



Does higher education protect against obesity? Evidence using Mendelian randomization



Petri Böckerman^{a,b,*}, Jutta Viinikainen^c, Laura Pulkki-Råback^{d,e}, Christian Hakulinen^e, Niina Pitkänen^f, Terho Lehtimäki^g, Jaakko Pehkonen^c, Olli T. Raitakari^h

^a Turku School of Economics, Labour Institute for Economic Research, Helsinki, Finland

^b IZA, Bonn

^c Jyväskylä University School of Business and Economics, Jyväskylä, Finland

^d Helsinki Collegium for Advanced Studies, University of Helsinki, Finland

^e Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Finland

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

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

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

ARTICLE INFO

Article history:

Received 16 March 2017

Received in revised form 14 June 2017

Accepted 17 June 2017

Available online 21 June 2017

Keywords:

Education
Schooling
Obesity
Body weight
BMI
Waist-hip ratio

ABSTRACT

Objectives. The aim of this explorative study was to examine the effect of education on obesity using Mendelian randomization.

Methods. Participants ($N = 2011$) were from the on-going nationally representative Young Finns Study (YFS) that began in 1980 when six cohorts (aged 30, 33, 36, 39, 42 and 45 in 2007) were recruited. The average value of BMI (kg/m^2) measurements in 2007 and 2011 and genetic information were linked to comprehensive register-based information on the years of education in 2007. We first used a linear regression (Ordinary Least Squares, OLS) to estimate the relationship between education and BMI. To identify a causal relationship, we exploited Mendelian randomization and used a genetic score as an instrument for education. The genetic score was based on 74 genetic variants that genome-wide association studies (GWASs) have found to be associated with the years of education. Because the genotypes are randomly assigned at conception, the instrument causes exogenous variation in the years of education and thus enables identification of causal effects.

Results. The years of education in 2007 were associated with lower BMI in 2007/2011 (regression coefficient (b) = -0.22 ; 95% Confidence Intervals [CI] = $-0.29, -0.14$) according to the linear regression results. The results based on Mendelian randomization suggests that there may be a negative causal effect of education on BMI (b = -0.84 ; 95% CI = $-1.77, 0.09$).

Conclusion. The findings indicate that education could be a protective factor against obesity in advanced countries.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Higher education has been associated with a lower risk of obesity in most developed countries (Cohen et al., 2013; Marmot, 2015). However, it is unclear whether the relationship is causal or not (Eide and Showalter, 2011). Unobserved confounders may influence both person's education and body weight later in life (Björklund and Salvanes, 2011; Clark and Roayer, 2013). This may bias results in a non-experimental settings where confounders, such as the ability to delay gratification (a low level known to be associated with lower

education and higher body weight), are very difficult to account for using standard covariates in observational studies (Clark and Roayer, 2013).

There are only few studies that have examined the causal relationship between education and BMI. These studies, which exploited changes in compulsory schooling laws to identify the causal effect, have either found that education has a protective effect on BMI (Spasojević, 2010) or that the effect is zero (Clark and Roayer, 2013) or inconclusive (Arendt, 2005). There is only one concurrent methodological study that used a genetic instrument to estimate the effect of education on obesity (van Kippersluis and Rietveld, 2017).

The current study examines the relationship between education and BMI using Mendelian randomization, which is based on the random assignment of genotypes at conception (Tyrrell et al., 2016; Gupta et al.,

* Corresponding author.

E-mail address: petri.boeckerman@labour.fi (P. Böckerman).

2017; Davey Smith et al., 2017). This randomization by nature enables the use of a genetic score as an instrument for education and to detect a causal relationship between education and BMI. An instrument induces changes in the explanatory variable (i.e. education) but has no independent effect on the dependent variable (BMI). The genetic score is based on variants that genome-wide association studies (GWASs) have found to be significantly associated with the number of years of education in an extensive population sample (Okbay et al., 2016).

2. Methods

2.1. Study design and sample

Cardiovascular Risk in Young Finns Study (YFS) (see <http://youngfinnsstudy.utu.fi/studydesign.html>) is an on-going epidemiological study that examines risk factors for coronary heart disease. The YFS began in 1980 when 4320 participants in six age cohorts (ages 3, 6, 9, 12, 15 and 18 years) were randomly chosen from five Finnish university regions using the national population register (Raitakari et al., 2008). A total of 3596 people participated in the study in 1980, and seven follow-up studies have been conducted. For the current study, 3182 participants had complete data on adult education and adult obesity measures and 2443 participants were genotyped. The final study sample consisted of 2011 participants.

2.2. Measures

The measures of obesity originate from professional health examinations conducted at local health centres. BMI was calculated as weight (kg) divided by height (m²). Weight was measured in light clothing without shoes to the nearest 0.1 kg. Waist-hip ratio (WHR) was calculated as the waist circumference (in cm) divided by hip circumference (in cm). We avoided bias related to self-reported measures (Stommel and Schoenborn, 2009). To mitigate idiosyncratic variation, the preferred specification uses the average value of the measured BMI (WHR) in 2007 and 2011. To maximize the sample size, if either the BMI (WHR) value from 2007 or 2011 was missing the average was calculated based only on one year value. We also estimated models for the annual outcomes.

To obtain comprehensive register-based information on education, we linked the YFS to the Finnish Longitudinal Employer-Employee Data (FLEED) of Statistics Finland (SF) using personal identifiers, which covers information on the educational level of all Finnish citizens. We measured education using the number of years of education in 2007, because GWAS findings are based on this parameter. Degrees were converted to the number of years of education using the official estimates of SF for the time to graduate with a specific degree. At the time education was measured (using FLEED in 2007) the youngest participants were 30 years of age and therefore the number of those who were still in school was very low. For those who were still enrolled, the years of education are based on the highest completed degree from FLEED. SF has made the data comparable across all cohorts.

We used the genetic risk score (GRS) as an instrument for the number of years of education. A recent GWAS identified 74 single nucleotide polymorphisms (SNPs) associated with the number of years of education (Okbay et al., 2016). The loci were identified in GWAS with 293,723 individuals in the discovery and additional 111,349 individuals in the replication sample (Okbay et al., 2016). The association of individual SNPs was not tested in the YFS population, since generally the effects of individual SNPs are modest and may not reach statistical significance in small study populations due to lack of power.

Genotyping ($n = 2443$) was performed using the Illumina Bead Chip (Human 670 K), and the genotypes were called using the Illumina clustering algorithm (Teo et al., 2007). Genotype imputation was performed using SHAPEIT v1 and IMPUTE2 software (Delaneau et al., 2012), 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). A GRS with 74 variants associated with educational attainment 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 of zero and unit standard deviation (see Table Appendix A1).

The GRS has three key advantages. First, the GRS is more powerful than any individual single loci, because the GRS explains more variation in the number of years of education. Second, the GRS is more valid because it significantly reduces the risk that any individual loci will bias the Instrumental Variable (IV) estimates via an alternative biological pathway (pleiotropy) (Palmer et al., 2012). Third, the use of the GRS is appropriate in our context, because education is a highly polygenic trait.

To account for observable differences in parental background and assortative mating within educational groups, we linked the YFS/FLEED data to the Longitudinal Population Census (LPC) of SF from the year 1980. We use the indicator for the parents' university-level education and the log of total family income from LPC. We also adjusted for birth month, birth year effects and gender (being female) in all models.

2.3. Mendelian randomization

The use of genetic instruments in IV estimation is called Mendelian randomization in the medical literature (Conley, 2016). This method uses genetic instruments to estimate causal effects (Tyrrell et al., 2016; Böckerman et al., 2017; Gupta et al., 2017). The method exploits Mendel's law of independent assortment according to which trait is inherited independently from other traits at conception. This natural randomization causes exogenous variation in the exposure variable (e.g., education), which can be used to identify causal effects. The motivation for the use of IV estimation arises from the influence of confounding factors that correlate with both the exposure and outcome variable. The IV estimator avoids the bias of the OLS (Ordinary Least Squares) estimator if three conditions are fulfilled (Von Hinke et al., 2016): (1) the genetic instrument must be informative; (2) the genetic instrument must be exogenous; and (3) the instrument and confounder must be independent.

2.4. Statistical methods

First, we ran OLS models to replicate the standard observational studies that show an association between education and BMI. Second, we estimated IV models in which the number of years of schooling was instrumented using the GRS for education.

3. Results

More highly educated individuals have significantly lower BMIs, which is 26.7 units in average in the group with low education versus 25.6 in average in the group with high education (Table 1). WHR had values of 0.91 and 0.87 for low versus high education (Table 1). Fig. A1 illustrates the correlation between BMI and the years of education (left-hand panel), and the years of education and GRS (right-hand panel).

The baseline OLS estimates showed that the years of education are statistically significantly associated with a lower BMI. The point estimate reveals that one additional year of education is associated with a 0.2-point lower BMI (Table 2, Panel A).

Three important patterns stand out from the preferred IV estimates. First, in the first stage of IV, the F-statistic on the instrument exceeds the minimum standard of 10 (Staiger and Stock, 1997). This finding shows that the GRS is a powerful instrument for the number of years of education, supporting the assumption of the IV model that the instrument is informative. Second, the IV point estimate is statistically significant at the 10% level in the preferred specification in which we use the average value of the BMI in 2007 and 2011 as the outcome variable. Third, we

Table 1
Descriptive statistics.

	All mean (SD)	Below median years of education	Above median years of education
Average BMI (2007, 2011)	26.1 (4.5)	26.7 (4.6)	25.6 (4.4)
BMI (2007)	25.9 (4.4)	26.4 (4.6)	25.3 (4.2)
BMI (2011)	26.3 (4.6)	26.9 (4.6)	25.8 (4.6)
Average WHR (2007, 2011)	0.89 (0.09)	0.91 (0.08)	0.87 (0.08)
WHR 2007	0.88 (0.09)	0.90 (0.09)	0.87 (0.08)
WHR 2011	0.90 (0.09)	0.92 (0.09)	0.88 (0.09)
Education years (2007)	13.8 (2.7)	11.6 (1.1)	16.0 (1.8)
Education GRS	0.00 (0.99)	−0.07 (0.98)	0.07 (1.00)
Female (%)	54.4 (49.8)	47.4 (50.0)	61.4 (48.7)
Average age (2007), years	37.5 (5.0)	37.8 (5.0)	37.2 (5.0)
Average age (2011), years	41.5 (5.0)	41.8 (5.0)	41.2 (5.0)
Parental education high (%) (1980)	13.0 (34.0)	6.4 (24.4)	20.4 (40.3)
Log of family income (1980)	9.33 (0.84)	9.18 (0.92)	9.47 (0.71)

Notes: The table values represent the mean and standard deviation (in parentheses). WHR refers to the waist-hip ratio, and GRS refers to the genetic risk score. The indicator for parental education high equals one if at least one of the parents has obtained at least some university education (based on the LPC data from 1980).

observed that the IV estimate is substantially larger than the OLS estimate. The point IV estimate indicated that one additional year of education leads to a 0.84-point decrease in BMI (Table 2, Panel B).

The additional estimates use the waist-hip ratio as an alternative outcome variable in the regressions. The OLS estimation reveals a negative association between education and the waist-hip ratio, but the IV estimates are not statistically significant in these models at the conventional level of significance (Table 3). However, the point estimate remains intact using the average WHR.

Appendix Tables A2 and A3 replicate the IV models with no additional controls and with controls for gender (being female), birth month and birth year effects. The results remain robust. These estimation results suggest that any possible confounding between the years of education and BMI is minimal when using the Mendelian randomization approach.

4. Discussion

The current study used data from the Young Finns Study to examine whether the association between educational attainment and BMI is causal using Mendelian randomization. The results showed that higher

education is associated with a lower BMI – and that this association is likely to be causal. We used GRS as an instrument for education, which accounts for potential confounders. However, we did not find statistically significant effect of education on waist-to-hip ratio.

The current study findings are in line with previous studies in which higher education has been associated with lower BMI (Cohen et al., 2013). The results additionally support the contention that there may be a causal effect between a higher education level and lower BMI. Therefore, high education potentially has a protective effect against obesity and obesity-related diseases. A recent study showed that a high body mass index (BMI) (>25 kg/m²) has an important causal role in the development of several non-communicable diseases and all-cause mortality (Global BMI Mortality Collaboration, 2016), emphasizing the importance of tackling obesity as an approach for improving public health and longevity.

Educational attainment is known to have a strong environmental component and only little is explained by individual SNPs. The candidate genes for educational attainment identified by the GWAS are ones with elevated expression levels in the central nervous system and potential functions in stages of neural development and cognition related phenotypes suggesting they are biologically relevant (Okbay et

Table 2
Baseline results: education and BMI.

	Panel A: OLS	Panel B: IV
	Coefficient (Robust SE)	Coefficient (Robust SE)
Average BMI (2007, 2011)		
Per additional year of education	−0.22*** [−0.29, −0.14] (0.04)	−0.84* [−1.77, 0.09] (0.47)
N	2011	2011
BMI 2007		
Per additional year of education	−0.21*** [−0.29, −0.13] (0.04)	−0.64 [−1.73, 0.45] (0.56)
N	1853	1853
BMI 2011		
Per additional year of education	−0.22*** [−0.30, −0.13] (0.04)	−0.97* [−2.01, 0.08] (0.53)
N	1650	1650

Notes: Panel A reports OLS estimates and Panel B IV estimates. All models include (unreported) controls for birth month, birth year effects, gender (being female), family education (1980) and log of family income (1980). The instrument used in the IV models is the genetic risk score for education based on genetic markers. The 95% confidence intervals for the parameter estimates are reported in square brackets. Heteroscedasticity-robust standard errors are reported in parentheses with significance at *10% **5% and ***1% levels.

Table 3
Baseline results: education and the waist-hip ratio (WHR).

	Panel A: OLS	Panel B: IV
	Coefficient (Robust SE)	Coefficient (Robust SE)
Average WHR (2007, 2011)		
Per additional year of education	−0.004*** [−0.005, −0.003] (0.001)	−0.004 [−0.017, 0.009] (0.007)
N	2011	2011
WHR 2007		
Per additional year of education	−0.004*** [−0.005, −0.003] (0.001)	0.005 [−0.012, 0.023] (0.009)
N	1853	1853
WHR 2011		
Per additional year of education	−0.004*** [−0.006, −0.003] (0.001)	−0.008 [−0.024, 0.007] (0.008)
N	1650	1650

Notes: Panel A reports OLS estimates and Panel B IV estimates. All models include (unreported) controls for birth month, birth year effects, gender (being female), family education (1980) and log of family income (1980). The instrument used in the IV models is the genetic risk score for education based on genetic markers. The 95% confidence intervals for the parameter estimates are reported in square brackets. Heteroscedasticity-robust standard errors are reported in parentheses with significance at *10% **5% and ***1% levels.

al., 2016). Okbay et al. (2016) used several methods of analysis designed to elucidate the biological mechanisms through which the identified loci might influence educational attainment. The list of candidate genes implicated most consistently across various analyses includes *TBR1*, which codes a neuron-specific transcription factor of the T-box family involved in neuronal migration and axonal projection (Notwell et al., 2016). Another candidate gene is *MEF2C* (myocyte enhancer factor 2C), which codes a transcription factor involved in early neuroprogenitor development (Paciorkowski et al., 2013). *De novo* mutations in both *TBR1* and *MEF2C* are linked to autism spectrum disorder and intellectual disability whereas mutations in other candidate genes, such as *CELSR3* and *SBNO1*, have been linked to schizophrenia.

Our approach has limitations. First, there are problems that arise from using BMI as a measure of obesity in the main specifications. BMI misses the fact that the numerical BMI values of two otherwise identical individuals with different muscle mass and fat contents are the same (Burkhauser and Cawley, 2008; Rothman, 2008; Johansson et al., 2009; Davillas and Benzeval, 2016). Second, we used the years of education to measure educational attainment, which does not account for potential nonlinear effects associated with different levels of education (Böckerman and Maczulskij, 2016). Third, the number of observations in the linked data was 2011. A larger data set is needed to provide more power and obtain more tightly estimated effects of education on BMI in the IV models (Palmer et al., 2012). Fourth, the sample size is not large enough to estimate separate IV models by gender. Fifth, education may affect BMI via higher income or because of other reasons such as peer effects (i.e., peers influence each other health behaviours, Trogdon et al., 2008). However, this study did not specifically contribute to the debate on the mechanisms why education impacts BMI.

The main strengths of the current study are the nationally representative sample, the use of register data to account for parental background, accurate administrative measures of education, BMI and WHR measured by health-professionals, and the use of GRS for the years of education.

In conclusion, the results show the value of using genetic instruments to obtain new insights into the causal effects. The estimates indicate that higher education could be a causal protective factor against obesity in advanced countries.

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; and Diabetes Research Foundation of Finnish Diabetes Association. The Palkansaaja Foundation supported the use of linked data. Böckerman thanks the Strategic Research Council funding for the project Work, Inequality and Public Policy (293120). Jutta Viinikainen and Jaakko Pehkonen acknowledge financial support from the Yrjö Jahnsson Foundation (grants 6664 and 6646).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.ypmed.2017.06.015>.

References

- 1000 Genomes Project Consortium, 2010. A map of human genome variation from population-scale sequencing. *Nature* 467 (7319), 1061–1073.
- Arendt, J.N., 2005. Does education cause better health? A panel data analysis using school reforms for identification. *Econ. Educ. Rev.* 24, 149–160.
- Björklund, A., Salvanes, K.G., 2011. Education and family background: mechanisms and policies. In: Hanushek, E.A., Machin, S., Woessmann, L. (Eds.), *Handbook of the Economics of Education*. vol. 3. Elsevier, Amsterdam, pp. 201–247.
- Böckerman, P., Maczulskij, T., 2016. The education-health nexus: fact and fiction. *Soc. Sci. Med.* 150, 112–116.
- Böckerman, P., Viinikainen, J., Vainiomäki, J., et al., 2017. Stature and long-term labor market outcomes: evidence using Mendelian randomization. *Econ. Hum. Biol.* 24, 18–29.
- Burkhauser, R.V., Cawley, J., 2008. Beyond BMI: the value of more accurate measures of fitness and obesity in social science research. *J. Health Econ.* 27, 519–529.
- Clark, D., Roayer, H., 2013. The effect of education on adult mortality and health: evidence from Britain. *Am. Econ. Rev.* 103, 2087–2120.
- Cohen, A.K., Rai, M., Rehkopf, D.H., Abrams, B., 2013. Educational attainment and obesity: a systematic review. *Obes. Rev.* 14, 989–1005.
- Conley, D., 2016. Socio-genomic research using genome-wide molecular data. *Annu. Rev. Sociol.* 42, 275–299.
- Davey Smith, G., Paternoster, L., Relton, C., 2017. When will Mendelian randomization become relevant for clinical practice and public health? *J. Am. Med. Assoc.* 317, 589–591.
- Davillas, A., Benzeval, M., 2016. Alternative measures to BMI: exploring income-related inequalities in adiposity in Great Britain. *Soc. Sci. Med.* 166, 223–232.
- Delaneau, O., Marchini, J., Zagury, J.F., 2012. A linear complexity phasing method for thousands of genomes. *Nat. Methods* 9, 179–181.
- Eide, E.R., Showalter, M.H., 2011. Estimating the relation between health and education: what do we know and what do we need to know? *Econ. Educ. Rev.* 30, 778–791.
- Global BMI Mortality Collaboration, 2016. Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. *Lancet* 388, 776–786.
- Gupta, V., Walia, G.K., Sachdeva, M.P., 2017. 'Mendelian randomization': an approach for exploring causal relations in epidemiology. *Public Health* 145, 113–119.
- 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.
- Johansson, E., Böckerman, P., Kiiskinen, U., Heliövaara, M., 2009. Obesity and labour market success in Finland: the difference between having a high BMI and being fat. *Econ. Hum. Biol.* 7, 36–45.
- van Kippersluis, H., Rietveld, C.A., 2017. Pleiotropy-robust Mendelian randomization. *Int. J. Epidemiol.* <http://dx.doi.org/10.1093/ije/dyx002> (forthcoming).
- Marmot, M., 2015. *The Health Gap: The Challenge of an Unequal World*. Bloomsbury Publishing, London.
- Notwell, J.H., Heavner, W.E., Darbandi, S.F., et al., 2016. *TBR1* regulates autism risk genes in the developing neocortex. *Genome Res.* 26, 1013–1022.
- 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.
- Paciorkowski, A.R., Traylor, R.N., Rosenfeld, J.A., et al., 2013. *MEF2C* Haploinsufficiency features consistent hyperkinesia, variable epilepsy, and has a role in dorsal and ventral neuronal developmental pathways. *Neurogenetics* 14, 99–111.
- Palmer, T.M., Lawlor, D.A., Harbord, R.M., et al., 2012. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat. Methods Med. Res.* 21, 223–242.
- Raitakari, O.T., Juonala, M., Rönnemaa, T., et al., 2008. Cohort profile: the cardiovascular risk in Young Finns study. *Int. J. Epidemiol.* 37, 1220–1226.
- Rothman, K.J., 2008. BMI-related errors in the measurement of obesity. *Int. J. Obes.* 32, S56–S59.
- Spasojević, J., 2010. Effects of education on adult health in Sweden: results from a natural experiment. In: Slotte, D., Tchernis, R. (Eds.), *Current Issues in Health Economics*. Emerald Group Publishing Limited, pp. 179–199.
- Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. *Econometrica* 65, 447–486.
- Stommel, M., Schoenborn, C.A., 2009. Accuracy and usefulness of BMI measures based on self-reported weight and height: findings from the NHANES & NHIS 2001–2006. *BMC Public Health* 9 (1), 421.
- Teo, Y.Y., Inouye, M., Small, K.S., et al., 2007. A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics* 23, 2741–2746.
- Trogdon, J.G., Nonnemaker, J., Pais, J., 2008. Peer effects in adolescent overweight. *J. Health Econ.* 27, 1388–1399.
- Tyrrrell, J., Jones, S.E., Beaumont, R., et al., 2016. Height, body mass index, and socioeconomic status: Mendelian randomisation study in UK Biobank. *Br. Med. J.* 352, i582.