

Accepted Manuscript

Does education protect against depression? Evidence from the Young Finns Study using Mendelian randomization

Jutta Viinikainen, Alex Bryson, Petri Böckerman, Marko Elovainio, Niina Pitkänen, Laura Pulkki-Råback, Terho Lehtimäki, Olli Raitakari, Jaakko Pehkonen



PII: S0091-7435(18)30257-3
DOI: doi:[10.1016/j.ypmed.2018.08.026](https://doi.org/10.1016/j.ypmed.2018.08.026)
Reference: YPMED 5473
To appear in: *Preventive Medicine*
Received date: 12 March 2018
Revised date: 9 August 2018
Accepted date: 21 August 2018

Please cite this article as: Jutta Viinikainen, Alex Bryson, Petri Böckerman, Marko Elovainio, Niina Pitkänen, Laura Pulkki-Råback, Terho Lehtimäki, Olli Raitakari, Jaakko Pehkonen , Does education protect against depression? Evidence from the Young Finns Study using Mendelian randomization. Ypmed (2018), doi:[10.1016/j.ypmed.2018.08.026](https://doi.org/10.1016/j.ypmed.2018.08.026)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Does education protect against depression? Evidence from the Young Finns Study using Mendelian Randomization

Jutta Viinikainen*, Alex Bryson**, Petri Böckerman***, Marko Elovainio****, Niina Pitkänen*****, Laura Pulkki-Råback#, Terho Lehtimäki##, Olli Raitakari###, Jaakko Pehkonen####

* Corresponding author. University of Jyväskylä, School of Business and Economics, Jyväskylä, Finland. Phone: +350-40-5767804. Fax +358 14 617 194. E-mail: jutta.viinikainen@jyu.fi

** University College London, NIESR, London, United Kingdom and IZA, Bonn, Germany.

*** University of Jyväskylä, School of Business and Economics, Jyväskylä, Finland, Labour Institute for Economic Research, Helsinki, Finland and IZA, Bonn, Germany

**** Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland and National Institute for Health and Welfare, Helsinki, Finland

***** Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland

#Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland#

Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland.

Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku and Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland.

University of Jyväskylä, School of Business and Economics, Jyväskylä, Finland.

Word count, abstract: 241/ Word count, article: 3488

Abstract

Using participants (N=1,733) drawn from the nationally representative longitudinal Young Finns Study (YFS) we estimate the effect of education on depressive symptoms. In 2007, when the participants were between 30 and 45 years old, they reported their depressive symptoms using a revised version of Beck's Depression Inventory. Education was measured using register information on the highest completed level of education in 2007, which was converted to years of education. To identify a causal relationship between education and depressive symptoms we use an instrumental variables approach (Mendelian randomization, MR) with a genetic risk score as an instrument for years of education. The genetic risk score was based on 74 genetic variants, which were associated with years of education in a genome-wide association study (GWAS). Because the genetic variants are randomly assigned at conception, they induce exogenous variation in years of education and thus identify a causal effect if the assumptions of the MR approach are met. In Ordinary Least Squares (OLS) estimation years of education in 2007 were negatively associated with depressive symptoms in 2007 ($b = -0.027$, 95% Confidence Interval (CI) = $-0.040, -0.015$). However, the results based on Mendelian randomization suggested that the effect is not causal ($b = 0.017$; 95% CI = $-0.144, 0.178$). This indicates that omitted variables correlated with education and depression may bias the linear regression coefficients and exogenous variation in education caused by differences in genetic make-up does not seem to protect against depressive symptoms.

Keywords: Education; depression; Mendelian randomization; instrumental variables

Introduction

Approximately 6.8% of the world's population suffer from depression (Layard et al., 2013). The global direct and indirect costs of mental health conditions were about US\$ 2.4 trillion in 2010 and the estimated loss in economic output due to mental health problems in 2010 was nearly US\$ 16 trillion (Bloom et al., 2011). The high economic and societal costs of mental health conditions reinforce the need for policy measures that would prevent or alleviate these problems.

Several studies find an association between higher education and better health (see Madden, 2016) and improved mental health (e.g. Bauldry, 2015; Chevalier and Feinstein, 2007).

Higher education can lead to better health outcomes in two ways. First, by enabling individuals to make lifestyle choices that are healthier in terms of diet and exercise. Second, higher education can improve individuals' marginal health returns from a given health input. For example, education may improve ability to comprehend advice from a doctor (Grossman, 1972; Grossman and Kaestner, 1997).

Education may also improve mental health indirectly through its impact on mediators such as higher earnings, thus providing financial means for individuals to pursue better mental health (Cutler and Lleras-Muney, 2010) and reducing the risk of unemployment and work-related adverse life events which have been associated with mental distress (Audhoe et al., 2010).

More educated individuals also tend to obtain jobs where control over job tasks is high and where there is a fair balance between effort and reward (Siegrist, 1996). These job characteristics help to reduce the mental strain associated with the job tasks (Karasek, 1979; Calnan et al., 2004). Additionally, education may improve the sense of mastery over life,

improve one's relative position in society and help to create larger social networks, all of which are associated with better mental well-being (Dalgard et al., 2007; Marmot et al., 1997; Kawachi and Berkman, 2001).

Although education may improve mental health, it is an open question whether the correlation between education and depression is causal. Unobserved confounders influencing educational attainment and depressive symptoms, or reverse causality, may explain the observed correlation. To account for reverse causality and/or unobserved variables the literature has exploited data on twins, used statistical matching approaches and instrumental variables (IV) methods. The twin design accounts for shared family background and confounding genetic factors, which may otherwise cause omitted variable bias. Based on twin data, McFarland and Wagner (2015) and Mezuk et al. (2013) found that higher education was related to lower levels of depressive symptoms. Fujiwara and Kawachi (2009), who also use twin design, however, did not find such a relationship. Bauldry (2015) uses propensity score matching to evaluate the link between education and depression and finds evidence for the protective effects of higher education against depressive symptoms. The method assumes that selection is based only on observed characteristics and thus omitted variable bias or the possibility of reverse causality cannot completely be ruled out.

Only a few studies aim to identify the causal effect of education on depression using an IV approach. Chevalier and Feinstein (2007) use longitudinal data from the British National Child Development Survey (NCDS) to estimate the connection between education and mental health. They use two different measures for mental well-being: a malaise score, which uses 24 items to assess depression and a self-reported measure of currently feeling sad or depressed. As instruments, Chevalier and Feinstein (2007) use teacher assessments

concerning the benefits the child would get from post-compulsory schooling and the child's smoking behaviour at age 16. The latter is assumed to be a proxy for the discount rate (a higher discount rate implies lower investments in education). Based on the IV results education has a protective effect on mental health. The effect seems to be larger for women and for those with mid-level qualifications. Crespo et al. (2014) exploit schooling reforms in several European countries as instruments for educational attainment. They find that education has a large protective effect on mental health. However, other studies come to different conclusions. Kamhöfer et al. (2015) use variation in college availability and student loan regulations in Germany as instruments for higher education but do not find effects on mental health.

This paper examines whether higher education protects against depressive symptoms. Our contribution to the literature is to use a new instrument, a genetic-based risk score for years of education to identify the causal effect. Over the past decade this method, called Mendelian randomization (MR), has become an established tool for achieving causal inference in observational research (Pingault et al., 2018). Using rich, longitudinal population-based data which combines survey data with administrative register information on educational attainment and parental background we do not find causal support for the hypothesis that education would protect against depressive symptoms.

Methods

Data

The Cardiovascular Young Finns Study (YFS) is a longitudinal, nationally representative study of 4,320 individuals in six age cohorts (aged 3, 6, 9, 12, 15, 18) who were randomly chosen from five Finnish university hospital regions (Raitakari et al., 2008). In 1980 3,596 subjects participated in the study and since then seven follow-up studies have been conducted.

In the YFS, depressive symptoms were evaluated in 2007 using a revised version of the Beck's Depression Inventory (BDI) (Beck and Steer, 1993; Elovainio et al., 2015; Rosenström et al., 2013). Depressive symptoms were measured using 21 items on a 5-point scale. The total depressive symptoms score is the average of all 21 items (Rosenström et al., 2012). The Beck's depressive symptoms scale is not a measure of clinically recognized psychiatric disorder nor does it indicate the chronicity or severity of depression (Elovainio et al., 2015).

To obtain information on the highest completed level of education, the YFS was linked to the Finnish Longitudinal Employer-Employee Data (FLEED) of Statistics Finland (SF) using unique personal identifiers. The matching is exact. The genetic risk score (GRS) we use as an instrument refers to years of education, so we used official guidelines from Statistics Finland to convert the education levels into years of education. For those who were still at school in 2007 (2.3% of N=1,733) years of education are based on the highest obtained degree.

The GRS is based on 74 single nucleotide polymorphisms (SNPs), which have been associated with the number of years of education of 293,723 individuals in a genome-wide

association study (GWAS) (Okbay et al., 2016). As an instrument, the GRS has two key advantages over individual SNPs. First, the GRS accounts for more variation in years of education, which increases its statistical power. Second, the GRS reduces the risk that any individual SNP would bias the IV estimates via an alternative biological pathway (pleiotropy) (Palmer et al., 2012).

Genotyping was implemented by using the Illumina Bead Chip (Human 670K) and the genotypes were called using the Illumina clustering algorithm (Teo et al., 2007). SHAPEIT v1 and IMPUTE2 software (Delaneu et al., 2011) were used for genotype imputation and the 1000 Genomes Phase I Integrated Release Version 3 (March 2012 haplotypes) was used as a reference panel (Howie et al., 2009; 1000 Genomes Project Consortium, 2010). The GRS for educational attainment was based on 74 variants, which were associated with years of education (Okbay et al., 2016). Okbay et al. (2016) explored the mechanisms through which the candidate genes may affect education and found that many of these genes were related to brain development particularly during prenatal stages. The GRS was calculated as a sum of genotyped risk alleles or imputed allele dosages carried by an individual. The GRS was standardized to have a mean zero and standard deviation of one. Testing for the Hardy–Weinberg equilibrium (HWE) was conducted using the SNPTEST program (Marchini et al., 2007). Considering multiple testing, all 74 SNPs were in HWE ($p > 0.001$).

All models are adjusted for five residential regions in 2007, the birth year effects, gender and parental education. The data on parental education were drawn from Statistics Finland's Longitudinal Population Census (LPC) from the year 1980. It was linked to YFS-FLEED using unique identifiers. The indicator variable for high parental education equals one if at least one of the parents has obtained some university education.

Because of missing information on some variables the estimation sample is smaller than the total sample size. We tested the randomness of attrition with two-sample test of proportions (Table A1 in Supplementary Appendix). The results indicated that participants who were dropped from the analyses are more likely men and less educated. In terms of birth cohort or parental education there were no significant differences between the total YFS-FLEED-LPC sample and our final estimation sample.

Statistical methods

We first run Ordinary Least Squares (OLS) models to replicate standard observational studies. Because of potential confounders and reverse causality, standard linear regression coefficients may be biased. To identify the causal effect of education on depressive symptoms we estimate two-stage least squares (2SLS) IV models which use GRS as an instrument for education (Tyrrell et al., 2016; Gupta et al., 2017). The method is called Mendelian randomization (MR) in the medical literature (Conley, 2016). It exploits Mendel's law of independent assortment, which states that each trait is inherited independently from other traits at conception. The randomization by nature causes exogenous variation in the exposure variable – in our case education – enabling the identification of causal links between education and depressive symptoms. The IV estimator avoids the bias of the OLS estimator if the following four conditions are satisfied (Von Hinke et al., 2016): (1) the genetic instrument is correlated with the exposure (education) variable (strong instrument assumption); (2) the genetic instrument is not associated with any confounder of the exposure-outcome relationship i.e. the instrument is as good as randomly assigned (independence assumption); (3) the genetic instrument is exogenous i.e. the instrument does

not affect the outcome, except possibly via its association with the exposure (exclusion restriction); and (4) the instrument must have a monotonic effect on education.

A potential threat to the instrument validity is that the frequency of the genetic variants is different in different subpopulations. Other potential threats are that a) the instrument affects the outcome variable either directly or through other pathways than the exposure variable (pleiotropy); b) the exposure variable is time-varying; c) there are gene-environment interactions; d) the exposure variable is measured with error; e) there is reverse causation; f) there are other genetic markers in the same chromosome that affect the outcome and are correlated with the marker used in the analysis (Burgess and Thompson, 2015; VanderWeele et al., 2014).

We address these potential identification violations in the following ways. First, the Finnish population is ethnically very homogenous, which reduces the possibility that there are differences in allele frequency among subgroups. We also test whether the distribution of various covariates differs across the instrument distribution (Von Hinke et al., 2016) and use family controls to account for assortative mating. Second, we conduct Sargan's over-identification test using individual SNPs as instruments for education. Failure of the identification test would suggest that one or more of the genetic instruments are invalid. Third, we use PhenoScanner to detect potential alternative pathways through which SNPs in our GRS may affect depression. Fourth, to minimize measurement error and problems associated with time-varying exposure the information on education is drawn from the official register maintained by SF when the youngest participants were 30 years old and thus the number of those who were still at school is very low (2.3%). Fifth, to abstract away potential biases from violations *b*, *c*, *d*, and *e*, as a robustness check, we estimate a

specification in which the outcome variable is explained by the GRS (i.e. reduced-form model), as suggested by VanderWeele et al. (2014).

Results

In our estimation sample, the mean value of BDI score and years of education in 2007 are 2.059 (SD 0.659) and 13.9 (SD 2.67), respectively (Table 1, Panel B; for the distribution of the BDI score see Supplementary Appendix, Figure A1). Fifty-nine percent of the sample were women, the average age in 2007 was 38 years and 13% had highly educated parents (Table 1, Panel A). According to descriptive statistics, less educated individuals tend to have more symptoms of depression (Table 1, Panel A; see also Figure A2 in Supplementary Appendix). There is also a 0.16 SD difference ($p = 0.0018$) in the GRS value between the more and less educated which supports the strong instrument assumption (Table 1, Panel A and Figure A3 in Supplementary Appendix).

Panel B of Table 1 compares individual differences by the instrument value. Among high GRS participants, the share of men and highly educated parents is slightly higher. Therefore, we control for gender and parental background in all models. The earlier literature identifies connections between personality characteristics and depression (Klein et al., 2011). To detect whether this causes pleiotropic effects we compared four components of Type-A behavior (aggression, leadership, responsibility and eagerness-energy) between low and high GRS individuals. These characteristics were measured for three age cohorts at age 15 using Hunter-Wolf A-B Rating Scale (Wolf et al., 1982). The results did not indicate that pleiotropy through personality characteristics would violate the MR assumptions. To detect potential alternative pathways through which the SNPs in the education GRS may affect

depression we also used PhenoScanner which is a publicly available database that provides summary results from genetic association studies (Staley et al., 2016). Of the SNPs linked to education, some were also associated with body mass index (BMI), obesity, waist-hip ratio and height. Although we do not observe differences ($p > 0.526$) in these attributes between low and high GRS individuals (Table 1, Panel B) as an extension we control for the genetic risk score for BMI (Speliotes et al., 2010), waist-hip ratio (Heid et al., 2010) and height (Allen et al., 2010) in robustness analyses. Two SNPs (rs2431108 and rs62379838) were also associated with depressive symptoms which may violate the exclusion restriction. For this reason, in robustness checks we excluded these two SNPs from the GRS.

The baseline OLS estimates (Table 2, Column 1) show that years of education are associated with lower levels of depressive symptoms, suggesting that education protects against depression. Based on the point estimate, one additional year of education is related to a 0.027-point lower level of depressive symptoms ($p < 0.001$). The result is consistent with the prior empirical literature that has documented that education protects against depression (Bauldry, 2015; McFarland and Wagner, 2015; Crespo et al., 2014; Mezuk et al., 2013; Chevalier and Feinstein, 2007).

The IV results in Table 2 reveal two important findings. First, the first stage F-statistics in the baseline MR model (Table 2, Column 2) is 10.19 and the instrument is significant in the first stage ($b = 0.185$; 95% CI = 0.071, 0.300), which indicates that the excluded instrument (i.e. GRS) is related to education. In the first stage, the partial R-squared of the GRS is 0.0074, which implies that the risk score explains approximately 0.7% of the variation in education. Second, in contrast to the OLS estimates, the MR estimates do not support the conclusion that higher education leads to lower levels of depressive symptoms. By contrast, the point

estimate turns positive, suggesting that higher education increases depressive symptoms, although the point estimate is not statistically significant. This remains the case even if we control for GRS's for high BMI, waist-hip ratio and height (Table 2, Column 3) and also if two SNPs which were associated with depressive symptoms were excluded from the GRS instrument (Table 2, Column 4). The implication is that there may be confounding factors or reverse causality which bias the OLS coefficient. The MR point estimate is statistically non-significant ($p = 0.835$), in part because the standard errors in the IV estimation are somewhat larger than in the OLS results. A Hausman test for the difference in the coefficients between the OLS and the IV estimates did not reject the hypothesis that the two estimates are similar to one another ($p = 0.599$). However, the Hausman test has relatively low power and because we have good reasons to argue that education might be endogenous, the IV estimator is retained.

Table 2 here

As a formal statistical test of instrument validity we conducted Sargan's over-identification test by using individual SNPs as instruments for education. Sargan's test supported the null hypothesis that all SNPs can be considered exogenous ($p=0.845$) lending support to the instrument's validity. However, it is technically possible that Sargan's test fails to reject the null hypothesis due to low statistical power (Glymour et al., 2012). As a robustness check, we also estimated a reduced-form model, where the BDI score was regressed on the GRS. This approach does not rule out the possibility that the exclusion restriction is violated but it abstracts away potential biases resulting from time-varying exposure, gene-environment interactions, measurement error in the exposure variable, and reverse causation (VanderWeele et al., 2014). However, it only identifies the effect of the exposure on the

outcome – not the actual size of the effect of interest. The findings from the reduced-form model are consistent with our MR results: the GRS coefficient is positive ($b = 0.003$) and insignificant (95 % CI: -0.026, 0.033) lending further support to the conclusion that exogenous variation in education caused by genetic differences does not reduce depressive symptoms. As a final robustness check, we also used the weighted GRS as an instrument. The weights for individual SNPs are based on the contribution of each SNP on years of education in a meta-analysis. The results remained intact (see Table A3 in Supplementary Appendix).

Discussion

Using longitudinal data for Finland we find that according to linear regression results higher education is related to lower level of depressive symptoms. However, the MR results which use a GRS as an instrument for years of education suggest that this association is not causal. Thus, there may be omitted variables affecting both education and depressive symptoms or reverse causality that bias the OLS results.

Our empirical approach has limitations. First, a potential explanation for the null effect in the MR setting is low statistical power due to an insufficient sample size (Freeman et al., 2013). To approximate the minimum required sample size to ensure sufficient power we utilized an online MR power calculator provided by Brion et al. (2013). Using the OLS and MR estimates provided in Table 2, the observed variances of the education and depression score and the explanatory power of the genetic instrument in our data (0.0074 i.e. ~ 0.7%), the required sample size for a definite test of no causal relationship with 0.7 power and 0.05 type I error is approximately $N = 2,867,364$. Thus, the result of no causal effect could reflect the inability to reject the null hypothesis due to a relatively small sample size. However, our IV

point estimate was positive lending support to the main conclusion. Second, the MR approach identifies causal effects only if the instrument is valid. One can never prove the null hypothesis of instrument validity in a MR setting or more generally in any IV study. However, the tests that we performed and information we have is consistent with instrument validity. Third, the relatively small sample size reduces the possibility to estimate the effects for relevant subgroups, such as for men and women, separately. Fourth, the local average treatment effect (LATE) identified by IV estimation captures the average effect of education on depressive symptoms among compliers (Swanson and Hernán, 2013) i.e. those whose years of education are raised via the impact of genetic variation used in the instrument. The variation in years of education due to factors other than genetic variance could have a different impact on depressive symptoms. It is also possible that social environment moderates the genetic predisposition through gene-environment interactions, which may affect our results.

Contrary to some prior studies, our results provide suggestive evidence that education does not protect against depressive symptoms. There are a number of reasons why our results may differ from the findings which indicate that education protects against depression. First, previous research has used different dependent variables to capture mental ill-health. Second, the earlier results are based on different populations and different periods (e.g. the 1958 British National Child Development Survey). Third, the earlier studies used different identification strategies which each have their own merits and weaknesses. Twin studies control for shared genetic and family confounders, but do not account for other potential confounding factors or reverse causality which may significantly bias results. The same weaknesses apply to matching approaches, which only account for observed heterogeneity. The IV method is able to circumvent both reverse causality and omitted variables bias if

conditions concerning instrument strength and validity are satisfied. Previous research using IV strategies has typically used aspects of the schooling system as instruments (e.g. reforms affecting the costs/benefits of education) or individuals' propensity to invest in the long run (e.g. smoking). In the latter case, it is unclear whether the exclusion restriction is supported. In our setting there is a clear element of randomization, which is usually lacking in the absence of experiments. Also, schooling reforms may identify a different local average treatment effect (LATE) compared to MR.

Two alternative explanations for our conflicting OLS and MR results are possible. First, individuals with severe recurrent depression may find it difficult to complete high education. If individuals with depressive symptoms in adulthood were inclined to mental health problems already in youth, reverse causality could explain the discrepancy between the results. Second, there may be unobserved confounders which bias the OLS results. However, we cannot completely rule out the possibility that the insignificant MR results are spurious. As in all IV studies, instrument validity is an untestable assumption. Earlier evidence on the effect of education on mental health has been inconclusive. Our results based on a novel genetic instrument support the findings (Kamhöfer et al., 2015; Mezuk et al., 2013), which indicate that there is no causal relationship between education and mental well-being. This suggests that education policies are not viable to address the mental health problems.

Conflict of interest

The authors declare that there are no conflicts of interest.

Acknowledgements

The Young Finns Study has been financially supported by the Academy of Finland: grants 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research; Finnish Cultural Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; The Jenny and Antti Wihuri Foundation (grant to Laura Pulkki-Råback) Diabetes Research Foundation of Finnish Diabetes Association: and EU Horizon 2020 (grant 755320 for TAXINOMISIS). Jutta Viinikainen and Jaakko Pehkonen acknowledge financial support from the Yrjö Jahnsson Foundation (grants 6664 and 6646). Petri Böckerman thanks the Strategic Research Council funding for the project Work, Inequality and Public Policy (293120).

References

- 1000 Genomes Project Consortium, 2010. A map of human genome variation from population-scale sequencing. *Nature* 467 (7319), 1061-1073.
- Allen, H. L., Estrada, K., Lettre, G., et al., 2010. Hundreds of variants clustered in genome loci and biological pathways affect human height. *Nature* 467, 832-838.
- Audhoe, S. S., Hoving, J. L., Sluiter, J. K. et al., 2010. Vocational interventions for unemployed: effects on work participation and mental distress. A systematic review. *J. Occup. Rehabil.* 20(1), 1-13.
- Bauldry, S., 2015. Variation in the protective effect of higher education against depression. *Soc. Ment. Health* 5(2), 145-161.
- Beck, A.T., Steer, R.A., 1993. Manual for the Beck Depression Inventory. Psychological Corporation, San Antonio, TX.
- Bloom, D.E., Cafiero, E.T., Jané-Llopis, E. et al., 2011. The global economic burden of non-communicable diseases. World Economic Forum, Geneva.
- Brion, M. J. A., Shakhbazov, K., Visscher, P. M. 2012. Calculating statistical power in Mendelian randomization studies. *Int. J. Epidemiol.* 42(5), 1497-1501.

Burgess, S., Thompson, S. G., 2015. Mendelian randomization: methods for using genetic variants in causal estimation. CRC Press, Boca Raton, FL.

Calnan, M., Wadsworth, E., May, M. et al., 2004. Job strain, effort-reward imbalance, and stress at work: competing or complementary models?. *Scand. J. Soc. Med.* 32(2), 84-93.

Chevalier, A., Feinstein, L., 2007. Sheepskin or prozac: the causal effect of education on mental health. UCD Geary Institute Discussion Paper WP/15/2007.

Crespo, L., López-Noval, B., Mira, P., 2014. Compulsory schooling, education, depression and memory: new evidence from SHARELIFE. *Econ. Educ. Rev.* 43, 36-46.

Cutler, D. M., Lleras-Muney, A., 2010. Understanding differences in health behaviors by education. *J. Health. Econ.* 29(1), 1-28.

Conley, D., 2016. Socio-genomic research using genome-wide molecular data. *Annu. Rev. Sociol.* 42, 275–299.

Dalgard, O. S., Mykletun, A., Rognerud, M. et al., 2007. Education, sense of mastery and mental health: results from a nation wide health monitoring study in Norway. *BMC Psychiatry* 7(1), 20.

Delaneau, O., Marchini, J., Zagury, J.F., 2012. A linear complexity phasing method for thousands of genomes. *Nat. Methods.* 9, 179–181.

Elovainio, M., Jokela, M., Rosenström, T., et al., 2015. Temperament and depressive symptoms: What is the direction of the association?. *J Affect Disord.* 170, 203-212.

Freeman, G., Cowling, B. J., Schooling, C. M. 2013. Power and sample size calculations for Mendelian randomization studies using one genetic instrument. *Int. J. Epidemiol.* 42(4), 1157-1163.

Fujiwara, T., Kawachi, I., 2009. Is education causally related to better health? A twin fixed-effect study in the USA. *Int. J. Epidemiol.* 38(5), 1310-1322.

Glymour, M. M., Tchetgen Tchetgen, E. J., Robins, J. M. 2012. Credible Mendelian randomization studies: approaches for evaluating the instrumental variable assumptions. *Am J Epidemiol.* 175(4), 332-339.

Grossman, M., 1972. On the concept of health capital and the demand for health. *J. Polit. Econ.* 80(2), 223-255.

Grossman, M., Kaestner, R., 1997. Effects of education on health. In: Behrman, R. R., Nevzer, S. *The social benefits of education*, The University of Michigan Press, pp 69-124.

Gupta, V., Walia, G.K., Sachdeva, M.P., 2017. 'Mendelian randomization': an approach for exploring causal relations in epidemiology. *Public Health* 145, 113–119.

Heid, I.M., Jackson, A.U., Randall, J.C., et al., 2010. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat

distribution. *Nat Genet.* 42, 949–960.

von Hinke, S., Davey Smith, G, Lawlor, D.A., et al., 2016. Genetic markers as instrumental variables. *J. Health Econ.* 45, 131–148.

Howie, B.N., Donnelly, P., Marchini, J., 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 5, e1000529.

Kamhöfer, D., Schmitz, H., Westphal, M., 2015. Heterogeneity in marginal non-monetary returns to higher education. *Ruhr Economic Papers* No. 591.

Karasek, R., 1979. Job demands, job decision latitude and mental strain. *Admin. Sci. Quart.* 24(2), 286–308.

Kawachi, I., Berkman, L. F., 2001. Social ties and mental health. *J. Urban Health* 78(3), 458-467.

Klein, D. N., Kotov, R., Bufferd, S. J. 2011. Personality and depression: explanatory models and review of the evidence. *Annu Rev Clin Psychol.* 7, 269-295.

Layard, R., Chisholm, D., Patel, V., Saxena, S., 2013. Mental illness and unhappiness. In: Helliwell, J., Layard, R., Sachs, J. (Eds) *World Happiness Report 2013*, pp. 38-53.

Madden, D., 2016. Do schooling reforms improve long-term health?. *IZA World of Labor* 306.

Marchini, J., Howie, B., Myers, S. et al., 2007. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet.* 39(7), 906.

Marmot, M., Ryff, C. D., Bumpass, L. L. et al., 1997. Social inequalities in health: next questions and converging evidence. *Soc. Sci. Med.* 44(6), 901-910.

McFarland, M. J., Wagner, B. G., 2015. Does a college education reduce depressive symptoms in American young adults?. *Soc. Sci. Med.* 146, 75-84.

Mezuk, B., Myers, J. M., Kendler, K. S., 2013. Integrating social science and behavioral genetics: testing the origin of socioeconomic disparities in depression using a genetically informed design. *Am. J. Public Health* 103(S1), S145-S151.

Okbay, A., Beauchamp, J.P., Fontana, M.A., et al., 2016. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 533(7604), 539–542.

Palmer, T.M., Lawlor, D.A., Harbord, R.M., 2012. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat. Methods Med. Res.* 21, 223–242.

Pingault, J. B., O'Reilly, P. F., Schoeler, T., et al., 2018. Using genetic data to strengthen causal inference in observational research. *Nat. Rev. Genet.* 1.

Raitakari, O.T., Juonala, M., Rönönen, T., et al., 2008. Cohort profile: the cardiovascular risk in Young Finns Study. *Int. J. Epidemiol.* 37, 1220–1226.

Rosenström, T., Jokela, M., Hintsanen, M., et al., 2012. Longitudinal course of depressive symptoms in adulthood: linear stochastic differential equation modeling. *Psychol. Med.* 43, 933–944.

Rosenström, T., Jokela, M., Puttonen, S., et al., 2013. Pairwise measures of causal direction in the epidemiology of sleep problems and depression. *PLoS One* 7, e50841.

Siegrist, J., 1996. Adverse health effects of high-effort/low-reward conditions. *J. Occup. Health Psychol.* 1(1), 27.

Speliotes, E. K., Willer, C. J., Berndt, S. I., et al., 2010. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet.* 42, 937-948.

Staley, J. R., Blackshaw, J., Kamat, M. A., et al., 2016. PhenoScanner: a database of human genotype–phenotype associations. *Bioinformatics* 32(20), 3207-3209.

Swanson, S. A., Hernán, M. A., 2013. Commentary: how to report instrumental variable analyses (suggestions welcome). *Epidemiology* 24(3), 370-374.

Teo, Y.Y., Inouye, M., Small, K.S., et al., 2007. A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics* 23, 2741–2746.

Tyrrell, J., Jones, S.E., Beaumont, R., et al., 2016. Height, body mass index, and socioeconomic status: Mendelian randomisation study in UK Biobank. *Brit. Med. J.* 352, i582.

VanderWeele, T. J., Tchetgen, E. J. T., Cornelis, M., et al., 2014. Methodological challenges in Mendelian randomization. *Epidemiology* 25(3), 427.

Wolf, T. M., Sklov, M. C., Wenzl, P. A., Hunter, S. M., et al., 1982. Validation of a Measure of Type A Behavior Pattern in Children: Bogalusa Heart Study. *Child Dev.*, 53(1), 126-135.

Table 1

Descriptive statistics and comparison of observations by the instrument value.

Panel A: Descriptive statistics						
	All Mean (SD)	Above median years of education	Below median years of education	Difference	t-statistics	N
Beck's Depression Inventory (2007)	2.059 (0.659)	1.959 (0.655)	2.107 (0.656)	-0.147	-4.405***	1733
Education GRS	0.014 (1.006)	0.121 (1.003)	-0.039 (1.004)	0.160	3.125***	1733
Female (%)	0.586 (0.493)	0.627 (0.484)	0.565 (0.496)	0.062	2.468**	1733
Average age (2007), years	37.657 (5.016)	36.195 (4.976)	38.371 (4.881)	-2.176	-8.660***	1733
Parental education high (%) (1980)	0.126 (0.332)	0.243 (0.429)	0.070 (0.255)	0.173	10.489***	1773
Panel B: Comparison of observables by the instrument value						
	All Mean (SD)	Above median GRS	Below median GRS	Difference	t-statistics	N
Beck's Depression Inventory (2007)	2.059 (0.659)	2.068 (0.656)	2.049 (0.663)	0.020	0.624	1733
Education years	13.879 (2.671)	13.979 (2.752)	13.779 (2.585)	0.200	1.558	1733
Share of highly educated	0.330 (0.470)	0.348 (0.477)	0.312 (0.463)	0.037	1.618	1733
Female (%)	0.586 (0.493)	0.565 (0.496)	0.606 (0.489)	-0.041	-1.736*	1733
Average age (2007), years	37.657 (5.016)	37.519 (5.026)	37.794 (5.006)	-0.275	-1.143	1733
Parental education high (%) (1980)	0.126 (0.332)	0.142 (0.349)	0.111 (0.314)	0.031	1.944*	1733
Other genetic risk scores						
GRS for BMI	29.125 (3.377)	29.117 (3.272)	29.103 (3.461)	0.013	0.082	1662
GRS for waist-	15.185	15.199	15.125	0.074	0.633	1662

hip ratio	(2.380)	(2.431)	(2.353)			
GRS for height	179.777 (8.698)	179.660 (8.601)	179.701 (8.741)	-0.041	0.096	1662
Personality characteristics at age 15						
Aggression	3.769 (0.887)	3.725 (0.844)	3.812 (0.927)	-0.088	-1.352	746
Leadership	4.202 (0.887)	4.188 (0.894)	4.215 (0.862)	-0.027	-0.424	746
Responsibility	4.750 (1.056)	4.774 (1.062)	4.72 (1.052)	0.049	0.629	746
Eagerness- energy	4.669 (0.654)	4.680 (0.663)	4.658 (0.644)	0.022	0.457	746

Notes: Table reports means and standard deviations in parenthesis. Differences between groups were tested using two-sample t test. The unit of analysis is individual. The six cohorts under study (aged 30 and 45 in year 2007) are drawn from the Young Finns Study. GRS refers to the genetic risk score. The indicator for parental education high equals one if at least one of the parents has obtained some university education (based on the Longitudinal Population Census data from 1980). Statistically significant at *10% **5% and ***1% levels.

Table 2

Education and depressive symptoms. Results based on linear regression (OLS) and instrumental variables (IV, 2SLS) methods.

	(1) OLS	(2) IV 2SLS	(3) IV 2SLS	(4) IV 2SLS
	Coefficient [95% CI]	Coefficient [95 % CI]	Coefficient [95 % CI]	Coefficient [95 % CI]
Per additional year of education (2007)	-0.027*** [-0.040; - 0.015]	0.017 [-0.144; 0.178]	0.036 [-0.139; 0.210]	0.009 [-0.140; 0.157]
Additional GRS controls			x	
Results based on 72 SNP GRS				x
R ²	0.023
First stage F- statistic	..	10.19	8.62	11.88
N	1733	1733	1662	1733

Notes: Dependent variable is the revised version of Beck's Depression Inventory score in 2007. The instrument used in the IV models is the polygenic score for education years based on genetic markers. The unit of analysis is individual. The six cohorts under study (aged between 30 and 45 years in 2007) are drawn from the Young Finns Study. The models include (unreported) controls for indicators for age, gender (being female), five regions in 2007 and parental education in 1980. The additional controls in Column 3 are genetic risk scores for body mass index, waist-hip ratio and height. In Column 4 two SNPs which may violate the exclusion restriction have been removed from the GRS. The 95% confidence intervals for the parameter estimates are reported in square brackets. Statistically significant at *10% **5% and ***1% levels. Full estimation results are reported in Table A2 (Supplementary Appendix).

Highlights

Higher education has been associated with better mental health

Reverse causality and confounders may bias linear regression results

We use genetic information as an instrument for education to establish a causal relationship

The IV estimates suggest no causal effect between education and depressive symptoms

ACCEPTED MANUSCRIPT