolation. *Nat. Commun.* **9**, 2457. <https://doi.org/10.1038/s41467-018-04930-1> (2018).
86. Richardson, J. G. *Handbook of Theory and Research for the Sociology of Education.* (Greenwood Press, 1986).
87. Bourdieu, P. *Distinction: A Social Critique of the Judgement of Taste.* (Routledge, 2010).
88. 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. <https://doi.org/10.1136/jech.2009.104075> (2011).
89. Stringhini, S. *et al.* Association of socioeconomic position with health behaviors and mortality. *JAMA* **303**, 1159–1166. <https://doi.org/10.1001/jama.2010.297> (2010).

Acknowledgements

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 Consortium (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 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, 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 is specified at megastroke.org/acknowledgements.html.

Author contributions

S.Y. designed the study, conducted the systematic review, performed the statistical analyses, drafted the manuscript, contributed to the interpretation of the results and critical revision of the manuscript for important intellectual content and approved the final version of the manuscript. Y.X. conducted the systematic review, contributed to the interpretation of the results and critical revision of the manuscript for important intellectual content and approved the final version of the manuscript. M.M. contributed to the interpretation of the results and critical revision of the manuscript for important intellectual content and approved the final version of the manuscript. K.M. contributed to the interpretation of the results and critical revision of the manuscript for important intellectual content and approved the final version of the manuscript. S.C.L. designed the study, contributed to the interpretation of the results and critical revision of the manuscript for important intellectual content and approved the final version of the manuscript.

Funding

Open access funding provided by Uppsala University. Funding for this study came from the Swedish Research Council (Vetenskapsrådet; Grant Number 2019-00977), the Swedish Research Council for Health, Working Life and Welfare (Forte; Grant Number 2018-00123), and the Swedish Heart-Lung Foundation (Hjärt-Lungfonden; Grant number 20190247).

Competing interests

All authors declared no support from any organization for the submitted work; no financial relationships with any organization 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.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-83801-0>.

Correspondence and requests for materials should be addressed to S.C.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021