Economics and Human Biology 46 (2022) 101134

Contents lists available at ScienceDirect

Economics and Human Biology

journal homepage: www.elsevier.com/locate/ehb

Does better education mitigate risky health behavior? A mendelian randomization study

Jutta Viinikainen [a,*,1], Alex Bryson [b,c,d,2], Petri Böckerman [a,d,e,3], Jaana T. Kari [a,4], Terho Lehtimäki [f,g,h,i,5], Olli Raitakari [j,k,l,6], Jorma Viikari [m,n,7], Jaakko Pehkonen [a,8]

[a] University of Jyväskylä, Jyväskylä University School of Business and Economics, Jyväskylä, Finland
[b] University College London, Social Research Institute, London, United Kingdom
[c] National Institute of Economic and Social Research, London, United Kingdom
[d] IZA Institute of Labor Economics, Bonn, Germany
[e] Labour Institute for Economic Research LABORE, Helsinki, Finland
[f] Tampere University, Department of Clinical Chemistry, Tampere, Finland
[g] Fimlab Laboratories, Tampere, Finland
[h] Tampere University, Faculty of Medicine and Health Technology, Tampere, Finland
[i] Tampere University, Finnish Cardiovascular Research Center, Tampere, Finland
[j] University of Turku and Turku University Hospital, Centre for Population Health Research, Turku, Finland
[k] University of Turku, Research Centre of Applied and Preventive Cardiovascular Medicine, Turku, Finland
[l] Turku University Hospital, Department of Clinical Physiology and Nuclear Medicine, Turku, Finland
[m] University of Turku, Department of Medicine, Turku, Finland
[n] Turku University Hospital, Division of Medicine, Turku, Finland

ARTICLE INFO

JEL classification:
I12
I20
I26

Keywords:
Education
Health behavior
Mendelian randomization
Smoking
Diet
Abusive drinking

ABSTRACT

Education and risky health behaviors are strongly negatively correlated. Education may affect health behaviors by enabling healthier choices through higher disposable income, increasing information about the harmful effects of risky health behaviors, or altering time preferences. Alternatively, the observed negative correlation may stem from reverse causality or unobserved confounders. Based on the data from the Cardiovascular Risk in Young Finns Study linked to register-based information on educational attainment and family background, this paper identifies the causal effect of education on risky health behaviors. To examine causal effects, we used a genetic score as an instrument for years of education. We found that individuals with higher education allocated more attention to healthy habits. In terms of health behaviors, highly educated people were less likely to smoke. Some model specifications also indicated that the highly educated consumed more fruit and vegetables, but the results were imprecise in this regard. No causal effect was found between education and abusive drinking. In brief, inference based on genetic instruments showed that higher education leads to better choices in some but not all dimensions of health behaviors.

* Correspondence to: Jyväskylä University School of Business and Economics, University of Jyväskylä, P.O. Box 35, FI-40014 Jyväskylä, Finland.
E-mail address: jutta.viinikainen@jyu.fi (J. Viinikainen).
[1] ORCID: 0000-0002-4252-3147
[2] ORCID: 0000-0003-1529-2010
[3] ORCID: 0000-0002-5372-2985
[4] ORCID: 0000-0001-5205-7031
[5] ORCID: 0000-0002-2555-4427
[6] ORCID: 0000-0001-9365-3702
[7] ORCID: 0000-0001-6452-010X
[8] ORCID: 0000-0002-9684-7139

https://doi.org/10.1016/j.ehb.2022.101134
Received 1 March 2021; Received in revised form 14 March 2022; Accepted 15 March 2022
Available online 23 March 2022

1. Introduction

Health is an important determinant of human well-being. Individuals with poorer health have lower life satisfaction and weaker labor market attachment (Strine et al., 2008; Currie, 2009; Sánchez and Tassot, 2014; Almond et al., 2018). At the societal level, the economic burden of poor health is substantial, so much so that improving health and well-being is a key policy priority in most countries. The high economic and societal costs of adverse health conditions reinforce the need for policy measures that prevent or alleviate these problems.

Health disparities begin to develop early in life. Twin and genetic epidemiological studies have revealed that observed variations in health are partly explained by genetic makeup (see, e.g., Burton et al., 2005; Polderman et al., 2015). *In utero* environments are also crucial, with studies showing that maternal exposure to the influenza pandemic in 1918 (Almond and Mazumder, 2005), prenatal malnutrition during the Dutch Hunger Winter (Roseboom et al., 2011), exposure to Ramadan (Schoeps et al., 2018), and sudden reductions in air pollution due to plant closures during the fetal period (Chay and Greenstone, 2003) all have long-lasting impacts on health. Individuals' initial overall health stock depreciates with age, but health can be maintained and improved by making investments in health capital (Grossman, 1972). Such investments include the allocation of time for physical exercise, reading health-related information, and the consumption of market goods, such as healthcare services and diets. In contrast, detrimental, risky health behaviors, such as excessive alcohol consumption, unhealthy diet, and heavy smoking, accelerate the depreciation of health capital. Thus, an individual's state of health partly reflects the resources they have allocated to health production.

Because health behaviors are important determinants of health, policies that encourage individuals to engage in better health behaviors are also likely to promote health. In this paper, we study the effect of education on risky health behaviors. We focus on smoking, abusive alcohol consumption, diet (i.e., the consumption of fruits and vegetables), and the level of attention toward healthy habits. Previous studies on education and health behaviors aiming for causal inference have been mostly based on historical school reforms, which may limit their external validity. To identify the causal impact of education on healthy behaviors, we use a genetic score as an instrumental variable for education. Studies that use genetic endowments to identify the link between education and health behavior are still rare and, to the best of our knowledge, they are all based on UK Biobank data (Gage et al., 2018; Davies et al., 2019; Rosoff et al., 2021; Sanderson et al., 2019; Zhou et al., 2019). Our study extends this research by providing evidence outside the UK context using data from another highly economically developed country, Finland. Our data, which is rich in terms of covariates, allowed us to focus on several closely connected measures of health behaviors (attention toward healthy habits, smoking, abusive drinking, and consumption of fruits and vegetables) that provide a more comprehensive picture of the effect of education on health behaviors.[9] A novelty of our empirical approach is that we formally evaluated the sensitivity of our estimates to the exclusion restriction of the Mendelian randomization method, following an approach proposed by van Kippersluis and Rietveld (2018).

2. Previous work

Several studies have found that higher education is related to better health (for a meta-analysis, see Furnee et al., 2008). A potential explanation for this relationship is that education leads to better health because the highly educated have better access to healthcare services and higher disposable incomes, thus enabling healthier lifestyle choices (e.g., Cutler and Lleras-Muney, 2006). Education may also increase the productive and allocative efficiencies of health production, leading to better health (Grossman, 2006). Assuming that higher education increases productive efficiency, the more educated are able to attain better health from a given level of input. They may, for example, better understand a doctor's advice. According to the allocative efficiency hypothesis, in turn, more educated individuals are more likely to choose inputs that promote health than the less educated. Higher education has also been linked to better health literacy skills (van der Heide et al., 2013) and an increased willingness to delay gratification (Perez-Arce, 2017). Therefore, a better understanding of the consequences of various health behaviors (e.g., with regard to diet and smoking) and higher patience levels among highly educated individuals may support their ability to make healthy choices.

Education may improve health, but causality may also run from better health to higher educational attainment or, alternatively, common factors such as parental background may explain the relationship. Studies using statistical methods for causal inference have been inconclusive in terms of the link between education and health. In their meta-analysis, Hamad et al. (2018) focused on studies that utilized historical changes in compulsory schooling laws as an exogenous source of variation in education. The results indicated that education had mixed but largely beneficial effects on a range of health outcomes. However, Xue et al. (2021) concluded that after correcting for publication bias, the effect of education on health outcomes was essentially zero. Furthermore, studies ignoring endogeneity tended to overstate the link between education and health. Thus, the evidence on the relationship between education and health is inconclusive.

Previous studies have also found a significant negative correlation between education and measures of risky health behavior. Among highly educated individuals, the prevalence of smoking and binge drinking has been found to be much lower, and higher education levels have also been shown to be related to better dietary choices (e.g., Cawley and Ruhm, 2011; Cutler and Lleras-Muney, 2010; De Irala-Estevez et al., 2000). However, as in the case of education and health, a key challenge in such empirical studies is that the negative correlation between education and risky health behaviors may also be driven by unobserved characteristics. For example, instead of being an outcome of higher education, differences in time preferences may reflect initial differences in attitudes toward postponed utility, hence affecting both education choices and risky health behaviors (Farrell and Fuchs, 1982). In addition, reverse causality may explain the correlation between education and health behaviors. For example, excessive drinking during the teenage years may lead to lower educational attainment in adulthood (Cook and Moore, 1994; Renna, 2007). If the association between education and health behaviors reflects omitted variables or reverse causality, this implies that investments in education would not reduce poor health by altering or preventing unhealthy behaviors.

Empirical studies have typically found a negative association between education and smoking status, but the results based on exogenous changes in schooling laws have been inconclusive, indicating either a negative (Arendt, 2005; Jensen and Lleras-Muney, 2012; Jürges et al., 2011) or zero effect (e.g., Braakmann, 2011; Clark and Royer, 2013; Dilmaghani, 2021; Kemptner et al., 2011; Li and Powdthavee, 2015). Similarly, higher education has been related to higher alcohol consumption (e.g., Cutler and Lleras-Muney, 2010), but this relationship may not be causal (Braakmann, 2011; Clark and Royer, 2013; Fletcher, 2015). However, education may alter drinking patterns by reducing binge drinking (Li and Powdthavee, 2015). The results concerning fruit and vegetable consumption have also been inconclusive. Previous studies have typically found a positive correlation between education and fruit and vegetable consumption (Braakmann, 2011; Clark and Royer, 2013; Cutler and Lleras-Muney, 2010; Li and Powdthavee, 2015),

[9] Obesity is potentially an important health-related outcome that may be related to education. However, an earlier paper (Böckerman et al., 2017) using the same data investigated the relationship between education and BMI/obesity using the Mendelian randomization method. Hence, we do not use BMI as an outcome variable in this paper.

but causal evidence suggests both positive (Li and Powdthavee, 2015) and zero (Braakmann, 2011; Clark and Royer, 2013) effects.

Another strand of instrumental variable studies has used genetic scores as an instrument for education. This method, called Mendelian randomization (MR), uses the random inheritance of genetic material from one's mother or father to isolate the causal impact of inherited traits, such as the propensity for more education, on outcomes such as healthy behaviors. Findings from such MR studies have suggested that higher education reduces smoking (Gage et al., 2018; Davies et al., 2019; Sanderson et al., 2019) and may change drinking patterns. Higher education appeared to increase alcohol intake (Davies et al., 2019; Zhou et al., 2019) because of increased drinking frequency (Rosoff et al., 2021; Zhou et al., 2019), but binge drinking was shown to be less prevalent among the highly educated (Rosoff et al., 2021). Thus, the higher disposable income of more educated individuals may increase their overall alcohol consumption, but a higher awareness of risks related to binge drinking or social norms may reduce the number of drinks per drinking day. Rosoff et al. (2021) also investigated the relationship between education and alcohol dependency. They did not find a relationship between education and alcohol dependency measures, such as the inability to cease drinking or the frequency of needing a morning drink.

3. Data and methods

3.1. Data

The Cardiovascular Risk in Young Finns Study (YFS) is a longitudinal study of 3596 participants who were randomly chosen from five Finnish university regions in 1980 (Raitakari et al., 2008). The participants represent six age cohorts (ages 3, 6, 9, 12, 15, and 18 years in 1980), and since 1980, several follow-ups have been conducted. The YFS includes comprehensive information on genetic markers, that are required to calculate the polygenic score (PGS), which we used as an instrument for years of education.

To obtain information on YFS participants' educational attainment, the YFS was linked to the Finnish Longitudinal Employer-Employee Data (FLEED) of Statistics Finland using unique personal identifiers. The matching was exact; that is, there were no misreported identification (ID) codes. Data on parental education was drawn from Statistics Finland's Longitudinal Population Census (LPC) from 1980 and were linked to the YFS-FLEED using personal identifiers.

3.2. Methods

3.2.1. Statistical model

To replicate standard observational studies of the literature, we first used an Ordinary Least Squares (OLS) estimation where a health behavior (y_{hi}, where h refers to a health behavior and i to an individual) was regressed on years of education ($educ$), and a vector of predetermined control variables (X): cohort, sex, region of residence, and parental education (Eq. 1). Because random sampling of YFS was conducted at the individual level, there was no need to cluster standard errors.

$$y_{hi} = \beta_0 + \beta_1 educ_i + \beta_2' X_i + \varepsilon_i \quad (1)$$

To identify causal effects, we used the instrumental variables estimation method, which employed a PGS for years of education as an instrument for education. In the first stage, the variable reflecting years of education was regressed on the instrument (PGS) and predetermined controls (X; Eq. 2). In the second stage, the education variable in Eq. (1) was replaced with the predicted values and estimated with OLS (Eq. 3):

$$\widehat{educ}_i = \delta_0 + \delta_1 PGS_i + \delta_2' X_i \quad (2)$$

$$y_{hi} = \alpha_0 + \alpha_1 \widehat{educ}_i + \alpha_2' X_i + \varepsilon_i \quad (3)$$

As discussed below, under suitable conditions, this method—called MR—identifies the local average treatment effect (LATE) on compliers (i.e., those whose years of education were higher because of genetic inheritance) and avoids biases related to the OLS results (von Hinke et al., 2016).

Earlier research has identified differences in health behaviors between females and males. For example, males have been shown to be more likely to smoke or engage in binge drinking than females (Cawley and Ruhm, 2011). To account for this, we also estimated the models separately for females and males.

3.2.2. MR assumptions for identification of causal effects

According to Mendel's law of segregation (first law) and independent assortment (second law), alleles segregate randomly when they are passed from one generation to the next, and each trait is inherited independently from other traits at conception. In the MR method, this exogenous variation is used to identify the causal links between the exposure—in our case, education—and the outcome variable. The rapid decline in genome sequencing costs has led to an increase in the number of genome-wide association studies (GWASs) that attempt to find relationships between genetic variants (a single nucleotide polymorphism, SNP) and expressed phenotypes. By summing up the number of SNPs that have been associated with the phenotype of interest in a GWAS, researchers have created PGSs that indicate a genetic risk of developing a disease or some other trait. In addition to raw frequency PGSs, weighted PGSs—in which each SNP is weighted based on its effect size in the GWAS—have also been used as instruments in the MR setting to capture exogenous variation in the exposure variable.

The MR estimator avoids biases related to the OLS estimator (von Hinke et al., 2016) under the four key assumptions as follows: 1) independence: the PGS is not associated with any confounder of the education–health behavior relationship—that is, the PGS is as good as randomly assigned; 2) relevance: there is a nonzero effect of the instrument on treatment—that is, the PGS is statistically significantly associated with education (strong instrument/relevance assumption); 3) exclusion: the PGS affects health behavior only via its association with education; and 4) monotonicity: higher PGS values lead only to higher, not lower, educational attainment for each individual.

The independence assumption could be violated because of population stratification (allele frequencies differ between population subgroups), assortative mating (phenotypes affect partner selection), or dynastic effects (parental phenotypes directly affect offspring phenotypes). All participants in the YFS were Caucasian, which minimized the possibility that systematic allele differences between ethnic subgroups biased the results. Furthermore, the features of the Finnish education system minimize the potential threat that dynastic effects violate the independence assumption: education is free of charge at all levels, and the role of private schools in the national education system, which are also tuition free, is minimal. To address potential assortative mating, we used a variable for parental education, which separated between neither parent, one parent, and both parents having attained university-level education. In this study, we also evaluated the independence assumption by testing whether there were systematic differences in observed characteristics between individuals with different PGS levels by comparing the observable variables by the median value of the instrument (von Hinke et al., 2016). To further assess the possibility that observed characteristics are related to the education PGS, we performed a balance test by regressing the education PGS on exogenous variables (i.e., family background), while simultaneously controlling for sex, age cohort, and regional fixed effects.

The relevance assumption is satisfied if the PGS is robustly associated with years of education and if this correlation is strong, the instrument is considered to have a strong first stage. The association between the PGS

SNPs and years of education was examined in a GWASs by Okbay et al. (2016a). In our main analysis, we utilized a PGS that was obtained using a significance threshold $p < 0.01$. The most important advantage of this instrument was its strength due to its inclusion of many SNPs (the actual number of SNPs is unknown). High instrument strength limits the possibility for finite sample bias, which is expected to decrease as instrument strength increases. In addition, weak instruments tend to increase bias, stemming from violations of the other MR assumptions (McMartin and Conley, 2020). Staiger and Stock (1997) have suggested that the minimum standard for a strong first stage is that the first-stage F statistics exceed the value of 10. We used this general rule to evaluate the strong instrument assumption.

The exclusion restriction requires that the PGS affects health behaviors only via its association with education. This assumption could be violated if genetic variants related to education years also affect health behaviors either directly or through pathways other than education (pleiotropy) or if they are in linkage disequilibrium (co-inherited) with SNPs affecting health behaviors via other pathways. Earlier studies found that SNPs related to years of education may have also been associated with noncognitive skills (Demange et al., 2021), mental health (Okbay et al., 2016a; Lam et al., 2019; Lee et al., 2019), cognition, body mass index (BMI), and height (Okbay et al., 2016a). To provide the complementary results that may take into account the bias stemming from pleiotropy related to these traits, we augmented our baseline models with PGSs for the Big Five personality traits (openness, conscientiousness, extraversion, agreeableness, and neuroticism), depressive symptoms, bipolar disorders, schizophrenia, childhood IQ, BMI, and height.

As noted earlier, the advantage of the instrument we used in our main analysis is its strength. However, the drawback of using an instrument with numerous SNPs is that the risk of pleiotropy may increase. To address this issue, we also used an alternative PGS that included 74 SNPs, which were associated with years of education at $p < 5 \times 10^{-8}$ in Okbay et al. (2016a). For the 74 SNP PGSs, we had information on individual SNPs that allowed us to perform Sargan's test of overidentifying restrictions to assess the validity of the 74 SNP PGSs. We also performed Sargan's test using our baseline PGS and the PGS based on 74 SNPs as instruments. Finally, we used the pleiotropy robust MR method van Kippersluis and Rietveld (2018) for sensitivity analysis to determine how strong the violation of the exclusion restriction would have to be for the causal effect to be zero.

Finally, the monotonicity assumption may be violated due to gene–environment interaction (G×E), which occurs when genetic predispositions are expressed differently in different environments. For example, tuition fees may prevent poor students with high genetic predispositions toward education from obtaining higher levels of education. In Finland, tuition-free education and extensive state-provided financial support to students at the university level substantially reduce the possibility that financial constraints would hamper the opportunities for children from lower socioeconomic backgrounds to obtain higher levels of education. In addition, a geographically extensive higher education network in Finland reduces regional disparities in educational opportunities. However, to mitigate potential biases from G×E, we estimated a reduced-form model in which the outcome variable was explained by the PGS, as suggested by VanderWeele et al. (2014). Because reduced-form models do not exploit information on the exposure that may express itself differently depending on the environment, potential biases stemming from G×E are eliminated. The reduced-form estimates identify the effect of the exposure on the outcome but not the quantitative size of the effect of interest.

3.3. Measures

3.3.1. Outcome variables

We focused on the following four health behavior measures that were determined in 2011: 1) attention participants paid to their healthy habits, 2) abusive drinking patterns, 3) smoking, and 4) consumption of fruit and vegetables. These also constitute the key components of health behavior that have been analyzed in the literature. The amount of attention participants paid to their health habits was determined based on responses to the question, "To what extent do you focus on your health habits?" The question was assessed on a five-point scale (1 = I hardly pay any attention; 5 = I pay a lot of attention). The respondents' abusive drinking patterns were assessed using 11 self-rated questions (see Appendix 1). The responses were given in yes/no format (0 = no; 1 = yes), and the sum of the items was used to measure the severity of abusive drinking behaviors. Long-term smoking was measured using pack years—that is, by multiplying the number of packs of cigarettes smoked per day by the years the person has smoked. Cigarette pack years measure the cumulative lifetime consumption of cigarettes. For example, a person has a 10-pack-year history of smoking if he or she has smoked one pack of cigarettes daily for 10 years. Information on the consumption of fruits and vegetables was collected using a food frequency questionnaire, which has been developed and validated by the Finnish National Institute for Health and Welfare (Paalanen et al., 2006). The participants were asked to report the daily frequency and proportion size of selected food items over the past 12 months. Average daily intake (in grams) of specific food groups was calculated using the National Food Consumption Database, Fineli. The consumption of fruits was based on five food items (citrus, apple, other fruits, fruit preserves, and berries), and the consumption of vegetables was based on eight items (root vegetables, lettuce, cabbage, fruit vegetables, mushrooms, onions, vegetable preserves, beans, and peas). The group "fruit vegetables" includes vegetables that are technically fruits—for example, avocado, tomato, pepper, cucumber, pumpkin, and eggplant.

3.3.2. Endogenous variable: years of education

Information on educational attainment in 2011 was drawn from the FLEED. The level of education was converted into years of education using the official formula used by Statistics Finland (i.e., upper secondary education and postsecondary nontertiary education = 12 years; short-cycle tertiary education = 14 years; bachelor's level = 16 years; master's level = 18 years; and doctoral level = 21 years). In 2011, the YFS participants were between 34 and 49 years old. However, a small fraction (1.6% of n = 1719) were still studying, so we ascribed their years of education based on the highest degree they had obtained until that point.

3.3.3. Instrumental variables

To identify the causal links between education and health behaviors, we used two alternative PGSs for years of education. In our main models, we used a PGS that was calculated as a weighted sum of the genotyped risk alleles (or imputed allele dosages) associated with years of education ($p < 0.01$) in a GWAS by Okbay et al. (2016a). The weights were based on the effect sizes of each SNP to years of education. In the robustness tests, we used a weighted PGS associated with years of education at a significance level of $p < 5 \times 10^{-8}$ that included 74 SNPs (Okbay et al., 2016a). Further information on genotyping and calculation of the PGSs is provided in Appendix 2.

3.3.4. Control variables

In all models, we used the following baseline controls: parental education (0 = neither parent, 1 = one parent, and 2 = both parents have completed university-level education by the year 1980, from the LPC); the region of residence in 1980 (four indicator variables from the YFS: Southern Finland, Eastern Finland, Western Finland, and Northern Finland); and birth year (from the YFS).

We also used the following PGSs in robustness analyses: the number of cigarettes smoked per day (Furberg et al., 2010), the Big Five personality traits (de Moor et al., 2012), BMI (Locke et al., 2015), height (van der Valk et al., 2015), childhood IQ (Benyamin et al., 2014), schizophrenia (Schizophrenia Working Group of the Psychiatric

Genomics Consortium, 2014), bipolar disorders (Ruderfer et al., 2014), and depressive symptoms (Okbay et al., 2016b).

4. Results

4.1. Descriptive analyses

The average age of the participants in the sample (n = 1719) was 41.9 years in 2011, the share of females in the sample was 55.7%, and 13.0% of the participants had at least one university-educated parent. The bivariate analyses (Table 1) revealed that higher education (i.e., above-median years of education) was significantly associated with better health behaviors. We found that individuals with higher education reported that they allocated more attention to healthy habits, were less likely to smoke, displayed less abusive drinking patterns, and consumed more fruits and vegetables.

Table 2 compares the observed variables by the median value of the instrument. Consistent with the patterns in Table 1, individuals with a higher genetic propensity toward higher education were less likely to smoke and expressed that they paid more attention to healthy habits. Individuals with higher education PGSs also displayed less abusive drinking patterns and consumed more fruit and vegetables, but in these cases, the differences between the low and high PGS groups were not significant. The average age and the share of females did not differ between high and low PGS individuals, which was consistent with the independence assumption of MR. Furthermore, the completed years of education were higher among individuals whose PGS values exceeded the median level, which was consistent with the relevance assumption of the instrument.

Table 2 further shows that the share of the participants with high parental education was higher among high PGS individuals, and genetic variation was clustered between geographical areas. This highlighted the importance of controlling these variables to ensure that thereafter the instrument would not correlate with the error term and, thus, the exclusion restriction would not be violated. In addition, the average level of PGSs for the Big Five of conscientiousness, bipolar disorders, schizophrenia, and BMI differed between individuals with low and high genetic propensities for higher education. The results from the balance test where the education PGS was regressed on parental education, while simultaneously controlling for sex, age cohort, and regional fixed effects, indicated that family background was significantly correlated with the PGS for education (Appendix 3). This highlighted the importance of controlling for parental background in the analyses.

4.2. Main results

The OLS estimates (Table 3, Column 1) showed that higher education was associated with paying greater attention to healthy habits, less abusive alcohol use, lower levels of tobacco consumption, and increased consumption of fruit and vegetables. The results showed that a one-year increase in years of education was associated with a 0.07 point (0.074 SD) increase in attention paid towards healthy habits, a 0.07 point (0.026 SD) decrease in abusive alcohol consumption patterns, 0.404 fewer pack-years of tobacco use, and a 7.306 g (0.027 SD) increase in fruit and vegetable consumption per day. The MR results (Table 3, Column 2) supported the OLS findings that the link was causal in the case of attention paid to healthy habits and smoking. The MR point estimates were larger compared to the OLS estimates indicating that a one-year increase in years of education was associated with a 0.147 point (0.155 SD) increase in attention paid towards healthy habits and a 0.820 pack-year reduction in tobacco consumption. The MR point estimate for fruit and vegetable consumption also increased compared to the corresponding OLS point estimate ($\alpha_{MR} = 14.980$ vs. $\beta_{OLS} = 7.306$), but the coefficient was imprecisely estimated, making the coefficient only marginally ($p < 0.10$) significant. The MR results did not show a statistically significant link between education and abusive drinking patterns. The first-stage F statistics in the MR models exceeded the level of 10, supporting the relevance assumption.

Table 3, Columns 3 and 4, reports the results separately for females and males. Based on the MR findings, both females and males with higher education allocated more attention to healthy habits and smoked less. The point estimates for fruit and vegetable consumption were also positive, but as in Column 2, they were imprecisely estimated and, thus, statistically insignificant.

4.3. Robustness analyses

To assess the sensitivity of our findings, we first augmented our baseline model with PGSs for traits, which, according to the previous literature, may share common genetic components with years of education, namely the Big Five personality traits, depressive symptoms, bipolar disorders, schizophrenia, childhood IQ, BMI, and height. The results (Appendix 4, Column 1) showed that these additional controls had only a minimal effect on the point estimates.

Second, we estimated a reduced-form model, where health behaviors were regressed on the PGS. The results from the reduced-form models (Appendix 4, Column 2) were consistent with the MR results, indicating that individuals with a higher genetic predisposition toward higher education paid greater attention to healthy habits and were less likely to smoke.

Third, the results based on the 74 SNP PGSs (Appendix 4, Column 3) also indicated that individuals with higher education paid more attention to healthy habits and smoked less. However, contrary to previous model specifications, the point estimate indicating the link between education and fruit and vegetable consumption was negative, although imprecisely estimated ($p > 0.950$), with the first-stage F statistics below the general rule of 10. Sargan's overidentifying restrictions tests supported the null hypothesis that all 74 SNPs yielded the same MR estimate ($p > 0.478$), thus lending support to the instrument's validity. In addition, when the 74 SNP PGSs and the PGSs used in our baseline models were used as instruments, Sargan's tests did not reject the null hypothesis of instrument exogeneity (regarding all outcome variables,

Table 1
Comparison of health behaviors by completed years of education.

	All	Above-median years of education	Below- median years of education	Difference	t-statistics	N
Attention toward healthy habits (2011)	3.538 (0.947)	3.732 (0.870)	3.323 (0.981)	0.410	9.116 ***	1719
Abusive alcohol consumption (scale: 0–11; 2011)	3.147 (2.666)	2.920 (2.554)	3.403 (2.766)	-0.483	-3.689 ***	1655
Smoking (pack years; 2011)	1.445 (5.391)	0.499 (2.581)	2.720 (7.516)	-2.221	-7.444 ***	1620
Fruit and vegetable consumption (g/day; 2011)	448.554 (275.130)	472.191 (259.213)	421.902 (289.922)	50.289	3.509 ***	1468

Notes: The table reports the means and standard deviations in parentheses. The differences between groups were tested using a two-sample t-test. Statistically significant at the * 10%, ** 5%, and *** 1% levels.

Table 2
Comparison of observables by the instrument value.

	All	Above-median PGS	Below-median PGS	Difference	t-statistics	N
Attention toward healthy habits (2011)	3.538	3.639	3.437	0.202	4.446 ***	1719
	(0.947)	(0.930)	(0.953)			
Abusive alcohol consumption (scale 0–11; 2011)	3.147	3.131	3.163	-0.032	-0.245	1655
	(2.666)	(2.677)	(2.656)			
Smoking (pack years, 2011)	1.445	0.929	2.006	-1.077	-3.991 ***	1620
	(5.391)	(4.604)	(6.086)			
Fruit and vegetable consumption (g/day; 2011)	448.554	452.471	444.572	7.899	0.550	1468
	(275.130)	(266.542)	(283.721)			
Female (share)	0.557	0.556	0.558	-0.002	-0.070	1719
	(0.497)	(0.497)	(0.497)			
Age (2011)	41.904	41.848	41.960	-0.113	-0.465	1719
	(5.040)	(5.126)	(4.954)			
Additional PGSs						
Openness	-6.698	-6.437	-6.959	0.522	1.038	1719
	(10.422)	(10.724)	(10.111)			
Conscientiousness	-1.536	-0.927	-2.144	-1.217	2.201 **	1719
	(11.474)	(11.509)	(11.474)			
Extraversion	1.158	1.288	1.029	0.259	0.476	1719
	(11.282)	(11.589)	(10.973)			
Agreeableness	37.763	37.930	37.597	0.332	0.726	1719
	(9.494)	(9.457)	(9.534)			
Neuroticism	-6.012	-6.170	-5.855	-0.315	-0.458	1719
	(14.261)	(13.893)	(14.627)			
Depressive symptoms	0.029	0.022	0.035	-0.014	-0.463	1719
	(0.610)	(0.607)	(0.614)			
Bipolar disorders	-8.552	-8.350	-8.754	0.405	2.648 ***	1719
	(3.172)	(3.135)	(3.199)			
Schizophrenia	-4.302	-4.427	-4.178	-0.249	-1.754 *	1719
	(2.944)	(2.926)	(2.958)			
Childhood IQ	0.977	1.017	0.937	0.080	1.528	1719
	(1.091)	(1.097)	(1.085)			
BMI	0.968	0.907	1.029	-0.122	-3.479 ***	1719
	(0.729)	(0.707)	(0.745)	(0.035)		
Height	2.123	2.214	2.032	0.182	1.466	1719
	(2.569)	(2.568)	(2.567)			
Region of residence (1980)						
- Southern Finland	0.166	0.182	0.151	0.030	1.695 *	1719
	(0.373)	(0.386)	(0.358)			
- Western Finland	0.362	0.364	0.359	0.005	0.219	1719
	(0.481)	(0.482)	(0.480)			
- Eastern Finland	0.310	0.290	0.330	-0.040	-1.810 *	1719
	(0.463)	(0.454)	(0.471)			
- Northern Finland	0.162	0.164	0.159	0.005	0.273	1719
	(0.368)	(0.371)	(0.366)			
Education years (2011)	13.994	14.746	13.242	1.504	11.888 ***	1719
	(2.728)	(2.812)	(2.420)			
Parental education (scale 0–2; 1980)	0.181	0.244	0.117	0.127	5.312 ***	1719
	(0.500)	(0.566)	(0.414)			

Notes: The table reports the means and standard deviations in parentheses. The differences between groups were tested using a two-sample t-test. Parental education variable equals 0 if neither parent, 1 if one parent, and 2 if both parents had university-level education. Statistically significant at the * 10%, ** 5%, and *** 1% levels.

$p > 0.206$).

Fourth, we augmented the model for tobacco consumption with a PGS for cigarettes smoked per day. Controlling for a genetic predisposition toward smoking had only a very minor effect on the point estimate for years of education (baseline MR model, Table 3, Column 2: −0.820, $p < 0.01$; augmented model: −0.816, $p < 0.01$). YFS data did not contain a PGS for other outcome variables.

As a final robustness check, we followed van Kippersluis and Rietveld (2018) approach to assess the sensitivity of our results to the violation of the exclusion restriction. We found that any violation of the exclusion restriction would have to account for 43–44% of the first stage effect in order to reduce the education point estimates to zero in relation to healthy habits and smoking. In the case of fruit and vegetable consumption, violation of the exclusion restriction would have to account for 19% of the variance in the first stage to reduce the causal effect to zero. These results are shown and discussed in Appendix 5.

5. Discussion

Using longitudinal data that combined survey information on health behaviors with register-based information on education, we found that higher education was related to better health behaviors in terms of paying greater attention to healthy habits, less abusive drinking patterns, lower levels of tobacco consumption, and higher amounts of fruit and vegetable consumption. The MR results, which identified the causal effect of education on health behaviors, showed that this link was causal in the case of attention paid to healthy habits and smoking. The MR specifications implied that education increased fruit and vegetable intake, but these coefficients were imprecisely estimated. The MR results did not reveal a statistically significant link between education and abusive alcohol consumption.

Previous studies have shown a strong link between low education and a higher incidence of risky health behaviors, such as smoking, binge drinking, and dietary choices (e.g., De Irala-Estevez et al., 2000; Cawley and Ruhm, 2011). Our OLS estimates were consistent with these results. Previous studies that used historical changes in schooling laws to

Table 3
Education and health behaviors. Results based on linear regression (OLS) and Mendelian randomization (MR) methods.

	(1) OLS	(2) MR	(3) MR, female	(4) MR, male
Attention toward healthy habits	0.070 ***	0.147 ***	0.150 ***	0.139 ***
	(0.009)	(0.030)	(0.038)	(0.046)
	[0.053; 0.087]	[0.089; 0.205]	[0.075; 0.225]	[0.050; 0.229]
First-stage F statistics	.	162.20	103.71	64.16
R^2	0.093	.	.	.
N	1719	1719	958	761
Abusive alcohol consumption	-0.070 ***	-0.043	-0.092	0.031
	(0.025)	(0.083)	(0.104)	(0.132)
	[− 0.119; − 0.022]	[− 0.206; 0.120]	[− 0.297; 0.112]	[− 0.228; 0.290]
First-stage F statistics	.	155.13	96.41	64.01
R^2	0.069	.	.	.
N	1655	1655	918	737
Smoking (pack years)	-0.404 ***	-0.820 ***	-0.586 **	-1.069 ***
	(0.052)	(0.197)	(0.240)	(0.311)
	[− 0.506; − 0.301]	[− 1.206; − 0.435]	[− 1.056; − 0.116]	[− 1.677; − 0.460]
First-stage F statistics	.	153.09	88.32	69.79
R^2	0.068	.	.	.
N	1620	1620	913	707
Fruit and vegetable consumption (g/day)	7.306 ***	14.980 *	10.261	18.882
	(2.751)	(9.108)	(13.029)	(12.003)
	[1.911; 12.702]	[− 2.872 32.832]	[− 15.275; 35.797]	[− 4.644; 42.407]
First-stage F statistics	.	129.05	70.26	62.16
R^2	0.062	.	.	.
N	1468	1468	848	620

Notes: The table reports regression coefficients, their standard errors (in parenthesis), and 95% confidence intervals (in square brackets).The instrument used in the MR models is the polygenic score for education years based on genetic markers ($p < 0.01$). The unit of analysis is the individual. The six cohorts under study are drawn from the Young Finns study. The models include (unreported) controls for indicators for age, sex (Columns 1 and 2, being female), region of residence in 1980, and parental education in 1980. Statistically significant at the * 10%, ** 5%, and *** 1% levels.

identify causal effects arrived at mixed conclusions regarding the effectiveness of education on improving health behaviors. Typically, these studies have been less optimistic about the effectiveness of schooling than association studies reporting correlations between the variables of interest. Although some studies have found that education decreased smoking (Arendt, 2005; Jensen and Lleras-Muney, 2012; Jürges et al., 2011) and (binge) drinking (Li and Powdthavee, 2015), as well as increased consumption of fruit and vegetables (Li and Powdthavee, 2015), many studies have found no effect of education on these health behaviors (Braakmann, 2011; Clark and Royer, 2013; Dilmaghani, 2021; Kemptner et al., 2011; Fletcher, 2015). Studies that have used MR as an identification strategy have found that higher education reduced smoking (Gage et al., 2018; Davies et al., 2019; Sanderson et al., 2019) and increased alcohol intake (Davies et al., 2019; Zhou et al., 2019) because education increased drinking frequency (Rosoff et al., 2021; Zhou et al., 2019). Our results provided support for the view that education reduces smoking and possibly increases fruit and vegetable consumption. In accordance with Rosoff et al. (2021), we did not find a significant link between education and abusive alcohol consumption.

Four issues should be considered when interpreting the results. First, it is possible that individuals with higher education levels are more sensitive to providing socially desirable answers to questions related to health behaviors, which may have biased our results. Second, MR relies on the assumption that the instrument is valid. A potential threat to the validity assumption is pleiotropy—that is, that the SNPs in the PGSs affected health behaviors through pathways other than education. To evaluate the sensitivity of results to this issue we followed van Kippersluis and Rietveld (2018) approach to determine how strong the violation of the exclusion restriction would have to be for the causal effect to be zero. We found that in the case of fruit and vegetable consumption, a relatively mild violation of the exclusion would reduce the causal estimate to zero. Regarding smoking and attention to healthy habits, the results were less sensitive, but the possibility of zero effects also applied to these findings. Furthermore, our analyses indicated that the education PGS was associated with other PGSs that may also affect health behaviors. However, controlling for these additional PGSs had only a minor effect on the point estimates. It is not possible to prove instrument validity in an MR or, in general, in any instrumental variable study, but the robustness tests we performed were consistent with instrument validity. Third, the LATE that the MR model identified captured the average treatment effect among those whose years of education were higher because of genetic inheritance. Variation in years of education driven by other factors, for example, changes in schooling laws, could have different impacts on health behaviors. Fourth, in all but one case (abusive alcohol consumption), the MR estimates suggested a stronger link between education and health behaviors than the OLS estimates. This difference may stem from two sources. First, there may have been unobserved confounders that biased the OLS estimate downward. Second, although OLS identified the average treatment effect (ATE) if education was uncorrelated with the error term, the LATE, which the MR method identified, captured the ATE among compilers—that is, those whose years of education increased via the impact of our genetic instrument. Thus, these two methods may capture different treatment parameters in a setting where there are heterogeneous treatment effects.

Our findings suggest that the benefits of education are not restricted to better labor market outcomes, but that education also promotes less risky health behaviors. Finland is a Nordic welfare state with relatively low differences in health outcomes (e.g., life expectancy) between different socioeconomic groups (OECD, 2019) and a highly ranked comprehensive education system financed by the government, a system that is tuition-free for all students (OECD, 2020). Our findings suggest that even in a country where social equality has been a key policy priority for many decades, education may lead to meaningful differences in health behaviors and, thus, health outcomes in the long term. This implies that public and health policies should pay particular attention to low-educated individuals and to their investments in health capital. If socioeconomic differences in health behaviors accumulate to large differences in health outcomes in the long run, the health care system may not be able to mitigate these differences.

6. Conclusions

In our study, we showed that higher education led to better choices in terms of some dimensions of health behaviors. Thus, the benefits of higher education are not only restricted to monetary outcomes (i.e., higher employment, lower unemployment, and higher earnings), which have been extensively documented in the literature (e.g., Card, 1999), but they also apply to nonmonetary domains of human well-being.

Funding

The Young Finns Study has been financially supported by the Academy of Finland: grant numbers 322098, 286284, 134309 (Eye),

126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), 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 number X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research; Finnish Cultural Foundation; the Sigrid Juselius Foundation, Finland; Tampere Tuberculosis Foundation, Finland; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Jenny and Antti Wihuri Foundation; Diabetes Research Foundation of Finnish Diabetes Association; EU Horizon 2020, Belgium (grant number 755320 for TAX-INOMISIS); European Research Council, Belgium (grant number 742927 for MULTIEPIGEN project); Tampere University Hospital Supporting Foundation; and the Society of Finnish Clinical Chemistry. The use of the YFS-FLEED-LPC data has been supported by Palkansaajasäätiö and OP Group Research Foundation.

CRediT authorship contribution statement

Jutta Viinikainen Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing; **Alex Bryson** Conceptualization, Methodology, Writing – review & editing; **Petri Böckerman** Conceptualization, Methodology, Writing – review & editing; **Jaana T. Kari** Conceptualization, Writing – review & editing; **Terho Lehtimäki** Conceptualization, Writing – review & editing; **Olli Raitakari** Conceptualization, Writing – review & editing, Funding acquisition, Project administration (YFS); **Jorma Viikari** Conceptualization, Writing – review & editing; **Jaakko Pehkonen** Conceptualization, Writing – review & editing, Funding acquisition, Project administration (YFS-FLEED-LPC).

Declarations of interest

None.

Appendix 1. Items indicating abusive alcohol consumption

	Share of "yes" answers
Have you ever had the habit of drinking alcohol before going to a party?	56.80%
Do you have a habit of drinking a bottle of wine or a similar amount of beer or other alcoholic drinks over the weekend?	43.63%
Have you ever had a daily habit of drinking a small amount of alcohol to relax?	24.77%
Have you ever felt that you need to drink more alcohol than before to have the same effect?	16.86%
Have you ever had trouble drinking less than your friends?	8.76%
Have you ever fallen asleep after drinking a reasonable amount of alcohol without knowing how you went to bed?	24.47%
Have you ever had a bad conscience after drinking alcohol?	51.90%
Have you ever taken a drink to cure a hangover?	23.26%
Have you ever tried to avoid drinking for a certain period, for example, a week?	34.14%
Have you ever found it difficult to stop drinking once you have started?	13.60%
Has a relative or a friend been concerned about your drinking or suggested you cut down?	16.50%
The average number of "yes" answers	3.147

Notes: The questions were answered on a yes/no scale (0 = no; 1 = yes). The sum of items indicates the severity of excessive drinking behavior. n = 1655.

APPENDIX 2. Genotyping and calculation of the PGS

Genotyping in the YFS was implemented using the Illumina Bead Chip (Human 670 K) from 2442 YFS participants, including 546,677 SNPs, and the genotypes were called using the Illumina clustering algorithm (Teo et al., 2007). Quality control was performed using the Sanger genotyping QC pipeline, and individuals with possible relatedness were removed. SHAPEIT v1 and IMPUTE 2 software (Delaneau et al., 2012) were used for genotype imputation with the 1000 Genomes Phase I Integrated Release Version 3 (March 2012 haplotypes) as a reference panel (Howie et al., 2009; 1000 Genomes Project Consortium, 2010).

GeneticRiskScoreCalculator version 0.1.0c was used for calculating several PGSs in parallel. Independent effect SNPs for each summary-statistic file were identified by double-clumping, first using a 350-kb distance and then a 5-Mb distance with $R^2 = 0.1$ as a linkage-disequilibrium threshold. A weighted PGS was calculated by summing risk alleles for each independent SNP using its GWAS effect size (β or log (OR) from the GWAS study) as a weight. Five GWAS *p*-value thresholds ($p < 5 \times 10^{-8}$, 1×10^{-5}, 1×10^{-4}, 1×10^{-3}, and 1×10^{-2}) were used for constructing PGSs for each summary-statistic file. The human leukocyte antigen region (chr6:25,000,000–35,000,000) was omitted from the calculations. The PGSs were scaled between 0 and 2 for compatibility with the QTL-mapping pipeline (Võsa et al., 2021).

Appendix 3. Balance test

	Coefficient (SE)	95% confidence interval
Parental education	0.171 *** (0.021)	0.129; 0.212

Notes: Table reports the OLS results from regression models where the instrument for education was regressed on parental education, while simultaneously controlling for sex, age cohort, and regional fixed effects. Statistically significant at the * 10%, * *5%, and * ** 1% levels. n = 2412.

Appendix 4. Education and health behaviors: robustness analyses

	(1) MR Additional PGS controls	(2) OLS Reduced-form model	(3) MR Using 74 SNP PGS
Attention toward healthy habits	0.141 * **	0.214 * **	0.199 * *
	(0.031)	(0.043)	(0.101)
	[0.081; 0.201]	[0.130; 0.298]	[− 0.000; 0.398]
First-stage F statistics	151.18	.	15.73
R^2	.	0.071	.
N	1719	1719	1719
Abusive alcohol consumption	-0.056	-0.062	0.015
	(0.086)	(0.120)	(0.249)
	[− 0.225; 0.112]	[− 0.297; 0.174]	[− 0.473; 0.503]
First-stage F statistics	146.67	.	16.73
R^2	.	0.064	.
N	1655	1655	1655
Smoking (pack years)	-0.830 * **	-1.193 * **	-1.538 * *
	(0.204)	(0.289)	(0.711)
	[− 1.229; − 0.430]	[− 1.760; − 0.625]	[− 2.930; − 0.145]
First-stage F statistics	144.52	.	11.21
R^2	.	0.043	.
N	1620	1620	1620
Fruit and vegetable consumption (g/day)	15.851 *	20.076	-2.117
	(9.326)	(12.306)	(38.188)
	[− 2.427; 34.129]	[− 4.063; 44.215]	[− 76.964; 72.729]
First-stage F statistics	119.12	.	8.55
R^2	.	0.059	.
N	1468	1468	1468

Notes: The table reports regression coefficients, their standard errors (in parenthesis), and 95% confidence intervals (in square brackets). The reported MR coefficients for years of education are obtained from models which use the polygenic score (PGS) for years of education as an instrument. Column 1 utilizes a PGS that was obtained using a significance threshold $p < 0.01$ and Column 2 a PGS based on 74 SNPs. The models in Column 1 also include additional controls for PGSs for the Big Five personality traits, depressive symptoms, bipolar disorders, schizophrenia, childhood IQ, BMI, and height. Column 2 shows the OLS coefficients for PGS ($p < 0.01$) for years of education. The unit of analysis is the individual. The six cohorts under study are drawn from the Young Finns study. The models include (unreported) controls for indicators for age, sex, region of residence in 1980, and parental education in 1980. Statistically significant at the * 10%, * *5%, and * ** 1% levels.

Appendix 5. Pleiotropy-robust Mendelian randomization

The plausibly exogenous estimation method (Conley et al., 2012) is an instrumental variables approach that can be applied when the instrument validity is debatable. It provides a method for performing inference in a situation where the exclusion restriction does not necessarily hold precisely—i.e., the instrument is plausibly or approximately exogenous. van Kippersluis and Rietveld (2018) applied this method to the MR setting (pleiotropy-robust Mendelian randomization, PRMR) and used it as a sensitivity analysis to determine how strong the violation of the exclusion restriction would have to be for the causal effect to be zero.

In their example, van Kippersluis and Rietveld (2018) estimated the effect of educational attainment on BMI. We followed their approach (and STATA codes) to address the potential horizontal pleiotropy problem. Fig. A1 plots the results. Alcohol consumption was omitted from the analyses because our MR results did not imply a causal link between education and excessive alcohol consumption. In Fig. A1, lambda refers to the percentage of standardized effect of the PGS on years of education, which is the direct effect of the PGS on standardized health behavior outcomes. The estimate for lambda = 0 corresponds to the MR point estimate. Moving to the right on the x-axis implies a stronger violation of the exclusion restriction. The results showed that the causal effect of education on attention to healthy habits was estimated to be zero when lambda = 0.43. Thus, a violation of the exclusion restriction 43% of the first stage effect would decrease the point estimate to zero. Regarding smoking and fruit and vegetable consumption, the corresponding lambdas were 0.44 and 0.19, respectively. Furthermore, the 95% confidence intervals included zeros at the following levels of lambda: attention to healthy habits: 0.2, smoking: 0.2, and fruit and vegetable consumption: 0. In their study, van Kippersluis and Rietveld (2018) interpreted that "a relative mild violation" of the exclusion restriction of 29% based on the confidence interval produced "at best weak evidence" of causal effects. Our PRMR results regarding attention to healthy habits and smoking were comparable to those of van Kippersluis and Rietveld (2018), implying that it is important to consider the possibility that zero effect cannot be ruled out. In the case of fruit and vegetable consumption, even a milder violation of the exclusion restriction would reduce the causal effect to zero.

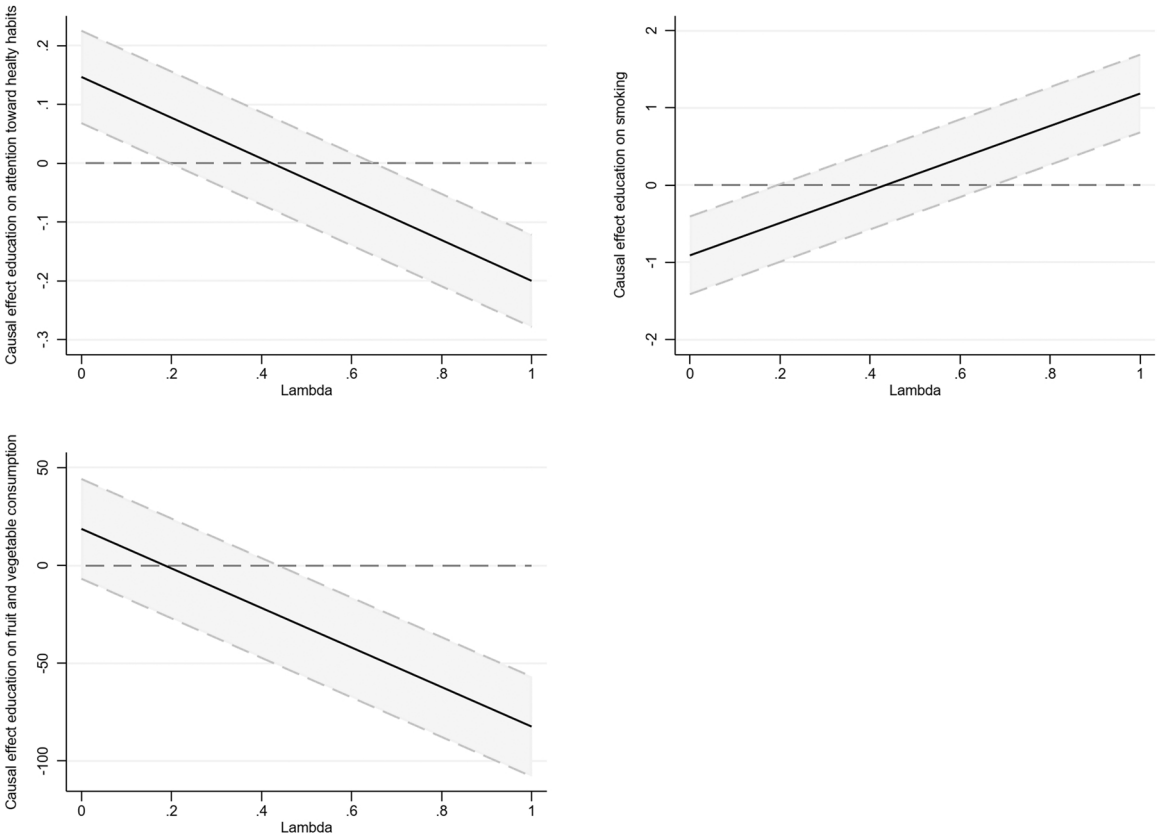

Fig. A1. The causal effect of education on healthy habits for varying values of lambda. Note: The gray area indicates the 95% confidence interval.

References

1000 Genomes Project Consortium, 2010. A map of human genome variation from population-scale sequencing. Nature 467 (7319), 1061–1073.

Almond, D., Mazumder, B., 2005. The 1918 influenza pandemic and subsequent health outcomes: an analysis of SIPP data. Am. Econ. Rev. 95 (2), 258–262.

Almond, D., Currie, J., Duque, V., 2018. Childhood circumstances and adult outcomes: Act II. J. Econ. Lit. 56 (4), 1360–1446.

Arendt, J.N., 2005. Education effects on health: a panel data analysis using school reform for identification. Econ. Educ. Rev. 24 (2), 149–160.

Benyamin, B., Pourcain, B., Davis, O.S., et al., 2014. Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. Mol. Psychiatry 19 (2), 253–258.

Böckerman, P., Viinikainen, J., Pulkki-Råback, L., Hakulinen, C., Pitkänen, N., Lehtimäki, T., Pehkonen, J., Raitakari, O.T., 2017. Does higher education protect against obesity? Evidence using Mendelian randomization. Prev. Med. 101, 195–198.

Braakmann, N., 2011. The causal relationship between education, health and health related behaviour: evidence from a natural experiment in England. J. Health Econ. 30 (4), 753–763.

Burton, P.R., Tobin, M.D., Hopper, J.L., 2005. Key concepts in genetic epidemiology. Lancet 366 (9489), 941–951.

Card, D., 1999. The causal effect of education on earnings. In: Ashenfelter, O.C., Card, D. (Eds.), Handbook of Labor Economics, Vol. 3A. Elsevier, pp. 1801–1863.

Cawley, J., Ruhm, C.J., 2011. The economics of risky health behaviors. In: Pauly, M.V., McGuire, T.G., Barros, P.P. (Eds.), Handbook of Health Economics, Vol. 2. Elsevier, pp. 95–199.

Chay, K.Y., Greenstone, M., 2003. The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession. Q. J. Econ. 118 (3), 1121–1167.

Clark, D., Royer, H., 2013. The effect of education on adult mortality and health: Evidence from Britain. Am. Econ. Rev. 103 (6), 2087–2120.

Conley, T.G., Hansen, C.B., Rossi, P.E., 2012. Plausibly exogenous. Rev. Econ. Stat. 94 (1), 260–272.

Cook, P.J., Moore, M.J., 1994. This tax's for you: the case for higher beer taxes. Natl. Tax. J. 47 (3), 559–573.

Currie, J., 2009. Healthy, wealthy, and wise: socioeconomic status, poor health in childhood, and human capital development. J. Econ. Lit. 47 (1), 87–122.

Cutler, D.M., Lleras-Muney, A., 2006. Education and health: evaluating theories and evidence. Natl. Bur. Econ. Res. Work. Pap. 12352.

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

Davies, N.M., Hill, W.D., Anderson, E.L., Sanderson, E., Deary, I.J., Smith, G.D., 2019. Multivariable two-sample Mendelian randomization estimates of the effects of intelligence and education on health. Elife 8, e43990.

De Irala-Estevez, J., Growth, M., Johansson, L., Oltersdorf, U., Prättälä, R., Martínez-González, M.A., 2000. A systematic review of socio-economic differences in food habits in Europe: consumption of fruit and vegetables. Eur. J. Clin. Nutr. 54 (9), 706.

de Moor, M.H., Costa, P.T., Terracciano, A., et al., 2012. Meta-analysis of genome-wide association studies for personality. Mol. Psychiatry 17 (3), 337–349.

Delaneau, O., Marchini, J., Zagury, J.F., 2012. A linear complexity phasing method for thousands of genomes. Nat. Methods 9 (2), 179–181.

Demange, P., Malanchini, M., Mallard, T., Biroli, P., 2021. Investigating the genetic architecture of non-cognitive skills using GWAS-by-subtraction. Net. Genet. 53, 35–44.

Dilmaghani, M., 2021. Education, smoking and health: evidence from Canada. Educ. Econ. 29 (5), 490–508.

Farrell, P., Fuchs, V.R., 1982. Schooling and health: the cigarette connection. J. Health Econ. 1 (3), 217–230.

Fletcher, J.M., 2015. New evidence of the effects of education on health in the US: Compulsory schooling laws revisited. Soc. Sci. Med. 127, 101–107.

Furberg, H., Kim, Y., Dackor, J., et al., 2010. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat. Genet. 42 (5), 441–447.

Furnee, C.A., Groot, W., van den Brink, H., 2008. The health effects of education: a meta-analysis. Eur. J. Public Health 18 (4), 417–421.

Gage, S.H., Bowden, J., Davey Smith, G., Munafo, M.R., 2018. Investigating causality in associations between education and smoking: a two-sample Mendelian randomization study. Int. J. Epidemiol. 47 (4), 1131–1140.

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

Grossman, M., 2006. Education and nonmarket outcomes. In: Hanushek, E., Welch, F. (Eds.), Handbook of the Economics of Education, first ed., Vol 1. Elsevier/North Holland, pp. 577–633.

Hamad, R., Elserb, H., Tranc, D.C., Rehkopfc, D.H., Goodman, S.N., 2018. How and why studies disagree about the effects of education on health: a systematic review and meta-analysis of studies of compulsory schooling laws. Soc. Sci. Med. 212, 168–178.

von Hinke, S., Smith, G.D., Lawlor, D.A., Propper, C., Windmeijer, F., 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 (6), e1000529.

Jensen, R., Lleras-Muney, A., 2012. Does staying in school (and not working) prevent teen smoking and drinking? J. Health Econ. 31 (4), 644–657.

Jürges, H., Reinhold, S., Salm, M., 2011. Does schooling affect health behavior? Evidence from the educational expansion in Western Germany. Econ. Educ. Rev. 30 (5), 862–872.
Kemptner, D., Jürges, H., Reinhold, S., 2011. Changes in compulsory schooling and the causal effect of education on health: evidence from Germany. J. Health Econ. 30 (2), 340–354.
Lam, M., Hill, W.D., Trampush, J.W., et al., 2019. Pleiotropic meta-analysis of cognition, education, and schizophrenia differentiates roles of early neurodevelopmental and adult synaptic pathways. Am. J. Hum. Genet. 105 (2), 334–350.
Lee, P.H., Anttila, V., Won, H., et al., 2019. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. Cell 179 (7), 1469–1482.
Li, J., Powdthavee, N., 2015. Does more education lead to better health habits? Evidence from the school reforms in Australia. Soc. Sci. Med. 127, 83–91.
Locke, A.E., Kahali, B., Berndt, S.I., et al., 2015. Genetic studies of body mass index yield new insights for obesity biology. Nature 518 (7538), 197–206.
McMartin, A., Conley, D., 2020. Commentary: Mendelian randomization and education–Challenges remain. Int. J. Epidemiol. 49 (4), 1193–1206.
OECD, 2019. Health at a glance 2019: OECD Indicators. OECD Publishing, Paris.
OECD, 2020. Education policy outlook: Finland. OECD Publishing, Paris (Accessed 1 November 2021). ⟨http://www.oecd.org/education/policy-outlook/country-profile-Finland-2020.pdf⟩.
Okbay, A., Beauchamp, J.P., Fontana, M.A., et al., 2016a. Genome-wide association study identifies 74 loci associated with educational attainment. Nature 533 (7604), 539–542.
Okbay, A., Baselmans, B.M., De Neve, J.E., et al., 2016b. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. Nat. Genet. 48 (6), 624–633.
Paalanen, L., Männistö, S., Virtanen, M.J., Knekt, P., Räsänen, L., Montonen, J., Pietinen, P., 2006. Validity of a food frequency questionnaire varied by age and body mass index. J. Clin. Epidemiol. 59 (9), 994–1001.
Perez-Arce, F., 2017. The effect of education on time preferences. Econ. Educ. Rev. 56, 52–64.
Polderman, T.J., Benyamin, B., De Leeuw, C.A., Sullivan, P.F., van Bochoven, A., Visscher, P.M., Posthum, D., 2015. Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat. Genet. 47 (7), 702.
Raitakari, O.T., Juonala, M., Rönnemaa, T., et al., 2008. Cohort profile: the cardiovascular risk in Young Finns Study. Int. J. Epidemiol. 37 (6), 1220–1226.
Renna, F., 2007. The economic cost of teen drinking: late graduation and lowered earnings. Health Econ. 16 (4), 407–419.
Roseboom, T.J., Painter, R.C., van Abeelen, A.F., Veenendaal, M.V., de Rooij, S.R., 2011. Hungry in the womb: what are the consequences? Lessons from the Dutch famine. Maturitas 70 (2), 141–145.
Rosoff, D.B., Clarke, T.K., Adams, M.J., McIntosh, A.M., Smith, G.D., Jung, J., Lohoff, F. W., 2021. Educational attainment impacts drinking behaviors and risk for alcohol dependence: results from a two-sample Mendelian randomization study with~ 780,000 participants. Mol. Psychiatry 26, 1119–1132.
Ruderfer, D.M., Fanous, A.H., Ripke, S., et al., 2014. Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. Mol. Psychiatry 19 (9), 1017–1024.
Sánchez, A.C., Tassot, C., 2014. An exploration of the determinants of the subjective well-being of Americans during the great recession. OECD Economics Department Working Papers, No. 1158. OECD Publishing, Paris. https://doi.org/10.1787/5jxzc9p57lq3-en. Accessed 1 November 2021.
Sanderson, E., Smith, G.D., Bowden, J., Munafò, M.R., 2019. Mendelian randomisation analysis of the effect of educational attainment and cognitive ability on smoking behaviour. Nat. Commun. 10 (1), 1–9.
Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511 (7510), 421–427.
Schoeps, A., van Ewijk, R., Kynast-Wolf, G., Nebié, E., Zabré, P., Sié, A., Gabrysch, S., 2018. Ramadan exposure in utero and child mortality in Burkina Faso: analysis of a population-based cohort including 41,025 children. Am. J. Epidemiol. 187 (10), 2085–2092.
Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 447–486.
Strine, T.W., Chapman, D.P., Balluz, L.S., Moriarty, D.G., Mokdad, A.H., 2008. The associations between life satisfaction and health-related quality of life, chronic illness, and health behaviors among US community-dwelling adults. J. Community Health 33 (1), 40–50.
Teo, Y.Y., Inouye, M., Small, K.S., Gwilliam, R., Deloukas, P., Kwiatkowski, D.P., Clark, T.G., 2007. A genotype calling algorithm for the Illumina BeadArray platform. Bioinformatics 23 (20), 2741–2746.
van der Heide, I., Wang, J., Droomers, M., Spreeuwenberg, P., Rademakers, J., Uiters, E., 2013. The relationship between health, education, and health literacy: results from the Dutch Adult Literacy and Life Skills Survey. J. Health Commun. 18 (sup1), 172–184.
van der Valk, R.J., Kreiner-Møller, E., Kooijman, M.N., et al., 2015. A novel common variant in DCST2 is associated with length in early life and height in adulthood. Hum. Mol. Genet 24 (4), 1155–1168.
van Kippersluis, H., Rietveld, C.A., 2018. Pleiotropy-robust Mendelian randomization. Int. J. Epidemiol. 47 (4), 1279–1288.
VanderWeele, T.J., Tchetgen, E.J.T., Cornelis, M., Kraft, P., 2014. Methodological challenges in Mendelian randomization. Epidemiology 25 (3), 427.
Võsa, U., et al., 2021. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat. Genet 53 (9), 1300–1310.
Xue, X., Cheng, M., Zhang, W., 2021. Does education really improve health? A meta-analysis. J. Econ. Surv. 35 (1), 71–105.
Zhou, T., Sun, D., Li, X., Ma, H., Heianza, Y., Qi, L., 2019. Educational attainment and drinking behaviors: mendelian randomization study in UK Biobank. Mol. Psychiatry 26, 4355–4366.